Published in final edited form as:

Mass Spectrom Rev. 2013 September ; 32(5): 386–398. doi:10.1002/mas.21369.

Purification & Characterization of Transcription Factors

LI Nagore^{2,*}, RJ Nadeau^{2,5,6,7,*}, Q Guo^{2,5,6,7}, YLA Jadhav^{1,4,5,6,7}, HW Jarrett^{2,5,6}, and WE Haskins^{1,2,3,4,5,6,7,8,9,}**

¹Pediatric Biochemistry Laboratory, University of Texas at San Antonio, San Antonio, TX, 78249

²Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, 78249

³Departments of Biology, University of Texas at San Antonio, San Antonio, TX, 78249

⁴RCMI Proteomics, University of Texas at San Antonio, San Antonio, TX, 78249

⁵Protein Biomarkers Cores, University of Texas at San Antonio, San Antonio, TX, 78249

⁶Center for Interdisciplinary Health Research, University of Texas at San Antonio, San Antonio, TX, 78249

⁷Center for Research & Training in the Sciences, University of Texas at San Antonio, San Antonio, TX, 78249

⁸Departments of Medicine, Division of Hematology & Medical Oncology, University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

⁹Cancer Therapy & Research Center, University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

Abstract

Transcription factors (TFs) are essential for the expression of all proteins, including those involved in human health and disease. However, TFs are resistant to proteomic characterization because they are frequently masked by more abundant proteins due to the limited dynamic range of capillary liquid chromatography-tandem mass spectrometry and protein database searching. Purification methods, particularly strategies that exploit the high affinity of TFs for DNA response elements on gene promoters, can enrich TFs prior to proteomic analysis to improve dynamic range and penetrance of the TF proteome. For example, trapping of TF complexes specific for particular response elements has been achieved by recovering the element DNA-protein complex on solid supports. Additional methods for improving dynamic range include two- and three-dimensional gel electrophoresis incorporating electrophoretic mobility shift assays and Southwestern blotting for detection. Here we review methods for TF purification and characterization. We fully expect that future investigations will apply these and other methods to illuminate this important but challenging proteome.

Keywords

transcription factor; capillary liquid chromatography; tandem mass spectrometry; protein database searching; response element; electrophoresis

^{**}Correspondence: William E. Haskins, Ph.D., Dept. of Biology, The University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas 78249-0662, WEH.scholar@gmail.com.

^{*}These authors contributed equally.

Nature of the TF Proteome

The transcription factor (TF) proteome is an attractive target for future studies to understand human health and disease. The Central Dogma of Biology is that genes, protein-coding stretches of DNA, are transcribed from DNA to RNA and translated from RNA to protein. Transcription is tightly regulated by a variety of mechanisms at the DNA, RNA, and protein levels to determine TF abundance and post-translational modifications (PTMs). TFs are sequence-specific DNA-binding proteins and represent 3–6% of the human genome (Venter et al., 2001), which act as transcriptional activators and repressors. To regulate transcription, TFs bind to specific DNA sequences, typically 6–12 base pairs (bp), called the response element (RE). Typically, parts of the TF protein insert into the major groove of the DNA double helix to recognize the RE sequence, though minor groove interactions are also known. Less than 100 of the 1500 genes encoding TFs in humans have been purified and characterized (Gadgil et al., 2001). The limited dynamic range of proteomic analysis with capillary LC/MS/MS and protein database searching, combined with the masking of low abundance TFs by more abundant proteins, is a major technological barrier to our quest for knowledge of the human proteome (Rusconi et al., 2002). Consequently, much of our current knowledge of TFs has been gleaned from genetic studies in genetically engineered model organisms, such as yeast, which allow n-hybrid interaction assays and facile gene knockout experiments. Many of these studies are not possible or ethical in mammals. Moreover, it has become clear that a detailed understanding of TFs must include PTMs, such as phosphorylation, due to the combinatorial nature of these interactions with DNA, RNA and other proteins. For example, multiple phosphorylation sites have been discovered in the c-Jun subunit of the activator protein 1 (AP1) TF (Behrens et al., 2000).

Here and elsewhere (Jarrett, 2012, Jiang et al., 2009) we review recent advances in the purification and characterization of TFs, with particular emphasis on purification strategies that exploit the high affinity of TFs for DNA response elements on gene promoters to enrich for low abundance TFs prior to proteomic analysis with capillary LC/MS/MS and protein database searching. We also present an introduction to TF families.

TF Families

TFs are grouped into families based on homologous protein domains that mediate binding to DNA or other proteins. In many cases, homo- or hetero-dimerization is required to control transcriptional activation or repression. Some of the major families of TFs are described below although the discussion is not comprehensive.

The Basic Helix-Loop-Helix and Basic Leucine Zipper Families

The basic helix-loop-helix (bHLH) TF family is highly conserved from yeast to humans, and it plays a pivotal role in developmental events such as sex determination, lineage commitment, and cell differentiation. This family of proteins contains an evolutionary conserved basic region that promotes binding to the deoxyribose phosphate backbone of DNA in addition to the ability to directly interact with other bHLH TFs (Murre et al., 1989). One important group in the bHLH family contains a basic alpha helix that binds the E-Box RE sequence (CANNTG) and a hydrophobic HLH that allows these proteins to form homo- or heterodimers. Members of this family include Myoblast determination protein 1 (Davis et al., 1987), c-Myc (Bishop, 1982), and Hypoxia inducible factor 1 alpha (Wang & Semenza, 1995). The bHLH proteins can also contain a leucine zipper domain of amphipathic alpha helices that homo- or hetero-dimerize to form coiled-coils to enable the TF to interact with the major groove of DNA in a sequence-specific manner. The phrase ‘leucine zipper’ derives from the occurrence of a leucine at every seventh amino acid residue position to place leucine residues on roughly the same side of the α -helix where they interact with

another leucine zipper during dimerization by hydrophobic interactions (Vinson et al., 1989). Activator protein 1 is a member of the bHLH family, and exists as a heterodimeric complex with members of c-Jun, c-Fos, ATF, and other protein families. This family of TFs regulates gene transcription in response to cytokines, growth factors, stress signals, and infection (Hess et al., 2004).

The related basic leucine zipper (bZIP) family of TFs are the largest family in eukaryotic cells and also contain the simplest known protein-DNA recognition motif (Fujii et al., 2000). These proteins contain a basic region and a leucine zipper region but not a bHLH motif. The bZIP basic region is a highly conserved region spanning fourteen to sixteen amino acids and mediates sequence-specific DNA binding. The leucine zipper is characterized by the several leucine residues, which are necessary for dimerization by forming a chopstick like structure around the major groove of palindromic sequences (Fujii & Shimizu & Toda & Yanagida & Hakoshima, 2000). For example, C/EBP is a bZIP protein which is made of a single, long alpha-helix. One region of that helix inserts into the major groove of DNA, and nearby is a basic region which binds the deoxyribose-phosphate DNA backbone. Elsewhere this long helix has several heptad leucine repeats that allow it to dimerize. In C/EBP, these regions are in the order discussed from N- to C-terminus. This family of TFs regulates cell development, differentiation, as well as environmental stresses (Hu et al., 2002).

The Winged-Helix Family

The winged-helix family is a subfamily of the helix-turn-helix (HTH) family of TFs. This group is classified by three β -strands, three α -helices and two “wings” (Gajiwala & Burley, 2000). The wings are found at the C-terminus along with three strands forming a twisted antiparallel β -sheet whereas the N-terminus is mostly helical. The forkhead-box (FOX) members of this family are characterized by their winged-helix DNA-binding domain of 80–100 amino acid residues (Myatt & Lam, 2007). However, they differ in other domains involved in transactivation and transrepression. Although the FOX members do have some differences from the original winged-helix family, this nomenclature is still accepted in the scientific community (Kaestner et al., 2000). Members of this family of TFs modulate a variety of cellular functions such as cell cycle progression, differentiation, DNA damage repair, and programmed cell death (Accili & Arden, 2004).

The Homeobox Family

The homeobox family is another subfamily of the helix-turn-helix (HTH) family of TFs. The homeobox is a 180 bp DNA sequence that encodes a 60 amino acid residue protein domain referred to as the homeodomain. The most famous examples in this family are the well-studied Hox TFs that govern development, including development of larval segments into specific body parts in *Drosophila* (Lewis, 1978). In mammals, Hox genes have redundant functions that regulate anterior-posterior axis formation and are grouped into four clusters, based on their chromosomal location. A recent proteomic investigation identified the Paired related homeobox protein as a regulator of another TF, known as Osterix, during bone differentiation (Lu et al., 2011).

The Zinc Finger Family

Zinc finger TFs facilitate the binding of zinc by coordination to cysteine and histidine residues. These residues form the base of the ‘finger’. that will recognize a specific DNA sequence, the RE. The first zinc finger TF, protein TFIIIA, was identified in *Xenopus* oocytes (Miller et al., 1985). The zinc finger domain itself encompasses an independently folding region of the protein that utilizes one or more zinc ions to stabilize packing of the antiparallel β -sheet with the α -helix finger (Berg, 1988, Lee et al., 1989, Miller & McLachlan & Klug, 1985).

Transcriptional Complexes and Transcription Activation

For protein-encoding genes, transcription is carried out by RNA polymerase II (Pol II) and we restrict our discussion to only Pol II promoters. The promoter region of any gene contains multiple REs and bind TFs of at least two recognizable types, here called the specific and general TFs. The specific TFs bind to REs with very high affinity (typically nM) that include Specificity protein 1, AP1, CAAT enhancer binding protein (C/EBP) and many others, which have already been discussed. In current models of transcription, this binding occurs early and recruits the general transcription machinery. The general transcription complex involves many individual proteins and subunits. Also contained within these promoters are other DNA sequences that bind components within the general transcription complex. Promoters typically contain core promoter sequences, including: the ATATAA consensus sequence bound by the TATA-binding protein subunit within the general transcription complex, the TFIIB recognition element, the downstream promoter element, and the initiation start site where mRNA transcription begins. The pre-initiation complex assembly starts with binding at the TATA box. In some species, assembly also involves a novel conserved DNA element such as the GA element for *Saccharomyces cerevisiae* (Seizl et al., 2011). Current models involve binding of specific TFs to TFIID to assemble over a DNA region 40–60 bp upstream and downstream of the RNA transcriptional initiator element (Chen & Hahn, 2004, Rani et al., 2004). This assembly ultimately includes Pol II and the mediator complex properly positioned at the transcriptional start site, and it is referred to as the pre-initiation complex (PIC). The TFII complex depends upon relatively weak interactions (lower than nM affinity) among core promoter elements, the TFII protein components, the specific TFs, and Pol II. Altogether, there are at least 60 proteins and subunits involved in the formation of the PIC. The convergence of these proteins leads to hyperphosphorylation within the C-terminus of the largest subunit of pol II by protein kinases incorporated in the PIC. Subsequent partial dissociation of mediators in PIC generates what is known as an ‘open-complex’. The open complex allows rNTPs to bind RNAPII and catalyze the formation of phosphodiester bonds to link rNTPs together as linear chain, driving transcription at the transcriptional initiation (or start) site. At this point, most of the TFII complex remains at the promoter, is referred to as the ‘scaffold complex’, and recruits more Pol II, TFIIB and TFIIF to re-initiate transcription (Spangler et al., 2001). A recent review on Pol II machinery highlights the importance of this complex (Chen & Hahn, 2004).

TF Purification

There are several methods available to enrich for low abundance TFs, including TFs present at less than 10^5 copies per cell. Because all purifications rely on a specific assay for the protein of interest, the assays used must be first discussed. The most popular method to determine whether a TF binds to particular DNA sequence is the electrophoretic mobility assay (EMSA), also called a gel shift assay. This procedure uses either radio- or fluorescence-labeled DNA, incubated with pure or crude (nuclear) extract prior to resolution with native (non-denaturing), polyacrylamide gel electrophoresis (PAGE), to detect TF-DNA complexes. If a TF is present and binds the labeled DNA, then the corresponding TF-DNA complex will be detected from its slower migration on the gel (i.e., a supershift) (Fried & Crothers, 1981, Lane et al., 1992).

The labeled DNA can also be used for detection with Southwestern blotting (SWB), where an electrophoresis gel is blotted and TFs partially renature and bind to labeled DNA (Jiang et al., 2009). TFs are the only nuclear proteins that have ever been reported to bind DNA with nM or less affinity. Therefore, the concentration and sequence of DNA used in the incubation and washing steps of SWB can be tailored to the affinity of target TFs. For

example, a low concentration of a DNA oligonucleotide with the sequence CAAT specifically targets C/EBP (Jiang & Jia & Zhou & Jarrett, 2009). SWB is demonstrated in Figure 2C where an electro-blotted PVDF membrane was blocked with 5% milk to reduce non-specific binding and then probed with radio-labeled C/EBP promoter DNA. A single spot for purified C/EBP demonstrated the specificity of the method for this example. However, SWB has limitations due to the potential inability of a protein to fully renature, and other limitations related to oligomerization of TFs such as the inability to detect TFs that bind as hetero-/homo-dimers. Figure 1C was treated in the same manner as Figure 2C however it was probed with radio-labeled human telomerase reverse transcriptase (hTERT) promoter DNA and later visualized with autoradiography. Each of the spots represents different hTERT-binding TFs.

The presence of particular TFs can be determined with one- or two-dimensional Western blotting (1D-WB or 2D-WB) with many commercially available antibodies specific for TFs present in the PIC (specific TFs, general transcription machinery and Pol II subunits). Major limitations arise from the limited availability of antibodies specific for each TF of interest. The recovery of TFs with various purification strategies can be assessed using protein stains (Coomassie, silver, and others). An example of 1D-WB is shown in Figure 1E. First, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to resolve proteins on a 12% gel by M_r (gel not shown). Second, the SDS-PAGE gel was electro-blotted onto a PVDF membrane, blocked with 5% milk to reduce non-specific binding, and probed with specific antibodies. The success of the purification experiment was verified by the use of positive and negative controls loaded into different lanes of the SDS-PAGE gel. Bands indicated TF components of DNA-TF complexes that were purified with promoter trapping. Absent or dim bands observed with anti-TBP-antibodies for WCL and NE, compared to the dark bands for hTERT PT Elute and c-jun PT-Elute, provided strong evidence for TF purification with promoter trapping (described below).

DNA Affinity Chromatography

In one version of this method, duplex DNA that contains several identical REs (a concatemer) is prepared and covalently immobilized to cyanogen-bromide activated Sepharose beads (Sepharose 4B or CL-2B). TF purification is carried out in the presence of a competitor DNA in the mobile phase, such as poly dI:dC (Kadonaga, 1991), to increase the specificity of binding to the concatemer on the stationary phase. Partially purified or crude nuclear extract passes through the column while TFs bind to the stationary phase. Gradient elution of TFs is performed by disrupting DNA-TF interactions with increasing concentrations of salt (e.g., NaCl). DNA binding activity of the protein fractions is monitored with EMSA. A major obstacle of DNA-affinity chromatography is the co-purification of non-specific proteins that bind to either the TF of interest or its respective RE. One study showed that a concatemer is no more effective than a single copy of a RE (Gadgil et al., 2001). To address non-specific protein binding, the sample can be pre-cleared with a non-specific DNA column made by attaching fragmented salmon or herring sperm genomic DNA or mutated RE DNA. Pre-cleared samples have been successfully used for the purification of C/EBP, the lac repressor, and various nuclear factors (Gadgil et al., 2001, Gadgil & Taylor & Jarrett, 2001). Sequential purification steps have yielded a 1000-fold improvement in TF purity (Yang, 1998). For example, SP1 can be enriched 500–1000 fold with 90% homogeneity and 30% yield with just two successive purifications. Lastly, tandem DNA affinity chromatography, with at least two different REs, has successfully purified SP1 and CAAT-binding TFs from the same sample (Kadonaga & Tjian, 1986, Kerrigan & Kadonaga, 2001).

DNA Affinity Chromatography with Biotin-Avidin

Avidin- or streptavidin- solid supports can be employed to capture biotinylated DNA as a stationary phase to capture DNA-bound TFs. The affinity of biotin for avidin is extremely high and allows for stringent washing conditions (i.e., high salt and/or urea). A nuclear extract that contains biotinylated DNA is bound to the (strept)avidin, the sample is washed, and TFs are eluted with a high ionic strength buffer (Gadgil & Taylor & Jarrett, 2001). A major limitation of this technique is that avidin is highly charged and might non-specifically bind other proteins, decreasing TF purity (Franza et al., 1987, Gadgil & Taylor & Jarrett, 2001, Jarrett & Foster, 1995). Alternatively, Neutravidin, a de-glycosylated form of avidin with a near-neutral isoelectric point, can be selected to reduce non-specific binding problems (Chen et al., 2008).

Trapping

Trapping methods have been developed by our laboratory to allow more efficient purification of TFs at more physiologically relevant concentrations (Figure 1). The specific TFs typically bind to their respective REs with nM-pM affinity *in vivo*, and a more specific approach to purification is to mimic these concentrations *in vitro*. In DNA affinity chromatography, immobilized REs present at μ M or higher concentrations, encourage undesirable non-specific binding. In the trapping method, duplex DNA, such as an oligonucleotide RE or an entire promoter that contains a single stranded poly(GT)₅ tail is incubated at nM concentrations with nuclear extract or a crude cell lysate. The specific TF-RE complex is allowed to form in solution. Next, the TF-RE complex is 'trapped' with a poly(AC)₅ Sepharose column which anneals with the poly(GT)₅ present on the duplex DNA (Gadgil & Jarrett, 2002, Jiang et al., 2008, Moxley & Jarrett, 2005). In order to reduce non-specific binding on trapped material, modifiers that diminish non-specific binding (heparin, poly dI:dC, T18, and non-ionic detergents) are added to the mixture. Trapped TF-DNA complexes are washed with low salt, and proteins are eluted with the addition of high salt buffer that disrupts TF-DNA interactions. Conditions need to be optimized for each oligonucleotide or promoter used in the trapping method. Compared with DNA affinity chromatography, trapping yields TFs of higher purity (Gadgil & Jarrett, 2002, Jiang & Zhou & Moxley & Jarrett, 2008, Moxley & Jarrett, 2005), and in some cases only requires a one-step purification from crude nuclear extract.

Oligonucleotide Trapping—Trapping with an oligonucleotide RE will result in less non-specific binding than trapping with an entire promoter based on size limitations of the duplex region. Typically, trapping experiments are performed with oligonucleotides between 18–40 bp in length. This method has been used to successfully purify rat liver C/EBP, *Xenopus* B3 TF (Gadgil & Jarrett, 2002, Moxley & Jarrett, 2005), MafA (Matsuoka et al., 2003) and a complex that binds to a novel RE present in the c-Jun promoter (Jiang & Zhou & Moxley & Jarrett, 2008). Figure 2A shows oligonucleotide trapping where the reaction mixture consist of bacterial crude extract, radio-labeled C/EBP oligonucleotide DNA, binding buffer (1X TE, 50 mM NaCl, 10 mM HEPES, 10 mM MgCl₂, 1 mM EDTA, 50 μ M ZnSO₄, 1 mM DTT, 0.1% Tween, pH 7.5) and poly dI:dC. The use of poly dI:dC reduces the degree of non-specific binding. The reaction mixture incubated at room temperature for 30 minutes to allow assembly of DNA-TF complexes. DNA-TF complexes were revealed by their higher molecular mass (M_r) on the EMSA gel (top band). Characterization on this band was performed with a variety of methods such as liquid chromatography-tandem mass spectrometry, Southwestern Blots, and Western Blots. (data shown in Figures 2B–2D).

Promoter Trapping—Oligonucleotide trapping uses short DNA sequences (typically 18–40 bp) whereas promoter trapping uses longer DNA sequences (100–1000 bp). Because of the increased length and relative number of REs present on a promoter sequence, there exists

a greater chance for non-specific binding; thus, modifying conditions will need to be more carefully optimized. This method has been used to successfully purify the PIC and specific TFs for the c-Jun promoter (Jiang et al., 2006). The PIC and specific TFs that bind to the c-Jun promoter were successfully purified and characterized with the c-Jun minimal promoter sequence (−200 to +81). The modifying conditions involved optimization of the concentration of both promoter and nuclear extract. In our experience, the c-Jun promoter trapping technique is good starting point for promoter trapping of other TFs. An example of promoter trapping is shown in Figure 1A. A reaction mixture containing nuclear extract (cell line HEK293), human telomerase reverse transcriptase (hTERT) promoter DNA containing poly(GT)₅ tail, binding buffer (1X TE, 50 mM NaCl, 10 mM HEPES, 10 mM MgCl₂, 1 mM EDTA, 50 μM ZnSO₄, 1 mM DTT, 0.1% Tween, pH 7.5), and poly dI:dC. The mixture is incubated at room temperature for 30 minutes to allow for assembly of DNA-TF complexes. DNA-TF complexes are captured by annealing (GT)₅ tails to a (CA)₅-Sephacrose column. Unbound proteins are then washed with 20 column volumes of low salt buffer (1X binding buffer) followed by elution of DNA-TF complexes with 5 column volumes of high salt buffer (1X TE 0.5M NaCl pH 7.5).

1-DE

Sodium-dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), isoelectric focusing (IEF), or EMSA gels can be used for further purification of TFs by size, charge, and DNA-binding, respectively. Furthermore, these one-dimensional methods can be combined to provide multiple dimensions of separation.

2-DE

Two-dimensional electrophoresis (2-DE) combines isoelectric focusing (1st dimension) with SDS-PAGE (2nd dimension) to separate proteins by charge and size (O'Farrell, 1975). This technique has been utilized for the MS identification of proteins and also has been adapted for EMSA. Two-dimensional electrophoretic mobility shift assay (2D-EMSA), is a variant of this method useful for TFs or other nucleic acid-binding proteins from crude extract (Stead & McDowall, 2007). In this method, partially purified or crude extract was separated with 2-DE. The isoelectric point (pI) and the relative molecular mass (M_r) for a protein can be estimated. Examples of 2DE-gels are shown in Figures 1B and 2B. Both samples were first resolved by their isoelectric point (pI). Isoelectric focusing was performed with a 7cm, pH 3–10 IPG strip. The second dimension resolved the TFs on a 12% SDS-PAGE by separating the proteins by their relative molecular mass (M_r). Corresponding silver-stained spots on the 2-DE can be excised, electro-eluted, re-natured, and assayed with EMSA. An alternative to 2D-EMSA is to replace the IEF dimension of 2-DE with EMSA. The TF-RE complex can be resolved from free RE due to its slower electrophoretic mobility. SDS-PAGE is used as the second dimension and TFs that were bound to the RE are observed as spots on the gel. These spots are excised, digested and subjected to MS analysis for protein identification. For example, the *Streptomyces coelicolor* TF AtrA was successfully purified and characterized from crude bacterial extract with this method (Stead et al., 2006, Stead & McDowall, 2007). This technique, unlike SWB, has the advantage that renaturation after spot excision is unnecessary.

EMSA-3-DE

The successful development of 2D-EMSA has allowed our laboratory to develop another technique, three-dimensional EMSA (EMSA-3-DE) (Figure 2). EMSA-3-DE combines non-denaturing EMSA with traditional 2-DE to purify TFs. The EMSA gel is electroblotted to PVDF, and the specific TF-RE complex band is eluted and applied to isoelectric focusing and SDS-PAGE for additional 2-DE separation. The advantage of EMSA-3-DE is that it combines the specificity of EMSA with the resolution of 2-DE and allows for purification of

TFs from a crude extract. This method was used to successfully isolate GFP-C/EBP fusion protein constructed to bind the C/EBP canonical sequence CAAT from a crude bacterial extract and C/EBP from a nuclear extract of HEK293 cultured cells (Jiang et al., 2011).

Characterization

Methods for TF identification and quantification with capillary LC/MS/MS and protein database searching are described below.

Identification

TF identification is typically performed, following TF purification, with femtomole (10^{-15} mol = fmol) amounts of low-abundance TF peptides injected on-column in microliter (μ L) volumes at nanomolar (10^{-9} M = fmol/ μ L) concentrations. However, low-abundance protein identification has been achieved with attomole (10^{-18} mol = amol) amounts of peptide injected on-column at picomolar (10^{-12} M = amol/ μ L) concentrations (Haskins et al., 2001, Thakur et al., 2011). Careful attention to each stage of identification, from sample preparation to validation, is essential to achieve the low mass- and concentration-limits of detection needed to identify TFs. The major steps involved in TF identification are described below and illustrated by the examples shown in Figure 1D.

Sample Preparation—*In vitro* proteolysis is performed in gel-, in-solution or on-blot to generate protease-specific peptides amenable to capillary LC/MS/MS and protein database searching, such as tryptic peptides with C-terminal lysine or arginine residues. For TFs, high recovery of their corresponding low-abundance TF peptides, particularly large and/or hydrophobic peptides that adsorb to surfaces such as microcentrifuge tubes and autosampler vials, is critical for success. Recovery can be maximized by being mindful of sample-handling surfaces, sample-storage buffers, temperature, and additives to minimize adsorptive losses. Just as critical is the removal of more-abundant peptides that originate from more-abundant nuclear proteins such as heterogeneous nuclear ribonuclear proteins, a problem encountered even when TFs are purified with affinity tagging (Grosveld et al., 2005, Sebastiaan Winkler et al., 2002) or antibody-based methods such as immunoprecipitation, where masking of TFs by tags or antibodies can also be problematic. Thus, it is important to maximize both the absolute and relative amount of TF peptides recovered during sample preparation. Pooling peptides from multiple different sample preparation strategies, including various proteases and/or chemical techniques such as cyanogen bromide cleavage, is an important strategy for low abundance protein identifications (Labugger et al., 2003).

Capillary LC—The capacity of capillary LC columns is typically on the order of picomole (10^{-12} mol = pmol) amounts of peptides injected on-column. Because the majority of these peptides are high-abundance peptides from masking proteins, a dynamic range of 3–6 orders of magnitude is required to identify fmol-amol (10^{-15} – 10^{-18}) amounts of low-abundance peptides from TFs. Miniaturization of columns, stationary phase particles, and integrated nanoelectrospray emitters for capillary LC/MS/MS at low flow rates (~10–400 nL/min) improves dynamic range through improved separation of low-abundance peptides from high-abundance peptides and other molecules (e.g., salts, detergents, metabolites, etc.) that suppress ionization. Improved separation and ionization efficiency also provide more tandem mass spectrometer time for gas-phase fragmentation of closely eluting peptides.

MS/MS—Comprehensive gas-phase fragmentation of multiply-charged peptide precursor ions into complementary N- and C-terminal sequence-specific product ions, cleavage of bonds between every amino acid residue along the peptide backbone, enables unambiguous

assignment of amino acid sequences from MS/MS spectra (Novak et al., 2005). Various fragmentation mechanisms, once considered uncommon, are now standard practice in the modern mass spectrometry laboratory, including: collision-induced dissociation (CID), electron transfer dissociation (ETD) and high-energy C-trap dissociation (HCD). The advantages and disadvantages of these mechanisms are beyond the scope of this review. However, ETD and other non-ergodic mechanisms should be seriously considered if labile PTMs, such as phosphorylation, are of interest. Pooling sequence-specific product ions from multiple different fragmentation mechanisms is a valuable means to bolster one's confidence in low-abundance protein identifications (Wu et al., 2007). Figure 1D depicts a spot excised from the 2-DE gel shown in Figure 1B. The gel plug was further processed by in-gel digestion with trypsin. Capillary LC/MS/MS was performed by injecting 1 μ g of the extracted peptides onto a 50 μ m-i.d. column packed with 7 cm of 3 μ m-o.d. C18 particles and an integrated nanoelectrospray emitter with a flow rate of 350 nL/min. Gas-phase fragmentation of the most abundant peptides were performed with a hybrid linear ion trap-Fourier-transform tandem mass spectrometer via high-energy C-trap dissociation (HCD) to generate MS/MS spectra. Examples of the spectra that can be acquired through this technique are shown in Figures 1D and 2D.

Protein Database Searching—Several transcription factor protein databases, relatively small compared to whole proteome databases, enable rapid, automated assignment of TF sequences from MS/MS spectra by protein database searching algorithms (Fogel et al., 2005, Matys et al., 2006, Messina et al., 2004, Murali et al., 2011, Pfreundt et al., 2010, Schaefer et al., 2011, Wilson et al., 2008, Zhang et al., 2011, Zhang et al., 2011). Pooling results from multiple different databases, including our own in-house human TF database at <http://proteomics.utsa.edu>, and/or multiple database search algorithms is a particularly valuable means to maximize true-positive and minimize false-positive protein identifications (Kwon et al., 2011, Searle, 2010). Protein database searching was performed to identify putative TFs that bind the hTERT promoter in Figure 1D, albeit much about the hTERT promoter remains unknown. For example, the low abundance protein high-mobility group protein B1 (HMGB1), an important transcriptional activator and repressor (Agresti et al., 2003, Grasser et al., 2007, Mouri et al., 2008, Naghavi et al., 2003, Riuzzi et al., 2012, Ueda et al., 2004) was identified (IPI00419258). HMGB1 was the 24th-ranked protein family with a Mascot protein score of 99 and sequence coverage of 37%, while the most abundant, 1st-ranked protein, heterogeneous nuclear ribonucleoprotein A2/B1 (HNRNPA2B1), was identified with a Mascot protein score of 2030 and sequence coverage of 70%. The amino acid sequence assignment of product ions for the top-ranked tryptic peptide from HMGB1, IKGEHPGLSIGDVAK, spanning amino acid residues 113–127, yielded an expectation value of 0.00118 (i.e., probability of a random match in the IPI human protein database). A representative MS/MS spectrum for this peptide is shown in Figure 1D, where b- and y-type ion annotations designate charge retention on N- or C-terminal sequence-specific product ions, respectively. Validation was performed by accurate mass measurement. The observed monoisotopic mass of the unmodified, doubly-charged precursor ion for IKGEHPGLSIGDVAK at 760.9258 m/z was 1519.8371 Da, and the theoretical value of 1519.8358 Da was in excellent agreement, with less than 1 ppm mass error. Likewise, product ions were assigned with less than a 1 ppm (rms) mass error from predicted values. Confidence in ion assignments was also reflected by 36/253 sequence-specific product ions and assignment of the same amino acid sequence to other MS/MS spectra from the triply- and quadruply-charged precursor ion. As another example, protein database searching also identified transcription factor CP2 (TFCP2) (IPI00037599) (Cunningham et al., 1995, Lim et al., 1992, Ramamurthy et al., 2001). The amino acid sequence assignment of product ions for the top-ranked tryptic peptide from TFCP2, LHDETLTYLNQGQSYEIR, spanning

amino acid residues 81–98, yielded an expectation value of 0.0000034. A representative MS/MS spectrum for this peptide is also shown.

Validation—Validation of low-abundance TF identifications by accurate mass measurements is an important means to improve dynamic range because it reduces the probability of false-positive assignment of amino acid sequences from MS/MS spectra with protein database searching (Lu et al., 2008). *De novo* sequencing, synthetic peptides, and MS/MS/MS are also strong tools for validation. For example, if the 1st-ranked *de novo*-derived amino acid sequence matches the 1st-ranked database-derived amino acid sequence, then it is highly unlikely that the 1st-ranked protein identification is a false-positive (Bern et al., 2007, Johnson & Taylor, 2000, Lu & Chen, 2003, Ma et al., 2003, Sheng et al., 2000, Standing, 2003, Zhang & McElvain, 2000). Alternatively, the peptide can be synthesized to compare MS/MS and MS/MS/MS spectra, known as spectral library searching (Zhang et al., 2011). A comparison of several of these methods to validate low-abundance protein identifications has been shown previously (Haskins et al., 2001). Finally, combining different sequence-specific proteases for *in vitro* proteolysis and/or combining different gas-phase fragmentation methods, as described above, are also important means to validate TF identifications.

Quantification

Determination of the relative and absolute abundance of proteins has received unprecedented attention in recent years with the advent of a plethora of quantitative gel-free proteomic techniques, including label-free, isotopic labeling, and isobaric labeling. For an example of isotopic labeling, the absolute quantification of c-Fos was determined to be only 6000 copies per HeLa cell with stable isotopic labeling of amino acids in culture (SILAC) (Zeiler et al., 2011). For an example of isobaric labeling, the relative phosphorylation of STAT3 was determined in a cancer stem cell model exposed to a kinase inhibitor by a workflow that incorporates phospho-protein/-peptide enrichment and tandem mass tagging (TMT) (Nilsson et al., 2010). In another example of isobaric labeling, quantification of SP1, c-MYC, and p53 in tissue specimens from cancer patients was achieved with isobaric tagging for relative and absolute quantification (iTRAQ) (Cai et al., 2012). For an example of label-free quantification method, FOXO3A interactions with 14-3-3 proteins were quantified from peptide intensity profiles in sequentially diluted samples (Rinner et al., 2007). It is beyond the scope of this review to speculate as to which technique has the largest dynamic range for TF characterization at this time. Based on our experience with several of these methods, TF identification remains more sensitive than TF quantification unless specific TFs are targeted for analysis by selected reaction monitoring. Despite the improved sensitivity of targeted analyses (Stergachis et al., 2011) and with the exception of subunit stoichiometry within a well-characterized TF complex, it is usually of interest to discover previously unknown binding partners in TF complexes. Therefore, targeted quantification is expected to be uncommon in future studies, whereas unbiased quantification is expected to become of more interest.

Promising New Strategies

Sample preparation, instrumentation and computer algorithms for capillary LC/MS/MS and protein database searching continue to evolve at a phenomenal rate. Microwave-assisted digestion strategies for sample preparation are particularly appealing for TF analysis. *In vitro* proteolysis in the microwave requires seconds instead of hours and limits time-sensitive adsorptive losses of low-abundance TF peptides (Lill et al., 2007, Sandoval et al., 2007). Improvements in ion transmission to, and the dynamic range of, hybrid tandem mass spectrometers (Hossain et al., 2011, Saba et al., 2009), now ubiquitous, are also very promising. Likewise, improvements in computer algorithms for data analysis, particularly

algorithms for validation of protein database searching results, as described above, are expected to enable more unexpected isoform measurements, including alternative splicing variants and PTMs, from characteristic mass shifts at specific amino acid residue(s). The dramatic evolution of top-down proteomics for capillary LC/MS/MS and protein database searching of low-abundance PTMs of histones (Kelleher, 2004, Parks et al., 2007) and other small-to-medium-sized proteins suggests that top-down proteomics might also be viable for TF characterization. In top-down proteomics, gas-phase fragmentation of intact proteins is performed without *in vitro* proteolysis, eliminating the ambiguity inherent when inferring protein information from peptide information. Finally, stable isotope standards and capture with anti-peptide antibodies (SISCAPA) is an attractive alternative to enrich for low-abundance TF peptides after *in vitro* proteolysis (Anderson et al., 2009).

Competing Technologies

Several platforms that incorporate antibodies conjugated to micrometer- or nanometer-sized beads have been developed for high-throughput, multiplexed analysis of low-abundance proteins, including TFs. For example, kits are available to simultaneously quantify ~40 TFs in a 96-well plate format, where the identity of each TF is deduced from the emission wavelength of dye(s) encapsulated within a luminescent bead conjugated to a corresponding anti-TF antibody. In nanoparticle-based immunoPCR, and variations thereof, the identity of each TF is deduced from the oligonucleotide sequence bound to a corresponding anti-TF antibody-nanoparticle conjugate (Malou & Raoult, 2011, Nam et al., 2003). The major advantage of these competing technologies and others (e.g., reverse-phase protein arrays) is improved dynamic range and sample throughput relative to capillary LC/MS/MS and protein database searching. However, just as the case for immunoblotting, the major disadvantage is that non-cross-reacting antibodies must be developed for each TF of interest, including isoforms that arise from alternative splicing and PTMs. 'Catch and release' technologies that can leverage the incredible improvements in massively parallel DNA microarray and high-throughput DNA sequencing, seek to quantify TFs by measuring the abundance of TF-bound DNA probes. However, akin to antibody technologies, these suffer from unknown specificity and cross-reactivity, and PTMs cannot be investigated.

Conclusions

Notable Achievements

Tremendous progress towards the illumination of the TF proteome has been made in recent years. For example, with 2DE-SWB and on-blot digestion, C/EBP was identified directly from an unfractionated nuclear extract (Jiang, 2009), whereas AP1 subunits and C/EBP were both identified directly from a nuclear extract when combined with EMSA-3-DE (Jiang, 2011). AP1 is a complex TF composed of homo- and heterodimers of c-Jun, c-Fos, ATF family members, and others. Previously, c-Jun had not been identified with proteomic analysis with the exception of one recent paper where biotin-avidin chromatography proved adequate for identification with MALDI-TOF-TOF-MS/MS and protein database searching (Yaneva & Tempst, 2003). The TF AP1 components are some of the least-abundant TFs and successful characterization has proved especially challenging. Our recent EMSA-3-DE identification of c-Jun required not only sophisticated electrophoretic methods for purification, but also state-of-the-art tools for protein identification with capillary LC/MS/MS and protein database searching.

Future Progress

Synergy among the TF purification and characterization methods described herein has partially overcome the dynamic range limitations of proteomic analysis. However, further

improvements in dynamic range are needed to enable widespread investigation of the TF proteome, particularly for scarce and heterogeneous clinical specimens. Innovative methods must be harnessed in order to reveal the subtle but important changes in TFs involved in human health and disease. Given the importance of transcriptional regulation to the overall central dogma of Biology and the progress on TF proteomics thus far, TFs are fully expected to serve as prognostic/predictive biomarkers and novel therapeutic targets in future studies.

Acknowledgments

We thank the RCMI program at UTSA (NIH 5G12RR013646-12) and NIH grant 1R01GM043609 (to H. W. Jarrett) for support.

List of Abbreviations

2-DE	two-dimensional electrophoresis
AP1	activator protein 1
bp	base pair
bHLH	basic helix-loop-helix
bZIP	basic leucine zipper
C/EBP	CAAT enhancer binding protein
CID	collision-induced dissociation
EMSA	electrophoretic mobility shift assay
ETD	electron-transfer dissociation
FOX	forkhead-box
GFP	green fluorescent protein
HCD	high-energy C-trap dissociation
HEK293	human embryonic kidney cell line 293
HMGB1	high-mobility group protein B1
HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1
hTERT	human telomerase reverse-transcriptase
HTH	helix-turn-helix
IEF	isoelectric focusing
ITRAQ	isobaric tags for relative and absolute quantification
LC/MS/MS	liquid chromatography-tandem mass spectrometry
Mr	relative molecular mass
MS	mass spectrometry
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PolII	RNA polymerase II
PIC	pre-initiation complex
PTM	post-translational modification

RE	response element
rNTP	ribonucleotidetriphosphate
SDS-PAGE	sodium-dodecyl sulfate-polyacrylamide gel electrophoresis
SILAC	stable isotope labeling of amino acids in culture
SWB	southwestern blotting
TF	transcription factor
TFCP2	transcription factor CP2
TMT	tandem mass tagging
WB	western blot

Literature Cited

- Accili D, Arden KC. FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell*. 2004; 117:421–426. [PubMed: 15137936]
- Agresti A, Lupo R, Bianchi ME, Muller S. HMGB1 interacts differentially with members of the Rel family of transcription factors. *Biochemical and biophysical research communications*. 2003; 302:421–426. [PubMed: 12604365]
- Anderson NL, Jackson A, Smith D, Hardie D, Borchers C, Pearson TW. SISCAPA peptide enrichment on magnetic beads using an in-line bead trap device. *Molecular & cellular proteomics : MCP*. 2009; 8:995–1005. [PubMed: 19196707]
- Behrens A, Jochum W, Sibilio M, Wagner EF. Oncogenic transformation by ras and fos is mediated by c-Jun N-terminal phosphorylation. *Oncogene*. 2000; 19:2657–2663. [PubMed: 10851065]
- Berg JM. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 1988; 85:99–102. [PubMed: 3124104]
- Bern M, Cai Y, Goldberg D. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem*. 2007; 79:1393–1400. [PubMed: 17243770]
- Bishop JM. Retroviruses and cancer genes. *Advances in cancer research*. 1982; 37:1–32. [PubMed: 6763842]
- Cai GM, Huang DH, Dai YZ, Liu Y, Pi LM, Tan HL, Liu LJ, Lv YX, Yu CY, Tan PQ, Tian YQ, Chen GZ, Zhang X. Analysis of transcriptional factors and regulation networks in laryngeal squamous cell carcinoma patients with lymph node metastasis. *Journal of proteome research*. 2012; 11:1100–1107. [PubMed: 22070577]
- Chen AK, Behlke MA, Tsourkas A. Efficient cytosolic delivery of molecular beacon conjugates and flow cytometric analysis of target RNA. *Nucleic acids research*. 2008; 36:e69. [PubMed: 18503086]
- Chen HT, Hahn S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell*. 2004; 119:169–180. [PubMed: 15479635]
- Cunningham JM, Vanin EF, Tran N, Valentine M, Jane SM. The human transcription factor CP2 (TFCP2), a component of the human gamma-globin stage selector protein, maps to chromosome region 12q13 and is within 250 kb of the NF-E2 gene. *Genomics*. 1995; 30:398–399. [PubMed: 8586452]
- Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*. 1987; 51:987–1000. [PubMed: 3690668]
- Fogel GB, Weekes DG, Varga G, Dow ER, Craven AM, Harlow HB, Su EW, Onyia JE, Su C. A statistical analysis of the TRANSFAC database. *Bio Systems*. 2005; 81:137–154.

- Franza BR Jr, Josephs SF, Gilman MZ, Ryan W, Clarkson B. Characterization of cellular proteins recognizing the HIV enhancer using a microscale DNA-affinity precipitation assay. *Nature*. 1987; 330:391–395. [PubMed: 2825023]
- Fried M, Crothers DM. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic acids research*. 1981; 9:6505–6525. [PubMed: 6275366]
- Fujii Y, Shimizu T, Toda T, Yanagida M, Hakoshima T. Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nature structural biology*. 2000; 7:889–893.
- Gadgil H, Jarrett HW. Oligonucleotide trapping method for purification of transcription factors. *J Chromatogr A*. 2002; 966:99–110. [PubMed: 12214709]
- Gadgil H, Jurado LA, Jarrett HW. DNA affinity chromatography of transcription factors. *Anal Biochem*. 2001; 290:147–178. [PubMed: 11237318]
- Gadgil H, Taylor WL, Jarrett HW. Bi-column method for purification of transcription factors. *J Chromatogr A*. 2001; 905:133–139. [PubMed: 11206779]
- Gadgil H, Taylor WL, Jarrett HW. Comparative studies on discrete and concatemeric DNA-sepharose columns for purification of transcription factors. *J Chromatogr A*. 2001; 917:43–53. [PubMed: 11403491]
- Gajiwala KS, Burley SK. Winged helix proteins. *Current opinion in structural biology*. 2000; 10:110–116. [PubMed: 10679470]
- Grasser M, Christensen JM, Peterhansel C, Grasser KD. Basic and acidic regions flanking the HMG-box domain of maize HMGB1 and HMGB5 modulate the stimulatory effect on the DNA binding of transcription factor Dof2. *Biochemistry*. 2007; 46:6375–6382. [PubMed: 17487980]
- Grosveld F, Rodriguez P, Meier N, Krpic S, Pourfarzad F, Papadopoulos P, Kolodziej K, Patrinos GP, Hostert A, Strouboulis J. Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry. *Annals of the New York Academy of Sciences*. 2005; 1054:55–67. [PubMed: 16339652]
- Haskins WE, Wang Z, Watson CJ, Rostand RR, Witowski SR, Powell DH, Kennedy RT. Capillary LC-MS2 at the attomole level for monitoring and discovering endogenous peptides in microdialysis samples collected in vivo. *Anal Chem*. 2001; 73:5005–5014. [PubMed: 11721892]
- Haskins WE, Wang Z, Watson CJ, Rostand RR, Witowski SR, Powell DH, Kennedy RT. Capillary LC-MS2 at the attomole level for monitoring and discovering endogenous peptides in microdialysis samples collected in vivo. *Analytical chemistry*. 2001; 73:5005–5014. [PubMed: 11721892]
- Hess J, Angel P, Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. *Journal of cell science*. 2004; 117:5965–5973. [PubMed: 15564374]
- Hossain M, Kaleta DT, Robinson EW, Liu T, Zhao R, Page JS, Kelly RT, Moore RJ, Tang K, Camp DG 2nd, Qian WJ, Smith RD. Enhanced sensitivity for selected reaction monitoring mass spectrometry-based targeted proteomics using a dual stage electrodynamical ion funnel interface. *Molecular & cellular proteomics : MCP*. 2011; 10 M000062-MCP000201.
- Hu CD, Chinenov Y, Kerppola TK. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular cell*. 2002; 9:789–798. [PubMed: 11983170]
- Jarrett HW. Proteomic methodologies to study transcription factor function. *Methods in molecular biology*. 2012; 786:315–334. [PubMed: 21938635]
- Jarrett HW, Foster JL. Alternate binding of actin and calmodulin to multiple sites on dystrophin. *The Journal of biological chemistry*. 1995; 270:5578–5586. [PubMed: 7890677]
- Jiang D, Jarrett HW, Haskins WE. Methods for proteomic analysis of transcription factors. *J Chromatogr A*. 2009; 1216:6881–6889. [PubMed: 19726046]
- Jiang D, Jia Y, Jarrett HW. 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*. 2011; 1218:7003–7015. [PubMed: 21880322]
- Jiang D, Jia Y, Zhou Y, Jarrett HW. Two-dimensional southwestern blotting and characterization of transcription factors on-blot. *Journal of proteome research*. 2009; 8:3693–3701. [PubMed: 19388704]

- Jiang D, Moxley RA, Jarrett HW. Promoter trapping of c-jun promoter-binding transcription factors. *J Chromatogr A*. 2006; 1133:83–94. [PubMed: 16934821]
- Jiang D, Zhou Y, Moxley RA, Jarrett HW. Purification and identification of positive regulators binding to a novel element in the c-Jun promoter. *Biochemistry*. 2008; 47:9318–9334. [PubMed: 18690718]
- Johnson RS, Taylor JA. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Methods Mol Biol*. 2000; 146:41–61. [PubMed: 10948496]
- Kadonaga JT. Purification of sequence-specific binding proteins by DNA affinity chromatography. *Methods in enzymology*. 1991; 208:10–23. [PubMed: 1779831]
- Kadonaga JT, Tjian R. Affinity purification of sequence-specific DNA binding proteins. *Proc Natl Acad Sci U S A*. 1986; 83:5889–5893. [PubMed: 3461465]
- Kaestner KH, Knochel W, Martinez DE. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes & development*. 2000; 14:142–146. [PubMed: 10702024]
- Kelleher NL. Top-down proteomics. *Analytical chemistry*. 2004; 76:197A–203A. [PubMed: 14697051]
- Kerrigan LA, Kadonaga JT. Purification of sequence-specific DNA-binding proteins by affinity chromatography. *Curr Protoc Protein Sci*. 2001 Chapter 9:Unit 9 6.
- Kwon T, Choi H, Vogel C, Nesvizhskii AI, Marcotte EM. MSblender: A probabilistic approach for integrating peptide identifications from multiple database search engines. *Journal of proteome research*. 2011; 10:2949–2958. [PubMed: 21488652]
- Labugger R, Simpson JA, Quick M, Brown HA, Collier CE, Neverova I, Van Eyk JE. Strategy for analysis of cardiac troponins in biological samples with a combination of affinity chromatography and mass spectrometry. *Clinical chemistry*. 2003; 49:873–879. [PubMed: 12765982]
- Lane D, Prentki P, Chandler M. Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiological reviews*. 1992; 56:509–528. [PubMed: 1480106]
- Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*. 1989; 245:635–637. [PubMed: 2503871]
- Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*. 1978; 276:565–570. [PubMed: 103000]
- Lill JR, Ingle ES, Liu PS, Pham V, Sandoval WN. Microwave-assisted proteomics. *Mass spectrometry reviews*. 2007; 26:657–671. [PubMed: 17474122]
- Lim LC, Swendeman SL, Sheffery M. Molecular cloning of the alpha-globin transcription factor CP2. *Molecular and cellular biology*. 1992; 12:828–835. [PubMed: 1732747]
- Lu B, Chen T. A suboptimal algorithm for de novo peptide sequencing via tandem mass spectrometry. *J Comput Biol*. 2003; 10:1–12. [PubMed: 12676047]
- Lu B, Motoyama A, Ruse C, Venable J, Yates JR 3rd. Improving protein identification sensitivity by combining MS and MS/MS information for shotgun proteomics using LTQ-Orbitrap high mass accuracy data. *Analytical chemistry*. 2008; 80:2018–2025. [PubMed: 18275164]
- Lu X, Beck GR Jr, Gilbert LC, Camalier CE, Bateman NW, Hood BL, Conrads TP, Kern MJ, You S, Chen H, Nanes MS. Identification of the homeobox protein Prx1 (MHox, Prrx-1) as a regulator of osterix expression and mediator of tumor necrosis factor alpha action in osteoblast differentiation. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*. 2011; 26:209–219. [PubMed: 20683885]
- Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2003; 17:2337–2342. [PubMed: 14558135]
- Malou N, Raoult D. Immuno-PCR: a promising ultrasensitive diagnostic method to detect antigens and antibodies. *Trends in microbiology*. 2011; 19:295–302. [PubMed: 21478019]
- Matsuoka TA, Zhao L, Artner I, Jarrett HW, Friedman D, Means A, Stein R. Members of the large Maf transcription family regulate insulin gene transcription in islet beta cells. *Mol Cell Biol*. 2003; 23:6049–6062. [PubMed: 12917329]
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E.

- TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic acids research*. 2006; 34:D108–D110. [PubMed: 16381825]
- Messina DN, Glasscock J, Gish W, Lovett M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome research*. 2004; 14:2041–2047. [PubMed: 15489324]
- Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *The EMBO journal*. 1985; 4:1609–1614. [PubMed: 4040853]
- Mouri F, Tsukada J, Mizobe T, Higashi T, Yoshida Y, Minami Y, Izumi H, Kominato Y, Kohno K, Tanaka Y. Intracellular HMGB1 transactivates the human IL1B gene promoter through association with an Ets transcription factor PU.1. *European journal of haematology*. 2008; 80:10–19. [PubMed: 18173740]
- Moxley RA, Jarrett HW. Oligonucleotide trapping method for transcription factor purification systematic optimization using electrophoretic mobility shift assay. *J Chromatogr A*. 2005; 1070:23–34. [PubMed: 15861784]
- Murali T, Pacifico S, Yu J, Guest S, Roberts GG 3rd, Finley RL Jr. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic acids research*. 2011; 39:D736–D743. [PubMed: 21036869]
- Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell*. 1989; 56:777–783. [PubMed: 2493990]
- Myatt SS, Lam EW. The emerging roles of forkhead box (Fox) proteins in cancer. *Nature reviews. Cancer*. 2007; 7:847–859. [PubMed: 17943136]
- Naghavi MH, Nowak P, Andersson J, Sonnerborg A, Yang H, Tracey KJ, Vahlne A. Intracellular high mobility group B1 protein (HMGB1) represses HIV-1 LTR-directed transcription in a promoter- and cell-specific manner. *Virology*. 2003; 314:179–189. [PubMed: 14517071]
- Nam JM, Thaxton CS, Mirkin CA. Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science*. 2003; 301:1884–1886. [PubMed: 14512622]
- Nilsson CL, Dillon R, Devakumar A, Shi SD, Greig M, Rogers JC, Krastins B, Rosenblatt M, Kilmer G, Major M, Kaboord BJ, Sarracino D, Rezai T, Prakash A, Lopez M, Ji Y, Priebe W, Lang FF, Colman H, Conrad CA. Quantitative phosphoproteomic analysis of the STAT3/IL-6/HIF1alpha signaling network: an initial study in GSC11 glioblastoma stem cells. *Journal of proteome research*. 2010; 9:430–443. [PubMed: 19899826]
- Novak P, Haskins WE, Ayson MJ, Jacobsen RB, Schoeniger JS, Leavell MD, Young MM, Kruppa GH. Unambiguous assignment of intramolecular chemical cross-links in modified mammalian membrane proteins by Fourier transform-tandem mass spectrometry. *Analytical chemistry*. 2005; 77:5101–5106. [PubMed: 16097745]
- O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *The Journal of biological chemistry*. 1975; 250:4007–4021. [PubMed: 236308]
- Parks BA, Jiang L, Thomas PM, Wenger CD, Roth MJ, Boyne MT 2nd, Burke PV, Kwast KE, Kelleher NL. Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. *Analytical chemistry*. 2007; 79:7984–7991. [PubMed: 17915963]
- Pfreundt U, James DP, Tweedie S, Wilson D, Teichmann SA, Adryan B. FlyTF: improved annotation and enhanced functionality of the *Drosophila* transcription factor database. *Nucleic acids research*. 2010; 38:D443–D447. [PubMed: 19884132]
- Ramamurthy L, Barbour V, Tuckfield A, Clouston DR, Topham D, Cunningham JM, Jane SM. Targeted disruption of the CP2 gene, a member of the NTF family of transcription factors. *The Journal of biological chemistry*. 2001; 276:7836–7842. [PubMed: 10995745]
- Rani PG, Ranish JA, Hahn S. RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Molecular and cellular biology*. 2004; 24:1709–1720. [PubMed: 14749386]
- Rinner O, Mueller LN, Hubalek M, Muller M, Gstaiger M, Aebersold R. An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nature biotechnology*. 2007; 25:345–352.

- Riuzzi F, Sorci G, Sagheddu R, Donato R. HMGB1-RAGE regulates muscle satellite cell homeostasis through p38-MAPK- and myogenin-dependent repression of Pax7 transcription. *Journal of cell science*. 2012; 125:1440–1454. [PubMed: 22328527]
- Rusconi F, Guillonnet F, Praseuth D. Contributions of mass spectrometry in the study of nucleic acid-binding proteins and of nucleic acid-protein interactions. *Mass spectrometry reviews*. 2002; 21:305–348. [PubMed: 12645088]
- Saba J, Bonnell E, Pomies C, Eng K, Thibault P. Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear ion trap/Orbitrap mass spectrometer. *Journal of proteome research*. 2009; 8:3355–3366. [PubMed: 19469569]
- Sandoval WN, Pham V, Ingle ES, Liu PS, Lill JR. Applications of microwave-assisted proteomics in biotechnology. *Combinatorial chemistry & high throughput screening*. 2007; 10:751–765. [PubMed: 18478957]
- Schaefer U, Schmeier S, Bajic VB. TcoF-DB: dragon database for human transcription co-factors and transcription factor interacting proteins. *Nucleic acids research*. 2011; 39:D106–D110. [PubMed: 20965969]
- Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics*. 2010; 10:1265–1269. [PubMed: 20077414]
- Sebastian Winkler G, Lacomis L, Philip J, Erdjument-Bromage H, Svejstrup JQ, Tempst P. Isolation and mass spectrometry of transcription factor complexes. *Methods*. 2002; 26:260–269. [PubMed: 12054882]
- Seizl M, Hartmann H, Hoeg F, Kurth F, Martin DE, Soding J, Cramer P. A conserved GA element in TATA-less RNA polymerase II promoters. *PLoS one*. 2011; 6:e27595. [PubMed: 22110682]
- Sheng QH, Xie T, Ding DF. De Novo Interpretation of MS/MS Spectra and Protein Identification via Database Searching. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)*. 2000; 32:595–600. [PubMed: 12058170]
- Spangler L, Wang X, Conaway JW, Conaway RC, Dvir A. TFIIF action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98:5544–5549. [PubMed: 11331764]
- Standing KG. Peptide and protein de novo sequencing by mass spectrometry. *Curr Opin Struct Biol*. 2003; 13:595–601. [PubMed: 14568614]
- Stead JA, Keen JN, McDowall KJ. The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. *Mol Cell Proteomics*. 2006; 5:1697–1702. [PubMed: 16845145]
- Stead JA, McDowall KJ. Two-dimensional gel electrophoresis for identifying proteins that bind DNA or RNA. *Nat Protoc*. 2007; 2:1839–1848. [PubMed: 17703194]
- Stergachis AB, Maclean B, Lee K, Stamatoyannopoulos JA, Maccoss MJ. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nature methods*. 2011
- Thakur SS, Geiger T, Chatterjee B, Bandilla P, Froehlich F, Cox J, Mann M. Deep and highly sensitive proteome coverage by LC-MS/MS without pre-fractionation. *Mol Cell Proteomics*. 2011
- Ueda T, Chou H, Kawase T, Shirakawa H, Yoshida M. Acidic C-tail of HMGB1 is required for its target binding to nucleosome linker DNA and transcription stimulation. *Biochemistry*. 2004; 43:9901–9908. [PubMed: 15274644]
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huseon DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Chaurab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A,

Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science*. 2001; 291:1304–1351. [PubMed: 11181995]

Vinson CR, Sigler PB, McKnight SL. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science*. 1989; 246:911–916. [PubMed: 2683088]

Wang GL, Semenza GL. Purification and characterization of hypoxia-inducible factor 1. *The Journal of biological chemistry*. 1995; 270:1230–1237. [PubMed: 7836384]

Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. DBD--taxonomically broad transcription factor predictions: new content and functionality. *Nucleic acids research*. 2008; 36:D88–D92. [PubMed: 18073188]

Wu SL, Huhmer AF, Hao Z, Karger BL. On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *Journal of proteome research*. 2007; 6:4230–4244. [PubMed: 17900180]

Yaneva M, Tempst P. Affinity capture of specific DNA-binding proteins for mass spectrometric identification. *Analytical chemistry*. 2003; 75:6437–6448. [PubMed: 14640712]

Yang VW. Eukaryotic transcription factors: identification, characterization and functions. *The Journal of nutrition*. 1998; 128:2045–2051. [PubMed: 9808664]

Zeiler M, Straube WL, Lundberg E, Uhlen M, Mann M. A protein epitope signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP*. 2011

Zhang H, Jin J, Tang L, Zhao Y, Gu X, Gao G, Luo J. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic acids research*. 2011; 39:D1114–D1117. [PubMed: 21097470]

Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, Guo AY. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic acids research*. 2011

Zhang X, Li Y, Shao W, Lam H. Understanding the improved sensitivity of spectral library searching over sequence database searching in proteomics data analysis. *Proteomics*. 2011; 11:1075–1085. [PubMed: 21298786]

Zhang Z, McElvain JS. De novo peptide sequencing by two-dimensional fragment correlation mass spectrometry. *Anal Chem*. 2000; 72:2337–2350. [PubMed: 10857603]

Figure 1. TF purification with trapping combined with a variety of TF characterization methods
A). In this example, the full hTERT promoter containing single stranded (GT)₅ tails at both the 5' and 3' ends was used to purify TFs. DNA-TF complexes are purified by annealing (GT)₅ tails to a (CA)₅-Sephacrose column. The column was washed with 4 mL of low salt buffer (1X binding buffer), and DNA-TF complexes were eluted with 1 mL of high salt buffer (1X TE 0.5M NaCl pH 7.5). Characterization of TFs in these DNA-TF complexes was performed with a variety of methods (1B-1E). **B)** TF characterization with two-dimensional gel electrophoresis (2-DE). Isoelectric focusing was performed for the first dimension followed by separating the proteins by their relative molecular mass (*M_r*) with the

use of a 12% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) Silver stained spots indicate TF components of DNA-TF complexes and other proteins that that were purified with promoter trapping. **C)** TF characterization with two-dimensional Southwestern blotting (2-D SWB). The 2-DE gel shown in B) was electroblotted onto a PVDF membrane and hybridized with the radiolabeled hTERT promoter. Spots indicate TF components of the DNA-TF complexes that were purified with promoter trapping. **D)** TF characterization (identification) with capillary liquid chromatography-tandem mass spectrometry (LC/MS/MS) and protein database searching¹. Two separate spots were excised from the gel and in-gel trypsin digestion was performed. 1 µg of extracted peptide was injected onto a C18 column and characterized with a hybrid linear ion trap-Fourier-transform tandem mass spectrometer. MS/MS spectra were generated via high-energy C-trap dissociation (HCD). HMGB1 (upper spectra) was identified with a Mascot protein score of 99 and sequence coverage of 37%. The amino acid sequence assignment of product ions for the top-ranked tryptic peptide from HMGB1, IKGEHPGLSIGDVAK, spanning amino acid residues 113–127, yielded an expectation value of 0.00118. The observed monoisotopic mass of the unmodified, doubly-charged precursor ion for IKGEHPGLSIGDVAK at 760.9258 *m/z* was 1519.8371 Da. Transcription factor CP2 (TFCP2) (IPI00037599) was also identified (lower spectra) with an expectation value of the top-ranked tryptic peptide (LHDETLTYLNQGQSYEIR) was 0.0000034 (Cunningham & Vanin & Tran & Valentine & Jane, 1995, Lim & Swendeman & Sheffery, 1992, Ramamurthy & Barbour & Tuckfield & Clouston & Topham & Cunningham & Jane, 2001). **E)** TF characterization with one-dimensional Western blotting (1-D WB). A 1D-SDS-PAGE gel was electroblotted onto a PVDF membrane followed by probing with specific antibodies (anti-TBP, anti-PolIII, anti-USF2, anti-Sp1 and anti-beta(b)-actin). The different lanes of the SDS-PAGE gel were loaded with whole cell lysate (WCL); nuclear extract (NE); hTERT promoter trapping eluent (hTERT PT Eluate) and flow-through (hTERT PT FT) from 1A); c-Jun promoter trapping eluent (c-Jun PT Eluate) and flow-through (c-Jun PT FT) were from a previous experiment. Bands indicate components of DNA-TF complexes.

¹ Unpublished data were collected in the RCMi Proteomics & Protein Biomarkers Cores at UTSA.

Figure 2. TF purification with electrophoretic mobility shift assay-three dimensional electrophoresis (EMSA-3-DE)

A) In this example, the CAAT enhancer binding protein (C/EBP) promoter was ^{32}P -labeled at both the 5' and 3' ends. Once the DNA-TFs complexes were formed, TF purification was performed on a 3% EMSA gel by separating the Complex from the free DNA. Unbound proteins are not visualized since the radiolabel is only on the oligonucleotide.

Characterization of TFs in DNA-TF complexes revealed by their higher molecular mass (M_r) on the EMSA gel (top) was performed with a variety of methods (B-D). **B)** TF characterization with two-dimensional gel electrophoresis (2-DE). 2DE-SDS-PAGE was performed as previously described in promoter trapping. Silver stained spots indicate TF

components of DNA-TF complexes and other proteins that were purified with EMSA. The single spot that was visualized was in excellent agreement with the predicted pI of 8.97 and M_r of 42,569 for C/EBP however, further characterization was necessary to confirm (shown in figures 2C and 2D). **C)** TF characterization with two-dimensional Southwestern blotting (2-D SWB). The 2-DE gel shown in B) was electroblotted onto a PVDF membrane, blocked with 5% milk, and hybridized with the radiolabeled C/EBP promoter. The single spot indicates C/EBP that was purified with EMSA. **D)** TF characterization (identification) with capillary liquid chromatography-tandem mass spectrometry (LC/MS/MS) and protein database searching in a manner similar that described in Figure 1. C/EBP was identified, and the amino acid sequence assignment of product ions for the top-ranked tryptic peptide from C/EBP, GSHMASMTGGQQMGR, spanning amino acid residues 18–32, and a representative MS/MS spectrum are shown.