Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Chromatography B, 797 (2003) 269–288

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Methods for transcription factor separation

Robert A. Moxley, Harry W. Jarrett*, Suchareeta Mitra

Department of Biochemistry, 858 Madison Avenue, University of Tennessee Health Science Center, Memphis, TN 38163, USA

Abstract

Recent advances in the separation of transcription factors (TFs) are reviewed in this article. An overview of the transcription factor families and their structure is discussed and a computer analysis of their sequences reveals that while they do not differ from other proteins in molecular mass or isoelectric pH, they do differ from other proteins in the abundance of certain amino acids. The chromatographic and electrophoretic methods which have been successfully used for purification and analysis are discussed and recent advances in stationary and mobile phase composition is discussed.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Reviews; Transcription factors

Contents

1. Introduction ... 270
2. Overview of transcription factor homology family motifs ... 271
   2.1. Helix-turn-helix motif ... 272
   2.2. Zinc finger motif ... 272
   2.3. $Zn(II)_2Cys_6$ binuclear cluster motif ... 273
   2.4. Leucine zipper motif ... 274
   2.5. Basic helix-loop-helix motif ... 274
   2.6. β-Ribbon motif ... 274
   2.7. Other motifs ... 275
3. Classical protein purifications methods used in TF purification ... 275
   3.1. Nuclear extract ... 275
   3.2. Chromatographic methods ... 276
      3.2.1. Size exclusion chromatography ... 276
      3.2.2. Ion-exchange chromatography ... 276
         3.2.2.1. Heparin ... 276
   3.3. Affinity chromatography ... 276
      3.3.1. Lectins ... 276
      3.3.2. Procion dyes ... 277
      3.3.3. DNA-affinity chromatography ... 277
         3.3.3.1. Heterogeneous sequence ... 277
         3.3.3.2. Homogenous sequence ... 277
         3.3.3.3. Discrete oligonucleotide ... 278
         3.3.3.4. Concatemers ... 278
4. Recent developments in DNA-affinity chromatography ... 279
   4.1. Oligonucleotide trapping ... 279
   4.2. Catalytic chromatography ... 280

* Corresponding author. Tel.: +1-901-448-7078; fax: +1-901-448-7360.
E-mail address: hjarrett@utmem.edu (H.W. Jarrett).

1570-0232/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S1570-0232(03)00609-3

4.3. Bi-column method ... 280
4.4. Methods for selective elution of DNA columns ... 280
4.4.1. Salt elution ... 280
4.4.2. Ligand-specific elution ... 280
4.4.3. Temperature ... 281
5. Electromigration methods: TF assays ... 282
5.1. Electrophoretic mobility shift assay (EMSA) ... 283
5.2. Capillary electrophoretic mobility shift assay (CEMSA) ... 283
5.3. Southwestern blot ... 283
5.4. 2D SDS-PAGE and sequencing/proteomics (renaturation of 2D gels) ... 284
6. Conclusion ... 285
Acknowledgements ... 285
References ... 285

1. Introduction

Transcription factors (TFs) are an interesting and complex group of proteins. TFs bind to the promoter region of genes and either activate or repress transcription. The promoter region contains specific DNA sequences (called variously an enhancer or repressor "element" or "footprint") that are bound with high affinity and specificity to individual TFs and this complex of proteins either recruit or fail to recruit active RNA polymerase. This determines whether or not a messenger (or other type) of RNA is produced and ultimately expressed. Genetic regulation is the direct result of these interactions and this controls the cell cycle, differentiation, and ultimately the cell's biology and fate.

A recent search for "transcription factor" in the PubMed nucleotide database reveals over 40,000 entries. Only a few hundred of these have ever been purified, and a much smaller number purified to homogeneity and characterized at the protein level. The purifications that have succeeded were all arrived at empirically but certain strategies recur. Most have been purified from nuclear extracts and have used ammonium sulfate or other methods to concentrate the sample. Some TFs occur in large complexes and for these, separation by size exclusion chromatography (SEC) has proved useful. Ion-exchange chromatography (IEC), primarily cation exchange but also anion exchange, has often also been used. The combination of all of these methods though has usually resulted, at best, in a few hundred-fold purification. Since many of the TFs are present in only very low amounts in the cell, purification of tens-of-thousand-fold are usually required and so these methods alone give only very poor enrichment. Purification to homogeneity has usually required the application of affinity chromatography techniques.

The most obvious and widely used kind of affinity chromatography is DNA-affinity chromatography. The first purifications to use this technique used fragmented genomic DNA attached to supports to purify those proteins which bind the DNA. Later, oligonucleotide synthesis became available along with foot printing techniques which could be used to characterize the DNA element bound by TFs within the promoter region. These footprints could then be synthesized and attached to suitable chromatography supports to make a very selective, specific oligonucleotide DNA-affinity chromatography which was more or less specific for a single TF. This technique suffers from some unexpected problems. For one, while most TFs bind their footprint element with very high affinity (often in the picomolar range), they and most other DNA-binding proteins will bind any DNA sequence with somewhat lower affinity. The high concentration of DNA attached to supports encourages these lower affinity interactions and specific elution of only the protein of interest has proved challenging. To circumvent this problem, competitor DNA and other substances (detergents, heparin, moderate salt concentrations) are added to the mobile phase to lessen these weaker interactions. Alternative methods for eluting DNA-affinity columns have been developed which can also improve the selectivity of the chromatography. Also, protein binding of the DNA at low concentrations in solution and then "trapping" the DNA–protein complexes in various ways has also provided a way to increase the specificity of the method and the protein purity obtained.

Other kinds of affinity chromatography have also sometimes been useful. While most intracellular proteins are not glycosylated, apparently some TFs are and lectin affinity chromatography has been useful in these cases. Also, Procion dyes are known to bind with nucleotide binding proteins, including DNA-binding proteins, and affinity chromatography on Procion dye columns (e.g. Cibacron–Sepharose) has sometimes been used.

Here we give an overview of the TFs, their chromatography, and purification. We also investigated, using the extensive database of TF sequences, whether there are any properties unique to this group of proteins. This analysis reveals that as a group TFs are quite similar to other types of proteins but there are some differences.

Table 1
Transcription factor amino acid composition, relative molecular mass, and isoelectric point analysis

	Hydrolase			Alcohol dehydrogenase			Dayhoff			Transcription factor[a]			Protein	
	X	S.D.	*P*[b]	*X*	S.D.	*P*	*X*	S.D.	*P*	*X*	S.D.	*P*	*X*	S.D.
Ala	7.73	3.28	0.71	9.00	2.20	0.008	8.36	4.18	0.248	8.07	3.44	0.390	7.49	3.32
Arg	5.70	3.09	0.207	3.20	1.42	<0.001	3.91	2.82	0.043	5.35	1.32	0.363	4.99	2.43
Asn	4.36	2.28	0.848	5.02	1.62	0.125	4.46	3.00	0.969	4.26	2.19	0.661	4.44	2.08
Asp	5.54	1.81	0.005	4.82	0.97	0.275	5.00	2.28	0.232	4.47	1.61	0.920	4.51	1.76
Cys	2.07	1.65	0.193	2.06	1.49	0.184	2.57	3.54	0.094	1.80	1.06	0.537	1.62	1.81
Gln	3.87	2.01	0.704	2.09	0.76	<0.001	3.34	2.00	0.328	5.21	1.99	<0.001	3.72	1.82
Glu	6.04	3.13	0.975	5.13	2.08	0.050	5.79	3.34	0.657	5.88	2.15	0.705	6.06	2.58
Gly	8.20	2.46	<0.001	8.78	2.14	<0.001	8.80	3.18	<0.001	6.95	3.29	0.188	6.21	2.21
His	2.72	1.68	0.052	2.34	1.05	0.358	2.98	2.22	0.020	3.11	1.33	<0.001	2.11	1.36
Ile	5.70	2.43	0.088	7.35	2.28	0.100	4.21	2.51	<0.001	3.11	1.50	<0.001	6.55	2.51
Leu	9.47	2.86	0.594	8.30	1.61	0.002	8.43	3.61	0.043	8.40	2.70	0.016	9.78	2.93
Lys	4.76	2.52	0.004	6.78	1.63	0.586	8.42	4.15	0.012	5.43	1.45	0.040	6.49	3.30
Met	2.23	1.02	0.257	1.27	1.02	<0.001	2.11	1.90	0.248	2.29	1.40	0.495	2.47	1.09
Phe	3.83	1.40	0.158	3.94	0.99	0.226	4.05	2.17	0.433	3.03	1.17	0.001	4.42	2.57
Pro	4.85	1.88	0.697	4.78	1.06	0.500	4.76	2.94	0.643	7.99	2.42	<0.001	5.00	2.01
Ser	6.41	2.39	0.217	4.37	1.47	<0.001	6.51	4.00	0.432	10.71	2.20	<0.001	7.05	2.81
Thr	4.79	2.30	0.013	7.88	2.87	<0.001	5.40	2.64	0.294	5.64	1.90	0.505	5.91	2.10
Trp	1.61	1.16	0.132	1.28	0.88	0.990	1.30	1.12	0.904	0.89	0.64	0.029	1.28	1.03
Tyr	3.01	1.34	0.479	1.97	0.77	<0.001	3.06	1.83	0.649	2.37	1.33	0.004	3.21	1.53
Val	7.17	1.85	0.264	9.65	1.23	<0.001	5.96	3.28	0.190	5.04	1.39	<0.001	6.71	2.25
p*I*	6.63	1.87	0.032	7.31	1.19	0.466	7.69	1.72	0.657	7.72	1.69	0.433	7.53	1.83
MW	40593	29053	0.405	26415	10143	0.004	15367	8317	<0.001	66067	63330	0.096	47157	47292

[a] The TF data was chosen by searching the PubMed protein database for "TF NOT putative NOT partial". The total number of entries was then multiplied by 50 random numbers to select individual sequences. For the other proteins, the same search string was used except replacing "TF" with "hydrolase", "alcohol dehydrogenase", or "protein". The Dayhoff dataset is from [2]. Protein parameters were determined by sequence analysis using the online program PROT-PARAM.

[b] *X* is the mean ($n = 50$), S.D. the standard deviation, and *P* is the probability resulting from a two-tailed Student's *t*-test comparison to the random protein dataset.

2. Overview of transcription factor homology family motifs

One of the largest and most diverse classes of DNA-binding proteins are the TFs. TF DNA-binding specificity is conferred through distinct structural motifs present in the DNA-binding domains. These motifs also provide the basis for their classification. The characteristics of these motifs (the preponderance of α-helices in DNA-binding regions, the charge interaction with DNA and RNA polymerases, etc.) suggests that TFs may differ in p*I* or amino acid composition due to alkalinity or acidity required for DNA or polymerase binding. Variance in p*I* amino acid composition, or relative molecular mass could be utilized for TF purification. However, the data in Table 1 shows a comparison of four sets of 50 proteins, all randomly chosen from the database. One set is TFs, and another three sets are non-TFs of various types. These data suggest that as a group TFs are not significantly different from other proteins in charge or molecular mass. Table 1 shows that all of the protein families have distinct amino acyl residues with significant variance from random proteins. Much of the variance within each group can probably be explained by structural and functional attributes distinct to each group of proteins. The TF data shows 11 residues that differ significantly from the random protein data. Importantly, 92% of the TF data are from the zinc finger (see Section 2.2), helix-turn-helix (HTH) (see Section 2.1), helix-loop-helix (see Section 2.5), and leucine zipper (see Section 2.4) families, all of which are composed of highly α-helical structures. Some of the observed divergence between the two groups can be explained by structural attributes. For example, residues Ile, Leu, Phe, Trp, Tyr, and Val are bulky, hydrophobic groups that are found with a high probability in β-sheets [1], and are in low abundance in the TFs data. Also, pro and ser are found in high abundance in turns and are significantly higher in the TF data. Finally, other residue divergence could be explained by function, as Gln, His, and Ser are often involved in hydrogen bonding of TFs to DNA nucleotide bases and are significantly higher in the TF data perhaps due to the number of sharp turns in many of these structures. The Dayhoff dataset [2] was an early work that analyzed sequenced proteins. While this work was extremely important in developing proteomics, the data available at the time was limited. And, as shown, sequenced proteins were usually of low relative molecular mass and represented a small population of proteins relative to the data available today. The differences shown in the Dayhoff data are not easily explained by structural divergence as many of the proteins were hormones, toxins, and inhibitors with no pre-

dicted secondary structure. The Dayhoff data was included because of its foundation in modern proteomics and as a comparison to the random dataset used in comparing TFs, as no TFs were sequenced at that time. Also, the Dayhoff data acts as a comparison of residue divergence of random proteins from several families to the other protein datasets. Alcohol dehydrogenases were chosen as a highly homologous family of proteins for comparison. The alcohol dehydrogenases show 11 residues that are significantly different from the random protein data. Much of the divergence can be explained by structural differences. Many alcohol dehydrogenases are composed of large regions of β-sheets and are thus very different from most TFs. The observed divergence is indicative of the β-sheet structural composition. For example, Val and Thr are in high abundance in the alcohol dehydrogenase data and bulky hydrophobic residues are characteristic of β-sheets. Further charged residues are often found in α-helices and residues of this type would be expected in less abundance in β-sheet proteins, as indicated with glu, gln, and arg. Finally, the hydrolase proteins were included because the hydrolase enzymes are a very diverse group containing both helical and sheet structures. The expected result from this dataset would be that few residues would differ significantly from the random protein dataset. As expected, the hydrolase dataset has four residues that differ significantly from the random dataset, as compared to six residues in the Dayhoff dataset. In conclusion, all of the datasets tested display significant differences of some residues. Many differences can be explained by structure, and the TF data does not differ from the random protein dataset in relative molecular mass or p*I*.

Many families of TFs with distinct motifs have been identified and classified. Structural variety within families makes classification of TFs complex. The wide structural variety of the TF families is rather difficult to mentally envision, and to assist the reader, many of the TFs mentioned in this review, that have available structural data, have been listed with corresponding PDB files in Table 2. A thorough discussion of TF families is beyond the scope of this review, but several more detailed reviews of TFs and DNA-binding proteins are available, including [3,4]. However, many of the more thoroughly characterized TF motif families are briefly discussed in this review including helix-turn-helix [5]; Wintjens [14] (p. 741), zinc finger [6], zinc binuclear cluster [7], leucine zipper [8], helix-loop-helix [9], and β-ribbon motifs.

2.1. Helix-turn-helix motif

The helix-turn-helix, family of TFs was the first DNA-recognition motif discovered and was identified by X-ray crystallographic analysis. Some of the first HTH protein structures described were the lambda phage Cro protein, the *Escherichia coli* catabolite activator protein (CAP), and the lambda repressor [10–12]. Since their discovery, members of the HTH TF family have become the most thoroughly studied family of DNA-binding proteins and they participate in many metabolic activities including development. Hundreds of HTH proteins have been identified and are widely distributed in prokaryotes and eukaryotes [5,13]. In general, the HTH motif is composed of 20 or 60 amino acid residues containing two α-helices connected by a turn of about 1–20 amino acids. The helices are positioned at 120° to each other, and the carboxyl-terminal helix is the DNA-recognition helix and binds within the major groove of DNA. The HTH motif is a small domain contained within structurally diverse proteins. HTH motif proteins have been classified by the surrounding structural elements such as, α-helices, β-strands, or hair-pins that are spatially close to the HTH motif and close off the hydrophobic core. A more detailed discussion of the structurally diverse families of HTH TFs can be found [5,14]. A few of the HTH TFs are: the Cro and lambda repressors [15], *lac* repressor and purine repressor [14,16], histone 5 and hnf3γ [17], diptheria toxin receptor [14], catabolite activator protein, lexA, and birA [14], some homeodomain TFs [18], and heat shock factor 1 (HSF1) [19].

A major difference between some of the prokaryotic and eukaryotic HTH proteins is that prokaryotic HTH proteins are not able to interact with DNA as a monomer and the HTH motif is not independently stable and requires other structures to confer DNA-binding activity. Binding is often as a dimer but others, such as *lac* repressor bind as a tetramer [19]. By contrast, some eukaryotic HTH homeodomain proteins can bind DNA as a monomer and the HTH motif alone can bind to DNA without other structural support [3].

2.2. Zinc finger motif

Zinc finger proteins are among the most abundant proteins in eukaryotic genomes. Their functions are extraordinarily diverse and include DNA recognition, RNA packaging, transcriptional activation, regulation of apoptosis, protein folding and assembly, and lipid binding. The *Xenopus laevis* TF, TF IIIA, was the first zinc finger motif protein described [20]. Since that time, many other proteins have been identified as zinc finger TFs, including SP1, glucocorticoid receptor, zip268, p55, WZF1, and GATA1–GATA4 [21–24]. Zinc chelation is the common feature of zinc finger TFs and is required for DNA-binding activity. Zinc is chelated by various combinations of Cys and His residues in a tetrahedral coordination, while intervening amino acids form a finger-like protrusion responsible for DNA-recognition; thus, the name zinc finger motif. Different zinc finger TFs vary not only in sequence, but they also vary in structure and are classified into zinc finger TF families. A brief discussion of some identified zinc finger TF families will follow: the classic zinc finger, Cys–Cys–His–His, family contains SP1 and TF IIIA with a consensus sequence of Cys–$X_{2\text{ or }4}$–Cys–X_{12}–His–$X_{3–5}$–His [21]. TF IIIA contains nine finger sequences of about 30 residues each and other classic zinc finger proteins are similar, though others generally contain fewer of the mo-

Table 2
Transcription factor structure file list

Transcription factor	Family	PDB file[a]	MMBD[b]
Lambda Cro	HTH	4CRO, 1LRP	3151, 1574
CAP	HTH	2CGP, 1CGP	13307, 512
Lambda repressor	HTH	1LLI, 1LMB	1553, 1555
lac repressor	HTH	1L1M, 1JWL, 1LBI	19908, 17832, 4547
Histone 5	HTH	1HST	1258
LexA	HTH	1LEA, 1LEB	1517, 1518
BirA	HTH	1BIA, 1BIB	365, 366
Paired	HTH	1PDN	3805
Engrailed	HTH	1HDD, 2HDD, 3HDD	1136, 7633, 8560,
Antennapedia	HTH	9ANT, 1AHD	8759, 208
Heat shock factor	HTH	1HKT, 2HTS, 3HSF	1201, 2736, 4041
TF IIIA	Zinc finger	1TF3, 1TF6	6403, 8100
SP1	Zinc finger	1SP1, 1SP2	5777, 5776
Glucocorticoid receptor	Zinc finger	1GLU, 2GDA, 1LAT	1055, 2689, 4278
GATA1	Zinc finger	1GAU, 1GAT	1000, 1001
p55	Zinc finger	1HVO, 2ZNF	4938
GAL4	C_6 binuclear cluster	1D66, 1WA6	697, 7497
PPR1	C_6 binuclear cluster	1PYI	2002
HAP1	C_6 binuclear cluster	1PYC, 1HWT	4900, 10272
LAC9	C_6 binuclear cluster	1CLD	10684
PUT3	C_6 binuclear cluster	1ZME, 1AJY	8434, 6450
C/EBP	Leucine zipper	1HJB, 1IO4, 1CI6	15977, 15976, 15196
GCN4	Leucine zipper	1GK6, 1KD8	18650, 17965
AP-1, FOS–JUN	Leucine zipper	1FOS	3506
JUN	Leucine zipper	1JUN	4637
CREB	Leucine zipper	1DH3	14498
MyoD	bHLH	1MDY	1661
Myc	bHLH	1A93, 2A93, 1NKP	8749, 9293, 21886
Mad	bHLH	1NLW	21892
Max	bHLH	1NLW, 1NKP, 1E91	21892, 21886, 14566
Met J	β-Ribbon	1CMA, 1CMB, 1CMC	548, 549, 550
Arc	β-Ribbon	1ARQ, 1ARR	269, 270
Mnt	β-Ribbon	1MNT, 1QEY	1699, 10915
DNA-binding protein II	β-Ribbon	1NI8, 1HUU, 1HNR	22129, 9396, 3492
Integration host factor	β-Ribbon	1IHF	5137
Transcription factor 1	β-Ribbon	1EXE, 1WTU	14831, 5120
TBP	β-Ribbon	1VOK, 1PCZ, 1TGH	5684, 5296, 4896
MotA	Double wing	1KAF, 1I1S, 1BJA	17914, 15730, 9221

[a] PDB structural files are available at http://www.rcsb.org/pdb.
[b] Structural files are also available at http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml.

tif. Each contains two Cys residues at the amino terminal end and two His residues at the carboxy-terminal end. Another zinc finger type is the Cys–Cys–Cys–Cys family containing the glucocorticoid and estrogen receptors and GATA1. Members of this family contain a highly conserved DNA-binding domain that consists of about 70 residues with a consensus sequence of $C–X_2–C–X_{17}–C–X_2–C$ [24,25]. This family uses four Cys to bind zinc, and has two zinc finger regions. In the three-dimensional (3D) structure, these two zinc fingers are not separated into discrete units but are interwoven into a single globular domain with extensive interactions between the two finger units. Finally, the retroviral zinc finger family, Cys–Cys–His–Cys contains the retroviral protein p55. During the assembly and budding stages of the retroviral life cycle, a large polypeptide is produced and forms a complex with viral RNA. After the complex has been transported to the cell membrane for budding, a low relative molecular mass nucleic acid binding protein cleaves the complex. In many retroviruses, this protein contains either one or two regions with the following consensus sequence $Cys–X_2–Cys–X_4–His–X_4–Cys$ [26].

2.3. $Zn(II)_2Cys_6$ binuclear cluster motif

The DNA-binding domain of $Zn(II)_2Cys_6$, or C_6, binuclear cluster motif protein was first characterized in the *Saccharomyces cerevisiae* GAL4 protein [25]. The C_6 binuclear cluster DNA-binding domain has been identified exclusively in fungal proteins primarily as TFs [27]. This characteristic is distinct to this TF family as all other DNA-binding protein families are not restricted to the Fungal Kingdom. More than 100 known or predicted C_6 proteins have been identified. Some of the best characterized are

GAL4, PPR1, LEU3, HAP1, LAC9, and PUT3 [7,28–31]. This zinc chelating TF family is very different from the zinc finger family (see Section 2.2) with a conserved sequence of Cys–X_2–Cys–X_6–Cys–X_6–Cys–X_2–Cys–X_6–Cys [32]. The C_6 binuclear cluster proteins consist of six conserved cysteines, which form two α-helical structures that tetrahedrally coordinate two zinc ions to form a cloverleaf-turn-helix shaped structure. The two zinc atoms are coupled by two bridging cysteine residues into a binuclear cluster with a short distance between the metal atoms; thus the name $Zn(II)_2Cys_6$ binuclear cluster [28]. The C_6 binuclear structure directly participates in DNA recognition and binding [7,28]. The DNA-binding sites consist of conserved trinucleotide dyads, usually in a symmetrical configuration containing an internal variable sequence of defined length [33]. For example, GAL4 and LAC9 bind to $CGGN_{11}CCG$, and PPR1 and UAY bind to $CGGN_6CCG$ [27].

2.4. Leucine zipper motif

The leucine zipper, or bZIP, motif was first discovered in rat liver nuclear extracts as CAAT/enhancer binding protein, or C/EBP, by McKnight and co-workers [8]. Since the discovery of the bZIP motif, members have been identified in such diverse kingdoms as fungi, plants, and animals and include GCN4, Fos, Jun, CREB, and Ig/EBP [3,4,34,35]. Leucine zipper proteins contain two distinct domains: a leucine zipper region that mediates dimerization, and a basic region that confers DNA-binding. The bZIP family is characterized by a heptad repeat of leucines over a region of 30–40 residues that form α-helices with a periodicity of 3.6 amino acids per turn [4]. Leucines occur at every seventh residue, placing them on one side of the helix, and these α-helices allow dimerization of monomers by forming a coiled-coil where the leucines interact hydrophobically and zip the monomers together; hence, the name leucine zipper [36]. Dimerization results in the formation of a "Y"-shaped structure where the leucine zipper forms the stem of the "Y". The forked basic region is rich in highly conserved arginines and is usually about 30 residues long per monomer [37]. The two-fold symmetry of the "Y" fork basic region extends along the major groove and mediates DNA-binding and recognition [3]. Evidence indicates that the putative recognition sites are between 9 and 10 base pairs; meaning that in order for the more distal region of the basic domain to contact the DNA, the DNA helix must be bent [4]. As in many DNA-binding proteins, bZIP proteins must dimerize in order to bind DNA. Leucine zipper proteins not only form homodimers, but also form heterodimers. Heterodimerization plays several different roles in bZIP TFs such as regulating activity, as in CREB, and altering DNA-binding specificity [38]. For example, AP-1 consists of one Fos-family monomer and one Jun-family monomer which bind different sites as heterodimers than as homodimers [39].

2.5. Basic helix-loop-helix motif

The basic helix-loop-helix, or bHLH, family of TFs was first identified in murine proteins as E12/E47 [40]. Over 240 bHLH proteins have been identified and are widely distributed from fungi to humans [41]. In *S. cerevisiae*, bHLH TFs regulate many important metabolic pathways including phosphate uptake and phospholipid biosynthesis [42–44]. bHLH proteins are required in multi-cellular organisms for many developmental processes including neurogenesis, myogenesis, hematopoiesis, and pancreatic development [45–47]. A highly conserved bipartite structure of domains for DNA binding and protein–protein interaction is characteristic of the bHLH proteins [40,48]. The bHLH motif consists of two aliphatic α-helices joined by a random-coil loop structure of variable length. A basic region located amino terminal of the first helix mediates DNA binding [40,48]. The HLH region is comprised of many hydrophobic residues that are the site of protein–protein interaction. Due to the many bHLH proteins that have been identified, different classification schemes were devised. One scheme based on tissue distribution, dimerization abilities, and DNA-binding specificities was developed by Murre et al. [40,48], and a brief description of this classification model will follow: Class I bHLH proteins also known as E proteins include E12, HEB and daughterless. Class I proteins are expressed in multiple tissues and can form either heterodimers or homodimers and bind only to the E box site [40,49]. Class II proteins include MyoD, myogenin, and Atonal and are incapable of forming homodimers and preferentially form heterodimers with Class I proteins. Also, Class II proteins are expressed in a tissue-restrictive pattern [40]. Class III proteins include the Myc family, TFE3, and SREB1 and contain a leucine zipper adjacent to the HLH motif [50,51]. Class IV proteins include Mad, Max, and Mxi and can heterodimerize with the Myc proteins (Class III) or Class IV proteins, or homodimerize [41,52]. Class V proteins lack a basic region and are negative regulators of Class I and Class II proteins. Id1–4 and emc are examples of Class V proteins [53,54]. Finally, the Class VI proteins are defined by having a proline in their basic region and include Hairy and Enhancer of Split [55,56].

2.6. β-Ribbon motif

The β-ribbon motif family of TFs is unique in that a β-sheet is used to bind DNA while the other TF families bind DNA using α-helices. Generally, the family is simply classified as DNA-binding proteins that use anti-parallel β-sheets to interact with DNA [3]. There are three classes of β-ribbon motif proteins that are separated based on β-sheet composition, protein–protein interactions, and DNA-binding specificity abilities [57]. A more detailed discussion of β-ribbon TFs can be found [57], and a brief overview will follow: the Met J protein family contains Met J, Arc, Mnt, Tra Y,

and the F episome [57]. Met J proteins form dimers with a core composition of four α-helices, two from each subunit. An anti-parallel β-ribbon formed from the amino terminal end of each monomer protrudes from the core and interacts with the major groove of DNA. Also, all of the Met J proteins bind to specific element sequences. Another β-ribbon family member is the HU family. The HU family consists of HU, or DNA-binding protein II, integration host factor (IHF), and TF 1 (TF1) [57]. HU family proteins form dimers and have a core composed of two helices, one from each monomer at the amino terminal regions and a three-stranded anti-parallel β-sheet from the carboxy terminal regions followed by a short helix [57]. The β-sheet binds the minor groove of DNA, resulting in a relatively low DNA-binding specificity. A final member of the β-ribbon TFs is the TATA-box binding protein (TBP), which recognizes the TATA-box consensus sequence (TATA a/t A a/t) [57]. TBP is quite distinct from the other β-ribbon proteins as the structure consists of a 10-stranded anti-parallel β-sheet twisted into the shape of a "saddle" that interacts with DNA and sits on the DNA like a saddle on a horse [57].

2.7. Other motifs

Many additional families have been identified in studies of TFs. Due to recent technological advances in genome sequencing and proteomics, many more families will likely be discovered. This review has discussed the TF families that are the most thoroughly studied and commonly found. Other less studied or commonly found TF families that have been identified include the zinc-binding, cysteine rich LIM motif, bipartite homeodomain proteins with the POU domain, non-histone high-mobility group (HMG) proteins, and one of the most recently discovered is the MotA TF with an HK, or 'double wing' motif [58].

3. Classical protein purifications methods used in TF purification

3.1. Nuclear extract

The nuclear extract is a rich source of a variety of TFs, DNA-binding proteins and several enzymes involved in different stages of transcription, recombination, replication and repair [59–62]. Therefore, a good preparation of nuclear extract is often essential for purification of TFs.

Dignam and Roeder procedure [63] is widely used when the source of the preparation is from tissue-culture cells ranging from HeLa cells to B cells (BJAB/Namalwa) to prostate cancer cell lines (LNCaP). Briefly, the method is to suspend the cell pellet in a hypotonic medium that causes the cells to swell. The distended cells are then disrupted with a dounce homogenizer leaving the nuclei intact. An alternative method for cell disruption, but leaving the nuclei intact, is cell lysis with a low concentration of the non-ionic detergent NP-40. This is followed by low speed centrifugation, which separates the intact nuclei from remaining cellular and cytoplasmic debris. The transcriptional components are then extracted from the nuclei by resuspending the cell nuclei in a moderate salt buffer (0.38–0.42 M KCl or NaCl). These moderate salt concentration elute TFs bound to the nuclear DNA without eluting non-specific DNA-binding proteins such as histones. A higher concentration of salt may extract TFs more efficiently (and should be tried for new TFs) but contamination by other nuclear proteins may result (i.e. such as histone H1, which is typically released at 0.6 M KCl). Optimization of the extraction conditions is therefore always helpful for obtaining the highest specific activity for the TFs of interest.

After extraction, proteins are precipitated with ammonium sulfate and then dialyzed to reduce the salt concentration. Dignam and Roeder found 53% to be optimal concentration for precipitation of proteins. A different percentage may be optimal for other proteins.

Dignam and Roeder's procedure has not been applied to solid tissue extracts because it is difficult to generate single cell suspension necessary to efficiently distend the cells prior to homogenization. The classic method of preparation of nuclear extract from tissue is that of Gorski [64]. The approach in the method is to disrupt the tissue first by mincing with scalpel and scissors in cold isotonic buffer, followed by incubation in hypotonic buffer. Then homogenization with a motorized pestle is used and the extract is layered onto a glycerol step gradient and the nuclei are pelleted from cellular debris. The nuclei are then lysed, extracted with salt, and the transcription components are precipitated with saturated solution of ammonium sulfate. Many alternative methods have modified the Gorski's procedure [65,66].

Using in vitro TF binding assays (e.g. EMSA, see below), if the nuclear extract does not contain the desired DNA-binding activity, whole cell extracts or cytoplasmic extracts are tested. Some nuclear proteins leak into the cytoplasmic fraction during the nuclear extract preparation, or are naturally abundant in the cytosol and the DNA-binding activity may be more abundant in non-nuclear fractions. A protocol quite often used for whole cell extract preparation is that of Manley and co-workers [67]. While the procedure is highly recommended for HeLa cells, the method has also been applied to other tissue-culture lines. Briefly the HeLa cells are incubated in a hypotonic buffer and lysed by Dounce homogenization. In the same mixture, the nuclei are then lysed by addition of a lysis buffer and saturated ammonium sulfate. The chromatin is removed by high-speed centrifugation and the supernatant containing the soluble proteins is further precipitated with solid ammonium sulfate. Finally, the pellet is resuspended and dialyzed in a buffer compatible with further purification and assay.

3.2. Chromatographic methods

3.2.1. Size exclusion chromatography

Size exclusion chromatography (SEC) is one of the most effective conventional steps of TF purification [68,69]. It is useful because it separates proteins on the basis of a property (i.e. size) that is unlinked to the DNA-binding activity of the protein. Since many transcription complexes can be quite large (e.g. TF IID), this property can be exploited to gain relatively high purity.

SEC is ideal for identifying early steps in transcription complex assembly and has been successfully applied to studying activator-mediated binding of TF IIB to TF IID. In one example, TF IID and TF IIB were incubated in the presence and absence of GAL4-VP16, and the excluded volume was complemented by addition of the missing factors and nucleotides. It was shown that TF IIB bound to TF IID only in the presence of activator. Surprisingly, the activator promoted binding of TF IIB, but not the other general factors, even when TBP was substituted for TF IID [70,71].

Chromatography is also used for an immobilized template technique and can be used to study different stages of transcription complex assembly. The technique was first utilized to study factor addition in Pol III transcription complexes [72]. The basic scheme is that complexes are assembled from nuclear extract or fractionated factors on a plasmid DNA promoter template and then applied to a gel filtration column. The column separates the plasmid DNA and any attached factors away from free, unbound factors by molecular sieving. The excluded (void) volume containing the plasmid DNA and bound factors can then be assayed by immunoblotting or transcription complementation assays.

A disadvantage of SEC, however, is that large columns are generally required for high resolution since the sample size should be no more than a few percent of column volume.

3.2.2. Ion-exchange chromatography

Ion-exchange chromatography separates proteins on the basis of ionic charge; elution is typically with consecutive steps of progressively high concentrations of salt. Sometimes multiple steps of ion-exchange chromatography with different supports are used, with different ranges of KCl concentration for elution.

Ion-exchange columns can bind large amounts of protein (typically 20–30 mg/ml support), and this property often makes them desirable as an early step of purification. However, the disadvantage of IEC is many DNA-binding proteins elute from these supports under comparable conditions. Therefore, sometimes separation from other proteins is incomplete.

Phosphocellulose was one of the earliest ion-exchange support used for the separation of factors controlling transcription by RNA polymerase II from cultured human KB cells (KB S-100). KB S-100 extract is loaded onto a phosphocellulose column at 0.1 M KCl and then eluted with successive steps of higher salt concentration. Matsui and co-workers [73] isolated several fractions. Some of these fractions were inactive on their own but together recreated specific transcription. The fractions were TF IIA (0.1 M KCl eluate), TF IIB (0.35 M KCl), TF IIC (0.5 M KCl) and TF IID (1 M KCl). Reinberg and Roeder subsequently purified factors TF IIB, and TF IIE employing DEAE–cellulose and heparin–Sepharose [74].

The number of columns needed for purification of DNA-binding proteins prior to an affinity column is variable. Transcriptions factors like SP1, AP-1, and Ikaros required only one gel filtration or ion-exchange column prior to affinity chromatography [68,75,76]. However, other proteins including NF-κB required several [77]. In the case of NF-κB, its unusually low abundance may explain why additional purification steps were required to achieve sufficient purity.

3.2.2.1. Heparin. Heparin is a glycosaminoglycan consisting of alternating hexuronic acid (D-glucuronic acid or L-iduronic acid) and D-glucosamine residues. The polymer is heavily sulfated, carrying sulfamino (N-sulfate) groups at C-2 of the glucosamine units and ester sulfate (O-sulfate) groups in various positions. It is extracted from the native proteoglycan of intestinal mucosa. Heparin is linked or coupled to Sepharose by cyanogen bromide activation [78]. It is employed in one of the steps for purification of TFs. It is difficult to know sometimes what chromatographic mode heparin–Sepharose belongs to. The immobilized heparin acts as a cation exchanger in some cases due to its high content of anionic sulfate groups. However, heparin also mimics the similarly polyanionic structure of nucleic acids and may also function as a DNA analog; in this case, the chromatographic mode is affinity chromatography. Indeed, heparin- and DNA-binding to TFs is competitive as discussed later [79]. Salt gradient elution is most commonly used. Basal TFs and several DNA-binding proteins elute in the range of 0.2–0.5 M KCl (as shown in Table 3) and the proteins or factors inhibiting transcription do not bind to the column [80].

Several different types of DNA-binding proteins have been purified using heparin chromatography [81–85].

Besides using heparin–Sepharose, heparin is also used sometimes for eluting TFs from DNA-affinity columns [79] as heparin competes with the DNA for TF binding.

3.3. Affinity chromatography

Affinity chromatography is a method of separating and purifying proteins using the biochemical specificity towards their ligands. Here, we discuss various ligands that are used for purification of TFs.

3.3.1. Lectins

The rationale for conducting lectin affinity chromatography is that a number of RNA polymerase II TFs, in-

Table 3
Various salt concentrations used in heparin chromatography

Transcription factor	Source	Supports used	Elution at different salt concentrations	Reference
ERF1	MCF7 cell nuclear extract	Heparin–Sepharose	0.6 M KCl	[187]
TEF1	Rat kidney nuclear extract	Heparin–agarose	0.2–1.0, 0.4, 0.5 and 0.6 M KCl	[90]
NF1-L	Rat liver nuclear extract	Heparin–Sepharose	0.35 M KCl	[188]
UEF3	HeLa cell extract	Heparin–Sepharose	0.45 M KCl	[189]
GRIP170	HeLa S3 nuclear extract	Heparin–Sepharose	1.5 M KCl	[190]
NF-AT	Jurkat cell nuclear extract	Heparin–Sepharose	0.3 and 0.4 M KCl	[191]
H2TF1	HeLa cells	Heparin–agarose	0.1–1 M KCl	[192]
ASP	Mouse Y1 cell nuclear extract	Heparin–agarose	0.3 M KCl	[193]
TF IIIC	HeLa nuclear extract	Heparin–agarose	0.6 M KCl	[194]

cluding SP1, serum response factor and RNA polymerase II itself were shown to be glycosylated [86–89] and the *N*-acetylglucosamine sugar residues covalently attached to the protein are avidly bound by the lectin wheat germ agglutinin (WGA). TEF1, a TF that binds the human transforming growth factor-α promoter is also glycosylated and WGA affinity chromatography is used in one of the multiple steps of purification [90]. SP1 [87] and LR1 from B lymphocytes [91] have been purified using the combination of lectin affinity and DNA-affinity chromatography.

3.3.2. Procion dyes

Triazine dyes such as Cibacron Blue F3G-A and Procion Red dyes have been very successfully used in dye-ligand affinity chromatography for the purification of enzymes interacting with nucleotides and other heterocyclic compounds [92,93]. Cibacron Blue has been particularly suited for purification of NAD^+-dependent dehydrogenases whilst Procion Red is more selective for $NADP^+$-linked enzymes. Cibacron Blue chromatography has been useful too in purification of serum albumins [94], anti-DNA antibodies [95], DNA-binding proteins constituting chromatin [96], T7 RNA polymerase [97] and restriction enzymes [98]. TF IIIB purification from *S. cerevisiae* [99] utilizes Cibacron Blue F3GA–Sepharose in one of the multiple step of purification. Elution of proteins from Cibacron Blue and other Procion dyes is done by NaCl gradients or with sodium thiocyanate. Sodium or potassium thiocyanates are chaotropic agents and are necessary in some cases where the dye–protein interaction is too stable to be disrupted by gentler means. Dye-ligand affinity chromatography has not been widely used in purification of TFs.

3.3.3. DNA-affinity chromatography

For the purification of DNA-binding proteins, the DNA is either adsorbed or linked covalently to a chromatographic support and used for DNA-affinity chromatography. Table 4 shows some of the various categories of DNA-binding proteins purified by this method. This technique has been developing as more specific DNA sequences and more selective supports are used. Here, we discuss in detail the developments of the technique chronologically.

3.3.3.1. Heterogeneous sequence. DNA-affinity chromatography was originally developed with heterogeneous or non-specific DNA, such as fragmented calf thymus DNA or salmon- or herring-sperm DNA, attached to supports such as cellulose and eventually agarose. These non-specific DNA columns bind all DNA-binding proteins without much preference, usually offering little selectivity for any specific protein.

One of the earliest and most popular bioselective adsorbent containing nucleic acid used is DNA cellulose. DNA polymerase [100] and RNA polymerases [101] were purified by DNA adsorption to cellulose and are the first group of DNA-binding proteins to be purified. The DNA ligand here is either physically entrapped within the matrix of cellulose fibers or covalently cross-linked to it using ultraviolet light.

Polynucleotides have also been used as matrix bound ligands for affinity chromatography [102]. The most common polynucleotide affinity matrices are those containing poly dT or poly U [103–105] for isolation of eukaryotic mRNAs. These rely on complementary nucleotide interactions (i.e. annealing) between the matrix bound polymer and the 3′-terminal polyA sequence associated with eukaryotic mRNAs. A few years later, using poly(dG:dC), a matrix was prepared which contained the Z-form of DNA [106]. With this material, Z-DNA-binding proteins were isolated from a partially purified protein extract from *Drosophila* [107] or a mammalian tumor cell line [108]. Briefly, certain DNA-binding proteins like polymerases and histones, which do not have preference for a specific DNA sequence, can be purified using non-specific DNA columns. Whereas low abundance DNA-binding proteins having specificity towards a specific consensus sequence such as TFs are purified using specific DNA-affinity columns as discussed in the next section.

3.3.3.2. Homogenous sequence. TFs and restriction endonucleases are DNA-binding proteins that show a high affinity for a specific double-stranded DNA sequence; the affinity for any other DNA sequence is several orders of magnitude lower. This specific sequence is referred to as the 'footprint' region of TFs [109,110] or the restriction site in case of restriction endonucleases. The restriction site for var-

Table 4
List of various functions of proteins purified by DNA-affinity chromatography

Class of proteins	Binding affinity		Sequence specificity	Reference
	ssDNA	dsDNA		
Restriction endonucleases	−	+	High	[120,195]
Transcriptional enhancers/ promoter proteins	+	+	High	[68,69,87,118,119,127,196–212]
Transposition	−	+	High	[213,214]
Recombination	+	−	Low	[215]
DNA repair	+	−	Low	[216]
DNA replication	+	+	Low	[121,200,217–220]

ss: single-stranded; ds: means double-stranded; (−): not found; (+): found.

ious restriction endonucleases is well known. The footprint sequence bound by TFs can be determined by techniques such as DNAse I foot printing [89,111], methidiumpropyl EDTA–Fe(II) foot printing [112], and dimethyl sulfate methylation protection [113], and others, DNA-affinity chromatography utilizes this sequence linked to a suitable chromatographic supports such as Sepharose [68,114,115] or silica [116,117] for the purification of DNA-binding proteins. Since sequence-specific DNA-binding proteins are generally present in small amounts, and are difficult to purify to homogeneity by other modes of chromatography, DNA-affinity chromatography offers a good technique to isolate site-specific binding proteins from complex mixtures. Due to high selectivity of DNA binding and effective means for elution, high purification of proteins often results from a single step.

3.3.3.3. Discrete oligonucleotide. Specific DNA-affinity columns, using short, simple and discrete oligonucleotides, were the last to be developed for purifying sequence-specific DNA-binding protein. Several ways have been developed to link the sequence to the chromatographic support. In adenovirus major late TF (MLTF), the MLTF DNA-binding site was end-labeled with biotin, and protein was purified on the DNA bound to streptavidin agarose beads [118]. Purification of CAAT DNA-binding protein used an affinity column consisting of double-stranded specific oligonucleotide sequence covalently linked through a 10-nucleotide long 5′ overhang on one DNA strand coupled to CNBr-activated Sepharose-4B [119]; single-stranded DNA sequence regions couple more efficiently. The restriction enzyme, *Eco*RI, has also been purified using palindromic sequence of the enzyme coupled through a 5′ spacer arm with terminal thiol group [120].

A further progressive step of using discrete sequences in DNA-affinity chromatography was while purifying nuclear factor I (NF1) from S-3 HeLa cell nuclear extract. Here the affinity matrix was prepared using plasmid DNA that contains 88 copies of NF1 binding site from the adenovirus origin of replication [121]. This plasmid was constructed by means of a novel cloning strategy that generated concatenated NF1 binding sites arranged exclusively in the direct head to tail orientation. This method paved the way to the development of using of concatemers linked to the chromatographic support.

3.3.3.4. Concatemers. DNA concatemers have been widely used in the purification of TFs, some of which are shown in Table 5. Concatemers are tandem repeats of DNA sequences, ligated together. DNA columns with concatemers, free from extraneous plasmid sequences were first used for purification of SP1 TF [122].

Concatemers are made by ligation of single copy of DNA strands containing the footprint sequence and flanking overhangs. The overhangs help in ligating one copy to another and thus multimers of identical sequences are formed (Fig. 1). Some studies suggest that longer DNA sequences may function better than simple, discrete and short footprints for TF purification [68,69]. However, there are some disadvantages to the concatemeric method. Longer sequences in concatemers means additional sequences, which in this case are the inter-footprint sequences produced by ligation. These sequences could have potential sites for other DNA-binding proteins present in the crude extract. Binding of other proteins to these sites could block the binding site for specific proteins, thereby decreasing the yield and purity of specific DNA-binding protein. Sometimes ligation has to be repeated several times to obtain long concatemers. This is not only time consuming but also leads to loss of oligonucleotides. Also, ligation of footprints can result in circular DNAs, which may couple ineffectively or bind other proteins. Ligation also makes DNAs of different lengths, which can result in heterogeneous stationary phase.

Recently, monomeric, concatemeric and poly(A): poly(T) tailed DNA columns were compared. The study suggested that there is no distinct advantage in using columns made by concatemeric sequences and better purity can be obtained by using short, discrete oligonucleotides. For proteins which have lower affinity for discrete footprints, DNA extended with poly(A):poly(T) tail gives better resolution of proteins than either discrete or concatemeric columns [34].

Hence, sometimes columns having just the footprint region or the footprint region extended with a simple DNA sequence could be a better option for purification of TFs.

Table 5
DNA-affinity chromatography (using concatemers)

Transcription factor	Source	Salt concentration for elution	Competitor DNA	Reference
NRE	HeLa cells	1 M KCl	Salmon sperm DNA	[221]
BBF	Fibroblasts and HeLa cells	0.25, 0.6 and 1 M NaCl	–	[222]
Leu3	*Saccharomyces cerevisiae*	1 M NaCl	Salmon sperm DNA	[223]
NAF	Porcine intestinal	0.15–0.5 M KCl	Sheared salmon sperm DNA	[224]
43 kDa pets factor	Bovine spleens	1 M KCl	Calf thymus DNA	[225]
ASP	Mouse Y1 cell nuclear extract	0.5 and 1 M KCl	Poly(dI:dC), salmon sperm DNA, tRNA and GC box sequence	[193]
TF IIIA	HeLa cells	0.17–0.33 M $(NH_4)_2SO_4$	–	[226]
PCF	HrpG2 nuclear extract			[227]
TBPF	*Acanthamoeba* nuclear extract	0.4 M KCl	–	[228]
NF1-L	Rat liver nuclear extract	1.5 and 3.0 M KCl	Poly(dI:dC)	[188]
UEF3	HeLa cell extract	1 M KCl	Poly(dI:dC)	[189]
ERF1	MCF7 cell nuclear extract	0.8 M KCl	Mutant oligonucleotide	[187]
TEF1	Rat kidney nuclear extract	0.2–1.0 M KCl	Poly(dI:dC)	[90]

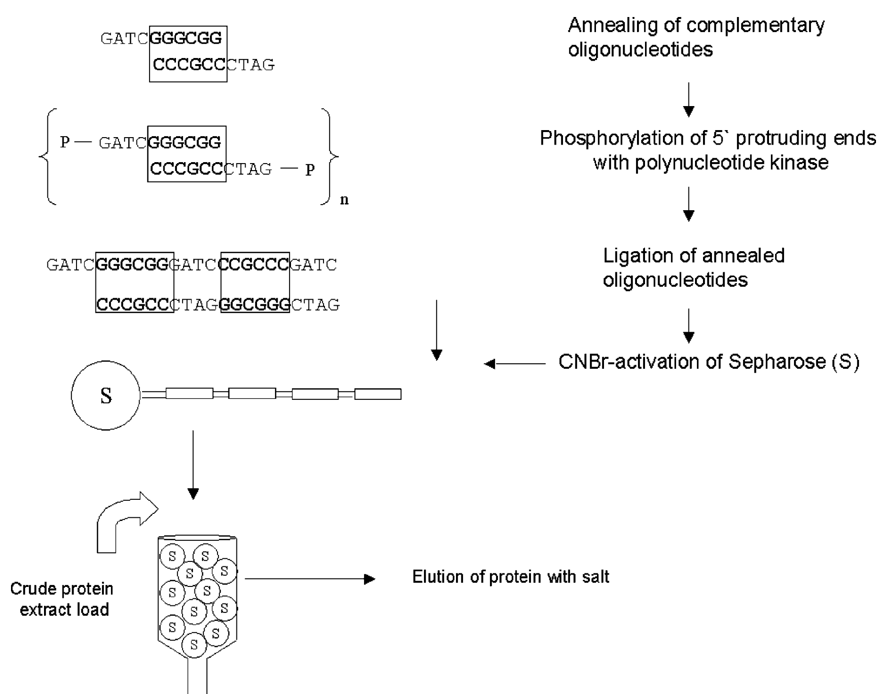

Fig. 1. Purification of sequence-specific DNA-binding proteins with DNA-affinity chromatography using concatemers of DNA sequence. The encircled "S" represents the support, in this case Sepharose.

4. Recent developments in DNA-affinity chromatography

DNA-affinity chromatography offers the highest selectivity of any chromatographic method and hence is widely used in purification of DNA-binding proteins. Here we discuss new techniques developed over the past few years. These developments could help in purification of DNA-binding proteins in a lesser number of steps and reduce the efforts required significantly.

4.1. Oligonucleotide trapping

This is a method in which a double-stranded DNA is constructed containing the footprint of a TF with a single-stranded $(TG)_5$ tail. The method involves incubation of the DNA strands at a very low concentration (as low as 50 nM) with crude protein extract containing TF for 30 min on ice. The DNA–protein complex forms in solution and is then trapped by loading this mixture onto a column with decameric oligonucleotide $(AC)_5$ attached to the support.

At 4 °C, the $(TG)_5$ region anneals with the column $(AC)_5$, as the sample DNA–protein complex passes through the column. Elution of the TF can then be achieved using high salt, which disrupt the DNA–protein interactions or by using moderate temperatures (25–37 °C) and low salt concentration (e.g. water) to melt the $(AC)_5$ and $(TG)_5$ hybrid and elute the DNA–protein complex. It was observed that when high salt is used, the TF free of DNA is eluted and when temperature/low salt is used, the TF bound to the oligonucleotides is eluted. Chimeric green fluorescent protein-CAAT enhancer binding protein (GFP-C/EBP) and TF B3 from *Xenopus* eggs have been highly purified with the trapping method [123]. Another related approach is to use biotinylated DNA and (strept)avidin columns to perform TF trapping [123]. When the two (biotin and tailed DNA) methods were compared, the tailed DNA method gave higher purity.

The only major drawback to the method is that the high concentration of DNA on the column can encourage non-specific binding of proteins that have a low affinity for DNA. Addition of several competitors such as heparin, mutant DNA footprints (μE3), non-specific DNA (fragmented salmon sperm DNA) and single-stranded DNA (T_{18}) can be added to the sample during trapping to improve purity.

4.2. Catalytic chromatography

The principle of the method is exploitation of both specific biological affinity and catalytic specificity to selectively purify enzymes. *Eco*RI restriction endonuclease and DNA polymerase of *E. coli* have been purified using this method [124]. In this chromatographic mode, the enzyme binds immobilized substrate coupled to a column support in the absence of a required cofactor. With the addition of missing cofactor (e.g. Mg^{2+} for *Eco*RI or dNTP for DNA polymerases) the enzyme converts substrate to product and selectively elutes from the column (Fig. 2). The advantage of this method is high resolution but the disadvantage is short column life. This is because the elution by catalysis alters the column and does not allow the column to be reused. A further step of development of this chromatographic mode was reusing the same column by coupling short adapter sequences to the chromatographic support (as discussed in the Section 4.1). In this approach, the substrate DNA is constructed with a $(GT)_5$ tail and annealed to the $(AC)_5$ column. After conversion to product, the product DNA can be stripped from the column by elevated temperature and low salt concentration and re-charged with fresh DNA. Catalytic chromatography is only useful for DNA-binding enzymes and is not generally applicable to TFs.

4.3. Bi-column method

In this method, two DNA-affinity columns are used—one a simple DNA column with the minimal footprint sequence of the TF, and the other a complex DNA column containing the footprint sequence with an additional homopolymeric tail (T_{18}:A_{18} tail) (Fig. 3). The method relies on the discovery that TFs elute from the simple DNA columns at lower heparin concentrations in the mobile phase than when more complex DNAs are used. Briefly, the crude extract containing the TF is first loaded onto the simpler DNA column; the column is thoroughly washed to remove any unbound protein and then the more complex column is connected to the outlet of the simple column. The TF from first column is eluted with a combination of low salt and low heparin concentration. At this low heparin concentration, the TFs elute from the first (simple) column but will still bind to the second (complex) column. The columns are then disconnected and the second, complex column is eluted with high salt resulting in highly purified TF devoid of heparin. This method has been useful for purification of *lac* repressor, C/EBP and *Xenopus* B3 TF [125]. This method can be used for the purification of any TF as long as the DNA footprint of the TF is known and two high specificity columns having different affinities for the TF can be generated.

4.4. Methods for selective elution of DNA columns

4.4.1. Salt elution

TFs on binding to DNA, displaces Na^+ and other counter ions from the DNA [126]. Thus, high Na^+ concentration diminishes the DNA–protein interaction. This displacement is common to all TFs and, thus salt gradient elution has only limited selectivity. Salt elution is the most common method for eluting proteins from DNA-affinity columns. A step gradient or a linear gradient with increasing salt concentration is normally employed; NaCl and KCl are the most common salts used. A linear salt gradient helps in separation of different proteins. Proteins that bind to the DNA with lower affinity elute at lower salt concentration. The protein of interest usually has the highest affinity for the specific coupled DNA and elute at high salt or later in the gradient. Salt elution has only moderate resolution but if combined with the property of temperature, it improves the purification (as discussed in detail in Section 4.4.3).

4.4.2. Ligand-specific elution

Certain TFs have specificity towards a specific ligand and when bound to it, lose the DNA-binding activity. For example the *lac* repressor proteins is unable to bind to DNA when it is bound by ligands such as IPTG or *lac*tose. In such cases, the TFs can be eluted from DNA-affinity columns using the specific ligands. Leblond-Francillard and co-workers achieved several fold enrichment of *lac* repressor complexed to operator using this approach [127].

Ligand-specific elution is highly selective and can yield highly purified protein. But it has not been so widely used since only a small number of TFs respond to specific ligands.

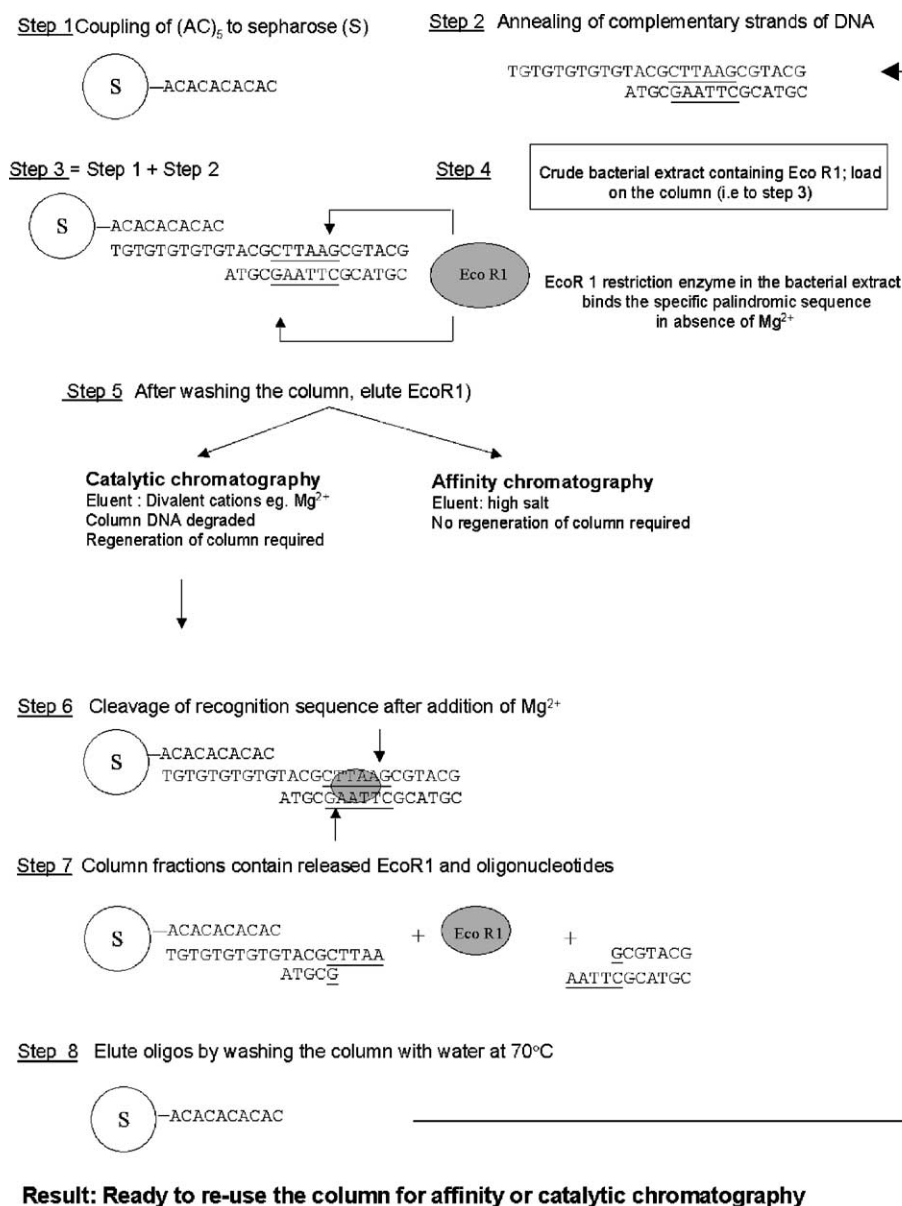

Fig. 2. Diagrammatic representation of catalytic chromatography with adapter approach. Shown are the steps required to perform purification of *Eco*RI.

As discussed earlier, the polyanionic structure of heparin resembles the negatively charged backbone of DNA. Many DNA-binding proteins are thought to bind heparin because of this resemblance. Hence, heparin can be used as a competitor to elute DNA-binding proteins from DNA-affinity columns. TFs C/EBP and *lac* repressor are eluted by heparin from their respective DNA-affinity columns [79]. The elution of these proteins is highly dependent on the nature and concentration of DNA on the column and the concentration of heparin in the mobile phase. Heparin can thus be used as a general ligand where no specific one is known.

4.4.3. Temperature

Protein–DNA interactions are generally temperature dependent [128–135]. This is usually true and is not surprising. Hydrophobic interactions are important to both DNA and protein structure. Both the base stacking within DNA double helix and the clustering of hydrophobic amino acids within the core of a protein are well appreciated. Also, to interact with DNA, amino acid side chains insert into the hydrophobic environment of the DNA major groove or, less frequently the minor groove. This is clearly seen in the structure of DNA–protein complexes that have been resolved, e.g. the yeast TF GAL4–DNA complex [136]. The interactions of DNA with proteins thus involve hydrophobic forces and because these forces derive from ordering of solvent water [137], this entropic component to binding is temperature dependent. Furthermore, the binding of TFs such as *lac*tose repressor (*lac* repressor) disp*lac*es counter cations from the polyanionic DNA and this may also account for a large in-

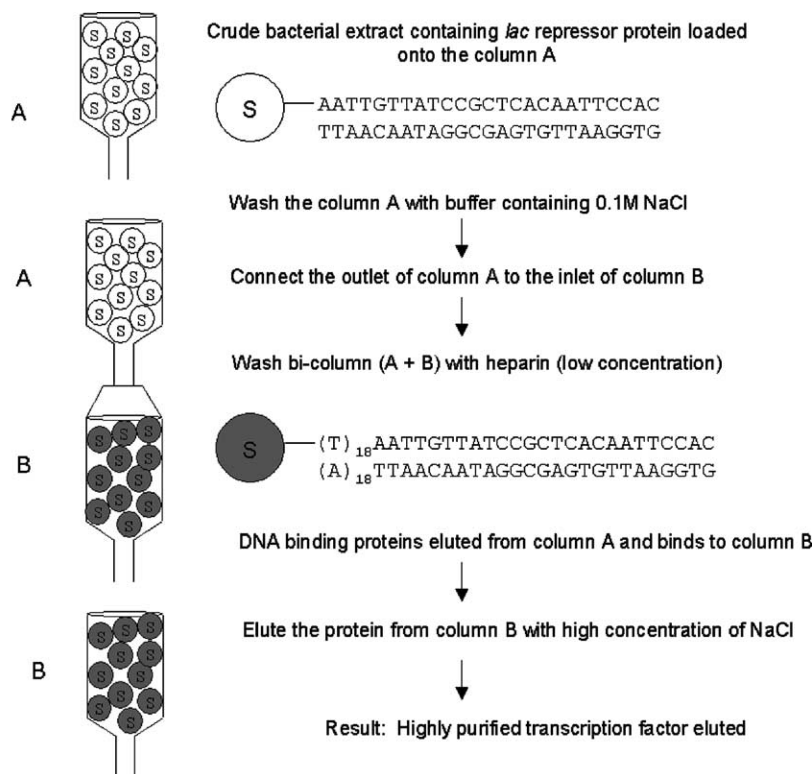

Fig. 3. Schematic diagram of purification of *E. coli lac* repressor protein by the bi-column method. Two different shades of "S" represents two different oligonucleotides coupled to the Sepharose.

crease in entropy [138]. Finally the heat capacity of the complex of the TFs such as *lac* or Cro repressor with DNA is markedly less than that of the individual components presumably because binding induces new conformations in the protein, DNA or both and this also makes binding strongly temperature dependent [128,134,135].

Recently, Jarrett and Taylor showed the effect of temperature on the elution of a fusion protein containing CAAT enhancer binding protein (C/EBP) sequences from DNA-Sepharose [139]. It was found that DNA-affinity chromatography gave higher purity than Ni^{2+}-agarose chromatography and chromatography on the same DNA-Sepharose column at two different temperatures resulted in the greatest purification. Further, when a temperature range (from 4 to 60 °C) was studied the salt concentration for elution decreased about two-fold.

In a study of three different TFs by Jarrett [140], the chromatography of all three was found to be temperature dependent and each had a distinctive behavior. When each protein was eluted from its specific DNA column with a salt gradient, the salt concentration at which elution occurs is highly temperature dependent. As temperature increased (in the range of 4–35 °C), less salt was required to elute C/EBP, more salt was required to elute *lac* repressor, while GAL4 showed a biphasic response with the amount of salt decreasing between 4 and 19 °C and then increasing above 19 °C. This temperature dependence is not due to protein or DNA unfolding but rather is a property of complex formation.

Besides the temperature, it was found that the more complex the DNA on the column, the higher the salt concentration required for elution. Thus, not only the temperature but DNA complexity also affects the elution pattern.

The role of temperature for elution of DNA-binding proteins and their purity is important in two ways. First, at a particular temperature, many different DNA-binding proteins present in a cell extract bind to the high concentration of DNA present in a typical column. Eluting such a column with salt at a different temperature increases the chances of highly purifying the desired TF because of its characteristic temperature-salt dependence. Second, in the "temperature jump" elution method [140], an abrupt temperature change causes elution. This latter method depends upon the peculiar temperature dependence of single TF–DNA complex and results in high purification.

Hence, temperature, well below what causes denaturation of either proteins or DNA, can also be used for effective elution. This temperature-dependent affinity chromatography provides an important new approach to TF purification.

5. Electromigration methods: TF assays

The identification and characterization of many distinct, dissimilar TFs within several families such as SP1, C/EBP, and AP-1, have been described [122,141,142]. Some char-

acterized TFs within these families are distinct but similar, for example SP1, SP2, SP3, and SP4 [143]. However, other TFs have not been fully characterized, and many more have not yet been identified. The first step of characterization is to purify the protein to adequate levels for sequencing and cloning. A prerequisite for the purification of any protein is a highly sensitive and specific assay for detection and identification. Clearly, having good detection methods is important for identifying TFs. TFs typically bind to their DNA element with affinities in the picomolar range, and bind to non-cognate DNA sequences with 10^3–10^5 lower affinity. For example, *lac* repressor binds the Op1 operator with a $K_d = 5 \times 10^{-14}$ M and to poly(AT) with a $K_d = 10^{-9}$ M [144], and Cro repressor binds its element with $K_d = 6 \times 10^{-13}$ M and to a non-cognate DNA with $K_d = 4 \times 10^{-8}$ M [135]. The high binding affinity of TFs has been utilized for the identification and characterization of TFs in the techniques described below.

5.1. Electrophoretic mobility shift assay (EMSA)

The gel shift assay, or electrophoretic mobility shift assay, provides a simple and rapid method for detecting TFs and other DNA-binding proteins [145,146]. The assay is based on the separation of protein–oligonucleotide complexes from single- and double-stranded oligonucleotides migrating through non-denaturing polyacrylamide gels. Protein–oligonucleotide complexes migrate more slowly than single- and double-stranded oligonucleotides; thus, when the TF binds the oligonucleotide, the complex is shifted upward in the gel and separated from the unbound oligonucleotides. The gel shift assay is performed by incubating a purified protein, or a complex mixture of proteins, with a labeled oligonucleotide containing the putative binding site. Gel shift assays typically use ^{32}P-labeled oligonucleotide probes but alternative labels are available including biotin, digoxygenin, and fluorochromes [147–149]. Since many DNA-binding proteins, including TFs, can bind DNA non-specifically, some unlabeled competitors may be used to minimize non-specific binding. Such competitors include poly(dI:dC) [68], single-stranded oligonucleotides (for example, dT_{18}), fragmented salmon sperm DNA, and heparin [123] (see Fig. 4). Further, to enhance DNA–protein interactions, bovine serum albumin or other proteins and non-ionic detergents may be included in the incubation mixture [150–153]. The incubation mixture is then loaded on a non-denaturing polyacrylamide gel and analyzed. The specificity of the DNA-binding protein for the labeled oligonucleotide is established by competition assays using unlabeled oligonucleotides containing the putative binding site for the protein of interest or other unrelated DNA sequences. Another test of specificity is the super shift assay. Binding a specific antibody to the TF of interest will result in shifting the protein–DNA band further upward on the gel, forming a super shifted band. Some techniques similar to EMSA have been developed including a form of high resolution gel filtration for the separation of protein–DNA complexes [154], and affinity coelectrophoresis (ACE) for determination of TF binding constants [155]. The major advantages of EMSA are high specificity, ease of operation, and multiple samples can be assayed simultaneously. Some disadvantages are relatively long assay time, and difficulty in kinetic parameter determination.

5.2. Capillary electrophoretic mobility shift assay (CEMSA)

Capillary electrophoresis (CE) [146], has become an important method for the study of biomolecular interactions. Capillary electrophoresis is a technique in which an electrophoretic separation takes place in a narrow capillary filled with ionic buffer or a gel matrix, usually polyacrylamide. Typical analysis of proteins by CE include mass determination [156,157], enzyme micro assays [158–160], separation of antigen–antibody complexes [161,162], and peptide mapping [163]. In addition to these wide applications, CE has been demonstrated to separate DNA–protein complexes from complex mixtures of DNA and protein with high resolution [164]. As in EMSA (see Section 5.1), the binding of protein to DNA results in a band shift. The resulting assay, CEMSA [165], combines EMSA and the high resolution, high speed, and full automation of CE. The CEMSA assay is performed by incubating a mixture containing purified protein, or a complex mixture of proteins, with a labeled oligonucleotide or other DNA containing the putative binding site. Fluorochrome labeled DNA is typically used for detection. Following incubation, the mixture is injected at one end of the capillary, either by electrokinesis or by osmosis, and an electric field is applied across the capillary. As the mixture migrates through the capillary due to the applied electric field, differing electrophoretic mobilities result in separation of each component into discrete bands based on charge to mass ratio. The separated bands are detected and quantified by laser-induced fluorescence (LIF) detection. CEMSA has many advantages, for example, high speed, high resolution, small sample quantity, reusability of capillaries, and the ability to automate. The major disadvantages are cost of equipment, only one sample can be run at a time, and samples are not used after electrophoresis for other assays.

5.3. Southwestern blot

Southwestern blotting was first used in screening complementary DNA libraries constructed in bacteriophage expression vectors for the detection of DNA-binding proteins [166,167]. The southwestern technique is a powerful tool for identifying and characterizing DNA-binding proteins by their ability to bind labeled oligonucleotide probes containing the putative binding site [168]. The southwestern blot assay is performed by separating whole cell extracts, nu-

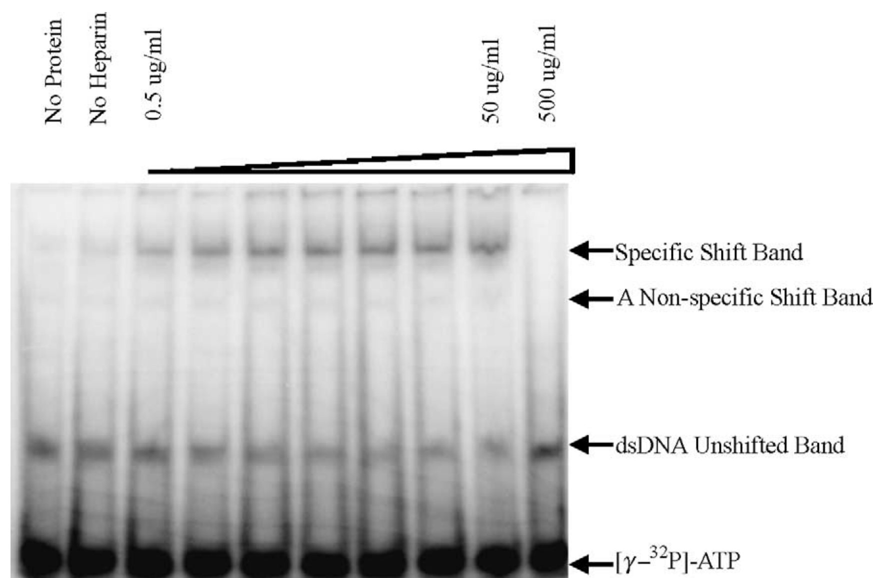

Fig. 4. Electrophoretic mobility shift assay. An electrophoretic mobility gel shift assay was performed in order to determine the maximum heparin concentration that does not interfere with the specific shift band. A BL21 (DE3) bacteria extract of induced GFP-C/EBP (13) was incubated with ^{32}P-labeled, annealed oligonucleotide, ACEP24 (5′-GCTGCAG **ATTG** CG **CAAT** CTGC AGC-3′), containing the C/EBP putative binding site (in bold) (13), poly(dI:dC), incubation buffer, and variable heparin concentrations. The highest heparin concentration that did not effect the specific shift band was 50 μg/ml. Also, this figure demonstrates, by comparing the lane with no heparin to heparin containing lanes, that heparin aids in specific protein–DNA interaction. But, at high concentrations, heparin interferes with GFP-C/EBP-ACEP24 interaction.

clear extracts, or purified proteins by SDS-polyacrylamide gel electrophoresis (PAGE) and then transferring to nitrocellulose for screening with oligonucleotide probes. Protein adsorbed onto nitrocellulose filters is probed with labeled, double-stranded oligonucleotide. Two procedures greatly increase the level of binding between oligonucleotide and protein [169]. First, nitrocellulose filters are processed through a denaturation/renaturation regimen using 6 M guanidine hydrochloride. Second, oligonucleotide is catenated extensively using DNA ligase, (see Section 3.3.3.4). Alternatively, double-stranded oligonucleotides may be labeled by using reverse transcriptase or DNA polymerase Klenow large fragment and α-^{32}P-nucleotides. The combination of the denaturation/renaturation, catenation or extension procedures leads to stronger detection signals. The specificity of the DNA-binding protein for the labeled oligonucleotide is established by competition assays using unlabeled oligonucleotides containing the putative binding site for the protein of interest or other unrelated DNA sequences. Radioactive signals are detected by autoradiography. A major advantages of using the southwestern blot assay is that specific DNA-binding signals can be detected on duplicate filters, filters can be washed and reused by repeating the cycle of denaturation/renaturation, and the relative molecular mass of specific DNA-binding signals can be determined using relative molecular mass markers. The usefulness of this technique is limited by problems with protein degradation during isolation, relatively long assay times, and difficulties in achieving efficient electrophoretic separation and transfer of a wide molecular size range of proteins.

5.4. 2D SDS-PAGE and sequencing/proteomics (renaturation of 2D gels)

Two-dimensional polyacrylamide gel electrophoresis separates proteins by displacement in two dimensions oriented orthagonally to one another [170,171]. Separation occurs by charge, using isoelectric focusing (IEF) in the first dimension, and by size, using SDS-PAGE in the second dimension [172]. High resolution 2D SDS-PAGE is considered the method with the highest resolution for the separation of complex protein mixtures permitting the simultaneous analysis of hundreds or even thousands of proteins. Further, 2D SDS-PAGE is used to detect alterations in gene expression using qualitative and quantitative comparisons and is used widely as an essential aspect of modern proteomics [171,173]. Although, an in depth discussion of 2D SDS-PAGE is beyond the scope of this review, more detailed reviews can be found [174–177]. The 2D SDS-PAGE is performed by loading purified protein, or a complex mixture of proteins onto IEF strips by in-gel rehydration or cup-loading for the first dimension [177]. Modern IEF has been greatly advanced with the introduction of immobilized pH gradients (IPG) which increase pH gradient stability and reproducibility [178] (see Fig. 5). IPG is based on the principle that the pH gradient is generated by 'immobilines' which are co-polymerized with the polyacrylamide matrix. IPG allows the generation of pH gradients of any desired range between pH 3 and 12 [179]. This large gradient range increases the resolution of each sample component allowing for an overview of patterns over a large range of pH or a

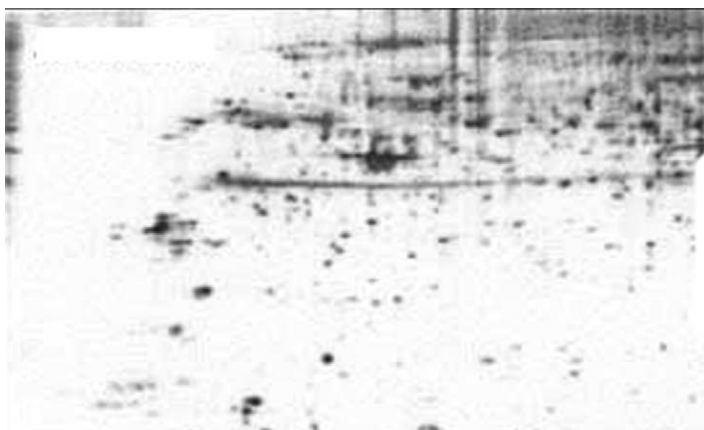

Fig. 5. 2D SDS-PAGE visualization. HeLa cells were lysed and approximately 30 μg of protein was desalted. The lysate was focused on a pH 4–7 IEP strip followed by 8–16% SDS-PAGE. The 2D SDS-PAGE gel was silver-stained. Figure borrowed from http://www.piercenet.com/products/browse.cfm?fldID=072BEB1B-0DA5-4FB9-BF9C-2F6EDBA8D4C8 with permission from Pierce Chemical Co.

narrow high-resolution fractionation of only a subset of isoelectric points. The second dimension of separation by 2D SDS-PAGE is performed most frequently using polyacrylamide gradient gels, resulting in spot formation of resolved proteins. Protein detection is performed by radiolabeling proteins, silver, copper, or zinc staining, or fluorescence [180]. Suspect spots are removed from the gel for further analysis. A very important method of facilitating protein identification in proteomics, following spot removal, is subjecting the protein to analysis by mass spectrometry. While many methods exist, only a few are commonly used in proteomics in conjunction with 2D SDS-PAGE spot removal, for example ES/FAB, QMS/QMS/QMS, MALDI/TOF, ES/TOF [181,182]. However, MALDI/TOF is considered the technique of choice in proteomics [183,184]. Analysis may involve exopeptidase/endopeptidase, or chemical degradation prior to mass spectral analysis [185,186]. Depending on the technique, liquid chromatographic (LC) fractionation may be used to fractionate the peptides. Undigested protein can also be analyzed. Proteins are identified based on fragment MW or sequencing data or both. Due to high costs of instrumentation, the mass spectrometric method of choice is likely dependent on readily available equipment. Another powerful method used in conjunction with 2D SDS-PAGE for the identification of DNA-binding proteins is spot removal followed by EMSA analysis (see Section 5.1). The 2D SDS-PAGE assay is highly desirable because of the potentially powerful resolving ability and has become a fundamental tool in proteomics analysis. Some disadvantages of this technique include technical difficulty in achieving high resolution, and available staining procedures are limited in sensitivity.

6. Conclusion

The chromatography and electrophoresis of TFs has been evolving. Beginning with fragmented nuclear DNA adsorbed to cellulose, the DNA-affinity chromatography has progressed to the point of using specific oligonucleotide sequences and the supports and chromatographs have greatly improved. Mobile phases have progressed from simple buffered salt solutions to more recent mixtures containing competitor DNAs, detergents, and other additives to diminish non-specific binding. New elution strategies have added further to selectivity. Electrophoretic methods in 2D or with capillaries and much more sensitive and selective detection have also improved analysis and characterization. Much further improvement can be expected over the coming years.

By analogy, much remains to be done. For example, antibodies are now routinely purified by protein A (or G)-Sepharose chromatography or using immobilized antigen columns. Dehydrogenases are routinely purified on Procion dye columns. These strategies are well known and so widely used that few would attempt a new purification without relying on these standard approaches. TF purification may one day reach this level of standardization but it has yet to do so.

Acknowledgements

This work was supported by the United States Public Health Service, National Institutes of Health, GM43609.

References

[1] P.Y. Chou, G.D. Fasman, Biochemistry 13 (1974) 211.
[2] M. Dayhoff, Atlas of Protein Sequence and Structure, The National Biomedical Research Foundation, Silver Spring, MD, 1972.
[3] C.O. Pabo, R.T. Sauer, Annu. Rev. Biochem. 61 (1992) 1053.
[4] S.C. Harrison, Nature 353 (1991) 715.
[5] S.C. Harrison, A.K. Aggarwal, Annu. Rev. Biochem. 59 (1990) 933.
[6] J.H. Laity, B.M. Lee, P.E. Wright, Curr. Opin. Struct. Biol. 11 (2001) 39.
[7] P. Schjerling, S. Holmberg, Nucleic Acids Res. 24 (1996) 4599.
[8] W.H. Landschulz, P.F. Johnson, S.L. McKnight, Science 240 (1988) 1759.

[9] M.E. Massari, C. Murre, Mol. Cell Biol. 20 (2000) 429.
[10] D.H. Ohlendorf, W.F. Anderson, R.G. Fisher, Y. Takeda, B.W. Matthews, Nature 298 (1982) 718.
[11] D.B. McKay, T.A. Steitz, Nature 290 (1981) 744.
[12] C.O. Pabo, M. Lewis, Nature 298 (1982) 443.
[13] C.O. Pabo, R.T. Sauer, Annu. Rev. Biochem. 53 (1984) 293.
[14] R. Wintjens, M. Rooman, J. Mol. Biol. 262 (1996) 294.
[15] C. Branden, J. Tooze, Introduction to Protein Structure, Garland, New York, 1991.
[16] M. Lewis, G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan, P. Lu, Science 271 (1996) 1247.
[17] N.S. Boutonnet, M.J. Rooman, M.E. Ochagavia, J. Richelle, S.J. Wodak, Protein Eng. 8 (1995) 647.
[18] A. Draganescu, J.R. Levin, T.D. Tullius, J. Mol. Biol. 250 (1995) 595.
[19] F.F. Damberger, J.G. Pelton, C. Liu, H. Cho, C.J. Harrison, H.C. Nelson, D.E. Wemmer, J. Mol. Biol. 254 (1995) 704.
[20] J. Miller, A.D. McLachlan, A. Klug, EMBO J. 4 (1985) 1609.
[21] V.A. Narayan, R.W. Kriwacki, J.P. Caradonna, J. Biol. Chem. 272 (1997) 7801.
[22] H. Berglund, H. Kovacs, K. Dahlman-Wright, J.A. Gustafsson, T. Hard, Biochemistry 31 (1992) 12001.
[23] N.P. Pavletich, C.O. Pabo, Science 252 (1991) 809.
[24] J.A. Lowry, W.R. Atchley, J. Mol. E 50 (2000) 103.
[25] Y. Oshima, in: J. Strathern, E. Jones, J.K. Broach (Eds.), Molecular Biology of the Yeast Saccharomyces. Vol. 11, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1982, p. 159.
[26] T.L. South, M.F. Summers, Protein Sci. 2 (1993) 3.
[27] R.B. Todd, A. Andrianopoulos, Fungal Genet. Biol. 21 (1997) 388.
[28] T. Pan, J.E. Coleman, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 2077.
[29] M.M. Witte, R.C. Dickson, Mol. Cell Biol. 8 (1988) 3726.
[30] R. Marmorstein, S.C. Harrison, Genes Dev. 8 (1994) 2504.
[31] K.M. Zhou, G.B. Kohlhaw, J. Biol. Chem. 265 (1990) 17409.
[32] K. Pfeifer, K.S. Kim, S. Kogan, L. Guarente, Cell 56 (1989) 291.
[33] R.J. Reece, M. Ptashne, Science 261 (1993) 909.
[34] H. Gadgil, L.A. Jurado, H.W. Jarrett, Anal. Biochem. 290 (2000) 147.
[35] W.D. Kohn, C.T. Mant, R.S. Hodges, J. Biol. Chem. 272 (1997) 2583.
[36] E.K. O'Shea, R. Rutkowski, P.S. Kim, Science 243 (1989) 538.
[37] T.G. Oas, L.P. McIntosh, E.K. O'Shea, F.W. Dahlquist, P.S. Kim, Biochemistry 29 (1990) 2891.
[38] N.S. Foulkes, E. Borrelli, P. Sassone-Corsi, Cell 64 (1991) 739.
[39] F.J.R. Rauscher, D.R. Cohen, T. Curran, T.J. Bos, P.K. Vogt, D. Bohmann, R. Tjian, B.R.J. Franza, Science 240 (1988) 1010.
[40] C. Murre, P.S. McCaw, H. Vaessin, M. Caudy, L.Y. Jan, Y.N. Jan, C.V. Cabrera, J.N. Buskin, S.D. Hauschka, A.B. Lassar, Cell 58 (1989) 537.
[41] W.R. Atchley, W.M. Fitch, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 5172.
[42] G. Berben, M. Legrain, V. Gilliquet, F. Hilger, Yeast 6 (1990) 451.
[43] D.K. Hoshizaki, J.E. Hill, S.A. Henry, J. Biol. Chem. 265 (1990) 4736.
[44] P. Henthorn, M. Kiledjian, T. Kadesch, Science 247 (1990) 467.
[45] G. Bain, E.C. Maandag, D.J. Izon, D. Amsen, A.M. Kruisbeek, B.C. Weintraub, I. Krop, M.S. Schlissel, A.J. Feeney, M. van Roon, Cell 79 (1994) 885.
[46] T.C. Lee, R.J. Schwartz, Nucleic Acids Res. 20 (1992) 140.
[47] C. Porcher, W. Swat, K. Rockwell, Y. Fujiwara, F.W. Alt, S.H. Orkin, Cell 86 (1996) 47.
[48] C. Murre, G. Bain, M.A. van Dijk, I. Engel, B.A. Furnari, M.E. Massari, J.R. Matthews, M.W. Quong, R.R. Rivera, M.H. Stuiver, Biochim. Biophys. Acta 1218 (1994) 129.
[49] C. Murre, D. Baltimore, The Helix-Loop-Helix Motif: Structure and Function, Cold Spring Harbor Laboratory Press, Plainview, NY, 1992.
[50] P.S. Henthorn, C.C. Stewart, T. Kadesch, J.M. Puck, Genomics 11 (1991) 374.
[51] G.Q. Zhao, Q. Zhao, X. Zhou, M.G. Mattei, B. de Crombrugghe, Mol. Cell Biol. 13 (1993) 4505.
[52] E.M. Blackwood, R.N. Eisenman, Science 251 (1991) 1211.
[53] R. Benezra, R.L. Davis, D. Lockshon, D.L. Turner, H. Weintraub, Cell 61 (1990) 49.
[54] H.M. Ellis, D.R. Spann, J.W. Posakony, Cell 61 (1990) 27.
[55] C. Klambt, E. Knust, K. Tietze, J.A. Campos-Ortega, EMBO J. 8 (1989) 203.
[56] I. Sanchez-Garcia, T.H. Rabbitts, Semin. Cancer Biol. 4 (1993) 349.
[57] S.E. Phillips, Annu. Rev. Biophys. Biomol. Struct. 23 (1994) 671.
[58] N. Li, E.A. Sickmier, R. Zhang, A. Joachimiak, S.W. White, Mol. Microbiol. 43 (2002) 1079.
[59] R.W. Carthew, L.A. Chodosh, P.A. Sharp, Cell 43 (1985) 439.
[60] T.M. Kristie, J.H. LeBowitz, P.A. Sharp, EMBO J. (1989) 4229.
[61] M. Sawadogo, J. Biol. Chem. 263 (1988) 11994.
[62] A.R. Krainer, T. Maniatis, B. Ruskin, M.R. Green, Cell 36 (1984) 993.
[63] J.D. Dignam, R.M. Lebovitz, R.G. Roeder, Nucleic Acids Res. 11 (1983) 1475.
[64] K. Gorski, M. Carneiro, U. Schibler, Cell 47 (1986) 767.
[65] R.C. Conaway, D. Reines, K.P. Garrett, W. Powell, J.W. Conaway, Methods Enzymol. 273 (1996) 194.
[66] K.J. Stuempfle, M. Koptides, A.M. Karinch, J. Floros, Biotechniques 21 (1996) 48.
[67] J.L. Manley, A. Fire, A. Cano, P.A. Sharp, M.L. Gefter, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 3855.
[68] J.T. Kadonaga, R. Tjian, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 5889.
[69] J.T. Kadonaga, Methods Enzymol. 208 (1991) 10.
[70] B. Choy, M.R. Green, Nature 366 (1993) 531.
[71] S.G. Roberts, B. Choy, S.S. Walker, Y.S. Lin, M.R. Green, Curr. Biol. 5 (1995) 508.
[72] M.F. Carey, S.P. Gerrard, N.R. Cozzarelli, J. Biol. Chem. 261 (1986) 4309.
[73] T. Matsui, J. Segall, P.A. Weil, R.G. Roeder, J. Biol. Chem. 255 (1980) 11992.
[74] D. Reinberg, R.G. Roeder, J. Biol. Chem. 262 (1987) 3310.
[75] W. Lee, P. Mitchell, R. Tjian, Cell 49 (1987) 741.
[76] K. Hahm, P. Ernst, K. Lo, G.S. Kim, C. Turck, S.T. Smale, Mol. Cell. Biol. 14 (1994) 7111.
[77] S. Ghosh, A.M. Gifford, L.R. Riviere, P. Tempst, G.P. Nolan, D. Baltimore, Cell 62 (1990) 1019.
[78] P. Cuatrecasas, J. Biol. Chem. 245 (1970) 3059.
[79] H. Gadgil, H.W. Jarrett, J. Chromatogr. A 848 (1999) 131.
[80] W.S. Dynan, R. Tjian, Cell 32 (1983) 669.
[81] V. Pirrotta, T.A. Bickle, Methods Enzymol. 65 (1980) 89.
[82] T.A. Bickle, V. Pirrotta, R. Imber, Nucleic Acids Res. 4 (1977) 2561.
[83] A.A. Farooqui, J. Chromatogr. 184 (1980) 335.
[84] P. Fusi, G. Tedeschi, A. Aliverti, S. Ronchi, P. Tortora, A. Guerritore, Eur. J. Biochem. 211 (1993) 305.
[85] F. Soncin, R. Prevelige, S.K. Calderwood, Protein Exp. Purif. 9 (1997) 27.
[86] S.P. Jackson, R. Tjian, Cell 55 (1988) 125.
[87] S.P. Jackson, R. Tjian, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 1781.
[88] A.J. Reason, H.R. Morris, M. Panico, R. Marais, R.H. Treisman, R.S. Haltiwanger, G.W. Hart, W.G. Kelly, A. Dell, J. Biol. Chem. 267 (1992) 16911.
[89] W.G. Kelly, M.E. Dahmus, G.W. Hart, J. Biol. Chem. 268 (1993) 10416.
[90] D. Wang, J.E. Kudlow, Biochim. Biophys. Acta 1449 (1999) 50.
[91] M. Williams, L.A. Hanakahi, N. Maizels, J. Biol. Chem. 268 (1993) 13731.
[92] P.D.G. Dean, D.H. Watson, J. Chromatogr. 1979 (1979) 301.

[93] C.R. Lowe, D.A.P. Small, A. Atkinsonb, Int. J. Biochem. 13 (1981) 33.
[94] E.C. Metc.alf, B. Crow, P.D.G. Dean, Biochem. J. 199 (1981) 465.
[95] K.M. Pollard, J. Webb, J. Immunol. Methods 54 (1982) 81.
[96] T. Kristensen, J. Holtlund, J. Chromatogr. 192 (1980) 494.
[97] M. Das, D. Dasgupta, Prep. Biochem. Biotechnol. 28 (1998) 339.
[98] K. Baksi, D.L. Rogerson, G.W. Rushizky, Biochemistry 17 (1978) 4136.
[99] G.A. Kassavetis, D.L. Riggs, R. Negri, L.H. Nguyen, E.P. Geiduschek, Mol. Cell Biol. 9 (1989) 2551.
[100] R.M. Litman, J. Biol. Chem. 243 (1968) 6222.
[101] B. Alberts, G. Herrick, Methods Enzymol. 21 (1971) 198.
[102] P.T. Gilham, Methods Enzymol. 21 (1971) 191.
[103] H. Nakazato, M. Edmonds, J. Biol. Chem. 247 (1972) 3365.
[104] H. Aviv, P. Leder, Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 1408.
[105] R. Sheldon, C. Jurale, J. Kates, Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 417.
[106] A. Nordheim, P. Tesser, F. Azorin, Y.H. Kwon, A. Moller, A. Rich, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 7729.
[107] E.M. Lafer, R. Sousa, B. Rosen, A. Hsu, A. Rich, Biochemistry 24 (1985) 5070.
[108] R.A. Fishel, K. Detmer, A. Rich, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 36.
[109] J.J. Hayes, L. Kam, T.D. Tullius, Methods Enzymol. 186 (1990) 545.
[110] W.J. Dixon, J.J. Hayes, J.R. Levin, M.F. Weidner, B.A. Dombroski, T.D. Tullius, Methods Enzymol. 208 (1991) 380.
[111] D.J. Galas, A. Schmitz, Nucleic Acids Res. 5 (1978) 3157.
[112] M.W. Van Dyke, P.B. Dervan, Nucleic Acids Res. 11 (1983) 5555.
[113] U. Siebenlist, W. Gilbert, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 122.
[114] J. Kohn, M. Wilchek, Biochem. Biophys. Res. Commun. 107 (1982) 878.
[115] P.S. Chockalingam, L.A. Jurado, F.D. Robinson, H.W. Jarrett, Methods Mol. Biol. 147 (2000) 141.
[116] T.A. Goss, M. Bard, H.W. Jarrett, J. Chromatogr. A 508 (1990) 279.
[117] T.A. Goss, M. Bard, H.W. Jarrett, J. Chromatogr. A 588 (1991) 157.
[118] L.A. Chodosh, R.W. Carthew, P.A. Sharp, Mol. Cell Biol. 6 (1986) 4723.
[119] A. Hatamochi, P.T. Golumbek, E. Van Schaftingen, B. de Crombrugghe, J. Biol. Chem. 263 (1988) 5940.
[120] R. Blanks, L.W. McLaughlin, Nucleic Acids Res. 16 (1988) 10283.
[121] P.J. Rosenfeld, T.J. Kelly, J. Biol. Chem. 261 (1986) 1398.
[122] M.R. Briggs, J.T. Kadonaga, S.P. Bell, R. Tjian, Science 234 (1986) 47.
[123] H. Gadgil, H.W. Jarrett, J. Chromatogr. A 966 (2002) 99.
[124] L.A. Jurado, J.T. Drummond, H.W. Jarrett, Anal. Biochem. 282 (2000) 39.
[125] H. Gadgil, W.L. Taylor, H.W. Jarrett, J. Chromatogr. A 905 (2001) 133.
[126] P.A. Whitson, J.S. Olson, K.S. Mathews, Biochemistry 25 (1986) 3852.
[127] M. Leblond-Francillard, M. Dreyfus, F. Rougeon, Eur. J. Biochem. 166 (1987) 351.
[128] J.-H. Ha, R.S. Spolar, M.T. Record, J. Mol. Biol. 209 (1989) 801.
[129] P.A. Whitson, K.S. Mathews, Biochemistry 25 (1986) 3845.
[130] C. Bracken, P.A. Carr, J. Cavanagh, A.G. Palmer, J. Mol. Biol. 285 (1999) 2133.
[131] F. Fogolari, A.H. Elcock, G. Esposito, P. Viglino, J.M. Briggs, J.A. McCammon, J. Mol. Biol. 267 (1997) 368.
[132] D. Foguel, J.L. Silva, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 8244.
[133] T. Lundback, J. Zilliacus, J.A. Gustafsson, J. Carlstedt-Duke, T. Hard, Biochemistry 33 (1994) 5955.
[134] R.S. Spolar, M.T. Record, Science 263 (1994) 777.
[135] Y. Takeda, P.D. Ross, C.P. Mudd, Natl. Proc. Acad. Sci. U.S.A. 89 (1992) 8180.
[136] R. Marmorstein, M. Carey, M. Ptashne, S.C. Harrison, Nature 356 (1992) 408.
[137] C. Tanford, The Hydrophobic Effect: Formation of Micelles and Biological Membranes, Wiley, New York, NY, 1973.
[138] R.B. Winter, O.G. Berg, P.H. von Hippel, Biochemistry 20 (1981) 6961.
[139] H.W. Jarrett, W.L. Taylor, J. Chromatogr. A 803 (1998) 131.
[140] H.W. Jarrett, Anal. Biochem. 279 (2000) 209.
[141] Z. Cao, R.M. Umek, S.L. McKnight, Genes Dev. 5 (1991) 1538.
[142] K.K. Perkins, A. Admon, N. Patel, R. Tjian, Genes Dev. 4 (1990) 822.
[143] A.R. Black, J.D. Black, J. Azizkhan-Clifford, J. Cell Physiol. 188 (2001) 143.
[144] S.Y. Lin, A.D. Riggs, Nature 228 (1970) 1184.
[145] M. Fried, D.M. Crothers, Nucleic Acids Res. 9 (1981) 6505.
[146] M.M. Garner, A. Revzin, Nucleic Acids Res. 9 (1981) 3047.
[147] L.B. Ludwig, B.J. Hughes, S.A. Schwartz, Nucleic Acids Res. 23 (1995) 3792.
[148] G. Suske, B. Gross, M. Beato, Nucleic Acids Res. 17 (1989) 4405.
[149] T. Onizuka, S. Endo, M. Hirano, S. Kanai, H. Akiyama, Biosci. Biotechnol. Biochem. 66 (2002) 2732.
[150] J. Hodgson, P.J. Enrietto, J. Virol. 69 (1995) 1971.
[151] X.Y. Zhang, C.K. Asiedu, P.C. Supakar, M. Ehrlich, Anal. Biochem. 201 (1992) 366.
[152] Z. Kozmik, P. Urbanek, V. Paces, Nucleic Acids Res. 18 (1990) 2198.
[153] A.J. Bannister, T. Kouzarides, Nucleic Acids Res. 20 (1992) 3523.
[154] A. Ozyhar, H.H. Kiltz, J. Chromatogr. 587 (1991) 11.
[155] W.A. Lim, R.T. Sauer, A.D. Lander, Methods Enzymol. 208 (1991) 196.
[156] A.S. Cohen, B.L. Karger, J. Chromatogr. 397 (1987) 409.
[157] K. Ganzle, K.S. Greve, A.S. Cohen, B.L. Karger, A. Guttman, N.C. Cooke, Anal. Chem. 62 (1992) 2665.
[158] A. Sancar, D. Wu, F.E. Regnier, Anal. Chem. 65 (1993) 2029.
[159] L.Z. Avila, Y.H. Chu, E.C. Blossey, G.M. Whitesides, J. Med. Chem. 36 (1993) 126.
[160] F.T. Chen, J.C. Sternberg, Electrophoresis 15 (1994) 13.
[161] K. Shimura, B.L. Karger, Anal. Chem. 66 (1994) 9.
[162] E.C. Rickard, M.M. Strohl, R.G. Nielsen, Anal. Biochem. 197 (1991) 197.
[163] H.T. Chang, E.S. Yeung, J. Chromatogr. B Biomed. Appl. 669 (1995) 113.
[164] M.A. Stebbins, A.M. Hoyt, M.J. Sepaniak, B.K. Hurlburt, J Chromatogr. B Biomed. Appl. 683 (1996) 77.
[165] N.H. Heegaard, F.A. Robey, J. Immunol. Methods 166 (1993) 103.
[166] B. Bowen, J. Steinberg, U.K. Laemmli, H. Weintraub, Nucleic Acids Res. 8 (1980) 1.
[167] H. Singh, J.H. LeBowitz, A.S.J. Baldwin, P.A. Sharp, Cell 52 (1988) 415.
[168] W. Knepel, L. Jepeal, J.F. Habener, J. Biol. Chem. 265 (1990) 8725.
[169] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[170] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[171] J. Klose, H. Spielmann, Biochem. Genet. 13 (1975) 707.
[172] E. Gianazza, F. Chillemi, M. Duranti, P.G. Righetti, J. Biochem. Biophys. Methods 8 (1983) 339.
[173] H.H. Rasmussen, J. Van Damme, M. Puype, B. Gesser, J.E. Celis, J. Vandekerckhove, Electrophoresis 12 (1991) 873.
[174] A. Gorg, W. Postel, A. Domscheit, S. Gunther, Electrophoresis 9 (1988) 681.
[175] A. Gorg, W. Postel, C. Friedrich, R. Kuick, J.R. Strahler, S.M. Hanash, Electrophoresis 12 (1991) 653.
[176] V. Brodard, B. David, A. Gorg, G. Peltre, Int. Arch. Allergy Immunol. 102 (1993) 72.
[177] A. Gorg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, W. Weiss, Electrophoresis 21 (2000) 1037.

[178] B. Bjellqvist, K. Ek, P.G. Righetti, E. Gianazza, A. Gorg, R. Westermeier, W. Postel, J. Biochem. Biophys. Methods 6 (1982) 317.
[179] A. Gorg, C. Obermaier, G. Boguth, A. Csordas, J.J. Diaz, J.J. Madjar, Electrophoresis 18 (1997) 328.
[180] T. Rabilloud, Anal. Chem. 72 (2000) 48A.
[181] A. Cappiello, M. Balogh, G. Famiglini, F. Mangani, P. Palma, Anal. Chem. 72 (2000) 3841.
[182] D.F. Hunt, H. Michel, T.A. Dickinson, J. Shabanowitz, A.L. Cox, K. Sakaguchi, E. Appella, H.M. Grey, A. Sette, Science 256 (1992) 1817.
[183] J.R.R. Yates, S. Speicher, P.R. Griffin, T. Hunkapiller, Anal. Biochem. 214 (1993) 397.
[184] R. Lovering, I.M. Hanson, K.L. Borden, S. Martin, N.J. O'Reilly, G.I. Evan, D. Rahman, D.J. Pappin, J. Trowsdale, P.S. Freemont, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 2112.
[185] R.M. Caprioli, Continuous-Flow Fast Atom Bombardment Mass Spectrometry, Springer-Verlag, Berlin, 1989.
[186] B.T. Chait, R. Wang, R.C. Beavis, S.B. Kent, Science 262 (1993) 89.
[187] L.A. McPherson, V.R. Baichwal, R.J. Weigel, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 4342.
[188] A.D. Adams, D.M. Choate, M.A. Thompson, J. Biol. Chem. 270 (1995) 6975.
[189] J. Berthelsen, J. Vandekerkhove, F. Blasi, J. Biol. Chem. 271 (1996) 3822.
[190] M. Eggert, C.C. Mows, D. Tripier, R. Arnold, J. Michel, J. Nickel, S. Schmidt, M. Beato, R. Renkawitz, J. Biol. Chem. 270 (1995) 30755.
[191] B. Corthesy, P.N. Kao, J. Biol. Chem. 269 (1994) 20682.
[192] D.A. Potter, C.J. Larson, P. Eckes, R.M. Schmid, G.J. Nabel, G.L. Verdine, P.A. Sharp, J. Biol. Chem. 268 1993 18882.
[193] N. Kagawa, M.R. Waterman, J. Biol. Chem. 267 (1992) 25213.
[194] R. Kovelman, R.G. Roeder, J. Biol. Chem. 267 (1992) 24446.
[195] G. Vlatakis, V. Bouriotis, Anal. Biochem. 195 (1991) 352.
[196] G. Heilgenthal, I. Grummt, J. Chromatogr. 587 (1991) 25.
[197] M.S. Kasher, D. Pintel, D.C. Ward, Mol. Cell Biol. 6 (1986) 3117.
[198] S.S. Broyles, L. Yuen, S. Shuman, B. Moss, J. Biol. Chem. 263 (1988) 10754.
[199] D.J. Gruol, K.A. Wolfe, Biochemistry 28 (1989) 2929.
[200] E.A. O'Neill, T.J. Kelly, J. Biol. Chem. 263 (1988) 931.
[201] K.J. Howard, C.W. Distelhorst, Biochem. Biophys. Res. Commun. 151 (1988) 1226.
[202] M.J. Hughes, H.M. Liang, J. Jiricny, J.P. Jost, Biochemistry 28 (1989) 9137.
[203] F.J. Calzone, C. Hoog, D.B. Teplow, A.E. Cutting, R.W. Zeller, R.J. Britten, E.H. Davidson, Development 112 (1991) 335.
[204] D. Staiger, F. Becker, J. Schell, C. Koncz, K. Palme, Eur. J. Biochem. 199 (1991) 519.
[205] F.V.J. Peale, Y. Ishibe, C.M. Klinge, S. Zain, R. Hilf, R.A. Bambara, Biochemistry 28 (1989) 8671.
[206] V. Moncollin, M. Gerard, J.M. Egly, J. Chromatogr. 510 (1990) 243.
[207] T.A. Gasiewicz, C.J. Elferink, E.C. Henry, Biochemistry 30 (1991) 2909.
[208] B.R.J. Franza, S.F. Josephs, M.Z. Gilman, W. Ryan, B. Clarkson, Nature (Lond.) 330 (1987) 391.
[209] G. Stetter, M.A. Braun, M. Montenarh, Oncogene 6 (1991) 389.
[210] J. Zhang, S.T. Jacob, Mol. Cell Biol. 10 (1990) 5177.
[211] O.S. Gabrielsen, E. Hornes, L. Korsnes, A. Ruet, T. Oyen, Nucleics Acid Res. 17 (1989) 6253.
[212] H. Kawaguchi, A. Asai, Y. Ohtsuka, H. Watanabe, T. Wada, H. Handa, Nucleic Acids Res. 17 (1989) 6229.
[213] L.K. Arciszewska, R.L. McKown, N.L. Craig, J. Biol. Chem. 266 (1991) 21736.
[214] D.C. Rio, G.M. Rubin, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 8929.
[215] J. Halbrook, K. McEntee, J. Biol. Chem. 264 (1989) 21403.
[216] S. Seki, S. Ikeda, S. Watanabe, M. Hatsushika, K. Tsutsui, K. Akiyama, B. Zhang, Biochim. Biophys. Acta 1079 (1991) 57.
[217] P. Alliger, W. Traut, E. Carstens, E. Fanning, Biochim. Biophys. Acta 951 (1988) 396.
[218] K.S. Sweder, P.R. Rhode, J.L. Campbell, J. Biol. Chem. 263 (1988) 17270.
[219] D.J. Arndt-Jovin, T.M. Jovin, W. Bahr, A.-M. Frischauf, M. Marquardt, Eur. J. Biochem. 54 (1975) 411.
[220] Z.F. Wang, J. Yang, Z.Q. Nie, M. Wu, Biochemistry 30 (1991) 1127.
[221] M. West, J. Mikovits, G. Princler, Y. Liu, F.W. Ruscetti, H. Kung, Raziuddin, J. Biol. Chem. 267 (1992) 24948.
[222] E.C. Ruteshouser, B. de Crombrugghe, J. Biol. Chem. 267 (1992) 14398.
[223] J. Sze, G.B. Kohlaw, J. Biol. Chem. 268 (1993) 2505.
[224] M. Munder, I.M. Herzberg, C. Zierold, V.E. Moss, K. Hanson, M. Clagett-Dame, H.F. DeLuca, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 2795.
[225] G.K. Fu, D.M. Markovitz, J. Biol. Chem. 271 (1996) 19599.
[226] B. Moorefield, R.G. Roeder, J. Biol. Chem. 269 (1994) 20857.
[227] P. Wen, J. Locker, Mol. Cell Biol. 14 (1994) 6616.
[228] F. Liu, E. Bateman, J. Biol. Chem. 269 (1994) 18541.