Published in final edited form as:

J Chromatogr A. 2011 September 28; 1218(39): 7003–7015. doi:10.1016/j.chroma.2011.08.023.

Transcription Factor Proteomics: Identification by a Novel Gel Mobility Shift-Three-Dimensional Electrophoresis Method Coupled with Southwestern Blot and HPLC-Electrospray-Mass Spectrometry Analysis

Daifeng Jiang, Yinshan Jia, and Harry W. Jarrett*

Department of Chemistry, University of Texas San Antonio, San Antonio, TX 78249

Abstract

Transcription factor (TF) purification and identification is an important step in elucidating gene regulatory mechanisms. In this study, we present two new electrophoretic mobility shift assay (EMSA)-based multi-dimensional electrophoresis approaches to isolate and characterize TFs, using detection with either southwestern or western blotting and HPLC-nanoESI-MS/MS analysis for identification. These new techniques involve several major steps. First, EMSA is performed with agents that diminish non-specific DNA-binding and the DNA-protein complex is separated by native PAGE gel. The gel is then electrotransferred to PVDF membrane and visualized by autoradiography. Next, the DNA-protein complex, which has been transferred onto the blot, is extracted using a detergent-containing elution buffer. Following detergent removal, concentrated extract is separated by SDS-PAGE (EMSA-2DE), followed by in-gel trypsin digestion and HPLC-nanoESI-MS/MS analysis, or the concentrated extract is separated by two-dimensional gel electrophoresis (EMSA-3DE), followed by southwestern or western blot analysis to localize DNA binding proteins on blot which are further identified by on-blot trypsin digestion and HPLC-nanoESI-MS/MS analysis. Finally, the identified DNA binding proteins are further validated by EMSA-immunoblotting or EMSA antibody supershift assay. This approach is used to purify and identify GFP-C/EBP fusion protein from bacterial crude extract, as well as purifying AP1 and CEBP DNA binding proteins from a human embryonic kidney cell line (HEK293) nuclear extract. AP1 components, c-Jun, Jun-D, c-Fos, CREB, ATF1 and ATF2 were successfully identified from 1.5 mg of nuclear extract (equivalent to 3×10^7 HEK293 cells) with AP1 binding activity of 750 fmol. In conclusion, this new strategy of combining EMSA with additional dimensions of electrophoresis and using southwestern blotting for detection proves to be a valuable approach in the identification of transcriptional complexes by proteomic methods.

Keywords

Electrophoresis; Transcription Factor; Proteomics; AP1; C/EBP

© 2011 Elsevier B.V. All rights reserved.

Address correspondence to Dr. Harry W. Jarrett, Department of Chemistry, One UTSA Circle, University of Texas San Antonio, San Antonio, TX 78249; Telephone: (210) 458-7053; FAX (210) 458-7428; harry.jarrett@utsa.edu.

Supporting Information Available: The supporting information contains one table, showing the matches of MS/MS sequencing of the peptide depicted in Fig. 1D, and one figure showing how EMSA for AP1 was systematically optimized.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

There is considerable interest in transcription factors (TFs) and their role in gene regulation. Gene expression is important to all fundamental biological processes and is regulated by sequence-specific DNA binding proteins. Prior to transcription initiation, a transcription complex forms at the promoter that contains TFs, DNA and co-regulators, of which the TFs are the core members. The individual transcription factors bind to discrete, specific DNA-sequences, called response elements, in the promoter region. The formed DNA-protein complex functions either to activate or repress the expression of a target gene by integrating signaling cues through protein-protein interaction and translating these cues into transcriptional regulation [1].

In humans, TFs compose the second largest group of proteins after the metabolic enzymes. They play a central role in many biological processes including cell cycle regulation [2], maintenance of the intracellular environment, cellular differentiation and development [3,4]. Abnormal TF activity leads to numerous diseases and developmental disorders [5,6]. Surprisingly little is known about TF proteins due to their difficulty of study. For example, how many transcription factors does the human genome encode? Based on annotations of DNA-binding domains of TF, the initial analyses of the human genome sequence estimated that there are 2,000 to 3,000 sequence-specific TFs [7]. The DBD database (<http://dbd.mrc-lmb.cam.ac.uk/DBD/index.cgi?Home>) predicts 1,508 human loci as potential TFs in human [8]. According to the characteristic DNA-binding domains, TFs are classified into different families by InterPro, the C2H2 zinc-finger (675 TFs), homeodomain (257 TFs), helix-loop-helix (87 TFs) and basic leucine zipper (bZIP, about 50 TFs) account for over 80% of the total TF repertoires. Among the 20 most highly cited transcription factors (TFs) in PubMed, the transcription factors Jun and Fos, which are the most common components of activator protein 1 (AP1) complex, rank third and fifth, respectively, of studies performed in humans and all other organisms [9]. c-Jun and c-Fos function to regulate various cellular behaviors, from the cell cycle, proliferation and development, the stress response and apoptosis. AP1 represents diverse homodimeric or heterodimeric combinations of members of the Jun family (JUN, JUNB and JUND), Fos (FOS, FOSB, FRA1 and FRA2), the closely related activating transcription factor (ATF and CREB) subfamily, the Maf subfamily and other bZIP TFs [10]. The dimeric combinations of AP1 mainly depends on the cell or tissue-specific expression patterns of individual proteins and their post-translational modifications in response to extracellular stimulation. The individual proteins dimerize with other partners and bind DNA via the bZIP domain [11]. For example, Jun proteins can form stable homodimers or form heterodimers with Fos that bind to the TPA response element (TRE, 5'-TGAC/GTCA-3') based on their ability to mediate transcriptional induction in response to the phorbol ester TPA. Conversely, Jun and ATF proteins form heterodimers that preferentially bind the cyclic AMP responsive elements (CRE, 5'-TGACGTCA-3') [12,13]. The different AP1 dimers bind to DNA with different affinities, resulting in different transactivation activity, protein stability and localization, and ultimately influencing the transcriptional repertoire of these proteins [14]. So identifying the protein composition of TF complexes and their post-translational modification becomes important to elucidate gene regulatory mechanism.

Currently, less than 5% of TFs have been experimentally characterized [15]. TF purification is the very first step for not only biochemical and structure study, but also for TF functional investigation in gene regulation. TF purification and characterization is complicated by the large numbers of TFs and their relatively low abundance in human cells, which is always less than 0.1% of the total nuclear protein [16]. Assuming 1 pmol of pure TF (50 ng for a 50 kDa. protein) is required for successful MS identification, it is estimated that 2×10^9 – 10^{11} cells are needed for a TF (10^3 – 10^5 molecules per cell) purification, supposing the overall

yield will be 50%. So purification of TFs often requires large amounts of starting material. Successful characterization requires sensitive analytical methods such as mass spectrometry (MS). The significant increase in mass accuracy and resolving power now allows for the identification of low abundant proteins as their spectral peaks become increasingly distinguishable from background noise in complex mixtures, especially when the MS is equipped with advanced mass analyzer such as the Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR) or the Orbitrap [1].

Since TFs have an affinity several orders of magnitude greater for their consensus binding site (response element) than for non-specific DNA, DNA affinity chromatography was developed to isolate sequence-specific-binding TFs. In this technique, double stranded oligonucleotides containing a consensus binding site for a TF is either absorbed or linked covalently to a chromatographic support, and TFs are purified by exploiting the inherent capacity of the TF to bind to DNA. Single-step DNA affinity chromatography results in high amounts of nonspecific co-purifying proteins that interfere or mask the identification of low-abundant TF [9]. Currently, TF purification often involves four or five different chromatographic steps that may include ion exchange, gel filtration, nonspecific and sequence-specific DNA affinity columns [15]. In some studies, the consensus site-mutated DNA was used to pre-clean prior to specific DNA affinity chromatography [17]. Recently, Mann et al combined one-step DNA affinity chromatography with stable isotope labeling with amino acid in cell culture (SILAC) quantitative proteomics to identify specific TFs that bind to the methyl-CpG site in a promoter. This strategy eliminated a large excess of copurifying background proteins by calculating the peptide isotope ratios between wild type and the methyl-CpG bait, and only the significantly expressed proteins were taken as methyl-CpG binding candidates [18]. Despite the success of this approach, it is laborious, and since it requires homogeneous tissue cultured cells, is not feasible for all research. Our research group has been endeavoring to develop simple and rapid techniques to improve TF purification for over a decade, developing systematically optimized oligonucleotide trapping (SOT) [19], promoter trapping (PT) [20] and two-dimensional southwestern blot (2D-SW) [21] to purify and identify c-Jun promoter binding proteins [20], CCAAT enhancer binding protein (C/EBP) [19] and MafA in cell nuclear extract [22]. However, in oligonucleotide or promoter trapping, the single stranded (GT)₅ tail at 3'-end of double stranded oligonucleotide or promoter DNA causes unspecific binding of heterogeneous nuclear ribonucleoprotein (hnRNP) family members and several other abundant nuclear proteins. To avoid interference of the single stranded tail, in this study, we explored the electrophoretic mobility shift assay (EMSA)-based purification using DNA without a tail and combining this with additional dimensions of electrophoresis.

Developed in 1981, EMSA is extensively used to monitor the ability of a protein binding to a segment of DNA and to quantify DNA binding activity [23,24]. The DNA used in EMSA is typically a double-stranded oligonucleotide of 20–25 bp containing a response element (RE) and it can be radio-, fluoro- or hapten-labeled. An image of the gel reveals the positions of the free DNA and TF-DNA complex. The formed TF-DNA complex migrates more slowly than the free DNA when subjected to non-denaturing polyacrylamide gel electrophoresis (PAGE). Previous studies have combined EMSA with proteomic techniques for TF identification. TFs were separated by two-dimensional gel electrophoresis (2-DE), which combines isoelectric focusing (IEF) in a first dimension with SDS-PAGE in a second dimension. Spots were cut from the gel for MS identification according to their pI and molecular mass and the proteins were also eluted and renatured for EMSA assay [16]. This method requires that the proteins be efficiently renatured from a denaturing (SDS-PAGE) gel for DNA binding activity analysis by EMSA. Only monomeric or homodimeric proteins binding with DNA can be characterized since heterodimeric proteins are dissociated during the denaturing gel separation. Another technique called 2D-EMSA [25] has also been

developed. A TF is separated by non-denaturing PAGE in either the presence or absence of DNA. The protein bands from the two conditions are then applied to a second dimension of SDS-PAGE. The protein showing a DNA-dependent electrophoretic shift was extracted from the corresponding SDS-PAGE gel for MS identification. This approach successfully identified recombinant AtrA in a bacterial lysate [26], but for low abundance TFs in a complicated cell nuclear extract, this technique requires improvement. Our group has also applied bands cut from EMSA gels for western blotting analysis, which helped to verify the identity of a TF [27].

Based on these previous studies, we developed a new EMSA-based purification approach and combined it with proteomic techniques for TF identification. The DNA-protein complex was extracted from EMSA and applied to additional dimensions of electrophoresis. In initial tests, these techniques enable successful purification and identifying GFP-C/EBP fusion protein from bacterial crude extract and allowed the characterization of the AP1 and CEBP complexes from a human cell nuclear extract. Overall, these techniques provide a powerful and robust method for investigating the low abundance transcription factor proteome.

2. Materials and Methods

2.1 Antibodies and reagents

Heparin (H-3393), poly dI:dC (deoxyinosinic–deoxycytidylic acid, P4929), sodium dodecylsulfate (SDS), dithiothreitol (DTT), iodoacetamide (IDA), trifluoroacetic acid (TFA), polyvinylpyrrolidone 40 (PVP40), zwittergent 3–16, acetonitrile (ACN), trypsin (sequencing grade), 2-mercaptoethanol and Tris were from Sigma-Aldrich (St. Louis, MO, USA). Tween-20, Triton X-100, NP-40 (Nonylphenyl-polyethylene glycol), DOC (Deoxycholic acid), CHAPS (3-[(3-Cholamidopropyl) dimethylamino]-1-propanesulfonate) and OGP (octyl β -D-glucopyranoside) were from Fisher Scientific (Houston, TX, USA), T4 polynucleotide kinase (10 000 U/mL) was from New England Biolabs (Beverly, MA, USA). Nitrocellulose (NC), polyvinylidene fluoride (PVDF) membranes, urea, thiourea, ampholytes, IPG strips and other gel electrophoresis reagents were from BioRad Laboratories (CA, USA). Antibodies including C/EBP, c-Jun, c-Fos, Jun-B, Jun-D, ATF1, ATF2, NF- κ B p65, and Lamin A/C were from Santa Cruz Biotechnology (CA, USA); SP1 antibody was from Upstate Inc. (Chicago, IL, USA), goat anti-rabbit or goat anti-mouse secondary antibody-HRP conjugate was from SouthernBiotech (Birmingham, Alabama, USA). All the oligonucleotides were synthesized from Integrated DNA Technologies (Coralville, IA, USA).

2.2 Systematic optimization of electrophoretic mobility shift assay (EMSA) condition

Oligonucleotides were labeled with [γ - ^{32} P] ATP using T4 polynucleotide kinase. HEK293 nuclear extract was prepared as previously reported [20]. EMSA was performed as previously described with little modifications [28]. Briefly, nuclear extract (5 mg/ml, 5 μ l) was mixed with 1.6 nM of radiolabeled oligonucleotide in 25 μ l 1 $\times$ EMSA buffer (20 mM HEPES, pH 7.9, 0.1 mM EDTA, 50 mM NaCl, 5 mM MgCl₂, 10% (w/v) glycerol, 1 mM DTT, 0.1% (v/v) Tween-20) containing 20 μ g/ml poly dI:dC. After DNA-protein complex formed at room temperature for 30 minutes, the complex was separated on a pre-electrophoresed 5% nondenaturing polyacrylamide gel (10 cm) in 0.25 $\times$ TBE buffer at 90 volts for 60 minutes and visualized by autoradiography. EMSA optimized condition was determined when the formation of DNA-protein complex was not diminished by added agents which diminish non-specific DNA binding. First, different concentration of [NaCl] from 0.05 to 0.4 M in 1 $\times$ EMSA buffer was assayed and 0.1 M [NaCl] was added in all subsequent EMSA reactions. Second, various detergents (0.1%), including Tween-20, Triton X-100, CHAPS, OGP, NP-40 and DOC were compared, and 0.1% of Tween-20 was added

to the EMSA reaction mixture. Third, a two-fold serial dilution of heparin from 20 to 0.32 ng/ μ l, a competitive inhibitor of DNA-binding [29], was added and the highest non-interfering concentration of heparin was determined as 2.5 ng/ μ l, and added to all subsequent EMSA reactions. Next, a two-fold serial dilution of T₁₈ (5'-TTTTTTTTTTTTTTTTTT-3') from 80 μ M to 1.25 μ M was added with poly dI:dC constant at 20 μ g/ml, and T₁₈ does not interfere with the specifically shifted band in this concentration range, so 10 μ M of T₁₈ was chosen for subsequent reactions. Additionally, a two-fold serial dilution of poly dI:dC from 160 to 2.5 ng/ μ l was added. In the range of 10 to 80 ng/ μ l of poly dI:dC there is no significant interference with complex formation and 20 ng/ μ l of poly dI:dC was chosen for economic reasons. Finally, CEBP mutant DNA (5'-TGCAGAGACTAGTCTCTGCA-3') was added in 2-fold dilution from 800 to 50 nM and 100 nM was selected as the optimal concentration. Taken together, the optimized AP1 EMSA conditions are as follows: HEK293 nuclear extract is diluted five fold in 1 $\times$ EMSA buffer (20 mM HEPES, pH 7.9, 0.1 mM EDTA, 100 mM NaCl, 5 mM MgCl₂, 10% glycerol, 1 mM DTT, 0.1% Tween-20) containing 2.5 ng/ μ l heparin, 10 μ M T₁₈, 20 ng/ μ l poly dI:dC and 100 nM of CEBP mutant DNA to enhance binding specificity.

2.3 AP1 EMSA cold competition and antibody supershift assay

In EMSA cold competition assays, 100-fold molar excess of unlabeled duplex AP1 (5'-ttccggctgagtcataagcg-3' annealed with its complement strand), CEBP (5'-tgcagattgcgcaatctgca-3', self complimentary), NF- κ B (5'-agttgaggggactttcccaggc-3' with its complement strand) or SP1 (5'-acgggcgggcccgcccatggcgggcccgcccgct-3', self complimentary) was added prior to adding radiolabeled AP1 DNA. In EMSA antibody supershift or blockade assay, nuclear extract (25 μ g, 5 μ l) was incubated with antibody (1.0 μ g) for 2 hours at 4°C prior to the addition of the labeled DNA.

2.4 Optimization of EMSA purification from a native gel or blot

To optimize the EMSA purification technique, pure GFP-C/EBP fusion protein, which was overexpressed and purified using Ni²⁺-NTA-agarose chromatography from BL21 crude extract containing GFP-C/EBP [30], was used as the model. Pure GFP-C/EBP (0.3 μ g/ μ l, 30 μ l) was incubated with 100 nM of unlabeled CEBP oligonucleotide and 1.6 nM of labeled CEBP DNA in 1.5 ml 1 $\times$ EMSA buffer containing 20 μ g/ml of dI:dC for EMSA. The formed DNA-protein complex was separated on three, one-well 5% non-denatured PAGE gels (0.5 ml/gel) at 90V for 60 min. One gel was dried for gel extraction; another two gels were electrotransferred to NC or PVDF membrane for use in blot extraction. After autoradiography of the native gel, PVDF or NC blot, the specific DNA-protein complex located on the gel or blot was cut and dried for extraction comparison: 1) *Gel extraction*: the dried gel band, with radiolabeled DNA-protein complex, was divided into 8 equal slices vertically and each slice was washed five times with water at room temperature. The pieces were then homogenized into smaller pieces and soaked twice in different extraction buffers at room temperature for two hours with gentle shaking, during which the samples were sonicated for ten minutes on ice using Fisher FS30 UltraSonic Cleaner; 2) *Blot extraction*: The radiolabeled DNA-protein complex band was divided into 8 equal slices vertically; each slice was cut into smaller pieces, soaked twice in different extraction buffers at room temperature for two hours with gentle shaking. The two extracts for each condition were pooled and detected by fluorescence using a TECAN Infinite M200 (Tecan Systems Inc., San Jose, USA) with an excitation wavelength of 398 nm and an emission wavelength of 512 nm, to detect GFP-C/EBP.

2.5 EMSA-Three-Dimensional Electrophoresis (EMSA-3DE) purification

For EMSA-3DE, there are four steps: 1) *EMSA*: EMSA for in EMSA-3DE includes modifications. For GFP-C/EBP, bacterial crude extract (17 μ g) containing GFP-C/EBP

fusion protein (1 μ g) was mixed with 20 nM of unlabeled CEBP DNA and 1.6 nM of labeled CEBP DNA in 1.5 ml 1 $\times$ EMSA buffer containing poly dI:dC (20 ng/ μ l) and incubated for 30 minutes. For AP1, HEK293 nuclear extract (5 mg/ml, 300 μ l) was incubated with all reagents to minimize non-specific binding (section 2.2) in 1.5 ml of EMSA buffer on ice for 10 minutes, and then 10 nM of unlabeled and 1.6 nM of radiolabeled AP1 oligonucleotides were added. EMSA reaction (1.5 ml) was separated on three one-well (10 cm) pre-electrophoresed 5% nondenaturing polyacrylamide gels and separated at 90V for 90 min. 2) *PVDF membrane blotting*. The three EMSA gels were electrotransferred to a PVDF membrane at 4 $^{\circ}$ C overnight in 20% (v/v) methanol, 25 mM Tris (free base) and 192 mM glycine at 100 mA. 3) *Protein extraction from PVDF blot*: After autoradiography, the DNA-protein complex was cut from the blot at the position that contains the specific complex based on autoradiography. Then the blotted band was divided into smaller slices that were soaked twice in 1 ml extraction buffer (50 mM Tris-HCl, pH 9.0, 50 mM DTT and 0.5% Tween-20) at room temperature for two hours with gentle shaking. 3) *Two-dimensional gel electrophoresis*: Because Tween-20 will interfere with separation of two-dimensional gel electrophoresis, it was removed using the Pierce Detergent Removal Spin column as instructed. Proteins in the extract were then concentrated using an Amicon spin column concentrator (10 kDa cut-off) and separated by two-dimensional electrophoresis as described previously [21].

2.6 EMSA-three-dimensional electrophoresis-southwestern blot (EMSA-3DE-SW) analysis

SW was performed as previously described [21] with minor modifications. EMSA-3DE was performed as described (section 2.5), then the EMSA-3DE gel was electrotransferred to PVDF membrane, blocked and renatured in 2.5% (w/v) non-fat milk in 1 $\times$ incubation buffer (20 mM HEPES, pH 7.9, 50 mM NaCl, 5 mM MgCl₂, 0.1 mM EDTA, 100 μ M ZnSO₄, 0.1% Tween-20, 1 mM DTT) at 4 $^{\circ}$ C overnight. After washing 20 minutes thrice with 1 $\times$ incubation buffer, the blot was probed with 1.6 nM of radiolabeled AP1 or CEBP oligonucleotides in 1 $\times$ incubation buffer containing 0.25% (w/v) BSA and 20 ng/ μ l of poly dI:dC at 4 $^{\circ}$ C overnight. The blot was washed and analyzed by autoradiography.

2.7 Western blotting analysis (WB)

Proteins were separated by EMSA-2DE or EMSA-3DE, then electroblotted onto 0.2 μ m pore PVDF membrane and incubated with antibody. Antibody dilutions for c-Jun, Jun-B, Jun-D, ATF1, ATF2, Lamin A/C or C/EBP were 1:1000; SP1 antibody dilution was 1:3000. Immunoreactive proteins were detected by enhanced chemiluminescence (ECL, SuperSignal® West Dura Extended Duration Substrate, Thermo Fisher Scientific, Rockford, IL) [21]. In some experiments, after SW analysis, the same blot was reacted with antibody for WB. Following ECL signal quenching, the same blot was subjected to autoradiography for southwestern (SW) analysis so that the protein spot or band from WB and SW could be aligned. In addition, the same blot was stripped with SDS stripping buffer (62.5 mM Tris-HCl, pH 6.8, 2% (w/v) SDS, 100 mM beta-mercaptoethanol) at 50 $^{\circ}$ C for 20 minutes and then reprobed with additional antibodies.

2.8 AP1 EMSA-Western blotting analysis

EMSA-western blotting was performed as previously described [31] with modifications. First, the AP1 DNA-protein complex was formed by incubating nuclear extract with radiolabeled AP1 oligonucleotide using optimized EMSA conditions (section 2.2) and the complex was separated by one-well (10 cm) pre-electrophoresed 5% nondenaturing PAGE gels at 90 V for 90 min, and then the complex was electroblotted to PVDF membrane in 20% (v/v) methanol, 25 mM Tris and 192 mM glycine. The one-well blot was cut into 10 equal slices vertically, where one piece was used for autoradiography, while the other nine pieces were reacted with antibody for western blotting as previously described.

2.9 In-gel and on-blot trypsin digestion

Concentrated proteins, following EMSA and extraction, were separated by SDS-PAGE and stained with 0.25% (w/v) Coomassie Brilliant Blue. The gel band was cut into ten equal slices vertically for in-gel trypsin digestion as previously described [28] with minor modifications. In general, the gel slice was cut into 1 mm × 1 mm small cubes, destained thrice with 25 mM NH₄HCO₃/50% (v/v) acetonitrile (ACN) for 30 minutes, dehydrated with ACN for 10 minutes and dried by speed-vacuum for another 10 minutes. Gel pieces were reduced in 150 µl 10 mM DTT in 100 mM NH₄HCO₃ at 56°C for 1 hour and then alkylated with 55 mM iodoacetamide in 100 mM NH₄HCO₃ at room temperature for another 1 hour in the dark. After washing and dehydrating, gel cubes were digested in 30 µl of trypsin (20 ng/µl in 25 mM NH₄HCO₃) at 37°C overnight. Similarly, EMSA extracted proteins analyzed by EMSA-3DE-SW were digested on-blot by trypsin as described [21] with minor modifications. To avoid interference of milk proteins in later MS analysis, when performing SW the PVDF blot was blocked into 0.5% (w/v) PVP-40 at room temperature for 30 min instead of blocking in 2.5% milk at 4°C overnight. After cutting the protein spot from EMSA-3DE-SW, it was wetted by methanol, washed 10 times by water. As previously described, proteins on the blot were reduced and alkylated by 10 mM DTT and 55 mM IDA sequentially. Then the blot was immersed into 25 µl of 25 mM NH₄HCO₃, heated at 95°C for 5 minutes to denature proteins, and 25 µl of trypsin solution (10 µl of 40 ng/µl trypsin, 5 µl of 5% Tween-20 and 10 µl of ACN) was added and incubated at 37°C overnight. The trypsin digested peptides were extracted from the gel or blot thrice with 40 µl of 5% (v/v) TFA/50% (v/v) ACN at room temperature for 1 h. The extract was dried by speed-vacuum and resuspended in 20 µl of 0.1% TFA for MS analysis.

2.10 Capillary HPLC-electrospray ionization-tandem mass spectrometry identification (HPLC-ESI-MS/MS)

Peptides were analyzed using a Thermo Finnigan linear ion trap mass spectrometer (LTQ-XLS, ThermoFisher, San Jose, CA) equipped with a nano-ESI source. On-line HPLC separation of the digests was accomplished with an Eksigent NanoLC reverse phase HPLC. The column was a PicoFrit™ (New Objective; 50 µm i.d.) column packed to 8 cm with C18 adsorbent (Vydac 218MS, 5 µm, 300 Å). The capillary LC gradient was 2–98% of 0.1% formic acid/acetonitrile over 60 min at a flow rate of 300 nL/min. Using data dependent mode, the seven most intense ions in the survey scan were fragmented at 35% relative collision energy by collision-induced dissociation (CID) with an isolation window for MS/MS of 2.

2.11 Database search

Raw files acquired by HPLC-ESI-MS/MS were searched against Swiss-Prot database 49.0 (268,833 sequences, 123,649,849 residues) fused with a local Transcription Factor database (2944 sequences, 1,671,525 residues) using Mascot (version 2.1.0) software. The exact sequence of the GFP-C/EBP fusion protein was added to the local transcription factor database. The searching parameters were as followings: trypsin as enzyme, two possible missed cleavages, monoisotopic mass values; peptide tolerance, 1.0 Da; MS/MS tolerance, 0.8 Da; instrument, ESI-TRAP; oxidized methionine and cysteine carbamidomethylation as variable modification. For MS/MS sequence information, each raw file was also opened with Xcalibur software (version 2.0.5) and manually sequenced. Results were scored using the probability-based Mowse score. The score threshold to achieve $p < 0.05$ is set by the Mascot algorithm and is based on the size of the database used in the search. Protein scores were derived from ion scores on a non-probabilistic basis. We accepted the protein identification when Mascot scores were above the statistically significant threshold ($p < 0.05$) and the proteins with two unique peptides matched were considered as high confidence. In some cases, one peptide with a very high ion score was also considered.

3. Results

In this study, we developed an approach, EMSA-three-dimensional electrophoresis (EMSA-3DE), to purify DNA binding proteins from the DNA-protein complex separated by EMSA. To verify this approach, GFP-C/EBP fusion protein was used as a model. GFP-C/EBP contains 34 amino acids providing a (His)₆ tag and other sequences at the N-terminus, followed by the 239 amino acid of enhanced green fluorescence protein (GFP) fused to 99 amino acids containing the -CAAT- motif binding sequence domain of rat liver C/EBP α (aa. 283–381). The CAAT-binding motif at the C terminal has the ability to bind as a dimer the CEBP oligonucleotides, and the enhanced GFP protein sequences at the N-terminal makes it possible to determine the abundance of GFP-C/EBP by determining the fluorescence with excitation at 398 nm and emission at 512 nm. The purification recovery can be calculated by comparing the fluorescence signal after EMSA purification with input GFP-C/EBP.

3.1 Optimizing extraction of DNA-protein complexes from EMSA

To be able to apply the DNA-protein complex to subsequent purification, it is necessary to recover the protein in high yield. To discover a way to accomplish this we used the Ni²⁺-NTA purified GFP-C/EBP fusion protein to determine EMSA purification recovery. The DNA-protein complex formed by the GFP-C/EBP fusion protein with CEBP oligonucleotide was first extracted from the native PAGE gel. After the DNA-protein complex was located by autoradiography, the corresponding gel slice was cut from the native gel and homogenized, and extracted by using different extraction buffers, as shown in Table 1. Among the five extraction buffer tested, the maximum recovery for gel extraction was 17%, which is a low yield. We then attempted to electrotransfer the DNA-protein complex from the native gel to nitrocellulose (NC) or polyvinylidene difluoride (PVDF) membrane for extraction. The DNA-protein complex was found to attach to both NC and PVDF, while free radiolabeled oligonucleotide was only detected on PVDF by autoradiography. The extraction recovery of GFP-C/EBP was compared between NC and PVDF using various extraction buffers, primarily by differing in the detergents used. As shown in Table 1, using the same extraction buffer, more GFP-C/EBP protein was extracted from PVDF than from NC. In addition, the best extraction buffer for our purposes was composed of 50 mM Tris-HCl, pH 9.0, 50 mM DTT and 0.5% Tween-20. This extraction buffer was able to recover 70% of the total GFP-C/EBP from PVDF, which is a significant improvement from gel extraction and was used for all the subsequent experiments. Slightly higher recovery (75% versus 70%) was obtained by including Zwittergent with Tween-20, however to simplify subsequent detergent removal, the use of Tween-20 alone was chosen, sacrificing yield only slightly. From these results using GFP-C/EBP, we can conclude that purifying DNA-protein complex by EMSA using PVDF blotting and extraction is a feasible, high-yield method.

3.2 Development of EMSA-3DE-SW

To further test this technique, we examined its robustness by using a more complicated starting material, bacterial crude extract containing GFP-C/EBP. In this experiment, the DNA-protein complex, formed by bacterial crude extract containing GFP-C/EBP with CEBP oligonucleotide, was electroblotted to PVDF and extracted with buffer containing 0.5% Tween-20 as described above. Comparing EMSA-based purification with Ni²⁺-NTA chromatography, Figure 1A proves that both techniques can isolate GFP-C/EBP, indicated by the arrow, from bacterial crude, however, the two methods show different protein bands after silver staining. The identity of this protein was confirmed by EMSA-3DE-SW as follows. A first dimension of EMSA after extraction was combined with a second dimension of IEF and third dimensional SDS-PAGE to further separate EMSA purified proteins. After blotting, the specific CEBP oligonucleotide binding protein was then detected using a

southwestern blot assay (SW). Figure 1B shows the EMSA-3DE-SW result of GFP-C/EBP in bacterial crude extract, indicating the protein spot that binds radiolabeled CEBP oligonucleotide have a position consistent with GFP-C/EBP (MW:43 kDa, pI:8.97). The protein spot was cut from the blot for MS identification. HPLC-nanoESI-MS/MS analysis demonstrated the protein spot to be GFP-C/EBP with sequence coverage of 35% and 14 peptides matched, as shown in Figure 1C. Typical ion spectra from MS/MS of the peptide GSHMASMTGGQQMGR of GFP-CEBP, corresponding to residues 18–32, is shown in Figure 1D. Collectively, these data proves that EMSA-3DE-SW coupled with HPLC-nanoESI-MS/MS is able to purify and identify a DNA binding proteins from a bacterial extract.

3.3 Characterization of the AP1 transcription factor from HEK293 cell nuclear extract

Next, we challenged this method to purify low abundance DNA binding proteins from more complicated samples, i.e. cell nuclear extracts. In this study, we used AP1 and CEBP oligonucleotide binding proteins as examples. To do this, the specific DNA-protein complex was first determined and the maximum DNA binding affinity in nuclear extract was estimated by Scatchard analysis.

The first step for EMSA purification was to determine the specific DNA-protein complex. The specific CEBP DNA-protein complex formed in HEK293 nuclear extract has been determined in a previous study [21]. In this study, we determined the specific AP1 DNA-protein complex by an EMSA competition assay. As shown in Figure 2A, the specific complex (C) formed by radiolabeled AP1 oligonucleotide with proteins in HEK293 nuclear extract can be competed away by unlabeled AP1 DNA, but not by unlabeled NF- κ B or SP1 DNA. Unlabeled CEBP DNA diminishes the complex by approximately 75% (based on densitometry), implying the AP1 complex contains some DNA binding proteins able to bind with CEBP oligonucleotide at some affinity. In order to determine the maximum AP1 DNA binding activity in HEK293 nuclear extract, a twofold serial dilution of radiolabeled AP1 oligonucleotide was incubated with the same constant amount of HEK293 nuclear extract as previously described [19], in Fig. 2B, and a Scatchard plot is shown in Fig. 2C. Fig. 2C provides an estimate of $K_d = 1.2$ nM and the maximum binding (B_{max}) of 0.5 nM for a 5-fold dilution. Thus, undiluted nuclear extract contained 2.5 nM AP1 binding activity.

Next, the conditions for EMSA reaction were optimized as shown in Supplemental Figure 1 by including agents known to diminish non-specific DNA-binding [19], including optimization of salt concentration in the EMSA buffer, detergent to increase specific complex formation, heparin (a competitive inhibitor of DNA-binding [32]), poly dI:dC, a duplex DNA and CEBP mutant DNA (a duplex oligonucleotide with seven base pairs mutated at CEBP consensus site) which also decreases non-specific duplex DNA binding, and T₁₈, which will reduce non-specific single stranded DNA binding. The concentration of each agent was selected at the highest concentration that did not significantly interfere with specific binding, thus EMSA purification was performed under optimized conditions. Once the AP1 EMSA reaction is optimized, all subsequent AP1 EMSA purifications are performed using these conditions: EMSA buffer containing 0.1 M NaCl, 0.1% Tween-20, 2.5 ng/ μ l of heparin, 10 μ M T₁₈, 20 ng/ μ l of dI:dC, and 100 nM of CEBP (MUT), 11.6 nM CEBP DNA, and 1 mg/ml HEK293 nuclear extract. This optimization was found to improve identification of AP1 components by MS/MS (data not shown).

To explore the purification obtained from EMSA, first HEK293 nuclear extract was applied to standard 2DE, without using EMSA and stained for protein (Fig. 3A, top panel) or HEK293 nuclear extract was first purified from EMSA using oligonucleotides containing consensus binding sites for either AP1 (Fig. 3A, middle panel) or C/EBP (lower panel) and separated by standard 2DE (EMSA-3DE). The EMSA-3DE protein stained profiles show

many fewer protein spots compared with HEK293 nuclear extract, showing that prior EMSA purification removes a majority of the nuclear proteins. While AP1 and CEBP EMSA-3DE have some spots in common, the majority of the protein spots are different between the two, implying the specificity of prior EMSA purification. In other words, EMSA purification is specific for different DNA-binding TFs.

In Fig. 3B is shown the southwestern blot of the AP1-EMSA-3DE (EMSA-3DE-SW) using radiolabeled AP1 oligonucleotide. Films were overlaid and inspected visually to carefully align the various blots. The results show that at least seven DNA-binding spots were detected of which two are indicated. Three of these spots were identified by western blotting with specific antibodies in Fig. 3C. One of the three spots identified with the c-jun antibody or the similar one identified with the ATF1 antibody seem to correspond to the spot enclosed in a rectangle (Fig. 3B). The JunD antibody (Fig. 3C) detects the encircled spot in Fig. 3B. ATF1 antibody detects two spots (Fig 3C) which may represent monomeric and multimeric forms of this TF and the slower migrating (higher molecular mass) one of these can also be observed on the SW blot (Fig. 3B). Although electrophoresis in SDS would be expected to dissociate multimers, it does not always do so [28]. Thus, Western and Southwestern blotting complement one another in identifying AP1 DNA-binding proteins.

In order to purify enough AP1 oligonucleotide binding proteins for successful MS identification, the EMSA reaction was enlarged and 1.5 mg of HEK293 nuclear extract (300 μ l) was used for optimized AP1 EMSA. This amount should contain 750 fmol AP1-binding activity as determined in Fig. 2C. Considering EMSA purification yield of about 70% and protein or peptides loss during MS sample preparation, this should still be sufficient to satisfy the detection limits of LTQ-MS/MS which is approximately 50 fmol. After the AP1 complex was extracted from PVDF, the extract that contains the AP1 components was analyzed by two approaches: EMSA-2DE (EMSA purification followed by SDS-PAGE) or EMSA-3DE, which is followed by trypsin digestion and MS identification. Dealing with EMSA-2DE first, optimized AP1 EMSA extract was separated by SDS-PAGE and stained by Coomassie brilliant blue. The MS data was searched against both SWISS-PROT and a database containing only known TFs. The proteins identified after EMSA-2DE from the TF database search are shown in Table 2. Among the identified candidates, two peptides matched c-Jun and Jun-D, four peptides matched ATF1 and five peptides matched CREB. As human ATF1 and CREB proteins share 66% amino acid sequence identity and they have similar C-terminal sequence, we identified two peptides in both candidates ATF1 and CREB, i.e. TLIEELK and VAVLENQNK. In addition, c-Fos and ATF2 only have one peptide match, however, these are also statistically significant results. Two zinc finger proteins: zinc finger protein 711 and Myelin transcription factor 1 like protein, have three peptides matched.

EMSA-3DE-SW was also performed with this higher amount of starting material. The EMSA extract was separated by 2-DE and detected by southwestern blot. The EMSA-3DE-SW profile of AP1 EMSA is shown in Figure 3B. There are four major protein spots that bound radiolabeled AP1 oligonucleotide with molecular mass between 26–50 kDa. The four spots were cut out from PVDF and digested on blot with trypsin and analyzed by HPLC-nanoESI-tandem MS as described above. One of the spots, indicated with circle in Figure 3B, was identified to be Jun-D; the other three protein spots are still in the process of identification. While not a high success rate the method clearly confirms the results presented in Table 2, which also identified Jun-D. Using a higher amount of protein for the analysis would likely increase this rate of success.

After EMSA purification and MS identification, the candidate proteins identified in the AP1 EMSA complex were validated by using EMSA-western blotting, EMSA-2DE-western

blotting and EMSA antibody supershift assays. EMSA-western blotting (Fig. 4A) was used in this study to clarify the nature of multimeric complexes formed in AP1 complex by using antibodies against DNA-binding proteins [31]. For EMSA-western blotting, EMSA was performed and the non-denaturing gel was blotted for detection with antibodies. The position of the AP1-DNA complex, located by autoradiography, is shown on the right. Since many proteins have been reported to bind the TRE site in AP1 oligonucleotides, it is not surprising that the complex formed by AP1 DNA is not a sharp band, but more like a smear, as observed in Figure 4A. c-Jun, c-Fos, Jun-D, ATF1 and ATF2 are found to be associated with AP1 EMSA complex formation while SP1 and lamin A/C are not bound with AP1 DNA as expected. Both antibodies were shown in other experiments to be effective at detection under these conditions (data not shown) and so in this case, the lack of staining is indicative of the absence of these proteins. In addition, c-Jun and ATF2 form two different complexes and the complex formed by Jun-B migrates further than the major AP1 complex.

Western blotting analysis, shown in Fig. 4B (EMSA-2DE-WB), shows that c-Jun, Jun-D, ATF1 and ATF2 were involved in AP1 complex formation. This agrees with the results in Fig. 3C and Table 2. In Figure 4B, EMSA-2DE-WB also demonstrates that ATF1 is detected as a monomer and multimer, and similar results were acquired in three independent experiments, implying SDS in sample buffer cannot completely dissociate multimers during electrophoresis, as has been previously reported [28]. While SP1 is not found in AP1 EMSA-2DE in agreement with Fig. 4A, application of nuclear extract to the same SDS-PAGE gel clearly detects SP1 and acts as a positive control. Both EMSA-3DE-WB of Figure 3C and EMSA-2DE-WB of Figure 4B confirm the compositions of the AP1 complex determined by MS in Table 2.

In Figure 4C, EMSA antibody supershift was performed to determine the components of the AP1 complex. The c-Jun or ATF1 antibodies significantly blocks the formation of AP1 complex; in both, the complex is diminished by 64% based on densitometry. The Jun-D (by 18%) or c-Fos (by 19%) antibody blocks AP1 complex formation to a lesser degree compared with the complex band formed with no antibody. The unrelated SP1 antibody had no effect showing that the effect is not due to the presence of antibody alone but rather to the antibody specificity and provides a negative control. These results further verify the result that c-Jun (64%), Jun-D (18%), ATF1 (68%) and c-Fos (19%) are AP1 complex components present in HEK293 cells. Interestingly, the NF- κ B antibody also partially blocked (14%) which may require further investigation since NF- κ B is not known to bind the AP1 DNA site.

3.4 Characterization of the C/EBP transcription factor from HEK293 cell nuclear extract

Finally, we also purified the EMSA complex formed by CEBP oligonucleotides in HEK293 nuclear extract using a similar strategy as the AP1 EMSA purification. The main difference being that EMSA in this case was not systematically optimized since we had shown in a previous paper that this TF could be identified by 2DE directly from gel blots of nuclear extracts [33]. CEBP EMSA extract was separated by SDS-PAGE (EMSA-2DE) or by 2-DE (EMSA-3DE) and detected by SW and WB. As shown in Figure 5A, EMSA-2DE-WB shows that CEBP EMSA extract contains two proteins bands that reacted with C/EBP beta polyclonal antibody with molecular mass between 34 kDa. and 50 kDa., and EMSA-2DE-SW in Figure 5A shows that there are at least two protein bands between 34–50 kDa in the CEBP EMSA extract that bind radiolabeled CEBP oligonucleotide by southwestern blot analysis (Fig. 5A). Using EMSA-3DE (Fig. 5B), the encircled spot bound the C/EBP antibody at the same position (Fig. 5B, lower panel) as the CEBP oligonucleotide probe (SW, upper panel). Furthermore, the molecular mass and pI of the encircled spot closely matches what would be expected for C/EBP. Thus, the CEBP EMSA-3DE purification successfully isolated C/EBP protein from HEK293 nuclear extract.

4.0 Discussion

In this study, we developed EMSA-based purifications, EMSA-2DE or EMSA-3DE, which combines native PAGE EMSA with SDS-PAGE (EMSA-2DE) or with 2-DE (EMSA-3DE) to isolate DNA binding proteins from two starting materials, bacterial extract or HEK293 nuclear extract. These techniques were then further coupled to southwestern and western blotting techniques to provide detection and aid in identification. These methods were further interfaced to LC-MS/MS to further characterize these proteins. Using this strategy, we demonstrated the purification and identification of chimeric GFP-C/EBP fusion protein from bacterial crude extract as a model to allow refinement of the methods and then characterized two specific TFs, AP1 and C/EBP, from HEK293 nuclear extract. Characterization of AP1 complexes has proved challenging in many studies and since our new methods worked in this case and for C/EBP, the methods can probably be applied to other transcription factors.

Our study is not the first time to combine EMSA with two-dimensional proteomics. Some researchers have utilized proteomic approaches to identify DNA-binding TFs using EMSA as a starting point, however, there are only two related techniques developed up to now. One technique requires previous knowledge of the isoelectric point (pI) and molecular mass (MW) of DNA binding protein since the candidate protein is cut out from 2-DE gel at the particular pI and MW for MS identification. In this approach, EMSA is coupled with IEF or SDS-PAGE to determine the pI or MW of DNA binding proteins [16]. Another technique is also called 2D-EMSA, which separates by EMSA in the presence or absence of added DNA as the first dimension and a second dimension of SDS-PAGE. The DNA-binding proteins are positioned by comparing the 2D-EMSA profile with or without added DNA and these bands are cut from 2D-EMSA gel for MS identification [25,26]. This 2D-EMSA clearly differs from our technique in that DNA-protein complex is not extracted from the EMSA gel or PVDF as we perform in this study where EMSA is used as a first step of purification. In addition, in this study, EMSA conditions are systematically optimized which significantly reduces background nonspecific binding. In unpublished data, MS identification also confirms that EMSA purification under optimized conditions removed the abundant TFs, i.e. transcription factor CP2/LSF, transcription factor CP2-like 1protein (LBP-9), upstream binding protein 1 isoform (LBP-1 α), pre-B-cell leukemia homeobox 1 or 2 (PBX1 or PBX2), Y-box binding protein 1 (YB-1), etc., (data not shown). Previous studies in our lab have also shown that systematically optimized oligonucleotide trapping (SOT), which uses some of the same agents used here, leads to dramatically higher purity and allowed a single step of purification to yield a single predominant protein band, the C/EBP- α p30 isoform, from rat liver nuclear extract [19].

Characterizing the AP1 transcription factor components has been challenging. Bernstein et al [34,35] used biotinylated AP1 DNA-binding to purify binding proteins from 54 mg of adenocarcinoma nuclear extract. Y-box-binding protein 1 (YB-1) and nucleolin were identified by nanospray MS/MS sequencing, but the low abundant AP1 components such as c-Jun and c-Fos were not. In our method, YB-1 and nucleolin were removed at the EMSA step (data not shown). Tempst et al also used biotinylated AP1 DNA to isolate binding proteins [17]. Starting with 1.7×10^9 leukemia cells, direct capture of AP1 DNA binding proteins from nuclear extract using magnetic beads was not successful due to several abundant proteins contaminants. Using an improved procedure called positive-negative-positive selection, where AP1 binding proteins were first captured on AP1 DNA, then negatively selected with a mutant AP1 DNA, followed again by selection with wild-type AP1 DNA, they identified c-Jun and c-Fos by MALDI-TOF-TOF. We also tried adding the same AP1 mutant DNA during our EMSA purification procedure and found that a five-fold molar excess of mutant AP1 DNA significantly diminished specific AP1 complex formation

(data not shown). We instead chose CEBP mutant DNA as competitor at high concentration (100 nM) which did not interfere AP1 complex formation and also efficiently removed several contaminants found in our early studies. We successfully identified known AP1 components including c-Jun, c-Fos, Jun-D, ATF1, ATF2 and CREB from nuclear extract representing only 3×10^7 HEK293 cells.

The AP1 DNA sequence used for EMSA purification in this study contains a TPA response element –tgactca–, which is considered as a consensus site for Jun and/or Fos binding. In our study, in addition to Jun and Fos, ATF1, ATF2 and CREB are also identified in the AP1 complex, and they are verified by EMSA-2DE-western blotting, EMSA-western blotting, EMSA supershift assay or southwestern blot, indicating the different AP1 compositions in HEK293 nuclear extract. This finding is consistent with previous studies showing that AP1 compositions is dependent on cell or tissue type [10]. Moreover, in addition to these known AP1 components, there are two other TFs identified by MS, i.e. myelin transcription factor 1-like protein (MYT1L) and zinc finger protein 711. Both are zinc finger proteins and have not been reported to bind AP1 DNA site, which will require further confirmation.

It should be pointed out that although we successfully identified AP1 components in this study, there are still several abundant proteins identified in the data from the isolated AP1 complex found when searched against public database such as SWISS-PROT, NCBI or IPI, but are excluded from the transcription factor database. This points out the value of the approach used, where searches against a more generalized database (e.g., SWISS-PROT) can be refined by further searches against more specialized databases such as our transcription factor database. Most of the abundant proteins found but excluded by this approach were RNA binding proteins or abundant DNA-binding proteins involved in other processes such as replication. It is likely they would be further diminished by further refining the EMSA purification step such as further increasing the amounts of T18 and poly dI:dC used.

Southwestern blotting can also diminish the contribution of abundant protein contaminants. In this study, we coupled EMSA-3DE with southwestern blotting to characterize DNA binding proteins. Our group has successfully established two-dimensional southwestern blot (2DE-SW) to identify DNA binding proteins in nuclear extract directly from 2-DE blots [21]. To do this, crude nuclear extract was separated by 2-DE, then the proteins in the gel were transferred to PVDF membrane, renatured and probed with radiolabeled DNA. The DNA binding proteins on 2-DE blot were cut out for on-blot digestion by trypsin and identified by MS. This technique has the ability to position specific DNA binding proteins on 2-DE blots without contaminations of non-specific proteins, which frequently migrate differently in 2-DE than the located transcription factor. Using EMSA-3DE-SW, Jun-D was successfully identified and C/EBP was characterized in HEK293 nuclear extract in this study. Compared with EMSA-2DE, EMSA-3DE-SW required an additional step and may require more starting material. This EMSA-3DE-SW approach is also a promising technique for TF purification and identification.

In our EMSA purification approach, the DNA-protein complex in a native gel was electroblotted to PVDF membrane and the protein extracted. Based on a previous study by our group [39], we tried to cut the EMSA complex band from the native gel and attached it to the top of second dimensional SDS-PAGE gel. The proteins will be electrotransferred from the native gel to SDS-PAGE and separated after electrophoresis, but recovery was low (data not shown). We also tried another commonly used technique to extract proteins from PAGE gels [40,41]. Homogenization of gel slices and sonication was used to release proteins from the gel matrix into extraction buffer. Although Dr. Alves's group could recover membrane and soluble proteins in up to 80% yield, we get less than 20% recovery

for nuclear proteins. Inspired by the EMSA-western blotting technique [31], we explored electrotransfer the DNA-protein complex in a native gel to NC or PVDF membrane for extraction. The commonly used recipe for protein extraction from PVDF is using Triton/SDS elution buffer (50 mM Tris-HCl, pH 9.0, 2% SDS, 1% Triton X-100) [42]. As detergents are often incompatible with subsequent procedures, in this study we optimized protein extraction from blots using different compositions. It was found that 0.5% Tween-20 included in extraction buffer efficiently elutes proteins from PVDF and the subsequent 2-DE and SW are not influenced by Tween-20 once it is removed by the Pierce detergent spin column without affecting protein recovery (data not shown).

Scatchard analysis of AP1 EMSA estimates K_d to be 1.2 nM in HEK293 nuclear extract. A previous study determined the affinity of AP1 binding to its consensus site yielded a K_d of approximately 10 nM using FITC-labeled DNA fixed on magnetic beads [43]. This represents reasonably good agreement and the difference may be caused by different experimental conditions and different materials. The maximum AP1 DNA binding activity in five fold-diluted HEK293 nuclear extract is estimated to be 0.5 nM. In order to ensure high yield and to select for high affinity binding, 10 nM AP1 DNA was added in EMSA reaction.

The results also allow us to back calculate the concentration of AP1-binding activity in HEK293 cells. HEK293 cells (1×10^8) have a packed cell volume of approximately 1.5 ml and yield 1 ml of the nuclear extract (5mg/ml) used here. This extract would contain 2.5 pmol of AP1-binding activity based on the results presented in Fig. 2C. This corresponds to AP1 binding activity being 1.7 nM in the HEK cells. Thus, AP1-binding is truly low abundance and yet can be successfully characterized by the methods presented.

Conclusions

EMSA purification coupled with southwestern blot and proteomics technique provides an alternative to feasibly purify a wide range of TFs bound with DNA response elements in a promoter, which is useful to elucidate the regulation mechanism of a gene expression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Maria Macias for excellent technical assistance. We are thankful to Dr. Robert Nadeau for his excellent work on manuscript correction. Also thanks Dr. William Haskins, Dr. Zhiwei Wang and the NIH funded (grant number 2G12RR013646) RCM Proteomics and Protein Biomarkers Core facility. This work was supported by the NIH grant (R01GM043609 to HWJ).

Abbreviations

TF	transcription factor
EMSA	electrophoretic mobility shift assay
RE	response elements
TRE	TPA response element
CRE	cyclic AMP responsive elements
C/EBP	CCAAT enhancer binding protein

hnRNP	heterogeneous nuclear ribonucleoprotein
PVDF	polyvinylidene fluoride
NC	nitrocellulose membrane
SOT	systematically optimized oligonucleotide trapping
PT	promoter trapping
IEF	isoelectric focusing
2-DE	two-dimensional gel electrophoresis
2D-SW	two-dimensional southwestern blot
WB	western blotting
EMSA-3DE	EMSA followed by IEF and SDS-PAGE
EMSA-3DE-SW	EMSA-3DE using southwestern blot detection
HPLC-ESI-MS/MS	high performance liquid chromatography-electrospray ionization-tandem mass spectrometry
FT-ICR	Fourier transform ion cyclotron resonance mass spectrometer

References

1. Simicevic J, Deplancke B. *Mol Biosyst.* 2010; 6:462. [PubMed: 20174675]
2. Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA. *Cell.* 2001; 106:697. [PubMed: 11572776]
3. Dynlacht BD. *Nature.* 1997; 389:149. [PubMed: 9296491]
4. Bain G, Maandag EC, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M, et al. *Cell.* 1994; 79:885. [PubMed: 8001125]
5. Furney SJ, Higgins DG, Ouzounis CA, Lopez-Bigas N. *BMC Genomics.* 2006; 7:3. [PubMed: 16405732]
6. Boyadjiev SA, Jabs EW. *Clin Genet.* 2000; 57:253. [PubMed: 10845565]
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chisoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. *Nature.* 2001; 409:860. [PubMed: 11237011]
8. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. *Nucleic Acids Res.* 2008; 36:D88. [PubMed: 18073188]
9. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. *Nat Rev Genet.* 2009; 10:252. [PubMed: 19274049]
10. Eferl R, Wagner EF. *Nat Rev Cancer.* 2003; 3:859. [PubMed: 14668816]
11. Lopez-Bergami P, Lau E, Ronai Z. *Nat Rev Cancer.* 2010; 10:65. [PubMed: 20029425]
12. van Dam H, Castellazzi M. *Oncogene.* 2001; 20:2453. [PubMed: 11402340]
13. Mechta-Grigoriou F, Gerald D, Yaniv M. *Oncogene.* 2001; 20:2378. [PubMed: 11402334]

14. Ramirez-Carrozzi V, Kerppola T. *Mol Cell Biol.* 2003; 23:1737. [PubMed: 12588992]
15. Gadgil H, Jurado LA, Jarrett HW. *Anal Biochem.* 2001; 290:147. [PubMed: 11237318]
16. Woo AJ, Dods JS, Susanto E, Ulgiati D, Abraham LJ. *Mol Cell Proteomics.* 2002; 1:472. [PubMed: 12169687]
17. Yaneva M, Tempst P. *Anal Chem.* 2003; 75:6437. [PubMed: 14640712]
18. Mittler G, Butter F, Mann M. *Genome Res.* 2009; 19:284. [PubMed: 19015324]
19. Moxley RA, Jarrett HW. *J Chromatogr A.* 2005; 1070:23. [PubMed: 15861784]
20. Jiang D, Moxley RA, Jarrett HW. *J Chromatogr A.* 2006; 1133:83. [PubMed: 16934821]
21. Jiang D, Jia Y, Zhou Y, Jarrett HW. *J Proteome Res.* 2009; 8:3693. [PubMed: 19388704]
22. Matsuoka TA, Zhao L, Artner I, Jarrett HW, Friedman D, Means A, Stein R. *Mol Cell Biol.* 2003; 23:6049. [PubMed: 12917329]
23. Fried M, Crothers DM. *Nucleic Acids Res.* 1981; 9:6505. [PubMed: 6275366]
24. Garner MM, Revzin A. *Nucleic Acids Res.* 1981; 9:3047. [PubMed: 6269071]
25. Stead JA, Keen JN, McDowall KJ. *Mol Cell Proteomics.* 2006; 5:1697. [PubMed: 16845145]
26. Stead JA, McDowall KJ. *Nat Protoc.* 2007; 2:1839. [PubMed: 17703194]
27. Jiang S, Galindo MR, Jarrett HW. *Proteomics.* 2010; 10:203. [PubMed: 19899074]
28. Jiang D, Zhou Y, Moxley RA, Jarrett HW. *Biochemistry.* 2008; 47:9318. [PubMed: 18690718]
29. Gadgil H, Jarrett HW. *J Chromatogr A.* 1999; 848:131. [PubMed: 10427753]
30. Jarrett HW, Taylor WL. *J Chromatogr A.* 1998; 803:131. [PubMed: 9604329]
31. Demczuk S, Harbers M, Vennstrom B. *Proc Natl Acad Sci U S A.* 1993; 90:2574. [PubMed: 8385336]
32. Gadgil H, Jarrett HW. *J Chromatogr A.* 1999; 848:131. [PubMed: 10427753]
33. Jiang D, Jia Y, Zhou Y, Jarrett HW. *J Proteome Res.* 2009; 8:3693. [PubMed: 19388704]
34. Samuel S, Twizere JC, Bernstein LR. *Biochem J.* 2005; 388:921. [PubMed: 15702969]
35. Samuel S, Twizere JC, Beifuss KK, Bernstein LR. *Mol Carcinog.* 2008; 47:34. [PubMed: 17626252]
36. Hung JJ, Wu CY, Liao PC, Chang WC. *J Biol Chem.* 2005; 280:36283. [PubMed: 16118214]
37. Cukier CD, Hollingworth D, Martin SR, Kelly G, Diaz-Moreno I, Ramos A. *Nat Struct Mol Biol.* 2010; 17:1058. [PubMed: 20711187]
38. Ting NS, Kao PN, Chan DW, Lintott LG, Lees-Miller SP. *J Biol Chem.* 1998; 273:2136. [PubMed: 9442054]
39. Jiang S, Galindo M, Jarrett HW. *Proteomics.* 2010; 10:203. [PubMed: 19899074]
40. Retamal CA, Thiebaut P, Alves EW. *Anal Biochem.* 1999; 268:15. [PubMed: 10036156]
41. Mirza UA, Liu YH, Tang JT, Porter F, Bondoc L, Chen G, Pramanik BN, Nagabhushan TL. *J Am Soc Mass Spectrom.* 2000; 11:356. [PubMed: 10757173]
42. Ursitti JA, Mozdzanowski J, Speicher DW. *Curr Protoc Protein Sci.* 2001; Chapter 10(Unit 10):7. [PubMed: 18429098]
43. Fukami T, Hirano K. *Nucleic Acids Symp Ser.* 2000:153. [PubMed: 12903314]

Figure 1. GFP-C/EBP is isolated and identified by EMSA-3DE-SW and HPLC-MS/MS from bacterial crude extract containing GFP-C/EBP

(A) Silver stain analysis of EMSA-2DE gel of GFP-C/EBP. Bacterial crude extract (3 μ g) containing 0.18 μ g GFP-CEBP was mixed with 20 nM of unlabeled and 1.6 nM of labeled CEBP oligonucleotide in a total of 1500 μ l for EMSA, separated by three 5% one-well native PAGE gels and then transferred to PVDF. The DNA-protein complex was extracted from PVDF using extraction buffer (500 μ l) containing 50 mM Tris-HCl, pH9.0, 50 mM DTT and 0.5% Tween-20). After Tween-20 was removed and proteins in the extract were concentrated to 50 μ l, 10 μ l of the EMSA extract was separated by 12% SDS-PAGE gel, along with GFP-C/EBP purified by Ni-NTA chromatography (Ni-NTA) and bacterial crude extract containing GFP-C/EBP, and stained by silver nitrate. The arrow shows the position of the GFP-C/EBP fusion protein. Molecular mass standards are shown on the left. (B) EMSA-3DE-SW analysis of GFP-C/EBP in bacterial crude extract. Concentrated EMSA extract (30 μ l) of GFP-C/EBP bacterial crude was separated by two-dimensional electrophoresis (2-DE), transferred to PVDF, renatured and probed with 1.6 nM radiolabeled CEBP DNA for southwestern blot (SW) analysis. The protein which binds radiolabeled CEBP DNA is indicated with a circle. The molecular mass of standards are shown on the left. The pH units of IPG strips are shown at the top. (C) Database search result of GFP-C/EBP in bacteria crude extract. The protein spot circled in Figure 1B was cut out from SW blot for on-blot trypsin digestion and HPLC-nanoESI-MS/MS identification. Database search shows the protein spot to be GFP-C/EBP with 35% sequence coverage and 14

peptides matched, which are shown with bold letters; (D) An MS/MS spectra of a typical peptide GSHMASMTGGQQMGR (768.613, 2+) identified from GFP-C/EBP on EMSA-3DE-SW blot of bacterial crude extract. A series of b- and y-fragment ions is indicated in the spectra.

Figure 2. Electrophoretic Mobility Shift Assay (EMSA) determines the specific AP1 complex and allows estimation of AP1 activity in HEK293 nuclear extract

(A) EMSA cold competition assay identifies the specific AP1 complex. AP1 oligonucleotide (final 1.6 nM) labeled with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ was incubated with 5 μl of HEK293 nuclear extract (25 μg) in 25 μl EMSA buffer containing poly dI:dC (40 $\mu\text{g}/\text{ml}$) for EMSA cold competition analysis. Unlabeled specific competitor AP1 and non-specific competitors CEBP, NF- κB and SP1 DNA were added individually at 100-fold molar excess over labeled AP1. (B) AP1 EMSA with serially diluted oligonucleotide determines AP1 binding activity in HEK293 nuclear extract. 5 μl of nuclear extract (5-fold total dilution) was mixed with 3.2 nM or a two-fold serial dilution of labeled AP1 in a 25 μl EMSA reaction. The arrow shows the position of the specifically shifted band. NP, no protein; NC, no competitor; C, specific shifted complex; F, free DNA. (C) Scatchard plot allows the estimation of maximum AP1 binding activity in HEK293 nuclear extract. The density of the specifically shifted band (C), the free DNA band (F), and the total density for each lane (T) was used to calculate the concentration bound ($= [\text{AP1}] \times \text{C}/\text{T}$) and the free DNA concentration ($= [\text{AP1}] \times \text{F}/\text{T}$). The slope is equal to $-1/K_d$ and the abscissa intercept is the maximal binding (B_{max}) for this nuclear extract dilution. Scatchard analysis of the gel shift yields an estimate of $K_d=1.2$ nM and the maximum binding (B_{max}) of 0.5 nM for a 5-fold dilution provides an estimate of 2.5 nM AP1 activity in the undiluted nuclear extract.

Figure 3. EMSA-3DE analysis of AP1 and CEBP DNA binding proteins in HEK293 nuclear extract
(A) 50 μ g HEK293 nuclear extract separated by 2DE is shown for comparison (top panel). For the other panels, HEK293 nuclear extract (100 μ l, 500 μ g) was incubated with 10 nM of unlabeled as well as 1.6 nM of radiolabeled AP1 or with the same amounts of CEBP oligonucleotide. The individual 500 μ l EMSA reactions were each loaded on one-well PAGE gels (10 cm), then electrotransferred to PVDF for complex extraction. The same amount of each concentrated extract, using AP1 or CEBP EMSA, was separated by two dimensional gel electrophoresis (pH 3–10 IEF and 12% SDS-PAGE). Silver stained gels for unfractionated HEK293 nuclear extract (top), AP1 (middle) and CEBP (bottom) are shown. **(B, C)** EMSA-3DE-SW and EMSA-3DE-WB analysis of AP1 DNA binding proteins in HEK293 nuclear extract. AP1 EMSA was performed under optimized conditions and concentrated extract was separated by two 2-DE gels and transferred to two PVDF membranes individually. One blot was renatured and probed with radiolabeled AP1 oligonucleotides for southwestern blot (panel B, SW: *AP1) analysis. Two of protein spots reacted with AP1 DNA are indicated with a circle and rectangle separately. Another PVDF blot was reacted with a specific c-Jun antibody for western blot (WB: c-Jun) analysis and the same blot was stripped and probed with ATF1 antibody, then stripped and probed with Jun-D antibody (panel C, from top to bottom). One protein spot reacted with Jun-D antibody is indicated with a circle which is at same position as the protein spot probed with radiolabeled

AP1 DNA encircled in panel B. This protein spot is identified to be Jun-D by MS/MS. The molecular masses of standards are shown on the left. The pH units of IPG strips are shown at the top.

Figure 4. Identification of components in AP1 complex purified by EMSA-based purification
(A) AP1 EMSA-Western blotting analysis. AP1 EMSA using 60 μ l HEK293 nuclear extract was performed under optimized conditions and the complex was loaded on a one-well PAGE gel. After transferring to PVDF, the blot was cut vertically into 10 equal strips. One strip was directly used for autoradiography to determine the position of specific AP1 DNA-protein complex, while the other strips were blocked and then reacted with antibodies against c-Jun, c-Fos, Jun-B, Jun-D, ATF1, ATF2, SP1 and Lamin A/C for western blotting analysis. An arrow shows the position of the specific AP1 DNA-protein complex. **(B)** EMSA-2DE-Western blotting (WB) analysis of the AP1 EMSA extract. AP1 EMSA was performed under systematic optimized conditions. Concentrated EMSA extraction (EMSA) as well as HEK293 nuclear extract (NE) was further separated by 12% SDS-PAGE and analyzed by WB against c-Jun, ATF1, Jun-D, ATF2 and SP1. The position of proteins reacted with their corresponding antibodies is indicated with arrows on the right. The molecular masses of standards are shown on the left. **(C)** AP1 EMSA antibody supershift or blockade analysis. HEK293 nuclear extract (5 μ l) was incubated with antibodies (1 μ g) against SP1, NF- κ B, ATF2, ATF1, Jun-D, Jun-B, c-Fos, or c-Jun as indicated. The results for adding no antibody (NA) or no nuclear extract (NP) are also shown. The nuclear extract and antibodies were incubated at 4 $^{\circ}$ C for 2 h before the addition of 1.6 nM of radiolabeled

AP1 oligonucleotide and EMSA. The arrows show the position of specific DNA-protein complex (C) and free DNA (F).

Figure 5. Characterization of C/EBP by EMSA-3DE analysis

CEBP EMSA was performed in 500 μ l EMSA buffer containing 100 μ l of HEK293 nuclear extract (NE), 1.6 nM of labeled and 10 nM of unlabeled CEBP oligonucleotide, and the formed complex was separated on a one-well 5% non-denature PAGE gel and then transferred to PVDF membrane for autoradiography. Extracted CEBP complex was analyzed by western blot and southwestern blot. (A) EMSA-2DE-WB and EMSA-2DE-SW analysis of CEBP DNA binding proteins in HEK293 nuclear extract. CEBP EMSA extract (CEBP EMSA) as well as HEK293 nuclear extract (NE) was separated by 12% SDS-PAGE gel and detected by western blotting using C/EBP beta polyclonal antibody (C-19, Santa Cruz Biotechnology). Arrow indicates the position of C/EBP protein. CEBP EMSA extraction was also separated by 12% SDS-PAGE and transferred to PVDF for southwestern blot, arrow shows the position of proteins binding radiolabeled CEBP oligonucleotide. (B)

EMSA-3DE-SW and EMSA-3DE-WB analysis of C/EBP in HEK293 nuclear extract. CEBP EMSA extract was separated by 2DE (pH 7–10 IEF and 12% SDS-PAGE) and firstly detected by southwestern blotting using radiolabeled CEBP oligonucleotide (upper panel) and then the same blot was probed with C/EBP beta polyclonal antibody for western blotting analysis (bottom panel), the protein reacted with C/EBP antibody and CEBP oligonucleotide is indicated with a circle.

Table 1

Comparison of EMSA extraction recovery using different methods

Matrix	Extraction buffer composition	Recovery (%) (mean±SD)
Gel	0.1 M NaOAc, 50 mM DTT, 0.1% CHAPS	10.85±0.21
	0.1 M NaOAc, 50 mM DTT, 0.1% SDS	16.96±1.85
	50% ACN+0.1% TFA	2.53±0.16
	Formic acid/water/isopropanol (1:3:2)	1.85±0.09
	0.5 M (NH ₄) ₂ CO ₃ /10 mM MgCO ₃ /1 mM EDTA/0.1% SDS	9.58±1.32
NC	50% ACN+0.5% Zwittergent	4.85±0.35
	50% ACN+0.1% TFA	3.75±2.19
	40% ACN+0.1M NH ₄ HCO ₃	9.00±2.26
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% Triton	25.90±0.71
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% DOC	14.20±1.99
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% SDS	24.50±2.40
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% CHAPS	18.95±1.21
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% Tween-20	38.80±6.46
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent	30.90±5.19
	50mM Tris, pH9.0, 50 mM DTT, 0.5% CHAPS	18.65±0.92
PVDF	50mM Tris, pH9.0, 50 mM DTT, 0.5% Tween-20	70.15±3.75
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Triton X-100	24.55±1.77
	50mM Tris, pH9.0, 50 mM DTT, 0.5% SDS	60.55±2.76
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% CHAPS	40.60±16.40
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% SDS	61.15±6.29
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% Triton	35.30±4.81
	50mM Tris, pH9.0, 50 mM DTT, 0.5% Zwittergent, 0.5% Tween-20	75.35±2.76

* AP1 EMSA was performed using pure GFP-C/EBP and separated by 5% non-denature PAGE gel and transferred to PVDF or NC membrane, the DNA-protein complex in PAGE gel or on blot is extracted by different extraction buffers, and the extraction recovery % is compared using the indicator of average±SD from three independent experiments.

Table 2

HPLC-MS/MS analysis of AP1 EMSA-2DE extraction

Protein name	Accession number	Mr. (kDa)	pI	Peptides matched	Sequence coverage	Protein score/best ion score	Peptides identified
c-Jun	NP-002219.1	35675	8.90	2	4%	49/27	LQALK LEEKVK
JUN-D	NP-005345.3	35174	6.86	2	6%	90/61	SQNTELASTALLR ALEDLHK
Activating transcription factor 1	NP-005162.1	29232	8.44	4	12%	122/37	DLSEDT TDDPOLKR VAVLENQNK TLIELK
Active transcription factor CREB	NP-604391.1	36688	5.46	5	20%	185/73	EILSRPSYR VAVLENQNK TLIELK ILNDLSSDAPGVPR TAPTSTIAPGVVMASSPALPTQPAEEAAR
c-Fos	NP-005243.1	40695	4.77	1	2%	21	AGVVKMTGGR
Activating transcription factor 2	NP-001871.2	54537	7.75	1	1%	22	DCPVTAMQK
zinc finger protein 711	NP-068838.3	84812	6.42	3	4%	80/31	TEDDDDDDVK LEHMGNTPLKIGSDGSQEDAK QQNELK
Myelin transcription factor 1 like protein	NP-055840.2	132858	5.38	3	2%	75/30	MEVDTEEK NNVPTTTPR LSGEQMLTIK

* AP1 EMSA was performed under optimized conditions and the DNA-protein complex was transferred to PVDF membrane, extracted with optimized extraction buffer. The extracted proteins from blot were separated by 12% SDS-PAGE (2D-EMSA), the gel lane was cut into ten equal pieces horizontally. The proteins in gel were digested by trypsin and analyzed by HPLC-nanoESI-MS/MS. After searching local transcription factor database, the identified proteins were shown in this table.