University of Tennessee Health Science Center
UTHSC Digital Commons

Theses and Dissertations (ETD) College of Graduate Health Sciences

6-2001

Advances in DNA Affinity Chromatography

William Himanshu S. Gadgil
University of Tennessee Health Science Center

Follow this and additional works at: https://dc.uthsc.edu/dissertations

Part of the Medical Biochemistry Commons

Recommended Citation

Advances in DNA Affinity Chromatography

Document Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Program
Biochemistry

Research Advisor
Harry W. Jarrett, Ph.D.

Committee
William L. Taylor, Ph.D. Satoru K. Nishimoto, Ph.D. Roderick T. Hori, Ph.D. Edwards A. Park, Ph.D.

DOI
10.21007/etd.cghs.2001.0102

Advances in DNA Affinity Chromatography

A Dissertation

Presented for

The Graduate Studies Council

The University of Tennessee

Health Science Center

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

From The University of Tennessee

By

Himanshu S Gadgil

June 2001

DEDICATION

Thisdissertation is dedicated to my wife

Shilpa

And to my parents

Alaka and Shashikant Gadgil

ACKNOWLEDGMENTS

ABSTRACT

Different aspects of DNA affinity chromatography such as DNA complexity heparin elution, the Bi-column method and the oligonucluotide trapping method were studied. The complexity (length) of a DNA sequence attached to an affinity chromatography column affects column retention, and the purity of transcription factors obtained. T_{18}: A_{18} tailed DNA affinity columns were better suited for purification of most of the transcription factors than either the discrete or concatemeric DNA affinity columns. A novel method using heparin for eluting transcription factors from DNA Sepharose columns was characterized. The amount of the lac repressor chimera which eluted from the column was shown to increase with increases in the mobile phase heparin concentration. The elution of the protein was also shown to be dependent on the amount of DNA coupled to the column and more protein eluted from columns containing lesser amounts of DNA. These data suggest that heparin and DNA compete for binding to the protein; this competition causes elution. Comparison of heparin- and salt-eluted protein demonstrated the heparin-eluted fraction of lac repressor was significantly purer than that eluted with salt and comparable to that obtained by elution with the specific ligand IPTG, a lactose analog. A novel Bi-column method was developed in which lac repressor is eluted from the Op1-Sepharose with a low heparin concentration and trapped on a $Op1T_{18}$-Sepharose column because of its higher affinity for the lac repressor protein. Elution of the latter column with buffer containing a high salt concentration gives significantly purer transcription

factor than the conventionally used single column methods and removes residual

heparin. Highly pure CAAT enhancer binding protein(C/EBP) and the B3

transcription factor are also obtained by using variants of this Bi-column method.

A new oligonucleotide trapping method in which a short oligonucleotide coupled

to Sepharose is used to trap a complex of the transcription factor and its

corresponding specific DNA sequence was developed. Highly purified

transcription factor B3 was obtained using the oligonucleotide trapping method.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACMVP	Acyclovir monophosphate
AP	Alkaline phosphatase
Ap1	Activator protein
C/EBP	CAAT enhancer binding protein
CM-Cellulose	Carboxymethyl cellulose
CNBr	Cyanogen bromide
DNA	Deoxyribonucleic acid
DNaes	Deoxyribonucleases
dNTP	Deoxynucleotide triphosphate
ECL	Enhanced-chemi-luminescence
EDTA	Ethylenediaminetetraacetic acid
GFP	Green fluorescent protein
HPLC	High Performance Liquid Chromatography
HSV-1	Herpes Simplex Virus-1
IPTG	Isopropyl-β-D-thiogalactopyranoside
KDa	KiloDalton
lac	Lac repressor
laciz	Lac repressor-β-galactosidase fusion protein
NBT	Nitroblue-tetrazolium
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCA	Trichloroacetic acid

TEA	Triethylamine
TEMED	N,N,N',N'-Tetramethylethylenediamine
UV	Ultraviolet

Chapter: 1. Transcription factor purification: A brief overview

1.1. Introduction

DNA binding proteins control cellular processes such as DNA replication, repair, and recombination, which allow an organism's development, embryogenesis, etc. Transcription factors belong to the larger family of DNA binding proteins and control gene expression. Timely regulation of genes is important in all living organisms. In higher eukaryotes regulation of genes is responsible for proper functioning of the cell cycle, proper control of metabolic pathways, and the ultimate type of cell produced at differentiation. Genetic regulation in prokaryotes is required for proper response to environmental conditions and their ultimate survival. Cellular and viral encoded transcription factors are also important for viral propagation. Transcription factors control gene expression by binding to specific DNA elements in the promoter region. In mammals transcription factors such as AP-1, cMyc, E2F, etc. are important regulators of cell cycle (1) (2) (3). In bacteria such as *Escherichia coli,* lac repressor protein controls the expression of the lactose operon genes in response to the presence of lactose in the medium (4) (5). In viruses such as bacteriophage λ, the counteractive effect of the transcription factors Cro and λ repressor determines whether the phage enters the lytic or the lysogenic phase (6). Improper functioning of transcription factors in humans contributes to diseases such as cancer, dwarfism and congenital severe combined immunodeficiency (7).

Purification of transcription factors from tissue extracts is challenging, as they are normally present at very low concentration. This problem is further

compounded by the fact that several purification steps are needed for the complete purification of a particular transcription factor and some amount of the transcription factor is lost at each step. In this chapter we will briefly discuss the classification, assay, and purification of transcription factors. The major focus of the chapter will be on sequence specific DNA affinity chromatography.

1.1.1 Brief overview of transcription factors

A detailed discussion of transcription factor structure, function and regulation is beyond the scope of this chapter, but are the subject of several books and review articles (8) (7) (9) (10) (11) (12). Transcription factors regulate gene expression by binding to a specific region of DNA, which is normally located upstream (5') of the gene in the promoter region. A classic example of a transcription factor is the lac repressor protein which, upon binding to its operator element (Op1), represses the expression of lac operon genes (13). Transcription factors can either repress the expression of the gene, as is the case for lac repressor or it can enhance gene expression. A good example of the latter is the CAAT enhancer binding protein (C/EBP) which binds specifically to DNA sequences containing a CAAT element. C/EBP participates in the regulation of expression of several genes in higher eukaryotes. The promoter region in both eukaryotes and prokaryotes usually contains multiple elements and binds more than one transcription factor. Thus, genetic regulation of a particular gene is the result of the interplay between multiple transcription factors and, ultimately, an RNA polymerase.

Transcription factors usually consist of two distinct domains, the activation domain and the DNA binding domain. The activation domain functions by recruiting components of the basal transcription complex such as RNA polymerase, factors associated with RNA polymerase, or other transcription factors at the promoter region. Activation domains may bind these proteins directly or may interact indirectly through a coactivator protein. Activation domains have been classified based upon their distinct amino acid content such as a predominance of acidic amino acids, proline and glutamine residues, etc. (9).

The DNA binding domain is responsible for binding to the specific DNA element, which to a greater or lesser degree, matches an ideal sequence called the consensus sequence. Each transcription factor usually recognizes only a specific consensus sequence and the consensus sequences vary from one transcription factor to another. For example, transcription factors belonging to the C/EBP family bind specifically to sequences containing the CAAT element while members of the Sp1 family bind to the GC box element, which has a high GC content. The specific interaction between transcription factor and the consensus DNA sequence is of a great biological importance as it allows a transcription factor to regulate expression of a certain set of genes whose promoter contains the specific response element while the expression of other genes remains unaffected. The match of the response element to the consensus sequence determines how strongly the corresponding transcription factor binds and, presumably, how large an effect it has on gene expression.

1.1.2 Transcription factor motifs

The distinct structural motifs present in the DNA binding domain are responsible for the binding specificity of the transcription factors. Transcription factors are frequently classified according to these structural features. These classes of transcription factors are briefly discussed below; for detail reviews refer to (14) (15).

1.1.2.1 Helix turn helix (HTH) family

Helix turn helix (HTH) was the first DNA binding motif to be recognized. Transcription factors belonging to this family include the lac repressor protein, the λ repressor and the *trp* repressor protein (15). Homeodomain proteins, which play an important role in development of higher organisms also, belong to this family (16) (17). The HTH motif consists of two α helices joined together by a short stretch of amino acids, which forms the turn (18). Proteins belonging to this family usually bind to the DNA as a dimer wherein each monomeric subunit recognize a half site in the DNA. There are, of course, exceptions such as lac repressor, which binds as a tetramer (13).

1.1.2.2 Zinc finger proteins

The first zinc finger motif was identified in the *Xenopus* transcription factor III (TFIIIA). Since then several transcription factors have been shown to contain this motif, e.g., Sp1, steroid and thyroid receptors, zif268 and many others. The TFIIIA zinc finger motifs contains the sequence pattern:

$Cys\text{-}X_{2\,or\,4}\text{-}Cys\text{-}X_{12}\text{-}His\text{-}X_{3\text{-}5}\text{-}His$ (15).

Although the sequence varies in the different zinc finger proteins, the common

5

feature is the presence of either four histidines or two histidines and two

cysteines. The cysteines and histidines tetrahedrally coordinate a zinc atom while

remaining amino acids form a finger like protrusion responsible for DNA

recognition; hence, the name zinc finger motif. Many of the proteins belonging to

this family contain several zinc finger motifs arranged in tandem arrays wherein

the α helical structures in the fingers can make continuous contact with the major

groove of DNA. For example, TFIIIA has 9 zinc finger motifs responsible for its

binding to DNA and also RNA.

1.1.2.3 The leucine zipper motif

The leucine zipper motif consists of two regions, the leucine zipper region,

which contains heptad repeats of leucine and a basic region, which contains

predominantly basic amino acids. Several transcription factors are known to have

this motif; the most prominent are C/EBP, Ig/EBP, CREB, *c-fos*, and *c-jun* (15).

The leucine zipper region has 30-40 residues, which are arranged in the form of

an α helix (with 3.6 amino acids per turn) such that the leucine residues

(occurring every seven residues) are arranged on the same face of the helix.

This specific arrangement of leucines brings about dimerization between two

monomers as the interstrand leucines intercalate and form tight contacts with

each other. The arrangement of the leucines on the two strands is like that of the

teeth on a zipper. Dimerization results in a "Y" shaped structure in which leucine

zipper regions of the two monomers are joined to form the stem while the basic

regions remain separate to form the fork. The basic regions form strong

electrostatic interactions with the negatively charged DNA phosphate backbone

(19).

1.1.2.4 Helix loop helix (HLH)

Like the leucine zipper proteins, HLH proteins are also comprised of two distinct regions, a basic region that is involved in binding to the DNA and a adjacent region that is involved in dimerization (20). Myo D protein, which is involved in muscle differentiation, belongs to this family.

1.1.2.5 Other motifs

Although many of the transcription factors can be assigned to one of the families mentioned above, there are other DNA binding motifs which have been identified but have been found less commonly. The most predominant among these is the β sheet motif. Proteins containing this motif use the antiparallel β sheets to bind to the DNA. Other motifs include the "Zinc cluster" motif found in Gal4, the LIM motif found in some of the homeodomain proteins, and the POU domain present in POU proteins (15).

1.1.3 Assay of transcription factors

Many families of transcription factors such as Sp1, C/EBP, and AP-1 have been identified and well characterized (21) (22) (23). There are other members to these families such as Sp1, Sp2, Sp3, and Sp4 that are similar but distinct proteins. There are also many other transcription factors which have not been fully characterized and others about which virtually nothing is known. The first step in the characterization of a transcription factor is to purify it sufficiently so that it can be sequenced and cloned.

A highly sensitive and specific assay is prerequisite for the purification of any protein. Hence, it is important to have good detection methods for identifying transcription factors. Transcription factors usually bind their cognate DNA sequence with affinities in the picomolar range and have a 10^3-10^5 higher affinity for this specific DNA sequence than for any other DNA sequences. This high binding affinity has been utilized for the assay of transcription factors in the different techniques described below.

1.1.3.1 Filter binding assay

The filter binding assay is one of the most popular assays for detecting transcription factors because of its ease of operation. It was first used for detection of the lac repressor protein (24). Since then, the filter binding assay has been widely used to detect and to analyze the thermodynamics and kinetics of DNA-protein interactions (25) (26) (27). Nitrocellulose membranes bind specifically to proteins without interacting with DNA. This property of nitrocellulose membrane forms the basis of filter binding assay. Transcription factor fractions are mixed with radiolabeled DNA containing the specific DNA element. The mixture is then filtered and, because of the protein binding, the transcription factor-DNA complex is retained on the filter. After thorough washing, the radiolabel remaining on the filter allows quantitation of the complex. For further details, see (25). The advantages of this technique are the ease and rapidity of analysis. The major disadvantage is the low specificity. In crude fractions, most of the protein bound by the filter is not the transcription factor of interest and competes with it for filter binding. Thus, filter binding assays are

8

seldom applied to crude extracts.

1.1.3.2 Electrophoretic mobility shift assay (EMSA)

Like the filter binding assay, the electrophoretic mobility shift assay
(EMSA) also utilizes radiolabeled specific oligonucleotide to detect transcription
factors. In this assay a radiolabeled oligonucleotide is incubated with a
transcription factor containing fraction. A non-labeled competitor DNA such as
poly(dI:dC) is added to lessen non specific binding. This mixture is then
subjected to nondenaturing polyacrylamide gel electrophoresis (PAGE). When
the transcription factor present in the extract binds to the probe it retards the
mobility of the probe (shifts it) on PAGE which can be easily detected after
autoradiography (28). When an antibody against a transcription factor is
available, a super shift assay can be performed. In this method EMSA is carried
out in presence of the antibody and an even greater retardation in the mobility is
observed because of formation of DNA-transcription factor-antibody complex.
Sometimes biotin- or digoxygenin-labeled probes are utilized in EMSA to provide
an alternative which uses no radiolabel (29) (30). Techniques similar to EMSA
which involve separation of the DNA-transcription factor complex by high
resolution gel filtration have also been used (31). Modifications of EMSA, such as
affinity coelectrophoresis (ACE), have been developed to measure DNA binding
constants (32). The major advantages of EMSA are high specificity and ease of
operation.

1.1.3.3 Capillary electrophoresis mobility shift assay (CEMSA)

CEMSA is a modification of EMSA in which instead of resolving DNA-protein complex on a gel it is resolved on the basis of charge to mass ratio with small uncoated capillaries (33). CEMSA is very rapid and can be used for precise determination of binding constants of DNA-protein interaction. It is also highly sensitive as it uses laser-induced fluorescence to detect fluorescein labeled oligonucleotides. Because of these advantages CEMSA is gaining in popularity (33) (34). The main disadvantages are the cost of the equipment and the sensitivity of most electrophoresis, and especially capillary electrophoresis, to salt concentrations.

1.1.3.4 Protection assay (DNA footprinting)

In footprinting experiments the ability of transcription factors to protect bound DNA from enzymes such as nuclease or methylase, or from chemical treatments is assayed. Upon binding to DNA a transcription factor shields it from the above mentioned modifications. The region of DNA that is protected (the "footprint") is often the minimal sequence required for DNA-transcription factor interaction. In DNAse1 protection assay, the first step is to radiolabel specifically only one strand. This DNA and the DNA-transcription factor complex is then treated with DNAse1, an endonuclease, which randomly cleaves the chain preferentially at pyrimidines. The area where the transcription factor is bound is protected from cleavage. By analyzing the reaction products on a DNA sequencing gel, next to one or more DNA sequencing reactions (to provide length and sequence markers), a region is found where no cleavages occurred in

the presence of transcription factor. This region is usually called the transcription

factor's "footprint". The reaction is repeated radiolabeling the other strand to

confirm location and to determine the boundaries of the footprint on each strand.

The DNAse I footprinting method is highly specific and hence is often used

in assaying transcription factors despite being technically more difficult than

either EMSA or filter binding assay. Other methods of footprinting involve the use

of small molecules such as the hydroxy radicals or even γ rays to cleave the

unbound DNA (35) (36). Principally these methods are similar to DNAse1

footprinting except for the agent used to cleave DNA. The major advantage of

using these agents is that better resolution can be obtained because of their very

small size.

Other sophisticated methods such as fluorescence anisotropy and surface

plasmon resonance spectroscopy have been used to detect DNA-protein

interaction (37) (38). These methods require expensive instrumentation and more

complex operation and hence are not yet routinely used to assay transcription

factors.

1.2. Purification of transcription factors

The number of transcription factors that have been purified to

homogeneity is growing continuously. Table 1.1 shows the different transcription

factors that have been purified in the last 8 years. A survey of the purification of

transcription factor homology families show that they are all purified in basically

similar ways. There is no technique that stands out as being especially applicable

Table 1.1. List of transcription factors purified in the last few years.

Transcription factor	Source	Protein sequence obtained	Steps used to accomplish purification	Additional strategies and components used	Reference
Transcription factor III C	HeLa nuclear extract	Yes	-Heparin-agarose -Phosphocellulose -Ion exchange chromatography -DNA cellulose -Sequence-specific DNA affinity column	-Concatemeric sequences	[39]
SpE2F	*Schizosaccharomyces pombe*	No	-Heparin-agarose -Sequence-specific DNA affinity column	-Concatemeric sequences	[40]
H2TF1	HeLa cells	Yes	-Ion exchange chromatography -Heparin-agarose -Sequence-specific DNA affinity column	-Both monomeric and concatemeric DNA affinity columns were effective for purification	[41]
PCF	HepG2 nuclear extract	Yes	-Heparin-agarose -Sequence-specific DNA affinity column	-Concatemeric sequences	[42]

Table 1.1 continued

Transcription factor	Source	Protein sequence obtained	Steps used to accomplish purification	Additional strategies and components used	Refe-rence
TFIIIA	HeLa cells	Yes	-Biorex70 -Sequence-specific DNA affinity column	-Monomeric DNA sequences.	[43]
NF-AT	Jurkart cell nuclear extract	Yes	-Ion exchange chromatography -Heparin-Sepharose -Octylamine agarose -Sequence-specific DNA affinity column	-Concatemeric DNA sequences -Mutated DNA affinity column.	[44]
TPBF	*Acantham oeba* nuclear extract	Yes	-Ion exchange chromatography -Sequence-specific DNA affinity column	-Concatemeric DNA sequences	[45]
TFIIIC	Discothelli um Discoidum .	No	-Ion exchange chromatography -Sequence-specific DNA affinity column	-Concatemeric DNA sequences	[46]
GRIP170	HeLa S3 nuclear extract.	Yes	-Ion exchange chromatography -Heparin Sepharose		[47]

Table 1.1 continued

Transcription factor	Source	Protein sequence obtained	Steps used to accomplish purification	Additional strategies and components used	Refe- rence
MtEBPs	Mammalian nuclear extract	No	-Ion exchange chromatography -Heparin-Sepharose -Sequence-specific DNA affinity column	-Monomeric DNA sequences	[48]
Stat3	Rat liver nuclear extract	Yes	-Ion exchange chromatography -Phenyl-Sepharose -Sequence-specific DNA affinity column	-Concatemeric DNA sequences	[49]
NF1-L	Rat liver nuclear extract	Yes	-Heparin-Sepharose -Calf thymus DNA affinity -Sequence-specific DNA affinity column	-Concatemeric DNA sequences - Repetitive use of DNA affinity chromatography	[50]
AF-1	MLA144 cells nuclear extract	Yes	-Phenyl-Sepharose -Bio-Rex70 -Superose 6 gel filtration		[51]

Table 1.1 continued

Transcription factor	Source	Protein sequence obtained	Steps used to accomplish purification	Additional strategies and components used	Reference
UEF3	HeLa cell extract	No	-Ioexchange chromatography -heparin Sepharose -Sequence-specific DNA affinity column	-Concatemeric DNA sequences -Mutant DNA affinity.	[52]
Inr-BP	Bovine testis nuclear extract	No	-Heparin agarose -All-gel blue -Sequence-specific DNA affinity column	-Monomeric DNA sequences	[53]
NF-Y	YNIH3T3 cell nuclear extract	Yes	-Sequence-specific DNA affinity column	-Concatemeric DNA sequences -Magnetic bead purification	[54]
SEF	HeLa cell extract	Yes	-Ion exchange -Phenyl-Sepharose -Sequence-specific DNA affinity column	-Concatemeric DNA sequences -Magnetic bead separation	[57]
TEF1	Rat kidney nuclear extract	No	-Heparin agarose -Wheat germ lectin -Sequence-specific DNA affinity column	-Concatemeric DNA sequences	[58]

Table 1.1 continued

Transcription factor	Source	Protein sequence obtained	Steps used to accomplish purification	Additional strategies and components used	Refe-rence
Recombinant C/EBP	Crude bacterial extract	Yes	-Sequence-specific DNA affinity column	-Monomeric DNA sequences -Temperature dependant elution of C/EBP	[59]
Reconbinant lac repressor		Yes	-Sequence-specific DNA affinity column	-Monomeric DNA sequences -Heparin elution of lac repressor	[60]
B3	*Xenopus* oocyte extract	Yes	-Sequence-specific DNA affinity column	-Monomeric DNA sequences -Bi-column method	[61]
C/EBP	Rat liver nuclear extract	Yes	-Sequence-specific DNA affinity column	-Monomeric DNA sequences -Bi-column method	[61]

to any one family of proteins. The DNA affinity chromatography is also basically the same. As mentioned before, purification of a transcription factor is an early step in its characterization.

Purification of transcription factors starts with identifying a suitable assay system, which has been discussed in section 1.2.1 of this chapter, and obtaining the DNA footprint sequence. The next step is to choose an appropriate tissue in the case of eukaryotic systems, and appropriate growth conditions in the case of prokaryotic systems. For example, C/EBP-α is more abundant in liver and hence rat liver nuclear extracts are often used in its purification (62). In eukaryotes transcription factors are predominantly present in the nucleus and hence nuclear extracts are often used, instead of whole cell extracts, for their purification. Use of nuclear extracts leads to10-100 fold enrichment of transcription factors and other DNA binding proteins. In the case of bacterial systems whole cell lysate is usually the starting material. In these systems it is important to identify growth conditions which give maximum expression of transcription factor of interest.

At least 4-5 different chromatographic steps are often required for the purification of transcription factors from such extracts. These steps include use of ion exchange, reversed phase (hydrophobic interaction chromatography), gel filtration, nonspecific and sequence specific DNA affinity columns (see columns 4 and 5, Table 1.1).

1.2.1 Non-affinity techniques

Ion exchange columns such as Mono S and Mono Q, DEAE cellulose, phosphocellulose or heparin-agarose are routinely used as a preliminary step in the purification of transcription factors. Heparin-agarose is anionic but is thought to function in a unique manner. The polyanionic sugar backbone of heparin is thought to form a structure that resembles the negatively charged DNA phosphate sugar backbone and hence is bound by most of the DNA binding proteins (and other proteins too). For this reason heparin-agarose is used almost ubiquitously in purification of transcription factors and other DNA binding proteins. Examples of gel filtration include the Sephacryl S300 used in purifying Sp1 (21) and reverse phase columns include the Aquapore butyl HPLC columns used for purification of C/EBP (62). Of all the methods mentioned above, DNA affinity chromatography has the highest selectivity and specificity and is often the most important step in purifications involving transcription factors and hence will be the major focus of the rest of this chapter.

1.2.2 DNA affinity chromatography

DNA affinity columns are made by coupling nonspecific or specific DNA to supports such as cellulose or Sepharose. Non specific DNA-cellulose columns were among the first DNA affinity columns used for purification of DNA binding proteins. Non specific columns are made from Salmon or herring sperm DNA. Nonspecific columns are not selective for a particular protein as almost all DNA binding proteins are able to bind to these columns. None-the-less, nonspecific

DNA-cellulose columns are commonly used as one of the many steps in purification of transcription factors (63) (64) (65) (66). Specific DNA affinity columns on the other hand are tailor made for a particular transcription factor of interest by coupling of the DNA footprint region (see section 1.2.1.4 for details on footprinting). The transcription factor of interest binds to its footprint region with a high affinity while most of the other DNA binding proteins have only a moderate affinity for this region and can be washed off from the columns with appropriate salt concentrations. Different aspects of DNA affinity chromatography such as the supports used and elution methods are discussed in the following sections.

1.2.2.1 The stationary phase

Many supports such as cellulose, Sepharose, latex and silica have been use to make DNA affinity columns. An ideal support is one that is inert and does not bind nonspecifically to proteins. Other desirable properties in a support are that it should be resistant to pressure and shearing, should be spherical and uniform, and should have a high mass transfer rate. No support reaches this ideal but some are quite good.

1.2.2.1.1 Cellulose

DNA cellulose was one of the first DNA affinity supports to be used in the purification of transcription factors (67). The major deficiencies of cellulose are that several proteins are able to bind to cellulose nonspecifically, it is made of irregular fiber lengths which do not pack uniformly in a column, it swells and shrinks in response to hydrostatic or osmotic pressures, and it is restricted to low pressure. While cellulose is still used, it has been superceded by better supports.

1.2.2.1.2 Sepharose

Sepharose is by far the most widely used support for making sequence specific DNA affinity columns (60) (21) (68). The major advantage of Sepharose is that it shows low nonspecific binding to proteins, has a wide range of coupling chemistries available for it, and is relatively cheap. The major disadvantages are that it is not resistant to high pressure and hence cannot be used in HPLC, and, like most soft gels, it shrinks and swells in response to osmotic and hydrostatic pressure, and has slow mass transfer.

1.2.2.1.3 Silica

In many ways silica is a nearly ideal support. It is mechanically rigid and does not shrink or swell in response to pressures, is available in precisely sized spheres of uniform porosity, and is available in a large number of surface coatings. Silicas have some of the highest mass transfer rates of any support, easily surpassing the soft gels. It is unstable at pH values less than about 3 or greater than about 7.5, but this has a larger effect on the availability of coupling methods than upon its chromatographic performance. DNA-silica is not widely used primarily because it was developed only about ten years ago (69) and the techniques of column packing and coupling chemistry are not widely known.

1.2.2.1.4 Latex beads

Latex beads are made of a styrene core and a glycidyl methacrylate (GMA) surface (70) (71). Latex beads are small in size and show high capacity for the DNA and moderately high mass transfer rates. Latex beads also show low nonspecific binding and are thought be a nearly ideal low pressure DNA affinity

20

support. As with silica though, this support has become available later than the soft gels and cellulose and is not widely used because of investigator unfamiliarity with the techniques involved.

1.2.2.2 Coupling chemistries

DNA can be coupled to the above-mentioned supports by either covalent or noncovalent means. Coupling chemistries for covalent attachment of DNA have been described for reaction occurring through the DNA bases and from endpoint attachment of DNA molecules to the support. The most widely used method for coupling DNA is to CNBr activated Sepharose (21) (60). In the methods where the end point attachment is not used, the inherent functional groups of nucleotide bases (e.g., the amino groups of A, C, and G or hydroxyl groups on the sugar phosphate backbone) are chemically coupled to the matrix.

The major disadvantage in coupling of inherent groups is that such reaction can modify DNA structure and affect the performance of the affinity columns. When end point attachment is used the major interaction is thought to take place between groups placed at the 5' or 3' end of DNA and the matrix. Therefore the end point attachment is preferred over the former method. But even with the end point attachment method some modification of DNA nucleotides does take place and could affect the performance of the column. In noncovalent coupling procedures the DNA sequence to be coupled is never exposed to any chemical modification and hence, may sometimes function better. The most commonly used noncovalent attachment method is immobilization of biotin labeled DNA onto streptavidin-coated supports (72) (73).

Other methods such as enzymatic synthesis (74) (75) are also available and have been discussed in details elsewhere (76).

1.2.2.3 Nature of DNA used for making sequence specific columns

The footprint region represents the minimum region of DNA required for interaction with its corresponding transcription factor. Ideally, the footprint region is the shortest sequence that can be used for making sequence specific DNA affinity columns. There are studies, which imply the use of longer DNA sequences, such as from plasmids, containing the footprint region, or concatemeric repeats of the footprint region, may function better than simple, discrete short footprints (77) (78). There is however a complete lack of comparative data to support such claims. None-the-less, the concatemeric method has been widely used and often successfully (see Table 1.1). The disadvantage of using longer sequences is that a longer sequence would contain additional DNA sequences, potentially bound by other proteins. For example, if a plasmid harboring the footprint region is used, only a small portion of the plasmid would represent the specific sequence and the rest of it would be nonspecific sequence which could be bound by other DNA binding proteins in the crude extract and hence decrease the selectivity of the column. Concatemerization on the other hand introduces new sequences which are absent in the original footprint region and again could decrease the selectivity of the column.

1.2.2.4 Elution of proteins from DNA affinity columns

1.2.2.4.1 Salt elution

When transcription factors bind DNA, they displace Na^+ and other counter-ions from the DNA (79). Thus, high Na^+ concentrations diminish DNA-protein binding. But this displacement is common to all transcription factors and thus the salt gradient is unlikely to be very selective. 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 for salt elution; NaCl and KCl are the most common salts used. Salt elution is indiscriminate and most of the proteins bound to the column are eluted with salt. Linear salt gradients may separate some of these proteins. Proteins that bind nonspecifically to the DNA may have a lower affinity for it and would presumably elute at low salt concentrations. The protein of interest should have the highest affinity for the coupled DNA and elute later in the gradient. But in practice salt elution has only moderate resolution and absolute separation is only rarely obtained with salt elution.

1.2.2.4.2 Ligand specific elution

Certain transcription factors lose their DNA binding activity when bound to specific ligands. For example, the lac repressor protein is unable to bind to DNA when it is bound by ligands such as IPTG or lactose. Such transcription factors can be eluted from DNA affinity columns by using specific ligands. Since ligands will elute only one specific protein from columns, ligand elution is highly selective and can yield highly purified protein. The major drawback of this method is that

only a small number of transcription factors are known to respond to ligands and hence ligand elution cannot be widely use. There are a very few examples of ligand specific elution in literature (72) (60).

1.2.2.5 Use of nonspecific (competitor) DNA in the mobile phase

The EMSA can be made specific for a particular transcription factor by using a nonspecific DNA sequence as a competitor to eliminate binding of extraneous DNA binding proteins. Similarly, competitor DNA can be added to the mobile phase used in DNA affinity chromatography to minimize nonspecific proteins binding to the DNA affinity column, as first proposed by Alberts and Herrick (80). Nonspecific competitor DNA has often been used in DNA affinity chromatography (see Table 1.1) to increase the purity but a systematic study on the use of competitor DNA has not been reported. It has been suggested that the highest concentration of competitor DNA that causes no inhibition in EMSA should be used as a reference and 5 times lower concentration than the reference concentrations should be used in mobile phase (78). While this provides a standard approach to choosing a competitor concentration, the concentration of specific oligonucleotide used in EMSA are low and in the range of 0.2-0.5 μM but the concentrations of DNA in affinity columns are often two orders of magnitude higher and EMSA-derived competitor concentrations may bear little on column chromatography. None-the-less, competitor DNA is widely used in DNA affinity chromatography and probably does increase the purity to a certain extent.

1.3. Conclusions

Sequence specific DNA affinity columns probably have greatest selectivity of all the methods available for purification of transcription factors. However homogenous transcription factor in only rarely obtained by using DNA affinity chromatography as the only step. Improvements in the technique such as optimization of DNA sequence used for making sequence specific DNA affinity columns, developing new methods for eluting transcription factors from these columns would help to increase the purification efficiency of DNA affinity columns.

Studies on the DNA length optimization, heparin elution and new purification techniques such as the bi-column and oligonucleotide trapping methods will be discussed in the subsequent chapters.

Chapter: 2. Comparative studies on chemically and enzymatically coupled DNA-Sepharose columns for purification of a lac repressor chimeric fusion protein

Part of this chapter was published in Journal of Chromatography (1999) 849: 403-12. Permission was obtained from the publisher to use our published article in the dissertation. Some variations in format were adapted for the dissertation.

2.1. Introduction

DNA affinity chromatography is frequently an important step in the purification of transcription factors and other DNA binding proteins. DNA affinity chromatography offers greater selectivity and hence is preferred over other methods. Affinity chromatography involves coupling of a DNA sequence to a solid support such as Sepharose, cellulose or silica. The different methods of coupling have been reviewed previously (78) (76).

Our lab developed protocols for template-directed enzymatic synthesis of DNA columns using DNA polymerase or reverse transcriptase (81) (82) (83). This procedure involves coupling of the 5' end of oligomeric (T)$_{18}$ to a solid support, using chemistry which does not result in the modification of thymidine bases. A template sequence containing a 3'-oligoadenylate tail is then hybridized with the bound sequence and the template specified sequence is copied enzymatically.

The only comparison of chemical and enzymatically produced columns was an unintended consequence of a study originally designed to improve upon the affinity purification of the FadR transcription factor (74). In that study, we reported that enzymatically prepared columns gave better yield of purified FadR protein than was obtained with a column prepared by conventional chemical coupling. This better yield was attributed to enzymatically synthesized DNA all being accessible to protein binding while chemical coupling may render some of the DNA inactive or inaccessible. In the same study we also reported that the protein eluted from the chemically coupled column appeared to be purer by gel

27

electrophoresis. Enzymatic synthesis requires an oligo-A: oligo-T primer region which was unnecessary and not used for chemical coupling. The increase in length makes enzymatically synthesized columns more complex. This region could have bound the extra proteins contaminating the FADR protein purified on the more complex enzymatically prepared support (74). Since this issue of complexity is poorly understood, we investigated it here and again addressed the issue of comparing columns produced chemically or enzymatically.

In this report, enzymatic and chemical coupling are compared using the lac operator DNA sequence and a lac repressor-β-galactosidase fusion protein. Lac repressor protein, which regulates the lac operon in *E. coli* , has been well characterized. Here, a lac repressor-β-galactosidase fusion protein (84) facilitated accurate assay of the protein, important for determining purity. This lac repressor-β-galactosidase fusion protein has DNA binding properties comparable to the native protein (85) (86). Differences in DNA complexity were negated by coupling the same DNA sequence using chemical and enzymatic coupling methods. Column dimensions and experimental conditions were also the same.

2.2. Methods

Unless stated otherwise, chemicals were of the highest purity available from Sigma Chemical Company (St. Louis, MO, USA).

2.2.1 Enzymatic synthesis

Enzymatic synthesis was essentially by the method described previously (74) except that Sequenase$^{®}$ 2.0 T7 DNA polymerase (Amersham Corp., Arlington Heights, IL, USA) which lacks 3',5'-exonuclease activity was used in place of the Klenow large fragment DNA polymerase. Briefly, 10 g of moist, suction dried Sepharose was washed thoroughly with water, 2 g of CNBr was added to it while stirring, and the mixture was maintained at pH 11 by addition of 5M NaOH until the reaction slowed. The activated Sepharose was then rapidly washed under vacuum on a coarse sintered glass funnel with 200 ml ice-cold water and then with 200 ml of 0.1 M $NaHCO_3$ pH 8.3, 0.5 M NaCl. Three grams of the activated Sepharose was reconstituted to 5 ml in the last buffer and 150 nmole of 5' NH_2-CH_2-CH_2-$(T)_{18}$ DNA was added. The mixture was mixed overnight on a tube rotator. The support was washed with 4 ml of the $NaHCO_3$ buffer and then blocked for 2 hours with 2 ml of 0.1M Tris-Cl, pH 8. The amount of $(T)_{18}$ coupled (30 nmole per g of Sepharose) was determined by the difference in the ultraviolet absorption of added DNA and that recovered from coupling in the wash fractions.

One gram of $(T)_{18}$-Sepharose was washed 3 times with 2 ml of Sequenase buffer (10 mM Tris, pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl) and resuspended in 2 ml of same buffer containing 45 nmole of $T_{18}Op1$ (5'GTGGAATTGTGAGCGGATAACAATT$(A)_{18}$). The mixture was heated to 95°C and allowed to cool slowly to room temperature with mixing. The support was then washed 5 times with 2 ml portions of Sequenase buffer. Washes included a

5 min. incubation on ice with mixing prior to centrifugation. The support was then

washed three times with 2 ml of Sequenase reaction mixture (300 μM dNTP, 5

mM DTT in Sequenase buffer). The Sepharose was resuspended in 2 ml of the

last buffer, 4 μl (52 units) of Sequenase 2.0 was added and the mixture was

incubated at 4°C for 5 min followed by incubation at 37°C for 2 hours. The

mixture was washed with 5 ml of Sequenase buffer, followed by 10 ml TE0.1 (0.1

M NaCl, 10 mM Tris, pH 7.5, 1 mM EDTA) containing 10 mM NaN_3 and stored at

4°C until needed. A portion of the support was eluted by washing repeatedly with

boiling water and the amount of complementary strand eluted was determined by

ultraviolet absorption. Support synthesized in this way had 12 nmol double

stranded DNA in the 0.56 ml bed volume column.

2.2.2 Chemical synthesis

Coupling of DNA to Sepharose was by the same protocol used for 5' NH_2-

CH_2-CH_2-$(T)_{18}$, described above, except in this case 50 nmol 5' Op1 T_{18}

(5'NH_2-CH_2-CH_2-$(T)_{18}$-AATTGTTATCCGCTCACAATTCCAC) was used per

gram activated Sepharose instead. After blocking with 0.1M Tris, pH 8, the

amount of DNA coupled was 33 nmole of Op1 T_{18}. After coupling the DNA-

Sepharose was washed with 10 ml TE0.1 (10 mM Tris, pH 7.5, 1 mM EDTA, 0.1

mM NaCl) and reconstituted in 2 ml of the last buffer. 50 nmole of Op1-$(A)_{18}$ was

added to the mixture and the mixture was heated to 95°C and allowed to cool

slowly to room temperature. It was then washed with TE0.1 containing 10 mM

NaN_3 and stored at 4°C until needed. As with the enzymatic column, a portion of

the support was eluted with boiling water and the amount of second strand eluted was determined by ultraviolet absorption. Support synthesized in this way had 11.8 nmol double stranded DNA in the 0.59 ml bed volume column used in. All oligos were synthesized by standard phosphoramidite chemistry using the university Molecular Resource Center DNA synthesis facility. 5' aminoethyl-oligos were synthesized in a similar manner except the last cycle utilizing the Amino Link reagent (Applied Biosystems) was included.

2.2.3 Production of lac repressor-b-galactosidase fusion protein

Lac repressor-β-galactosidase fusion protein was produced by growing clone BMH-72-19-1 which was a generous gift of Dr. David Levens (Laboratory of Pathology, National Cancer Institute, Bethesda Maryland). The clones were grown overnight in 2 l. superbroth (1.2% bactotryptone, 2.4% Yeast extract, 0.5% glycerol, 0.072M K_2HPO_4 and 0.028M KH_2PO_4) at 37°C and induced for 4 hours with 1 mM IPTG. The cells were pelleted by centrifugation at 8000 rpm for 30 min. in th Sorvall GS-3 rotor The pelleted cells were resuspended in 40 ml lysis buffer (4 mg/ml lysozyme (Boehringer Mannheim, Indianapolis, IN, USA), 5 mM NaH_2PO_4, 10 mM, Na_2HPO_4, 30 mM NaCl, 25 mM benzamidine, 10 mM 2-mercaproethanol, 10 mM EDTA, 1 mM PMSF and 0.2% Tween 20). The cells were lysed by sonication on ice, 30 sec. on followed by 30 sec. off, repeated three times at setting 12 using a VirSonic 50 sonicator with a microprobe (Gardiner, NY, USA). Cellular debris was removed by centrifugation at 15,000 rpm for 30 min in a Sorvall SS-34 rotor. The protein was dialyzed against 5 l. of

TE0.1. This crude protein preparation was stored at -85°C in 1 ml aliquots until needed.

2.2.4 Chromatography

Two columns, one from the chemically coupled support (0.59 ml bed volume) and the other from the enzymatically produced one (0.56 ml bed volume) were packed in 1 ml syringe columns and equilibrated in TE0.1. 500 μl of crude fusion protein was loaded onto the columns. The columns were washed with 15 ml TE0.1 and the proteins were eluted with a 20 ml linear gradient from TE0.1 to TE1.2 (10 mM Tris, pH 7.5, 1 mM EDTA, 1.2 M NaCl). One ml fractions were collected and the flow rate was maintained at 0.33 ml/min throughout.

2.2.5 Assay of lac repressor-β-galactosidase fusion protein

Lac repressor–β–galactosidase fusion protein was assayed for galactosidase activity using Buffer O (3 mM o-nitrophenyl-O–β–D-galactopyranoside, 0.1 M sodium phosphate, pH 7.5, 1 mM $MgCl_2$ and 45 mM 2-mercaptoethanol). 150 μl of buffer O was added to 50 μl of each sample to be assayed. The reaction was done in a microtiter plate and absorption at 405 nm monitored continuously at 25°C.

2.2.6 Definition of enzyme units

One unit of lac repressor–β–galactosidase fusion protein is defined here as that which causes a change of one absorbance unit per min at 25°C.

2.2.7 Gel electrophoresis

Sodium dodecylsulfate polyacrylamide gel electrophoresis was on 7.5% gels by the method of Laemmli (87) and stained with silver using a BioRad Laboratories (Richmond, CA, USA) kit.

2.2.8 Protein assay

Protein concentrations were determined by bicinchoninic acid method using the protocol provided by Pierce Chemical Co. Samples were precipitated with ice-cold 10% trichloroacetic acid and reconstituted in a reduced volume of 2% Na_2CO_3 in 0.1 M NaOH before assay.

2.3. Results

We had hypothesized previously that the complexity of a column attached DNA would affect the purity obtained in transcription factor purification (74). As a DNA sequence becomes more complex we reasoned that other sequences are produced which may be bound by other cellular proteins. Since a common practice in transcription factor purification is to use long concatemers (produced by ligating oligonucleotides to produce long stretches of DNA) containing multiple copies of the binding site (78) this issue of complexity could be quite important. Figure 2.1 shows that indeed DNA complexity does affect transcription factor binding but not quite in the way we envisioned.

The Op1 operator sequence used here and in Figure 2.2 is a 25-mer. It is shown in Figure 2.3. The complementary strand DNA (i.e., αOp1) was
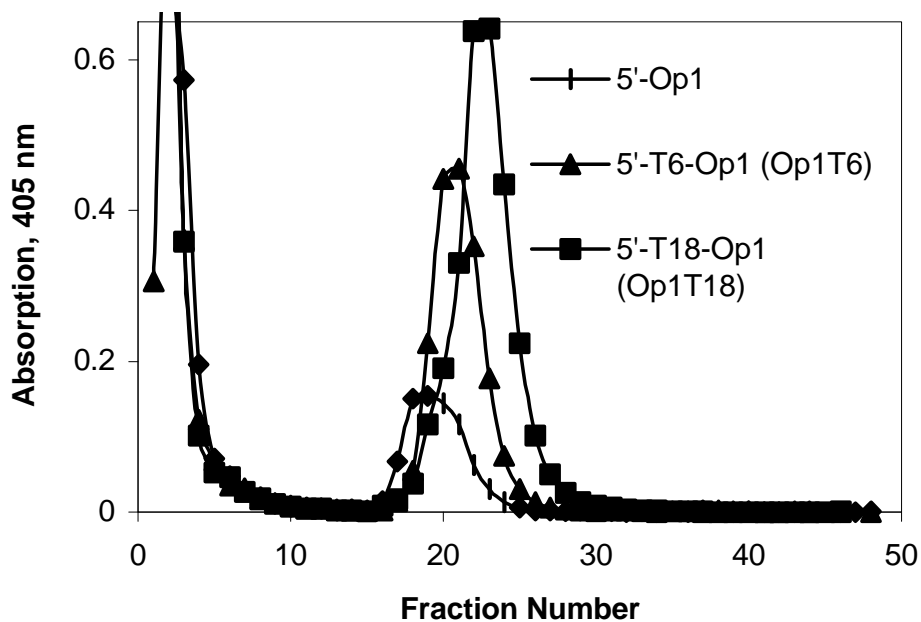
Figure 2.1. *The length of coupled DNA affects retention of laciz.*

Each of the three DNA's shown was chemically coupled to CNBr-activated-Sepharose and a 1 ml syringe column packed. Additions in each case were to the 3' end of the Op1 sequence (see Figure 2.3), at the same end of the double stranded DNA where the 5'-aminoethyl group on the complementary strand would be found. The amount of DNA coupled was 16 nmol/g Sepharose for Op1 and 32 nmol/g for the other two. For each column, 0.5 ml of the crude laciz bacterial extract was loaded. Elution was with 10 ml of constant TE0.1, followed by a linear gradient of 20 ml from TE0.1 to TE1.2, followed by 20 ml of constant TE1.2. One ml fractions were collected and assayed for β-galactosidase activity (absorption 405 nm).

34
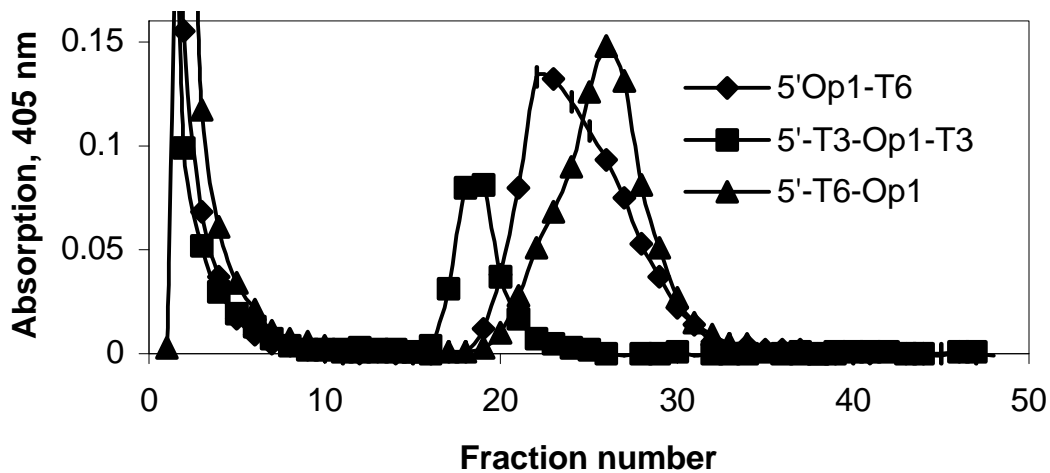
Figure 2.2. *The position of additional sequences also affects retention.*

For each of the DNA's, six additional adenylate residues were added, but the position was different as shown in the figure. The columns were all 1 ml and contained 26, 30, or 29 nmol DNA/g Sepharose for 5'-T3-Op1-T3, 5'-Op1-T6 and 5'-T6-Op1respectively. The columns were loaded and eluted as described for Figure 2.1.
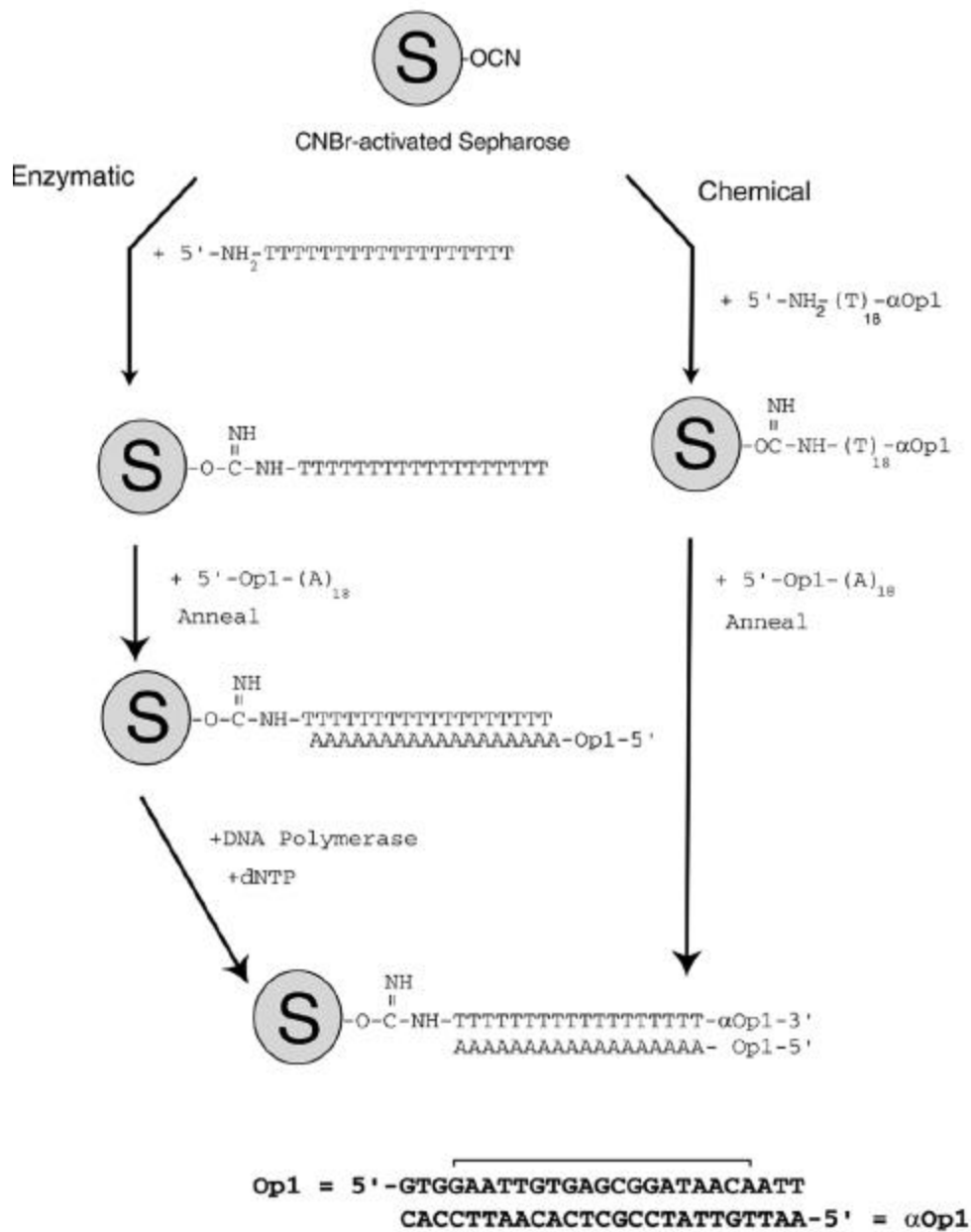
Figure 2.3. *Schematic representation of enzymatic and chemical DNA-Sepharose synthesis.*

synthesized with a 5'-aminoethyl group using the AminoLink reagent. It was directly coupled to Sepharose using CNBr activation, and annealed to the Op1 strand. Since this DNA lacks an oligoA: oligoT region, we have called it Op1 in the figure. The same DNA sequence was also synthesized containing either a six or an eighteen long 3'-oligothymidylic acid "tail" on the Op1 strand (and a complementary 5'-oligothymidylic acid tail on the other strand). These DNA's were also coupled and are called $Op1T_6$ and T_{18} in the figure. To test columns prepared from these DNA sequences, a crude bacterial extract containing a chimeric fusion protein (laciz) of lac repressor (i.e., laci) and β-galactosidase (i.e., lac z) was applied to each column and eluted using a salt gradient. The column fractions were then assayed for β-galactosidase activity in an assay, which results in an increased absorption at 405 nm for active fractions. While lac repressor would be expected to have high affinity for the Op1 sequence, it should have minimal affinity for the unrelated homopolymeric sequences and yet clearly these sequences do affect retention as shown in Figure 2.1. Since the presence of T_6 or T_{18} increases retention time, this sequence must increase the overall affinity of the repressor for the stationary phase DNA.

Furthermore, the position and distribution of homopolymeric sequences can also affect retention as shown in Figure 2.2. In this experiment, six thymidylate residues were added to one end or the other of the Op1 sequence, or three residues was added to each end. The corresponding number of adenylates were added to the complement strand. The results show that additions of three residues to each end give lower retention times than is

obtained by adding six residues to either end and that, furthermore, additions to the 3'-end of the Op1 sequence give the highest retention time of all.

Peak heights are also different in Figure 2.1 and Figure 2.2 but this is due to differences in the amount of DNA coupled to each column which affects the capacity of the columns for laciz. We have found that the amount coupled, while affecting capacity does not affect retention times for salt gradient elution though (data not included). These differences in retention are due to the DNA complexity and were consistently observed with columns containing different amounts of coupled DNA.

From these results it is clear that minor differences in sequence, even with unrelated homopolymeric regions, can affect retention. Thus, to find out if the differences previously observed for chemically and enzymatically produced columns (74) were due to the method of synthesis or to differences in sequence, we prepared columns by both procedures using the same DNA sequence. The scheme used for column synthesis is depicted in Figure 2.3. Columns prepared by chemical and enzymatic synthesis depicted in the figure would have identical double stranded DNA, despite differences in how they were synthesized.

Figure 2.4 shows the elution of laciz fusion protein from chemically and enzymatically synthesized columns. Both columns were the same dimensions and contained virtually identical amounts of the same double-stranded DNA sequence; they differ only in how they were produced. It can be seen that both columns have similar properties in terms of amount of fusion protein they bound and eluted. However, we also found that the repressor elutes at slightly lower salt

38
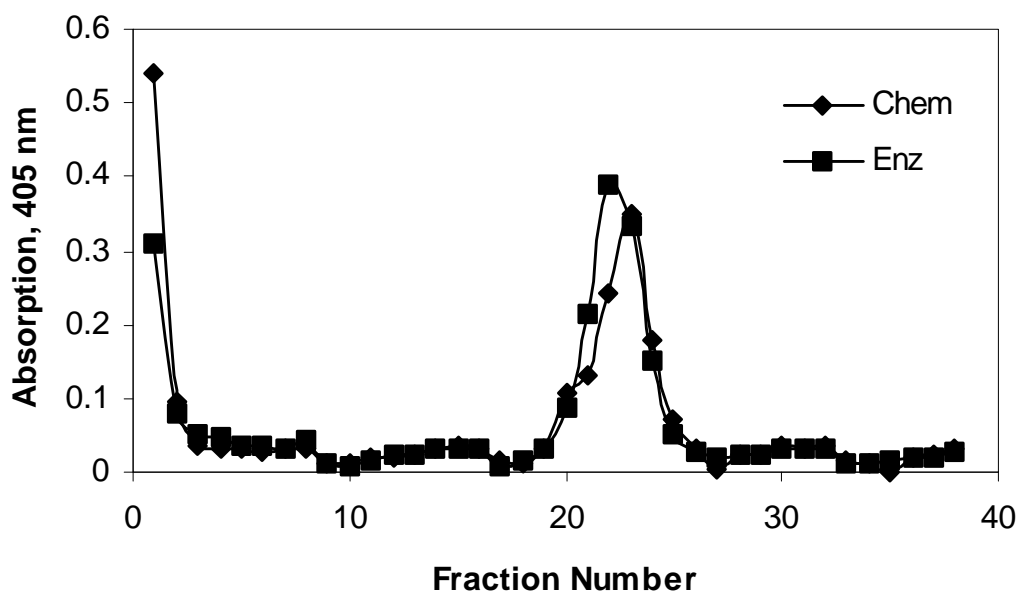
Figure 2.4. *Enzymatically and chemical synthesized supports show similar chromatography.*

Synthesis was as depicted in Figure 2.3 and described in Methods. The enzymatically prepared column was 0.56 ml bed volume and contained 12 nmol double stranded DNA; the chemically prepared column was 0.59 ml and contained 11.8 nmol double stranded DNA. Flow rate was 0.33 ml/min and 1 ml fractions were collected. The elution was with a gradient of constant TE0.1 for 90 min, a linear increase to TE2.0 (10 mM Tris, pH 7.5, 1 mM EDTA, 2 M NaCl) at 230 min, followed by 20 min of constant TE2.0.

concentration from the enzymatic column. The peak fraction elutes from the enzymatic column at 0.7 M NaCl (determined by conductivity) while the peak fraction from the chemical column elutes at 0.8 M. This result was reproducible in all our chromatographic runs and with columns made at different times and containing different amount of DNA coupled but its cause is unknown.

Figure 2.5 shows an acrylamide gel of fractions obtained from the chemical and enzymatic columns. A protein band corresponding to the expected molecular mass (84) of the lac repressor–β–galactosidase fusion protein (155 kDa) can be seen in both fractions and is indicated by the arrow. This band, and two prominent bands just below it (indicated by lines) all stain with an anti-lac repressor antibody (data not shown). Thus, the full length fusion protein and at least two truncated forms of it are bound by and elute from both columns. Only a single, minor band (indicated by an asterisk) is unique to the enzymatic column and the purity of both fractions is comparable.

Table 2.1 shows a balance sheet for the average purification and yield of protein eluted from the two columns. Yield and purity are virtually identical for the two columns. Since the columns were loaded with an excess of the fusion protein, the yield is a measure of column capacity in this experiment. Thus, column capacity is virtually identical for the two types of columns. The purification was repeated three times with each column and the results were averaged and compared statistically. There is no statistically significant difference between the yield or purity obtained with either column.
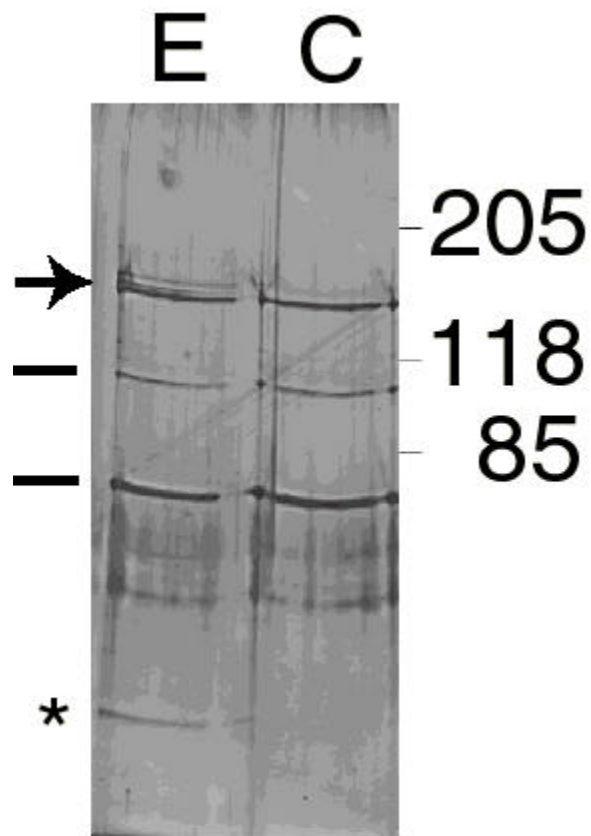
Figure 2.5. *Electrophoresis demonstrated that purity is similar for proteins purified on the chemical and enzymatic columns.*

Crude bacterial extract containing the laciz fusion protein was purified on the chemical (C) and enzymatically (E) produced column using the conditions in Figure 2.4. The peak eluted fractions were pooled, concentrated, and applied to a 7.5% sodium dodecylsulfate polyacrylamide gel [10]. The position of molecular weight markers, in kilodaltons, are shown to the left of the gel. Other symbols are described in the text.

Table 2.1. Balance sheet for purification on chemically and

enzymatically produced DNA-Sepharose columns.

Fraction	Total Activity, units	Total Protein, mg	Yield, %	Purification, fold
Crude	2.3 ± 0.1	5.2 ± 1.2	100	1
Chemical	0.62 ± 0.33	0.014 ± 0.004	33 ± 13	93 ± 20
Enzymatic	0.68 ± 0.44	0.020 ± 0.014	33 ± 13	82 ± 23
			P = 0.50	P = 0.28

[a] Shown are averages ± the standard deviation for three (N=3) different experiments on each column.

P= probability that the means for the chemical and enzymatic columns are the same.

2.4. Discussion

In a previous report from our lab (83) we had reported that a chemically synthesized column yielded purer FAD R protein while more protein could be recovered from the enzymatically synthesized column. However, the DNA sequence on the two types of columns was not the same. The enzymatic synthesis required a $T_{18}{:}A_{18}$ primer region not necessary or used for chemical coupling. Since the operator B sequence bound by FAD R in those experiments is a 30-mer, this difference amounted to coupling a 30-mer chemically or copying a 48-mer enzymatically. This makes the enzymatically produced column more complex. If this additional DNA sequence could bind other proteins, this could account for the lower purity. To answer this, here we chemically coupled the same DNA sequence used for enzymatic column production. Under these conditions, there is no difference in the purity obtained. In our studies with the purification of the lac repressor–β–galactosidase fusion protein, when identical DNA-Sepharose columns are produced by the two methods, both types of columns behaved almost identically and neither of the columns had an advantage in terms of yield or fold purification. The only difference found was that the protein eluted at a slightly lower salt concentration from the enzymatic column than from the chemical column. We do not understand the mechanism behind this behavior but it doesn't seem to be important to the chromatography.

While the method of synthesis was not important in this study, the exact sequence and complexity of DNA coupled was shown to be very important. Here, we show that even simple, homopolymeric sequences can have marked effects

43

on column performance and retention times. Why this is so is not known but could arise in at least two ways. Base pairing involves weak forces, primarily hydrogen bonding, stabilizing the double-stranded DNA structure. As DNA is heated, these weak forces are ultimately insufficient to resist thermal motion and at some temperature, DNA becomes single stranded. This "melting" temperature was never exceeded in our experiments. Even at lower temperatures though, the ends of a double-stranded DNA are not stabilized by as much hydrogen bonding as occurs mid-strand, base stacking is more solvent exposed, and localized "melting" or fraying of the ends can occur. If sequences near the end are important to transcription factor binding, this fraying could adversely affect binding. Footprinting with DNAase I (88) shows that lac repressor binds to the DNA shown in Figure 2.3 covering the entire sequence shown except for the last three base pairs at each end (indicated in the figure). By adding additional DNA to the ends, the fraying is distanced from the binding site and an intact binding site is maintained. Thus, the homopolymeric sequenced could prevent thermal denaturation of more distal regions of the operator DNA sequence.

The other potential explanation involves what is frequently referred to as "sliding" (26). It has been observed that lac repressor and operator 1 DNA associate at a rate of about 100-fold faster than three dimensional diffusion should allow (89). Lac repressor also binds "non-specific" (i.e., non-operator 1) DNA sequences with relatively high affinity. For example, under low ionic strength conditions, the dissociation constant for operator 1 is about 5×10^{-14} M while the constant for "non-specific" (alternating poly(AT) DNA) is about 10^{-9} M.

This led to the hypothesis that the repressor may bind from bulk solution (i.e., 3-dimensional diffusion driven) to any non-specific DNA sequence and then slide (i.e., 1-dimensional diffusion) along the helix until the operator is encountered and bound (90)]. In fact, any of several other mechanisms which restrict the dimensionality of diffusion would accomplish the same enhancement of rate; what is necessary is that binding be a two step process involving "non-specific" binding followed by a "sliding", "hopping", or "bridging" step which restricts diffusion to less than three dimensions (26) (27). The dependence of Op1-binding of lac repressor on both salt concentration and DNA length agrees with what would be predicted for such a sliding mechanism (27) (88) (91) (92) (93). While long DNA sequences (> 70 bp) do affect association rates in a way consistent with a sliding mechanism (93) any affect of shorter lengths is unclear and may not be measurable by current binding assay methods. Since lengths as short as six bp do affect our chromatography, chromatography may provide an even more sensitive method for measuring limited diffusion affects on protein-DNA binding.

Whether either of these possible explanations (melting stabilization or sliding) accounts for the current results is not currently known but is being investigated. However, it is interesting to note that enhanced lac-repressor binding to poly AT DNA had been noted using filter binding assays as far back as 1970 (94) is in agreement with the data presented here.

Another possible way that additions to the ends of DNA sequences could contribute to column performance is by acting as a spacer, distancing the

operator sequence from support surfaces which could sterically inhibit binding. This seems unlikely to explain the results here. Cyanogen bromide-activated Sepharose can couple directly to nucleic acid bases, presumably adenine, guanine, or cytosine (95). Here, we have introduced a 5'-aminoethyl moiety which should provide a favored coupling site but coupling elsewhere would presumably also occur. While this issue has not been directly investigated for CNBr-activated Sepharose, it was shown with activated ester coupling that coupling is predominantly through the aminoethyl but that some base modification also occurs (81). The DNA strand which was chemically coupled in the present report was always the αOp1 strand containing a 5'-aminoethyl group and any poly(T) sequences necessary to complement poly(A) regions on the complement strand. This was done since T is unreactive for coupling and so the added regions were not providing additional coupling sites. Thus, reaction can occur at the 5'-end or intrachain but not within the homopolymeric sequences introduced. An aminoethyl group would contribute about 5 Å spacing and each base pair in the DNA helix contributes 3.4 Å. The lac repressor binding site (DNAase I footprint) is three base pairs from either end. Thus, aminoethyl coupled Op1 would place the binding site about 15 Å from the Sepharose surface and, for example, $(T)_6$ would add an additional 20 Å. In Figure 2.2, the sequence 5'-T_6-Op1 has the homopolymeric region on the same end of the DNA as the aminoethyl moiety and the spacer length would be about 35 Å; while Op1-T_6-3' has the same sequence at the opposite end and a spacer length of about 15 Å. Either of these gave greater retention than the T_3-Op1-T_3 (spacer length about 25

Å). Thus, spacer length does not correlate with retention.

While the mechanism by which DNA complexity affects retention time is at present unclear, the effect is undoubtedly useful in transcription factor purification. Our previous study of FadR showed that purification on columns of different complexity impacted both yield and purity (74). Others have reported that highly complex DNA, composed of concatemers of ligated oligonucleotides are necessary for effective purification of some transcription factors (78) (77). Since complexity alters retention time (Figure 2.1 and 2.2), it also affects resolution of a specific transcription factor from other cellular DNA-binding proteins. Careful characterization of this complexity effect should greatly benefit our understanding of this affinity chromatography and improve protein purification.

Enzymatic synthesis has an inherent disadvantage in being more cumbersome than chemical coupling. Hence, chemical synthesis seems to be preferable for purification of lac repressor and probably other transcription factors. However, enzymatic synthesis does have other advantages. Techniques are available for producing columns enzymatically from either RNA (82) (83) or DNA templates (81) and the DNA support produced can be directly sequenced (83). Furthermore, since chemical coupling procedures modify nucleic acid bases while the milder enzymatic synthesis does not, it may prove more useful for the purification of proteins whose DNA binding affinity is more sensitive to chemical modification such as, for example, proteins involved in DNA repair.

Appendix

Robinson, F. D. who is also the co-author of paper on this work (75) has contributed data for the Figures 2.4 and 2.5 in this chapter.

Chapter: 3. Comparative studies on discrete and concatemeric DNA-Sepharose columns for purification of transcription factors

Part of this chapter was published in Journal of Chromatography (2001) in press. Permission was obtained from the publisher to use our published article in the dissertation. Some variations in format were adapted for the dissertation.

3.1. Introduction

Transcription factors and other DNA binding proteins are often present in cells in small amounts. Hence, purification of these proteins from cellular extracts is frequently difficult and involves several chromatographic steps. DNA affinity chromatography offers greater specificity and selectivity than any other chromatographic process and hence is frequently use in purification of these proteins (78) (74) (76).

Each transcription factor binds to a specific DNA element, which can be identified by endonuclease or chemical footprinting techniques. Specific sequence DNA affinity columns are made by covalent coupling of the specific DNA element to a variety of solid supports such as silica, Sepharose or cellulose (76) (78). Such columns are bound with greatest affinity by the protein of interest and can be used for its purification.

DNA columns containing tandem repeats of the footprint element, i.e., concatemers, were first used for purification of the Sp1 transcription factor (21) (77). Sp1 binds a GC-rich decanucleotide, the GCbox. The early promoter of the SV40 virus contains six tandem copies of the GCbox and Sp1 binding activates SV40 transcription. Thus, the use of concatemer GCbox for the purification of Sp1 followed rationally from the promoter structure. However, since the successful purification of Sp1, concatemer DNA columns have become widely used for purification of other transcription factors. Indeed, in 1991 at least fifty transcription factors had been purified using concatemers (78). Concatemers are usually made by ligating single copy DNA strands having a complementary

overhang sequence. Catenating DNA makes it more complex and can introduce new DNA sequences (that are not part of the element but are necessary for ligation) which may bind other, additional DNA binding proteins as shown in Figure 3.1. Furthermore, it is difficult to control the ligation reaction and thus concatemeric columns are, of necessity, a heterogeneous mixture of different lengths, which also could adversely affect chromatographic resolution. There is currently no report that critically investigates whether concatemer columns are better than simpler strategies for transcription factor purification.

Lac repressor protein is a regulatory protein, which controls the expression of the lac operon in *E. coli*. The lac repressor has been widely studied and the operator 1 (Op1) region of the DNA to which it binds has been well characterized (96). CAAT enhancer binding protein-α (C/EBP) is another widely studied protein, which is found ubiquitously in mammals; it binds to the CAAT element and regulates the expression of several developmental and viral genes (22) (66). In both these cases, the transcription factor binds well to discrete DNA columns and thus concatemers have not been used and any beneficial effects of concatemers on their purification is untested. However, it is known that adding simple homomeric sequences (e.g., T_{18}: A_{18} tails) increases retention of lac repressor (75) suggesting that more complex DNA sequences may prove beneficial.

In this paper, we have used green fluorescent protein-CAAT enhancer binding protein (GFP-C/EBP) and a lac repressor-β-galactosidase fusion protein (laciz) to compare the properties of discrete, concatemeric, and T_{18}: A_{18}-tailed
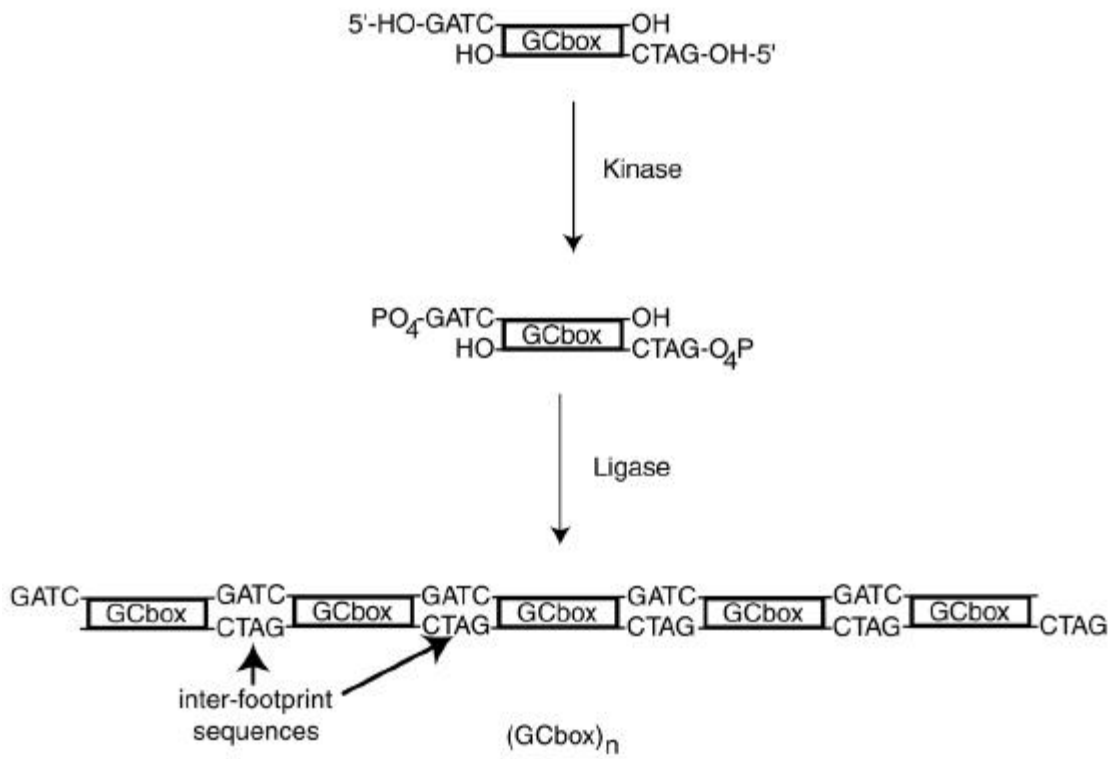
Figure 3.1. *Schematic representation of the strategy used for production of concatemeric DNA sequences.*

DNA columns. Both fusion proteins have DNA binding properties comparable to the native transcription factors (84) (97) (98). Here, we show that for some proteins such as GFP-C/EBP, discrete DNA columns give better purity and yield than the concatemeric columns. For other proteins such as laciz, which has a lower affinity for discrete columns, T_{18}: A_{18}-tailed DNA columns give better resolution than discrete or concatemeric columns. We also show that Sp1 can bind to discrete, concatemeric and T_{18}: A_{18}-tail columns with similar affinity.

3.2. Methods

3.2.1 DNA Sepharose preparation

The oligonucleotide shown in Table 3.1 were used for making DNA affinity columns. All strands having Aminolink were coupled to CNBr-preactivated Sepharose 4B (Sigma Chemical Co.). Coupling and end-capping were carried out according to the protocol provided by the manufacturer. Op1, Op1T_{18} and (Op1)$_4$ columns were made double stranded by adding the corresponding, complementary strand. The mixture was then heated to 95°C and allowed to cool slowly to room temperature. (EP18)$_5$ is self complementary and does not require the addition of a second strand.

Complementary strands of GCbox and GCboxT_{18} were annealed before coupling. The concatemeric (GCbox)$_n$ DNA sequence was made by kinasing both oligonucleotides and ligating them as described by Kadonaga (77). A schematic description of concatemer formation is given in Figure 3.1. A mixture of oligomers

Table 3.1. Oligonucleotides used for making DNA affinity columns.

Name	Sequence of strand which was coupled	Sequence of complementary strand
Op1	5'-NH$_2$AATTGTTATCCGCTCA**CAAT**TCCAC**	5'GTGGAATTGTGAGCGGATAA**C** **AAT**T**
Op1T$_{18}$	5'-NH$_2$-(T)$_{18}$AATTGTTATCCGCTCA**CAAT**TCCAC	5'GTGGAATTGTGAGCGGA**TAAC** AATT(A)$_{18}$
(Op1)$_4$	5'-NH2(AATTGTTATCCGCTCA**CAAT**TCCAC)$_4$	(5'GTGGAATTGTGAGCGGA**TAAC** AATT)$_4$
EP18	5'-NH$_2$GCAGATTGCGCAATCTGC	NA*
(EP18)$_5$	5'-NH$_2$-(GCAGATTGCGCAATCTGC)$_5$	NA*
GCbox	5'-NH$_2$GGGGCGGGGC	5'GCCCCGCCCC
GCboxT$_{18}$	5'-NH$_2$-(T)$_{18}$GGGGCGGGGC	5'GCCCCGCCCC(A)$_{18}$
(GCbox)$_n$	5'-GATCGGGGCGGGGC	5'-GATCGCCCCGCCCC

* NA stand for not applicable, EP18 and (EP18)$_5$ are self complementary and do not require the addition of a second strand.

** A CAAT element is present in both the strands of Op1 sequence.

"5'NH$_2$" represents aminoethyl group added on the last synthetic cycle with the Aminolink reagent (Applied Biosystems)

54

containing 1 to 20 copies of the original sequence was obtained as determined

by their mobility on agarose gel. This mixture was then used for coupling. The

amount of DNA coupled was determined by the difference in absorption of DNA

added before and recovered after coupling. For each gram of CNBr-activated

Sepharose, the amount (nmol) of DNA added, the amount which coupled per

gram of Sepharose, and the percent coupling are: EP18 (26.5 nmole, 21 nmol,

80%), $(EP18)_5$ (16.2, 4.9, 30%), Op1 (41.6, 19.5, 47%), $Op1T_{18}$ (42, 11, 26%),

$(Op1)_4$ (57, 6, 11%), GCbox (36, 14.9, 40%), and $GCboxT_{18}$ (20, 5.7, 30%). For

$(GCbox)_n$, molecular weight is not applicable; and per gram of Sepharose, 88 μg

DNA was added, 37 μg coupled, for a yield of 42%. In the same units, this is

comparable to the 43 μg and 49 μg coupled per gram of the GCbox and

GCboxT18, respectively. All columns were stored at 4°C in TE0.1 buffers (10 mM

Tris, pH 7.5, 1 mM EDTA, 0.1 M NaCl) containing 10 mM NaN_3 when not in use.

3.2.2 Production of proteins

Lac repressor-β-galactosidase fusion protein was produced as described

earlier (60) by growing clone BMH-72-19-1, which was the generous gift of Dr.

David Levens (Laboratory of Pathology, National Cancer Institute, Bethesda,

Maryland). The protein is used as a dialyzed crude extract (4.7 mg/ml) containing

1-1.5% of total protein as laciz. GFP-C/EBP was produced by growing *E. coli*

strain BL21 containing plasmid pJ22-GFP-C/EBP as described previously (98).

The crude bacterial extract (2.9 mg/ml, 1.4-2% of which is GFP-C/EBP)

was used for the purification experiments. The protein was also purified by Ni^{2+}-

NTA-agarose (Qiagen) as described (98).

HeLa cell nuclear extract (1.8 mg/ml) used for studies on Sp1 was obtained by the procedure in (77).

3.2.3 Chromatography

All columns were 1 ml bed volume syringe columns initially equilibrated in TE0.1 buffer. Crude preparations of laciz (bacterial extract, 4.7 mg/ml), GFP-C/EBP (bacterial extract, 2.9 mg/ml), Sp1 (HeLa cell extract, 1.8 mg/ml) or purified preparations of GFP-C/EBP were loaded onto the appropriate columns. Unless otherwise stated, all columns were then washed with 10 ml of TE0.1 and were eluted with a 20 ml gradient from TE0.1 to TE1.2 (10 mM Tris, pH 7.5, 1 mM EDTA, 1.2 M NaCl). The flow rate was 0.5 ml/min and 1 ml fractions were collected.

3.2.4 Assay of fusion proteins

GFP-C/EBP was assayed by measuring fluorescence as described earlier (98). Laciz was assayed for β-galactosidase activity by mixing 150 μl of Buffer O (3 mM o-nitrophenyl-β-D-galactopyranosidase, 0.1 M NaH_2PO_4, 1 mM $MgCl_2$ and 45 mM β-mercaptoethanol) with 50 μl of each fraction to be assayed. The reaction was carried out on microtiter plates and monitored continuously at 25°C for absorption at 405 nm.

3.2.5 Protein assay

Protein concentrations were determined by the bicinchoninic acid method using the protocol provided by Pierce Chemical Co. All samples were precipitated with 10% ice-cold trichloroacetic acid (TCA) and re-dissolved in 2% Na_2CO_3, 0.1 M NaOH before assay.

3.2.6 Blotting experiments

0.6 ml of each column fraction was applied to 0.45 μm pore nitrocellulose filter paper in a BioRad Slot-Blot apparatus, allowing the samples to slowly percolate through the filter under gravity. The membrane was then washed three times with 0.6 ml portions of TBS (20 mM Tris, pH 7.5, 0.5 M NaCl). The filter was blocked overnight with 10 mg/ml bovine serum albumin in TTBS (TBS that additionally contains 0.05% Tween 20). Next day, the membrane was washed three times with BSA/TTBS (1mg/ml bovine serum albumin in TTBS). The filter was incubated for 60 min in 1:3000 dilution of a polyclonal antibody specific for Sp1 ((PEP 2)-G, goat polyclonal IgG obtained from Santa Cruz Biotechnology). It was then washed three times with BSA/TTBS and incubated in a 1:3000 dilution of alkaline phosphatase rabbit anti-goat IgG conjugate supplied by Pierce. The membrane was washed three times with BSA/TTBS and then stained for alkaline phosphatase using a kit supplied by Promega Co. (Madison, WI, USA). For some experiments, the blots were quantified by densitometry. The blot was scanned on a Hewlett Packard ScanJet 6100 scanner and the resulting image densities calculated using NIH Image.

3.2.7 Polyacrylamide gel electrophoresis

All samples were concentrated using Centriplus centrifugal filter devices supplied by Milipore Corporation (Bedford, MA, USA). One fourth of each sample was applied to sodium dodecylsulfate polyacrylamide 4-15% Bio-Rad precast gradient gels using the method of Laemmli (87) and stained after electrophoresis with silver using the Bio-Rad Laboratory kit (Richmond, CA, USA).

3.3. Results and discussion

Figure 3.2 shows elution of highly purified GFP-C/EBP from the EP18 and the $(EP18)_5$ columns, the latter containing 5 copies of EP18 arranged as tandem repeats. The elution profile of the protein from the two columns looks similar and the protein is eluted at a NaCl concentration between 0.65-0.9 M. There is a small difference in the retention times and the proteins elutes at a slightly greater salt concentration from the $(EP18)_5$ column than from the EP18 column. This difference was consistently observed in all experiments. Both columns behave similarly and bind nearly the same amount of the purified GFP-C/EBP.

When a crude bacterial extract was used, the results are different in some respects. The purity and yield of GFP-C/EBP obtained from the EP18 and $(EP18)_5$ columns is compared in Table 3.2. Such crude extracts contain other fluorescence which prevents determining the amount of GFP-C/EBP in the extract, however, by loading the same amount on each column, the results of the two columns could be compared. The amount of GFP-C/EBP obtained from the EP18 column was significantly greater than that obtained from the $(EP18)_5$
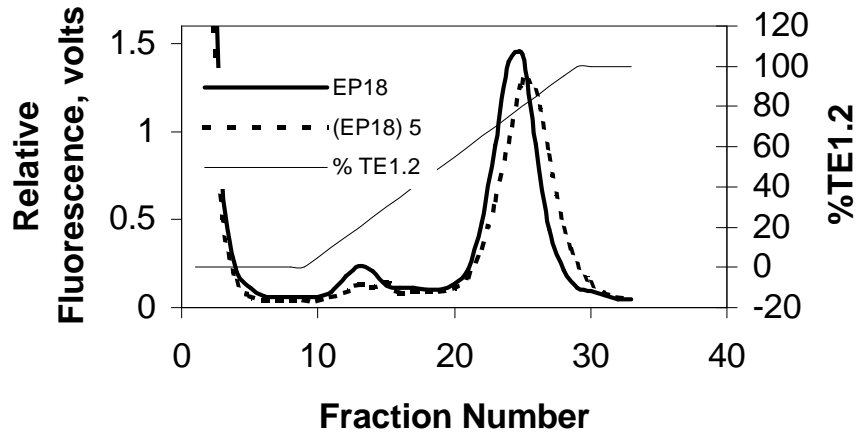
Figure 3.2. *Elution of GFP-C/EBP from EP18 and (EP18)$_5$ columns.*

100 μl of purified GFP-C/EBP was loaded on 1 ml EP18-Sepharose (solid line) or (EP18)$_5$ (dashed line) column. Chromatography was performed as described in section 3.2.3

Table 3.2. Balance sheet for purification of GFP-C/EBP from EP18

and (EP18)$_5$.

Sample	EP18	(EP18)$_5$	P
Total Fluorescence (relative fluorescence in volts)	6.4 ± 0.6^a	3.8 ± 0.8	0.008^b
Specific Activity (volts/mg of protein)	752 ± 76	596 ± 138	0.08

[a]The results of three experiments were averaged (n=3) and averages are reported for all columns. For each experiment 500 µl of crude bacterial extract containing GFP-C/EBP was loaded onto a 1 ml EP18 or (EP18)$_5$-Sepharose column. The columns were then washed and eluted as described in the section 3.2.3. Active fractions were pooled for assay.

[b]The probability (P) that the averages shown in the row are not different is given.

column with P= 0.008 (i.e., significantly different at the 0.8% confidence level). The protein obtained from the discrete EP18 column also had a higher specific activity though not significant with P= 0.081 (i.e., different at an 8% confidence level). The decreased yield for $(EP18)_5$ could be because of binding of other DNA binding proteins to the new sites (i.e., the DNA sequences between footprints in the concatemers, see Figure 3.1) that are created by oligomerization of the EP18 sequence.

Alternatively, the chemical synthesis of an 18-mer (EP18) should yield more homogeneous product while synthesis of a 90-mer such as $(EP18)_5$ may be more heterogeneous, containing fore-shortened sequences as a consequence of less than 100% efficiency at each step of synthesis. These fore-shortened sequences may have imperfect elements which function poorly or not at all.

Since DNA complexity had little effect on retention time (Figure 3.2), other ways of generating complex DNA (i.e., tailing) were not investigated.

As a consequence of characterizing columns prepared by enzymatic primer extension (75) we had shown before that binding of lac repressor protein to its operator is improved by addition of polyA:polyT tails of different length. In Figure 3.3, it can be seen that the Op1 column, containing discrete operator, binds weakly and has a lower retention time for laciz than the $(Op1)_4$ column containing four tandem repeats of Op1 or the $Op1T_{18}$ column which contains Op1 with a $T_{18}: A_{18}$-tail. The increased binding affinity of $(Op1)_4$ and $Op1T_{18}$ could be because of an effect of longer, more complex DNA on retention times as discussed in more details elsewhere (75). Interestingly, the more complex
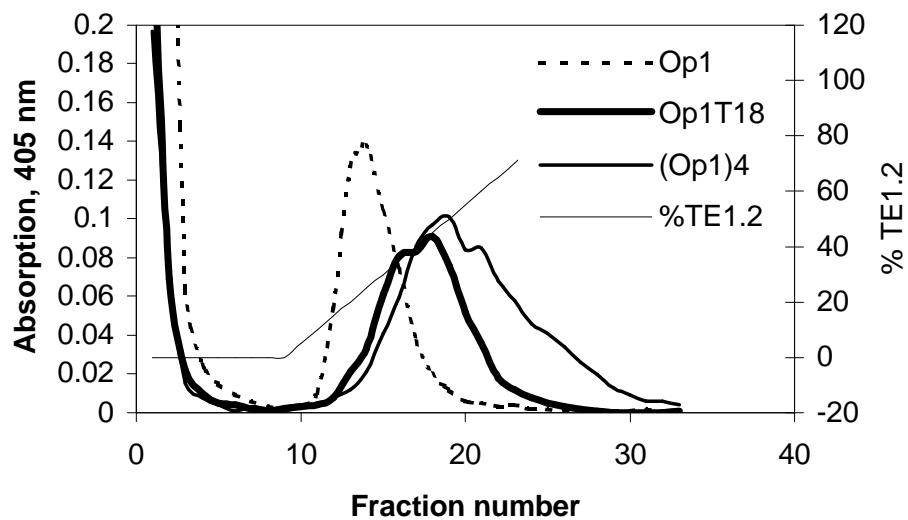
Figure 3.3. *Elution of laciz from the Op1, the $Op1T_{18}$ and the $(Op1)_4$*

columns.

100 μl of crude bacterial extract, containing laciz, was loaded onto a 1 ml

Op1, (dashed line), $Op1T_{18}$ (bold line), or $(Op1)_4$-Sepharose (solid line).

columns also appear to resolve more than one species of lac repressor fusion protein. Previously (75) we had shown this sample to contain several proteolyzed forms of laciz which may account for the peaks resolved on the $(Op1)_4$ or $Op1T_{18}$ columns.

The purity of laciz obtained from the three columns was assessed on a polyacrylamide gel. It can be seen from Figure 3.4 that laciz (indicated by the larger, darker arrow near the top of the gel) obtained from $Op1T_{18}$ column (lane3) has similar purity to that obtained from the concatemer column (lane 4) and higher purity than was obtained from the discrete, untailed column (lane 2). Another interesting observation is that all three columns behave differently and some of the contaminant bands (indicated by the lighter, smaller arrows) in the three column runs are different. Hence using some combination of these three columns in sequence would probably get rid of some of the contaminant proteins and give higher purity than any column alone would accomplish. The $T_{18}:A_{18}$ tail and catenation seem to function similarly by increasing the complexity of the DNA and this results in an increase in the retention time of laciz on the columns; increased retention time may allow weakly binding contaminants to wash from the column before the protein of interest elutes. Different purity of laciz was obtained from similar columns in other experiments (Chapters 2 and 4). These differences were observed because only the peak fractions were used in other experiments while all the active fractions were pooled here. Moreover variations in expression levels of laciz during production of crude extracts could also lead to differences in the purity obtained.
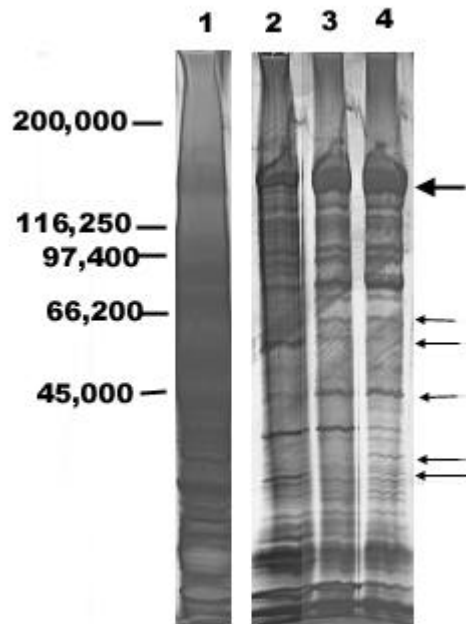
Figure 3.4. *Purity of laciz from the Op1, the Op1T$_{18}$ and the (Op1)$_4$ columns.*

500 μl of crude bacterial extract (same as in Figure 3.3) containing laciz was loaded onto a 1 ml Op1, Op1T$_{18}$ or (Op1)$_4$-Sepharose. Lane 1 shows 100 times diluted crude extract while lanes 2, 3 and 4 represent proteins obtained from Op1, Op1T$_{18}$ and (Op1)$_4$-Sepharose, respectively. The big arrow indicates the laciz protein The smaller arrows indicate contaminant proteins that are unique in different column runs. The numbers on the left indicate molecular weight (in Dalton) of standards run on the same gel.

Figure 3.4 shows that for Op1, and presumably for other DNA sequences as well, many different proteins can bind to the DNA. One way this could happen would be if two transcription factors bound elements that are both present in the column DNA. In fact, in our limited study of less than a half-dozen transcription factors, we have found two that bind to the Op1 DNA - C/EBP and laciz (60). Since we found a case in such a small group of proteins, this is likely to be a very common phenomenon. In this case, the Op1 sequence contains a CAAT element bound by C/EBP (see the bold sequence in the Table 3.1). To model the ability of simple and complex columns to separate such protein mixtures, we combined laciz and GFP-C/EBP. It can be seen from Figure 3.5A that the Op1 column shows poor resolution of the two proteins with R = 0.143. R, the resolution factor, is the ratio of the separation between the two peaks divided by the mean peak width. A value of R greater than 1 represents complete peak separation and values less than 1 indicates partial overlap with a value of 0 indicating complete overlap. The Op1 column has the biological DNA sequence that is specifically bound by lac repressor but surprisingly GFP-C/EBP binds more tightly to this column than laciz and is eluted at a higher salt concentration than laciz. As can be seen from the figure, it would be difficult to separate the two proteins with a salt gradient.

Figure 3.5B shows the resolution of GFP-C/EBP and laciz on the $(Op1)_4$ column. It can be seen that now laciz elutes later than GFP-C/EBP but the resolution of the two proteins is still poor with R = 0.133, primarily because the peaks are broad on this concatemer column. Broad peaks could result from a

Figure 3.5. *Resolution of laciz and GFP-C/EBP on different DNA-*

Sepharose columns.

100 μl of crude laciz (same as Figure 3.3) was mixed with 100 μl of

purified GFP-C/EBP and loaded onto Op1(panel A), the $(Op1)_4$, (panel B) or the

$Op1T_{18}$ column,(panel C). Fluorescence of GFP-C/EBP for early fractions is not

shown because of interfering fluorescence of crude cell constituents flowing

through the column.
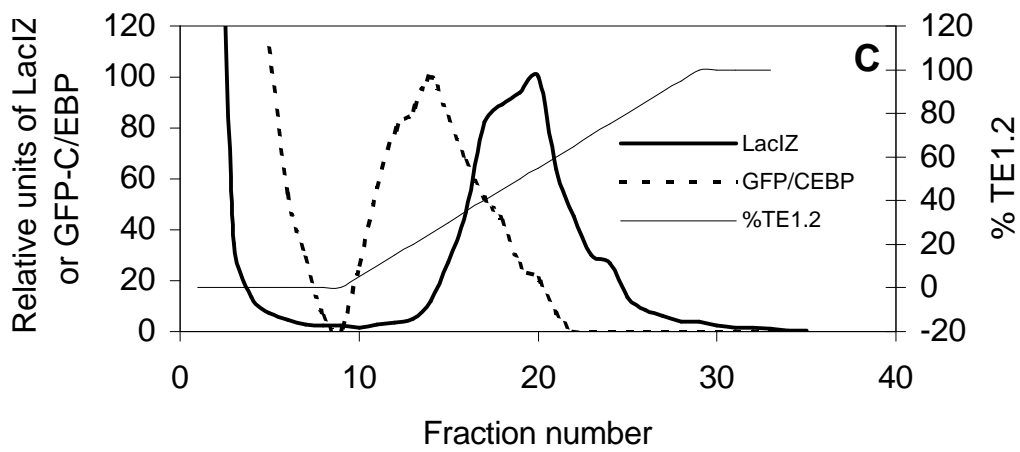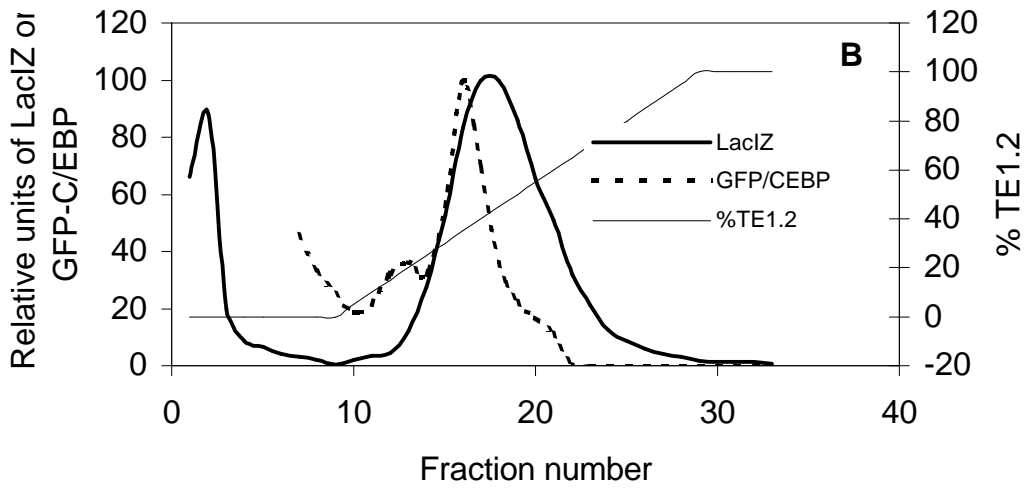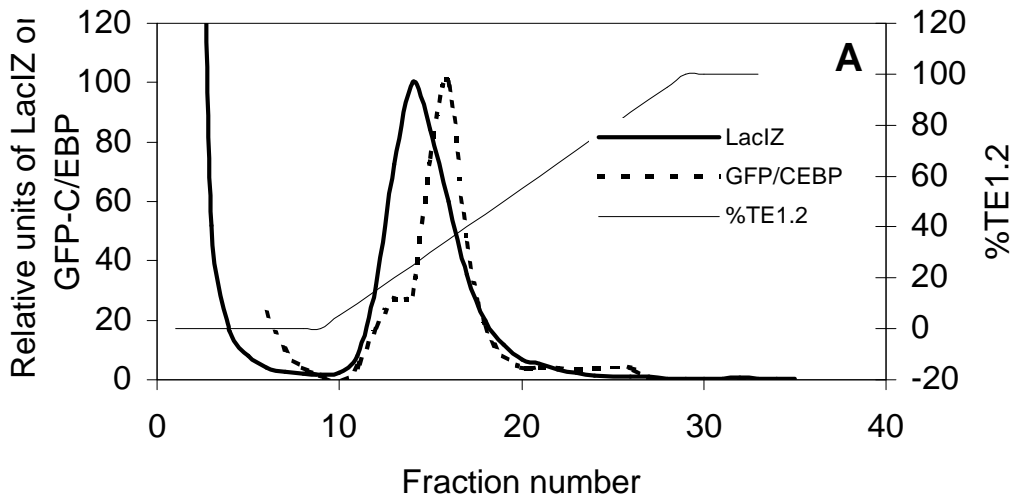
heterogeneous stationary phase which is to be expected for this 100-mer DNA, heterogeneous because of sequence fore-shortening as described above.

Of all the three columns $Op1T_{18}$ column gives the best resolution with $R = 0.4$ (Figure 3.5C). The better resolution is because of greater time differences in the retention time and because the peaks of laciz and GFP-C/EBP eluted from the $Op1T_{18}$ column are sharper than those eluted from $(Op1)_4$ column. The latter is probably a result of the more homogeneous DNA resulting from synthesis of this 43-mer. Thus, Figure 3.5 reveals the basis of the selectivity differences observed in Figure 3.4. As the length and complexity of the column attached DNA is altered, so is the retention of each of the proteins bound, the protein of interest as well as contaminants. Thus, a contaminant may co-elute or not with the protein of interest depending upon the exact DNA sequence used. Resolution is also affected by the homogeneity of the stationary phase. Tailed DNA because it is relatively short can be made chemically and in high yield. Concatemers are longer and their chemical synthesis yields shortened forms because of less than 100% efficiency of each synthesis step. Concatemers produced by ligation would also be heterogeneous mixtures. This contributes heterogeneity to the stationary phase and this probably accounts for the lower resolution of concatemers compared to tailed sequences.

Concatemeric columns were first used for the purification of the transcription factor Sp1. We investigated whether catenated $(GCbox)_n$ columns made specifically to bind Sp1 behaved differently than the corresponding discrete (GCbox) or $GCboxT_{18}$ columns. This would certainly be the case if, as previously

reported, Sp1 binds best to concatemer columns (77). To our surprise Sp1 binds to all three columns equally well (Figure 3.6) and can be seen eluting in fractions 13 to 21 (0.4 – 0.75 M NaCl) from all three columns. We do not find any apparent differences in the affinity of Sp1 for GCbox, $(GCbox)_n$ or $GCboxT_{18}$ columns and the protein eluted in the same fractions from all three columns. While Figure 3.6 shows that the elution behavior was similar for all three columns, Figure 3.7 shows that $GCboxT_{18}$ column gives a better yield of Sp1 than either GCbox or $(GCbox)_n$ columns. For this experiment the active fractions from the three columns were each pooled, two-fold serial dilutions of each pool were made and blotted onto nitrocellulose paper and Sp1 was detected with a specific antibody. It can be seen from the figure that a similar amount of Sp1 elutes from the GCbox and $(GCbox)_n$ columns but the amount eluted from the $GCboxT_{18}$ column is about four-fold higher. This $GCboxT_{18}$ column had 14-32% more DNA coupled to it (see the Methods section) than the GCbox or $(GCbox)_n$ columns, but that alone cannot account for the four-fold increase in protein obtained from $GCboxT_{18}$ columns. The higher yield obtained with the $GCboxT_{18}$ column could be because the T_{18} tail acts as an inert spacer, making more of the element accessible. Alternatively, the length of this DNA may facilitate binding by a sliding model mechanism we discussed previously (75). The $(GCbox)_n$ column, on the other hand, has several tandem repeats of the GCbox. Ligation introduces additional DNA sequences in the inter-footprint sequences used for ligation as can be seen from Figure 3.1. These additional sequences may be binding sites for other DNA binding proteins; and the binding of these proteins to the DNA
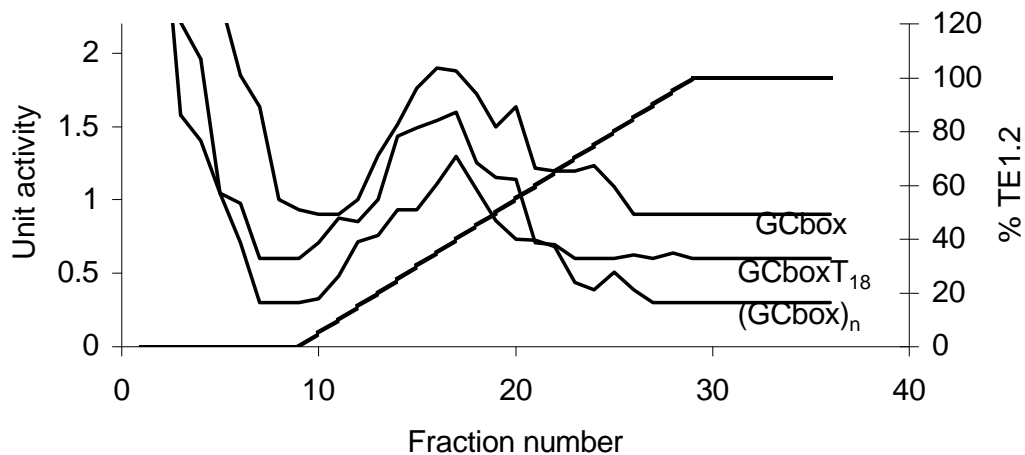
Figure 3.6. *Elution of Sp1 from different GCbox columns.*

500 μl of HeLa Cell nuclear extract was loaded onto a 1 ml GCbox, (Gc

box)$_n$, or GCboxT$_{18}$-Sepharose column. 600μl of each fraction (1 ml) was blotted

onto nitrocellulose paper using Biorad slotblot apparatus and probed with anti

Sp1 (PEP 2) antibody. Intensity of color developed in each fraction was

measured by densitometry. Values from each column run were normalized and

displaced by 0.3 units so that they could all be plotted on the same graph.

GCbox   (GCbox)$_n$ GCboxT$_{18}$ dilution factor

			2
			4
			8
			16
			32
			64
			128
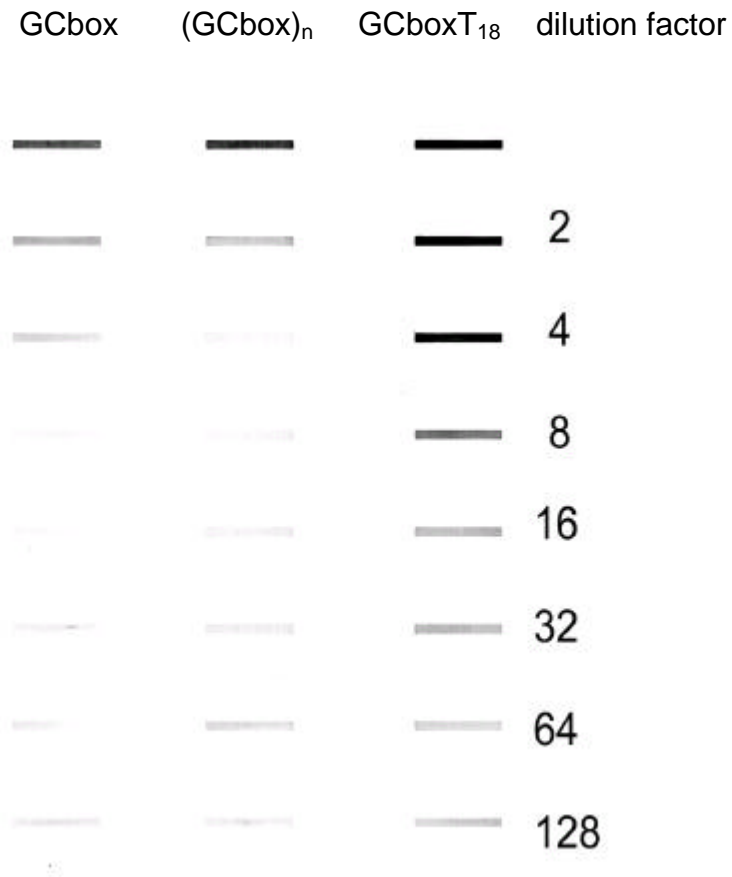
Figure 3.7. GcboxT$_{18}$-Sepharose yields more Sp1 than Gcbox and (Gcbox)$_n$-Sepharose.

Peak fractions from each of the columns showing Sp1 in Fig 4 were pooled and two-fold serial dilutions were made. 600 µl of each dilution was blotted on nitrocellulose paper and probed with anti Sp1 (PEP 2) antibody.

71

could block the binding site for Sp1 and decrease yield. Alternatively, the protracted procedure for preparing concatemers may have rendered some of the DNA inactive. We did not find any significant difference in the purity of Sp1 obtained from the three columns (data not shown) and hence there is no distinct advantage of using concatemeric columns for the purification of Sp1.

This comparison of transcription factor chromatography has resulted in some important conclusions. The length and complexity of the DNA attached to the columns affects their retention of transcription factors. As the length and complexity of the DNA is increased, retention time increases for all three transcription factors tested. The magnitude of the effect though is quite individual for each protein. Retention of C/EBP is shifted only slightly as column attached DNA is changed from an 18-mer to a concatemer five times as long (Figure 3.2). Contrasted to this is the case of lac repressor, which shows dramatic changes in retention (and resolution of proteolyzed repressors forms) as the Op1 25-mer is extended by only an additional 18 bases with a T_{18} tail. Further lengthening to a 100-mer concatemer increases retention a little more (Figure 3.3). Of these two types, Sp1 falls in the first group since it elutes from simple and complex columns with quite similar retention times (Figure 3.6). This comparison suggests that using two columns that differ in DNA complexity would be a prudent strategy for resolving difficult mixtures of DNA-binding proteins. Since the largest effect of complexity occurred with a modest increase in length, simply extending DNA with a T_{18} tail or not would be a good choice for these two columns.

This individual effect of DNA complexity on each protein makes resolution

of complex mixtures by affinity chromatography alone more likely. This is seen in Figure 3.4, Table 3.2, and in Figure 3.5. The contaminants found in C/EBP depends upon the column used (Figure 3.4). This results from this individual behavior. As column DNA becomes more complex, some proteins elute later while others are little affected. The protein of interest may be sensitive or insensitive to complexity, but in either case, the group of co-eluting (contaminant) proteins will differ on different DNA columns (Figure 3.4) and purity may be improved by altering the column DNA sequence (Table 3.2). This was shown most dramatically by investigating two proteins which bound to the same column (Figure 3.5). What happened was that on simple DNA columns these two proteins (C/EBP and lac repressor) co-elute. Increasing DNA complexity moved lac repressor to later retention times while C/EBP was still eluting early. This increased the likelihood the two could be resolved. This models what happens during purification: DNA columns bind multiple proteins that are resolved when their retention times or peak widths are altered favorably.

The resolution of chromatography is improved as the stationary phase support is made more homogeneous. Supports with uniformly coupled short DNA sequences would be the most homogeneous. As DNA is made more complex, either by chemical synthesis or ligation, the DNA is made more heterogeneous, because neither synthesis nor ligation is 100% efficient. There is also the potential that either could result in less of the DNA being active, another form of heterogeneity which also affects column capacity. Inefficient chemical synthesis results in differing lengths and some "footprint" binding regions being defective.

Ligation can result in circular DNAs and, by crowding footprints together on a DNA strand, steric crowding may make some inaccessible. Ligation also makes DNAs of very different lengths, which represents a major form of heterogeneity. The predicted outcome of such heterogeneity is seen in the data here. The more complex DNAs give somewhat broader peaks, which probably results directly from stationary phase heterogeneity (Figure 3.2, 3.2, and 3.4). Chromatographically inactive DNA is probably the cause of the decreased yield demonstrated in Table 3.2 and Figure 3.7. Thus, while DNA complexity can aid in purification, it has the harmful consequence of sometimes increasing peak width and decreasing column capacity.

The origin of concatemere DNA-affinity chromatography is the purification of Sp1, where concatemers mimic the repetitive GCbox of the early promoter of SV40 (21). However, our studies suggest that a non-concatemer column with a GCbox made more complex with a T_{18} tail should actually work as well or better. An unpublished study also showed that Sp1 binds to a non-concatemer column. The promoter of TFIIIA contains three elements (E1, E2, and E3), which bind transcription factors regulating TFIIIA expression in *Xenopus* oocytes. The proteins which bind these elements were tentatively named B1, B2, and B3 (99). E2 DNA was produced with a T18 tail coupled to Sepharose and this DNA-Sepharose was successfully used to purify B2. The purified protein turned out to be the *Xenopus* homolog of Sp1 (unpublished data, W.L. Taylor and W.T. Penberthy). Since the identity of B2 wasn't previously known, it is clear from these results that Sp1 not only binds tailed, discrete sequences but that this

binding is sufficiently specific to allow purification.

Furthermore, transcription factors and other DNA-binding proteins have been purified by others using discrete DNA-Sepharose without apparent difficulty (100) (101) (102) (103). Such discrete oligonucleotides are also effectively used in most electrophoretic mobility shift assays (104) (including Sp1, (21)) and in filter binding assays (89). Indeed, it is clear that transcription factors bind discrete DNA sequences very well, with high affinity and specificity. Since simple, discrete DNA sequences are sufficient for binding in all these cases, it is unlikely that chromatography alone would require concatemers.

All these results suggest that there is no clear advantage to concatemeric columns. For certain protein including C/EBP, columns made with just the discrete DNA element give protein with high purity and yield. For other proteins such as lac repressor which do not bind as well to discrete elements, columns having a footprint region extended with a simple sequence like T_{18}: A_{18} work better than concatemeric column.

Making concatemeric columns involves more work and resources. Sometimes it is hard to get the ligation to work in the first step and ligation has to be repeated several times (78). This is not only time consuming but also leads to loss of oligonucleotides during each step. Our study shows concatemeric columns don't have any distinct advantage for the three different transcription factors we studied including Sp1, the original justification for the concatemeric approach. Hence columns having just the footprint region or footprint region extended with a simple DNA sequence would be more suitable for the purification

of transcription factors.

Chapter: 4. Heparin elution of transcription factors from DNA-Sepharose columns

Part of this chapter was published in Journal of Chromatography (1999) 848: 131-8. Permission was obtained from the publisher to use our published article in the dissertation. Some variations in format were adapted for thedissertation.

4.1. Introduction

DNA-binding proteins such as transcription factors, DNA and RNA polymerases and exo- and endonucleases play important roles in cellular function, differentiation and regulation. Hence, purification and characterization of these proteins is of great importance. The low amount of these proteins which frequently occur in cells makes their purification challenging. DNA affinity chromatography has been widely used for purification of DNA-binding proteins (78) (77) (105). Specific or nonspecific DNA sequences are coupled to solid supports such as silica, Sepharose or cellulose. The different coupling procedures and supports have been previously reviewed (76) (78). After loading of proteins on these columns, the proteins are usually eluted with a salt gradient (76) (78). While DNA affinity chromatography is the most selective method currently available for the purification of these proteins, only rarely is a homogeneous transcription factor obtained using this method alone.

Heparin-Sepharose chromatography has also been widely used for purification of a large number of different proteins, including DNA-binding proteins (106) (107). For the latter proteins, the polyanionic structure of heparin presumably mimics the highly negatively charged backbone of DNA giving heparin the ability to bind the DNA-binding domains of protein (108). Conversely, some DNA sequences are known to bind the heparin-binding motif in thrombin (109). Heparin is typically coupled to Sepharose by cyanogen bromide activation (110) and a salt gradient is usually employed to elute proteins from heparin columns. Elution of proteins from heparin-Sepharose with heparin has also been

reported (111). Since heparin-Sepharose is also used in the purification of many other types of proteins, other protein types often co-eluted with DNA-binding proteins, affecting purity.

Of all the methods which can be used for purification of DNA-binding proteins, specific sequence DNA affinity chromatography offers the highest selectivity. In spite of this high selectivity, purification to homogeneity is seldom obtained with this technique alone. The specificity of DNA affinity chromatography could be better exploited by using it repetitively (78). However, repeating chromatography using the same or similar protocols would be of little use. What is needed is to elute the highly selective columns each time using different strategies based upon different principles. In practice, eluting protein from DNA affinity columns is currently limited to using high salt and, for a few proteins, elution by specific ligand; e.g., lac repressor elution with isopropyl β-D-thiogalactopyranoside (IPTG), or other lactose analogs. This limits the number of times that DNA affinity chromatography can be effectively used in protein purification protocols. Hence new elution methods, based upon different principles, should facilitate their purification.

Lac repressor, which regulates the lac operon in *E. coli* (112) and CAAT enhancer binding protein C/EBP (10) which binds the CAAT element in eukaryotic promoters, are both well-characterized transcription factors. The DNA sequences, which are specifically bound by these proteins, have been identified (96) (66) In this paper we have used a lac repressor-β-galactosidase fusion protein (laciz) and a Green Fluorescent Protein-CAAT enhancer binding protein

(GFP-C/EBP) fusion protein to characterize a new method for eluting DNA-binding proteins. Both laciz (84) (97) and GFP-C/EBP (98) have DNA-binding properties comparable to lac repressor and C/EBP which justifies their use in our model study. Here, we show that these proteins elute from DNA columns with heparin in a highly purified state.

4.2. Methods

4.2.1 DNA Sepharose preparation

For studies with laciz, $Op1_T6$ oligonucleotide (5'-NH_2-(T) $_6$-AATTGTTATCCGCTCACAATTCCAC), containing the operator 1 (Op1) DNA sequence, was coupled to CNBr activated Sepharose. For studies with GFP-C/EBP, EP18 (5'-NH_2-GCAGATTGCGCAATCTGC) was coupled. The "5'-NH_2" represents the aminoethyl group added on the last synthesis cycle with the Amino Link reagent (Applied Biosystems). $Op1T_6$- and EP18- Sepharose were prepared as described previously (74). Two grams of Sepharose 4B was washed thoroughly with water. 0.4 g of cyanogen bromide was added while stirring and the mixture was maintained at pH 11 by addition of 5 M NaOH until the reaction slowed. The activated Sepharose was then washed rapidly under vacuum on a coarse sintered glass funnel with 100 ml ice-cold water and then with 200ml ice-cold coupling buffer (100 mM $NaHCO_3$, pH 8, 500mM NaCl). The Sepharose as then mixed with 100 nmol of 5'-amino $OP1T_6$ or 5'-NH_2- EP18 to a final volume of 5 ml coupling buffer and mixed on a tube rotator overnight. The next day the

DNA-Sepharose was washed with 15ml 100 mM $NaHCO_3$ pH 8, 500mM NaCl.

The amount of DNA coupled, 22 nmol αOP1T6 and 29 nmol EP18 per gram of

Sepharose were determined by UV absorption spectroscopy of the DNA before,

and recovered after, coupling. The DNA-Sepharose was end-capped by adding

4ml of blocking buffer (100 mM Tris, 500 mM NaCl, pH 8) and incubating on a

tube rotator for 2 hours at 4°C. After incubation, the Sepharose was washed 2

times with 4 ml of TE0.1 buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 0.1 M NaCl).

$Op1T_6$ was made double stranded by adding 50 nmol of complementary strand

(5'GTGGAATTGTGAGCGGATAACAATTAAAAAA), the mixture was heated to

95°C, and allowed to cool slowly to room temperature. We refer to this double-

stranded product here as $Op1T_6$-Sepharose. The EP18 oligonucleotide is self-

complementary and a double-stranded column results directly from coupling. All

columns were stored at 4°C in TE0.1 buffer containing 50 mM NaN_3 when not in

use. The DNA concentration for different columns of Op1-Sepharose was

adjusted by diluting DNA-Sepharose with CNBr activated, and end-capped

Sepharose, which had beeen treated the same way except no DNA was used.

4.2.2 Production of proteins

Lac repressor-β-galactosidase fusion protein was produced by growing

clone BMH-72-19-1 which was the generous gift of Dr. David Levens (Laboratory

of Pathology, National Cancer Institute, Bethesda, Maryland). The clones were

grown overnight in 2 L. Superbroth (1.2% Bactotryptone, 2.4% Yeast extract,

0.5% glycerol, 0.072M K_2HPO_4 and 0.028M KH_2PO_4) at 37°C and then induced

for 4 hours with 1 mM IPTG. The cells were pelleted by centrifugation at 8000

rpm for 30 min in the Sorvall GS-3 rotor The pelleted cells were resuspended in

40 ml lysis buffer (4 mg/ml lysozyme (Boehringer Mannheim, Indianapolis, IN,

USA,), 5 mM NaH_2PO_4, 20 mM, Na_2HPO_4, 30 mM NaCl, 25 mM benzamidine,

10 mM 2-mercaptoethanol, 10 mM EDTA, 1 mM PMSF and 0.2% Tween20). The

cells were then lysed by sonication on ice for 30 sec, followed by 30 sec

incubation on ice without sonication; this was repeated 3 times at setting 12

using VirSonic50 sonicator with a microprobe.(Gardiner, NY, USA). Cellular

debris was removed by centrifugation at 15,000 rpm for 30 min in the Sorvall SS-

34 rotor. The protein was dialyzed against TE0.1 buffer and stored at -85°C until

needed.

GFP-C/EBP was produced by growing *E. coli* strain BL21 containing

plasmid pJ22-GFPC/EBP as described previously (98).

4.2.3 Chromatography

All columns were 1 ml bed volume syringe columns initially equilibrated in

TE0.1 buffer. Crude preparations of either laciz or GFP-C/EBP was loaded onto

the column and eluted with a gradient of NaCl, heparin or IPTG. Details of the

gradient are given in the figure legends.

4.2.4 Assay of fusion proteins

GFP-C/EBP was assayed by measuring fluorescence as described earlier

(98) Laciz was assayed for β-galactosidase activity by mixing 150 µl of Buffer O

(3 mM o-nitrophenyl-β-D-galactopyranosidase, 0.1 M NaH_2PO_4, 1 mM $MgCl_2$

and 45 mM mercaptoethanol) with 50 µl of each fraction to be assayed. The

reaction was done on microtiter plates and monitored continuously at 25°C for

absorption at 405 nm.

4.2.5 Definition of enzyme units

One unit of laciz fusion protein was defined as that which gives a change

in 1 absorption unit (at 405 nm) per min per ml of enzyme at 25°C.

4.2.6 Protein assay

Protein concentrations were determined by bicinchoninic acid method

using the protocol provided by Pierce Chemical Co. All samples were

precipitated with 10% ice-cold TCA and redissolved in 2% Na_2CO_3, 0.1 M NaOH

before assay.

4.2.7 Polyacrylamide gel electrophoresis

Peak fractions from each column run were used for gel electrophoresis.

Sodium dodecylsulfate polyacrylamide gel electrophoresis was carried out on

12% gels using the method of Laemmli (87) and stained with silver after

electrophoresis using the kit provided by the Bio-Rad Laboratory (Richmond, CA

USA).

4.3. Results

Figure 4.1 shows that GFP-C/EBP can be eluted from EP18 DNA-column with either a salt or heparin gradient. It was shown earlier that GFP alone does not bind to the DNA Sepharose column (98). The fusion protein does not bind to Sepharose with no coupled DNA (data not shown). This indicates that the binding of this fusion protein to the column was due to the specific interaction between C/EBP and the specific DNA sequence coupled to the column, an interaction disrupted by high salt or heparin concentrations.

Since this is the first demonstration of heparin elution in DNA affinity chromatography, we next investigated whether other transcription factors can be eluted in the same way.

The elution of laciz from the $Op1T_6$-Sepharose column is shown in Figure 4.2. The amount of laciz eluted is dependent on the concentration of heparin in the mobile phase; however, concentrations of more than 40 mg/ml have a high viscosity and were not used routinely. A negative control, β-galactosidase, does not bind to the DNA-Sepharose column. Laciz also failed to bind to the column containing only Sepharose indicating the specific interaction between laciz and lac operator sequence bound to the column provides the basis for this chromatography as well. Thus, elution is dependent upon the heparin concentration of the mobile phase.

Figure 4.1 and 4.2 also show considerable material flowing through the column unretained. The columns in these experiments were intentionally overloaded so that we could gauge the ability of salt or heparin to elute all of the
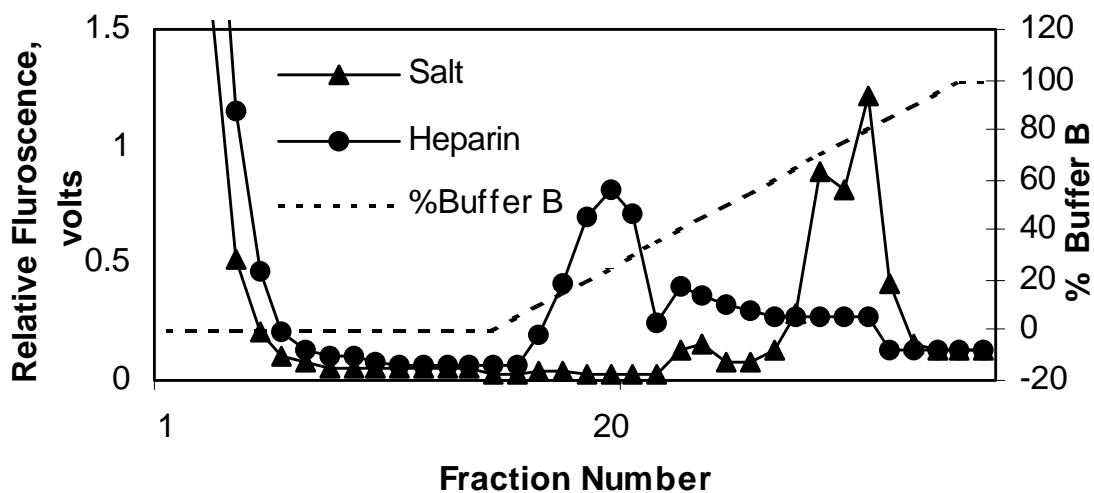
Figure 4.1. *Elution of GFP-C/EBP with heparin from EP18-Sepharose.*

A crude bacterial extract (500 µl) containing GFP-C/EBP was loaded on a

1 ml EP18-Sepharose column having 29 nmol DNA/g of resin. The column was

washed with 16 ml TE0.1 and was eluted either with a 20 ml heparin gradient

from 0-40 mg/ml heparin dissolved in TE0.1 or with a 20 ml salt gradient from

TE0.1 to TE1.2 (10 mM Tris, pH 7.5, 1 mM EDTA, 1.2 M NaCl). The flow rate

was 0.5 ml/min and 1 ml fractions were collected.
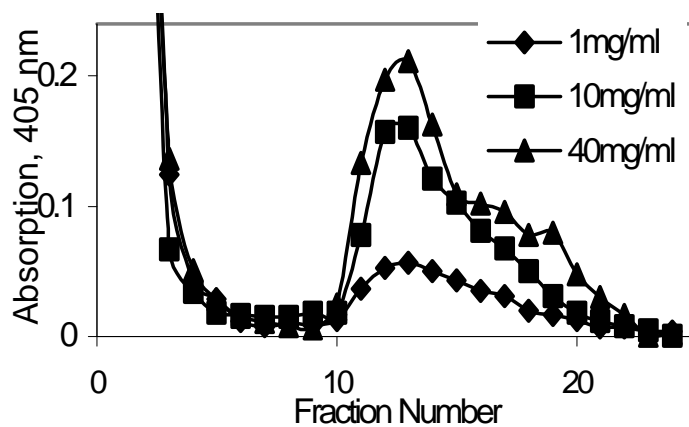
Figure 4.2. *Effect of heparin concentration on elution of laciz from Op1T$_6$-Sepharose.*

Bacterial extract (100 µl) containing laciz was loaded onto a 1 ml Op1T$_6$-Sepharose column having 4.4 nmol DNA/g of support. The column was washed with 10 ml of TE0.1 and eluted with either 1 mg/ml, 10 mg/ml or 40 mg/ml heparin dissolved in TE0.1. The flow rate was 0.5 ml/min min and 1 ml fractions were collected and assayed.

retained protein and to obtain large peaks which clearly show where elution occurred. However, the size of the unretained peak is also unrepresentative of the amount of the transcription factor which did not bind and overestimates it. Crude bacterial extracts contain multiple fluorescent compounds (Figure 4.1) and glycosidases (Figure 4.2; e.g., β-galactosidase) which are also present in these unretained fractions and increase the signal obtained. When these fractions are analyzed by sodium dodecylsulfate – acrylamide gel electrophoresis and other methods (data not shown), we find that most of the DNA-binding protein binds to the column and is then eluted by either salt or heparin. After washing away these contaminants, the signal (fluorescence or enzymatic activity) of the eluted peak does accurately reflect the amount of the fusion protein eluted.

Figure 4.3 shows that more laciz elutes with 10 mg/ml heparin from columns having lower amounts of DNA. Columns with lesser amounts of DNA coupled also bound slightly less of the applied protein and yet, even though less bound, more total laciz is obtained from the 4.4 nmol column elution. This is because even though more laciz binds to high DNA columns, most of it cannot be eluted by the fixed, 10 mg/ml heparin concentration used in the figure. Higher concentrations of heparin or salt would be required for complete elution.

While laciz has its highest affinity for its specific operator sequence, the lac repressor (and many other transcription factors) will also bind to non-specific DNA sequences, albeit with lower affinity. Figure 4.4 shows that laciz binds to EP18-Sepharose, a non-specific DNA in this context, and elutes with heparin. The EP18 sequence is not a normal consensus operator sequence for

Figure 4.3. *The effect of DNA concentration on elution of laciz from*

Op1T$_6$-Sepharose.

Crude bacterial extract (100 μl) containing laciz was loaded onto 1 ml

Op1T$_6$-Sepharose columns having either 4.4 nmol, 11 nmol or 22 nmol DNA/g of

Sepharose prepared by mixing 22 nmol/g DNA-Sepharose with control

Sepharose. The columns were washed with 10ml of TE0.1 and eluted with

10mg/mL heparin dissolved in TE0.1. The flow rate was 0.5 ml/min min and 1 ml

fractions were collected
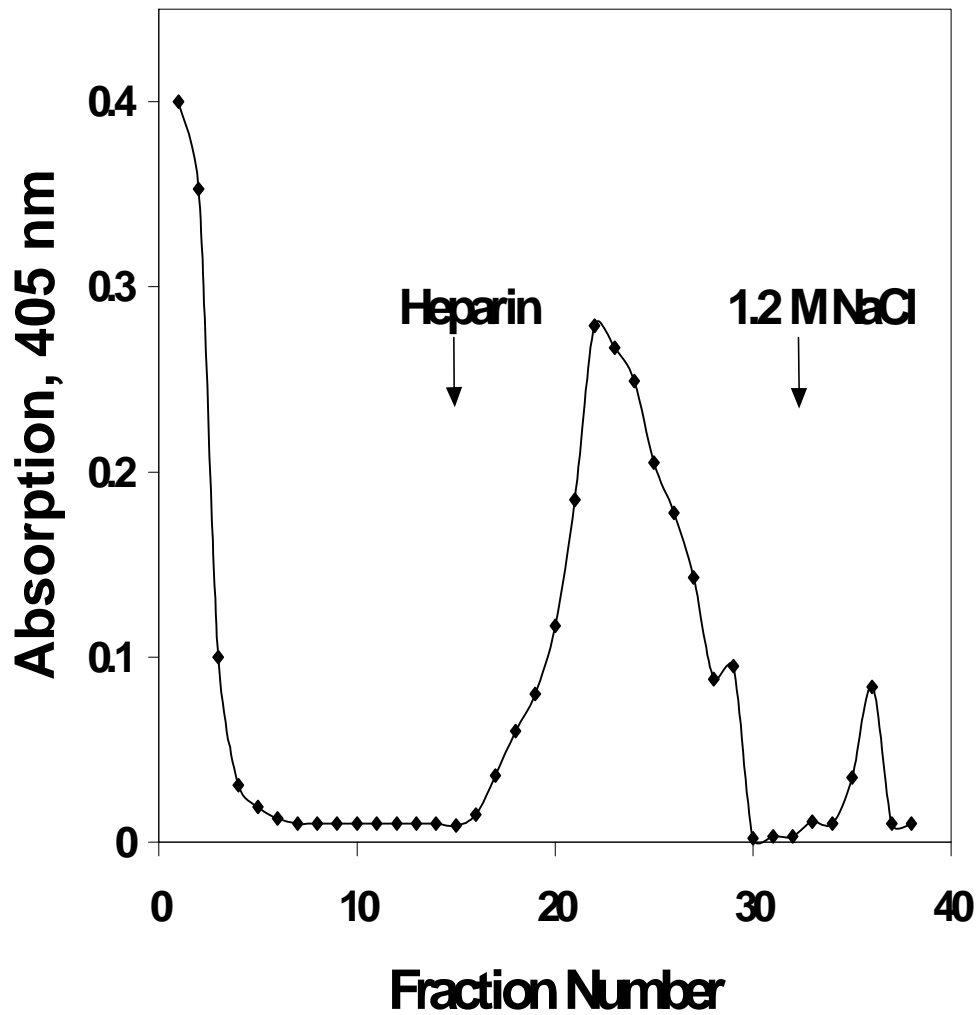
Figure 4.4. *Elution of laciz from the EP18 column.*

Crude bacterial extract (500 µl) containing laciz was loaded onto the
EP18-Sepharose column having 29 nmol DNA/g of resin. The column was
washed with 15 ml of TE0.1 and was eluted with a 20 ml heparin gradient from 0-
40mg/ml heparin in TE0.1. Any uneluted protein was then eluted with 5 ml of
TE1.2. The flow rate was 0.5 ml/min min and 1 ml fractions were collected.

the lac repressor; the two sequences only have the CAAT element in common between EP18 and Op1 DNA. However, large amounts of laciz bind to the EP18-Sepharose column (Figure 4.4). A negative control, β-galactosidase does not bind to EP18-Sepharose (data not included).

Figure 4.4 demonstrates that most of the bound protein is eluted by heparin and very little additional laciz can be eluted with high salt. This EP18 column contains a relatively high amount of DNA (29 nmol/g Sepharose) and yet heparin elutes it completely. Even when lesser amount of a specific DNA were used in Figure 4.3, complete elution was not obtained at these heparin concentrations. Thus, the specificity of the DNA-sequence used for the stationary phase also influences the heparin concentration required for elution. Thus, with two different transcription factors and with two different DNA-Sepharose columns, heparin is an effective eluent and elutes proteins which are specifically as well as nonspecifically bound. Heparin-elution is probably a general phenomenon which is widely applicable.

The purity and yield of protein obtained from the different elution methods is compared in Table 4.1. We found that the heparin pool was very pure as compared to the salt pool. The fold purification of the heparin pool was significantly greater than the salt pool with P= 0.014 (i.e., significantly different at the 1% confidence level). The lower yields obtained for both heparin and salt were because of loading a large excess of protein onto the small columns so that sufficient protein eluted to determine its amount accurately.

We couldn't directly compare ligand specific elution using IPTG under the

90

Table 4.1. Heparin eluted laciz is significantly purer than that eluted with salt.

Fraction	Total Enzyme, U.	Total protein, mg	Yield	Fold Purification
Crude	4.5	10.3	100	1
Heparin	0.56	0.006	10 ± 6	212 ± 84
Salt	0.87	0.040	16 ± 9	48 ± 11

[a] The results of three experiments were averaged (n=3) and averages are reported for all columns and with standard deviations for yield and purification. For each experiment, 2ml of a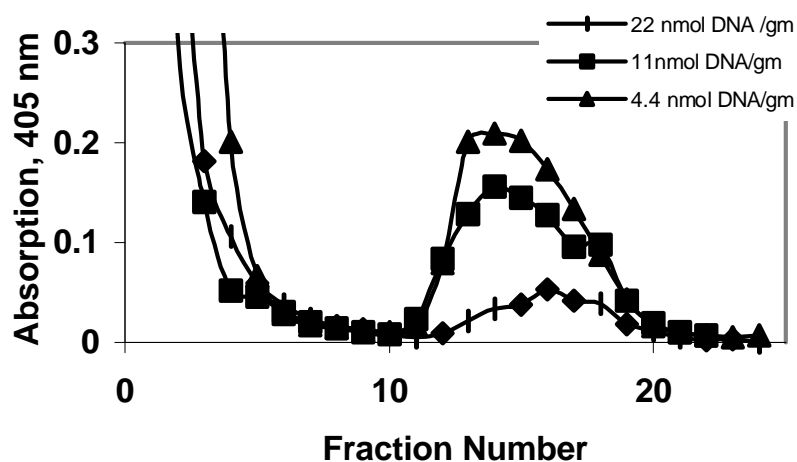 bacterial extract containing laciz was loaded onto a 1 ml. Op1T$_6$-Sepharose column having 11nmol DNA/g of Sepharose. The column was washed with 10 ml of TE0.1 and then eluted either with a 20 ml heparin gradient from 0-40mg/ml heparin in TE0.1 or with a 20 ml salt gradient from TE0.1 toTE1.2 (10 mM Tris, pH 7.5, 1 mM EDTA, 1.2 M NaCl). The flow rate was 0.5 ml/min. Active fractions were pooled together for assay.

same conditions used in Table 4.1 because very high concentration of IPTG were required for elution in the TE0.1 buffer and these concentrations interfered with the β-galactosidase assay, giving erroneous results. Dialysis to remove the IPTG was also unsuccessful since dialyzing the laciz protein at these high dilutions led to its inactivation. It was found that less IPTG is required for elution when the NaCl concentration of the mobile phase is increased to 0.3 M and β-galactosidase activity could be accurately measured in the presence of these lower IPTG concentrations (data not shown). Under these conditions the purity of protein eluted with IPTG was also greater than that of salt eluted protein (Table 4.2). However, when heparin elution was attempted at 0.3 M NaCl so little protein eluted that, despite concentrating the eluate and trying two different sensitive protein assays, it was not possible to get an accurate measure of protein concentration and, hence, the purity of the heparin eluted fraction. Thus, while heparin and IPTG could not be compared directly under the same conditions, the indirect comparison shown in Tables 4.1 and 4.2, show that IPTG elution yields purity which is similar to that obtained with heparin, and that any difference is not likely to be statistically significant.

Comparison of Tables 4.1 and 4.2 also demonstrates another important point – the purity of salt-eluted protein can be greatly improved by washing the column at an intermediate salt concentration. In Table 4.1, when the column was washed with TE0.1 prior to elution, the protein is only 1/3 as pure as in Table 4.2 where washing in TE0.3 was performed. Washing with higher salt removes additional contaminants.

Table 4.2. Comparison of heparin and IPTG eluted Laciz.

Fraction	Total Enzyme U	Total protein, mg	Yield	Fold Purification
Crude	2	2.7	100	1
IPTG	0.53	0.0025	24 ± 11	254 ± 38
Salt	0.67	0.007	35 ± 5	148 ± 66

[a]The results of the two experiments were averaged (n=2) and averages are reported for all columns and with standard deviations for yield and purification. For each experiment, 500 μl bacterial extract of laciz was loaded on 1ml OP1T$_6$ having 11nmol DNA/g of Sepharose. The column was washed with 10 ml of TE0.3 (10mMTris, pH 7.5, 1mM EDTA, 0.3 M NaCl) and eluted with 20 ml gradient of IPTG 0-0.1M in TE0.3 or with a 20 ml salt gradient from TE0.1 to TE1.2. The flow rate was 0.5 ml/min. All the active fractions were pooled together for assay.

Figure 4.5 shows the gel electrophoresis of IPTG, salt and heparin eluted fractions of laciz. It can be seen that the salt eluted fraction (in this case, from the column washed only with TE0.1) contains some contaminant proteins along with laciz fusion protein. However, the heparin fraction is very pure and only a single band corresponding to the laciz fusion protein (Mr 155 kDa.) is seen. As expected the IPTG fraction is also very pure.

4.4. Discussion

We have described an alternative method of eluting proteins from sequence specific DNA Sepharose columns. We have shown that GFP-C/EBP, a member of basic leucine zipper DNA-binding motif family of proteins (66) and laciz, having a helix-turn-helix motif (13) can both be eluted with heparin. It is very likely that protein having other DNA-binding motifs will also elute with heparin. We have also shown that protein (laciz) which is bound to a non specific sequence (EP18) can also be eluted with heparin (Figure 4.4). Our hypothesis is that the elution with heparin results from a direct competition between heparin and DNA for the same site on the protein. During elution a dynamic equilibrium is established in which the protein shuttles between heparin in the mobile phase and DNA which is bound to the column. This equilibrium is affected by the concentration of DNA bound to the column, the amount of heparin in the mobile phase, and the affinity of protein for both DNA and heparin. Elution with salt, on the other hand, presumably results from salt shielding the charge on DNA and protein and hence disrupting ionic interactions between them. Ligand specific
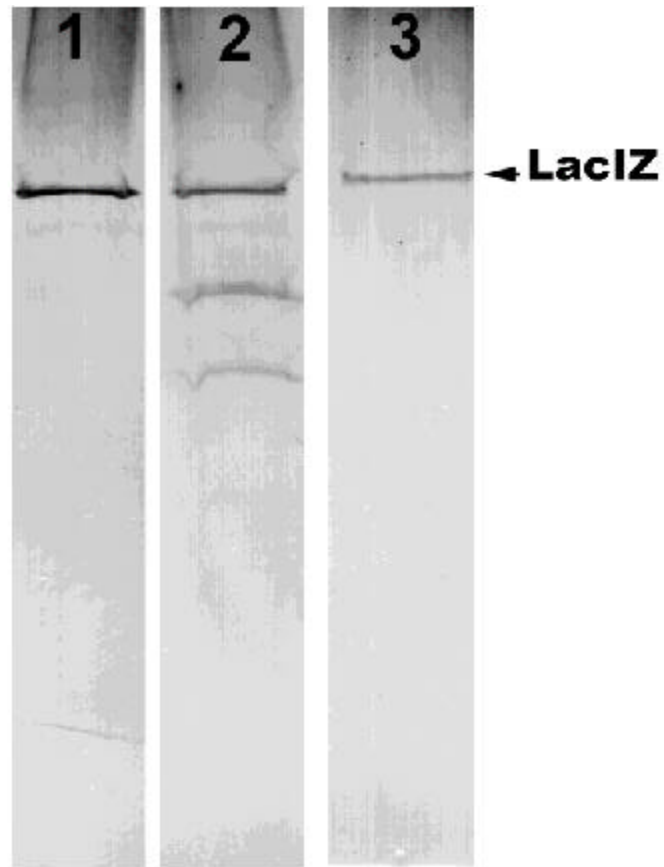
Figure 4.5. *Polyacrylamide gel of the salt-, IPTG- and heparin-eluted protein.*

Lanes 1, 2 and 3 represent heparin, salt, and IPTG eluted fractions, respectively. Lanes 1 and 2 are from the same gel; lane 3 was taken from a second gel and is shown for comparison.

elution of lac repressor is caused by the ligand (IPTG) binding to a site other than the DNA-binding site leading to a conformational change and diminished DNA-binding properties. Thus elution with heparin is not just a new but probably a mechanistically different type of elution than either salt or ligand based elution.

Our hypothesis about competitive elution is supported since elution of laciz from the column is dependent upon the concentration of both column DNA and mobile phase heparin. This competitive elution probably accounts for the purer protein eluted by heparin relative to that obtained with salt. IPTG elution also yields highly purified protein but very few transcription factors have a specific ligand with which to elute them. In any case, our data suggest that heparin eluted protein is comparable in purity to IPTG eluted protein and purer than obtained with salt under the same conditions, the most commonly used strategy to elute transcription factors from DNA columns. Protein eluted with heparin can be diluted and reapplied to the same DNA column (data not shown) allowing repetitive uses of same DNA column with different elution strategies. Another interesting observation was that when heparin was dissolved in TE containing 0.3M NaCl (TE0.3), the laciz activity eluted as a broader peak which contained so little protein it could not be accurately measured even after concentrating the protein and using sensitive protein assay techniques. This low protein concentration is also indicative of high purity but the purity could not be accurately measured. High salt concentrations have been shown to decrease the affinity of lac repressor for DNA (113) and we suspect high salt also decreases the affinity of this transcription factor for heparin. As both affinities are decreased,

a broad peak results.

Heparin represents a very flexible and controllable method for eluting transcription factors from DNA columns. Elution is affected by the heparin and salt concentration of the mobile phase (Figure 4.1 and 4.2, Table 4.1), the amount of DNA on the stationary phase (Figure 4.3 a and b), and its sequence (Figure 4.4). The purification obtainable is several hundred-fold and comparable to that obtainable when ligand-specific elution is feasible. In contrast, salt dependent elution is a much less flexible, cruder strategy. However, by washing columns at an intermediate salt concentration prior to elution, the method can be improved. By optimizing both salt and heparin elution of a particular protein, high purity of DNA-binding proteins should be more easily achieved.

A possible drawback of using heparin to elute was the lower yields obtained but this can be easily overcome by controlling the amount of DNA bound to the column. Another strategy to increase yield is to decrease the affinity of the column bound DNA sequence for the protein. This can be achieved by altering crucial base pairs of the DNA sequence or by using a nonspecific, or less specific, column. We have shown that laciz binds to EP18 column, a rather extreme case of an "altered DNA", and can be eluted with heparin. It was also observed that almost 90% of bound protein was eluted with heparin from this column while about 50% of bound protein is eluted from the specific $Op1T_6$ column under similar conditions and when the columns have comparable amount of DNA (data not included).

In a previous paper our laboratory has discussed temperature-dependent

elution of transcription factors (98). In this paper we have demonstrated heparin as specific eluent of two different transcription factors. DNA affinity chromatography is the most selective of any known method for the purification of transcription factors. Alternative methods for eluting proteins from DNA columns such as temperature-dependent elution and heparin elution, allow this high selectivity to be applied repetitively to aid in the purification of DNA-binding proteins.

Chapter: 5. Bi-column method for purification of transcription factors

Part of this chapter was published in Journal of Chromatography (2001) 905: 133-9. Permission was obtained from the publisher to use our published article in the dissertation. Some variations in format were adapted for thedissertation.

5.1. Introduction

Purification and characterization of transcription factors and other DNA binding proteins such as DNA polymerases, and endo- and exonuclease is important. These proteins control important cellular processes such as regulating the transcription of genes, DNA replication, recombination, cell division and differentiation. Moreover these proteins are frequently present in cells in small amounts and hence their purification is challenging.

A typical purification for transcription factor such as Sp1 involved purification through five steps (21) the last being DNA affinity chromatography. All these required steps make this purification time consuming. Furthermore, since no purification step yields complete recovery, the overall yield is often low. DNA affinity chromatography offers the highest selectivity and hence is widely used for purification of DNA binding proteins. Improvements in DNA affinity chromatography could allow purification in a lesser number of steps and significantly reduce the effort required.

DNA affinity columns are made by attaching either nonspecific (100) or specific (74) DNA sequences to a variety of supports such as cellulose, agarose, silica, etc. and has been discussed in more details elsewhere (76) (78). Non-specific DNA (e.g. salmon sperm or other fragmented genomic DNA) selects only for the property of DNA binding and hence nonspecific DNA columns are not selective for a particular protein. Specific columns on the other hand are highly selective and are made by coupling carefully designed sequences, which are specifically bound by the protein of interest. Although specific columns are highly

selective, more than one protein can often bind to these columns and contaminate the protein of interest.

Several approaches to improving the purity obtained have been reported. These include repeated use of DNA affinity columns (114), using nonspecific and specific sequence DNA affinity columns in tandem (115), using multimeric (concatemeric) DNA columns (77), washing the columns with a nonspecific DNA sequence, or pre-incubating extracts with nonspecific DNA sequences (21) (78). In spite of these efforts, homogenous protein is rarely obtained with DNA affinity chromatography as the only step and clearly new approaches are needed. We have previously shown that DNA binding proteins such as lac repressor and CAAT enhancer binding protein (C/EBP) can be eluted from DNA affinity columns by using heparin. Elution results from competition between the anionic polysaccharides (heparin and DNA) for transcription factor binding (60) and represents a novel elution method.

Lac repressor protein, which regulates the lac operon in *Escherchia coli,* has been studied extensively. A DNA operator sequence (Op1) which is bound by this protein has been identified (96). We have shown previously (75) that lac repressor binds more tightly to columns having Op1 with an additional $(dA)_{18}$: $(dT)_{18}$ tail than to columns having Op1 alone. Two other transcription factors have also been investigated. C/EBP binds to CAAT elements in some eukaryotic and viral promoters and regulates gene expression (22) (66). B3 is a developmentally regulated transcription factor, which regulates TFIIIA transcription in early stage oocytes (99).

In this paper we describe a new method which we call the Bi-column method for purification of transcription factors. This method involves the use of two columns having high specificity but different affinities for the transcription factor of interest. Proteins such as lac repressor-β-galactosidase fusion protein (laciz), are eluted with heparin from a column having high specificity but moderate affinity (Op1-Sepharose). The eluate is passed over a second column, connected in tandem, having both high specificity and higher affinity. The proteins that are bound by the second column are finally eluted with salt. While we developed this method originally as a fascile way to remove heparin from samples, the method has advantages for purification we did not anticipate. We have shown that this approach gives highly pure lac repressor, C/EBP and B3 proteins.

5.2. Methods

5.2.1 DNA Sepharose preparation

The oligonuucleotides shown in Table 5.1 were used for coupling to Sepharose. All strands having Aminolink were coupled to CNBr-preactivated Sepharose 4B (Sigma, St. Louis, MO, USA). Coupling and end-capping were carried out according to the protocol provided by the manufacturer. The columns were made double stranded by adding the corresponding complementary strand. The mixture was then heated to 95°C and allowed to cool slowly to room temperature. (EP18)$_5$ is self complementary and does not require the addition of

Table 5.1. Oligonucleotides used in for coupling to Sepharose.

Name	Sequence of strand which was coupled	Sequence of complementary strand
Op1	5'-NH$_2$-AATTGTTATCCGCTCACAATTCCAC	5'GTGGAATTGTGAGCGGATAACAATT
Op1T$_{18}$	5'-NH$_2$-(T)$_{18}$-AATTGTTATCCGCTCACAATTCCAC	5'GTGGAATTGTGAGCGGATAACAATTA$_{18}$
(EP18)$_5$	5'-NH$_2$-(GCAGATTGCGCAATCTGC)$_5$	NA*
(EP9)$_{10}$	5'-NH$_2$-(GCAGATTGC)$_{10}$	(5'GCAATCTGC)$_{10}$
E3	5'-NH$_2$-TGTGGTTACTAGGTTACAAATTACCCTAGCAACCATG	5'-CATGGTTGCTAGGGTAATTTGTAACCTAGTAACCACA
E3/2	5'-NH$_2$-AAATTACCCTAGCAACCATGCATT	5'-AATGCATGGTTGCTAGGGTAATTT

* NA stand for not applicable, EP18 and (EP18)$_5$ are self complementary and do not require the addition of a second strand.

"5'NH$_2$" represents aminoethyl group added on the last synthetic cycle with the Aminolink reagent (Applied Biosystems)

a complementary strand. The amount of DNA coupled was determined by the difference in UV absorption of DNA added and recovered after coupling. 115 μg of (EP18)$_5$, 95 μg of (EP9)$_{10}$,177 μg of Op1, 160 μg of Op1T$_{18}$, 230 μg of E3/2 and 253 μg of E3 was found to be coupled /g Sepharose.

5.2.2 Production of proteins

Lac repressor-β-galactosidase (laciz) fusion protein was produced as described earlier (60) by growing clone BMH-72-19-1 which was generous gift of Dr. David Levens (Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland).

TB1 cell extract was obtained by using the same protocol used for production of laciz the only difference was that *E. coli* strain TB1 containing plasmid pMalC (New England Biolabs, Beverly, MA, USA) was used in place of clone BMH-72-19-1.

Rat liver nuclear extract used for studies on C/EBP was prepared by the procedure described in (116).

Xenopus laevis oocyte extract used for B3 purification was made from stage I-II Xenopus oocytes as described in (99)

5.2.3 Chromatography

All supports were packed in 1 ml bed volume syringe columns initially equilibrated in TE0.1 buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 0.1 M NaCl). Details of elution and the gradient used are given in the figure legends.

5.2.4 Assay of laciz

Laciz was assayed for β-galactosidase activity by using the protocol described previously (60) and in Chapter 3 of thisdissertation.

5.2.5 Polyacrylamide gel electrophoresis

All samples were concentrated using Centriplus centrifugal filter devices supplied by Milipore Corporation (Bedford, USA). One fourth of each sample was applied to sodium dodecylsulfate 12% polyacrylamide gels using the method of Laemmli (87) and stained with silver using the Bio-Rad Laboratory kit (Richmond, CA USA).

5.2.6 Western blot analysis

Electrophoresis was carried out as above, proteins were then transferred to nitrocellulose filters and probed with appropriate antibodies as described by Towbin et al. (117). For lac repressor, a 1:5000 diluted rabbit anti-lac repressor polyserum (Stratagene, La Jolla, CA, USA) was used as primary antibody. For C/EBP a 1:5000 dilution of anti-C/EBP (14AA) supplied by Santa Cruz Biotechnology (Santa Cruz, CA, USA) was used as primary antibody. A 1: 5000 dilution of rabbit polyserum generated against purified B3 (HTI Bioproducts) was used as primary antibody for detection of B3. Immunoreactive proteins were visualized by using a 1:3000 diluted rabbit secondary antibody alkaline phosphatase conjugates and stained for alkaline phosphatase using a kit supplied by Promega (Madison, WI USA).

5.3. Results and discussion

Figure 5.1 shows elution of lac repressor-β-galactosidase (laciz) fusion protein from Op1 and Op1T$_{18}$ columns. It can be seen that laciz elutes from Op1 as a sharp peak at heparin concentration from 10-20 mg/ml with the peak fraction at 16 mg/ml. The peak of laciz eluted from Op1T$_{18}$ is fairly broad, with laciz eluting at heparin concentrations as low as 10 mg/ml, however, the bulk of laciz elutes at heparin concentrations between 25-40 mg/ml with the peak fraction at 32 mg/ml. These data are in agreement with our previous report where we had shown a similar difference in the salt elution of this protein from these two columns (75). The Op1T$_{18}$ column shows some heterogeneity in the eluted peak. We have found that some proteolysis occurs in the laciz protein. Since both β-galactosidase and lac repressor form tetramers, a large number of species could be formed with truncated forms of laciz and this probably accounts for the heterogeneity observed.

Figure 5.2 shows a diagrammatic presentation of the Bi-column method. Column A in the figure represents a column such as Op1-Sepharose, which has a moderate affinity for the protein to be purified. Because of the moderate affinity, proteins elute from this column at a low heparin concentration. Column B represents columns such as Op1T$_{18}$, which have a greater affinity for the protein. Although column A has a moderate affinity, it is by no means a nonspecific column as it includes either the entire 25 base pair Op1 sequence bound by lac repressor protein or a part of a known DNAse1 footprint (EP9 or E3/2) bound by C/EBP and B3, respectively. Column B on the other hand has the entire footprint
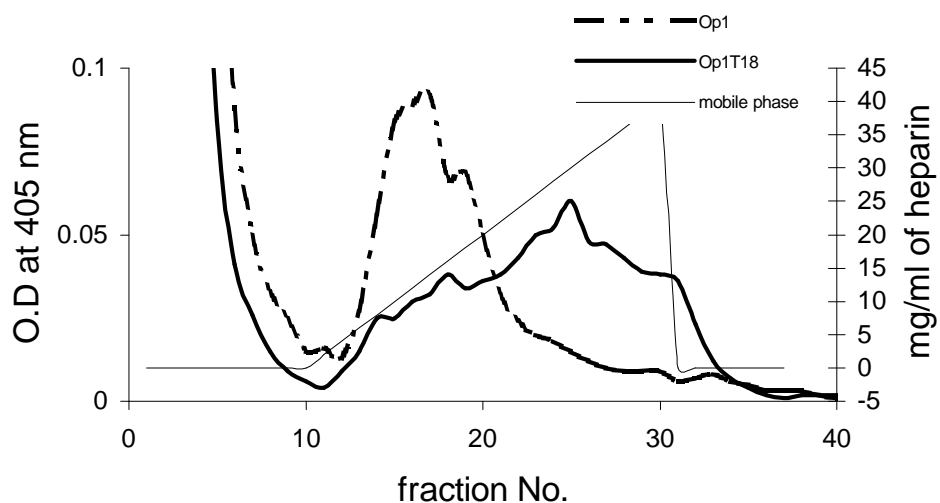
Figure 5.1. *Elution of laciz from Op1-Sepharose and Op1T$_{18}$-Sepharose with heparin.*

100 μl of crude bacterial extract containing laciz was loaded onto either Op1-Sepharose or Op1T$_{18}$-Sepharose, the columns were washed with 10 ml of TE0.1(10 mM Tris, pH 7.5, 1 mM EDTA, 0.1 M NaCl) and eluted with a 20 ml gradient of heparin from 0-40 mg/ml in TE0.1 followed by one more wash with 10 ml TE0.1. The flow rate was 0.5 ml/min and 1 ml fractions were collected.

1

Load crude extract on column A and wash with 10 ml of TE0.1

2

Wash columns A&B with 20ml gradient of heparin from 0-25mg/ml inTE0.1.

3

Elute column B with TE1.2
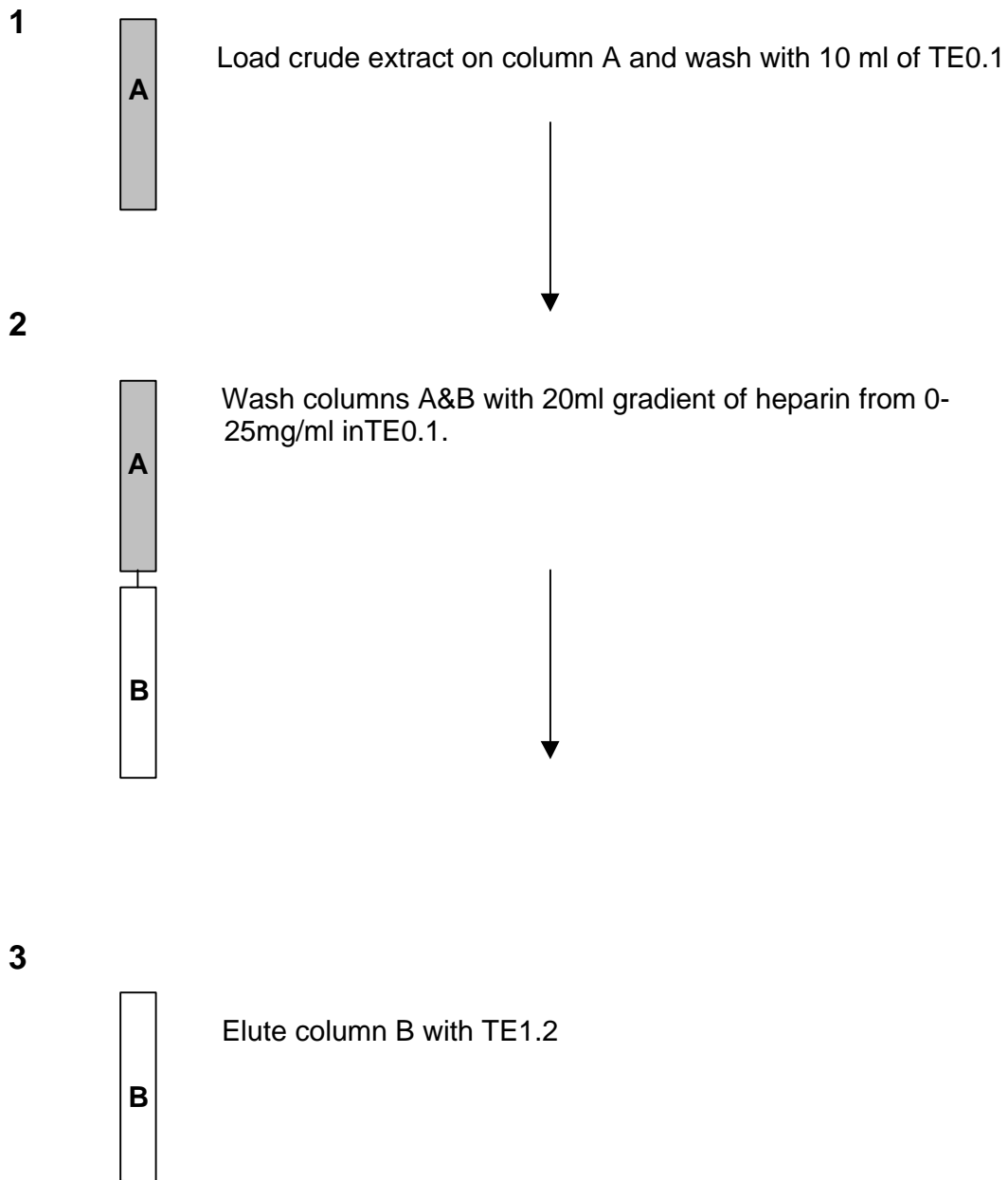
Figure 5.2. *Diagrammatic representation of Bi-column method.*

Column *A represents Op1, (EP9)$_{10}$ or E3/2-Sepharose. Column B represents Op1T$_{18,}$ (EP18)$_5$ or E3-Sepharose. TE1.2 is 10 mM Tris, pH 7.5, 1 mM EDTA, 1.2 M NaCl.*

region (EP18 and E3) and sometimes a footprint with additional sequences attached to it (Op1T$_{18}$) which increases binding affinity. Crude protein extract is loaded onto column A which is washed with TE0.1. Column B is then attached downstream of column A. A heparin gradient is applied to the Bi-column at concentrations which will elute the protein from column A but not from column B. Column B is then separated and eluted with buffer TE1.2 which contains 1.2 M NaCl to obtain the purified transcription factor free of heparin.

Figure 5.3 shows such a Bi-column experiment for lac iz. Op1-Sepharose and Op1T$_{18}$-Sepharose are used as columns A and B, respectively. Most of the laciz that is eluted from Op1 with heparin in step 2 is able to bind to the Op1T$_{18}$ column and very little passes through the column as can be seen in fractions 20-35 in Figure 5.3. Laciz then elutes as a sharp peak in fractions 40-45 from the Op1T$_{18}$-Sepharose column with TE1.2.

The Bi-column method also works with the lacI^q mutant of the lac repressor protein expressed from the low copy number pMalc plasmid. It can be seen from Figure 5.4 that highly purified lac repressor protein is obtained with the Bi-column method (lane1) when a crude bacterial extract is used as the starting material. It can also be seen from the figure that protein obtained from the same extract with salt elution alone of Op1T$_{18}$-Sepharose is impure and many bands of protein impurities can be seen on the gel (lane2). Lanes 3 and 4 show a western blot with specific lac repressor antibody of the proteins in lanes 1 and 2.

Figure 5.5 shows a Bi-column purification of C/EBP from rat liver nuclear extract. (EP9)$_{10}$-Sepharose which has only half of the binding site for C/EBP was
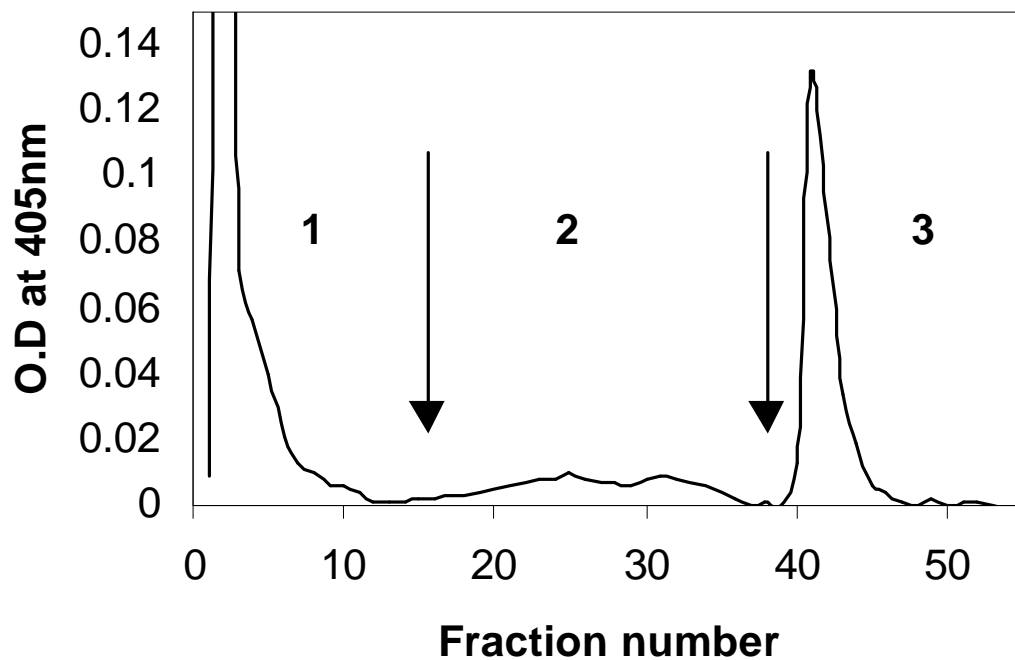
Figure 5.3. *Bi-column elution of laciz.*

300 μl of crude extract containing laciz was loaded onto Op1-Sepharose and the Bi-column elution was carried out as described in Figure 5.2. Op1-Sepharose was used as column A and $Op1T_{18}$-Sepharose was used as column B. The flow rate was 0.5ml/min and 1 ml fractions were collected.
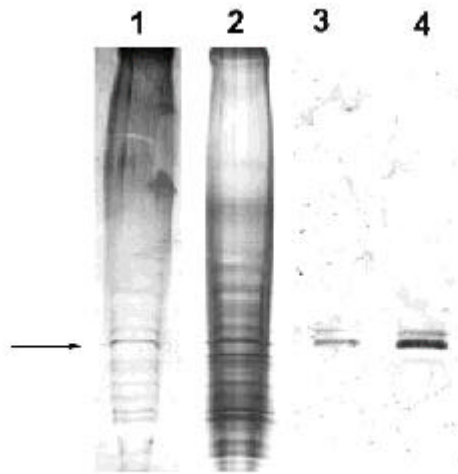
Figure 5.4. *The Bi-column method works better for purification lac repressor than salt elution.*

500 µl of crude extract containing lacI^q produced from a low copy number plasmid was loaded onto Op1-Sepharose and the Bi-column elution was carried out as in Figure 5.2 except a 10 ml wash with TE0.4 (10 mM Tris, pH 7.5, 1 mM EDTA, 0.4 M NaCl) was applied to column B (Op1T$_{18}$–Sepharose) before eluting the protein with TE1.2. For salt elution, 500 µl of crude extract containing lac repressor was loaded onto Op1T$_{18}$-Sepharose, the column was washed with 30 ml TE0.1 followed by a 10 ml wash with TE0.4, and eluted with 10 ml TE1.2. Lanes 1 and 2 are from a silver stained gel while lanes 3 and 4 are western blots of Bi-column purified and salt eluted lac repressor, respectively.

111
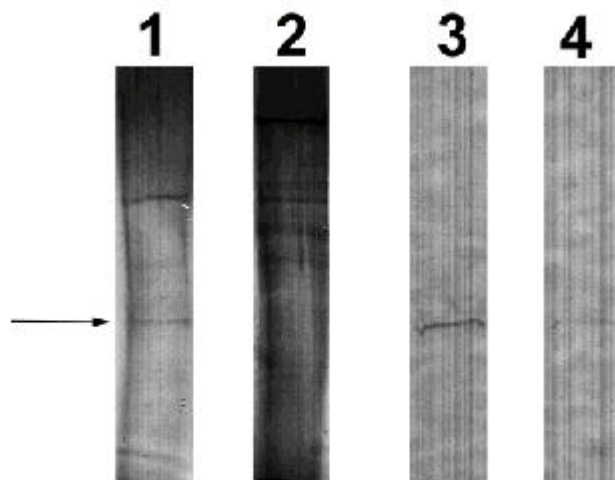
Figure 5.5. *Bi-column method and salt elution purification of C/EBP.*

500 µl of rat liver nuclear extract containing C/EBP was loaded onto

(EP9)$_{10}$ –Sepharose and the Bi-column experiment was carried out as described

in Figure 5.2. (EP9)$_{10}$ - and (EP18)$_5$ –Sepharose were used as columns A and B,

respectively. A 10 ml wash with TE0.4 was applied to column B before eluting the

protein with TE1.2. For salt elution, 500 µl of Rat liver nuclear extract containing

C/EBP was loaded onto (EP18)$_5$ –Sepharose. The column was washed with 30

ml TE0.1, followed by a 10 ml wash with TE0.4, and eluted with 10 ml TE1.2.

Lanes 1 and 2 represent a silver stained gel while lanes 3 and 4 are western

blots of C/EBP purified with Bi-column and salt elution, respectively.

used as column A and (EP18)$_5$-Sepharose which consists of 5 tandem copies of the entire DNA binding sequence was used as column B. We found that Green fluorescent protein-CAAT enhancer binding protein chimeric fusion protein (GFP-C/EBP) binds to (EP9)$_{10}$ but elutes at a lower heparin concentration than from (EP18)$_5$ (data not shown), hence these two columns were used. It can be seen from the figure that the Bi-column method gives highly purified protein and only two or three contaminant protein bands can be seen after silver staining (lane1). Protein purified with single step salt elution on the other hand looks very impure and several other contaminant protein bands can be seen along with that of C/EBP on the gel (lane2). Lane 3 and 4 show western blots of the samples used for lanes 1 and 2. Notice that so little C/EBP elutes with salt that it can be barely detected with antibody while with Bi-column method, it is a major constituent. Transcription factor B3 regulates TFIIIA transcription in early *Xenopus* oocytes. The DNA binding region of B3 consists of four dyads. It has been reported earlier (99) that B3 binds with a lower affinity to a sequence containing dyads 3 and 4 (E3/2) than to sequence containing all four dyads (E3). Hence we used E3/2-Sepharose and E3-Sepharose as column A and column B in Bi-column method, respectively. It can be seen from Figure 5.6 that the Bi-column method works very well for purification of B3 from Xenopus oocyte extract and a single band of B3 can be seen after silver staining (lane1). As for C/EBP and lac repressor, conventional salt elution gives an impure protein and several other protein bands are visible on the gel (lane 2). The western blot however confirms that salt elution gives higher yield.
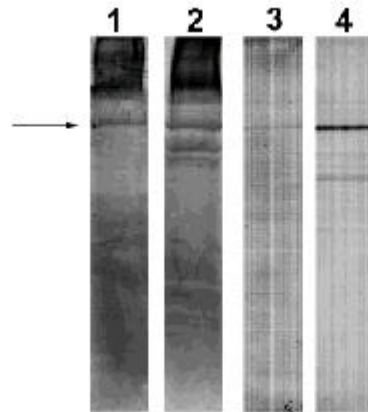
Figure 5.6. *Bi-column and salt elution for purification of B3.*

500 µl of *Xenopus* ocyte extract containing B3 transcription factor was loaded onto E3/2 –Sepharose and the Bi-column experiment was carried out as depicted in Figure 5.2. E3/2 -Sepharose and E3 –Sepharose were used as columns A and B, respectively. A 10 ml wash with TE0.4 was applied to column B (E3 –Sepharose) before eluting the protein with TE1.2. For salt elution, 500 µl of Xenopus oocyte extract containing B3 was loaded onto E3–Sepharose, the column was washed with 30 ml TE0.1, followed by a 10 ml wash with TE0.4, and eluted with 10 ml TE1.2. Lanes 1 and 2 represent silver stained gel while lanes 3 and 4 are western blots of B3 purified with the Bi-column and salt elution, respectively.

114

We have shown earlier that laciz eluted with heparin from a $Op1T_6$ columns is highly pure (60). Although heparin eluted protein is highly pure, heparin has to be removed later since it can interfere with the DNA binding activity of the protein. The Bi-column method was initially developed to remove heparin. We found out that the Bi-column methods not only performs this function but is also very effective way of purifying transcription factors.

Highly pure lac repressor, C/EBP and B3 can be obtained, from their respective sources, bacterial extract, rat liver nuclear extract and *Xenopus* oocyte extract by using the Bi-column method alone. The Bi-column method is probably effective because it combines the use of two different specific DNA affinity columns and two different elution methods. In the Bi-column method the proteins shuttles from DNA on Column A to heparin in mobile phase and back to DNA on column B. The protein to be purified has a higher affinity for column B and hence can be retained by it. Proteins, which are nonspecifically bound to column A on the other hand probably, have similar affinities for both columns and hence are not retained by column B.

5.4. Conclusions

The Bi-column method could be extended to purify any transcription factor as long as its DNA element has been, or can be, identified and two high specificity columns having different affinities for the transcription factor can be generated.

This Bi-column protocol yields higher purity of B3 that has ever been

115

reported, even using much more complex purification schemes involving several different chromatographic steps. The C/EBP obtained is also of a higher purity than has ever been obtained from such a simple technique as the Bi-column (62). These purifications are all from less than one ml of a crude extract containing only native or near native amounts and yet could be purified to levels, which would be sufficient for characterization.

Chapter: 6. An oligonucleotide trapping method for the

purification of transcription factors

6.1. Introduction

DNA affinity chromatography is one of the most widely used techniques for the purification of transcription factors and other DNA binding proteins (100) (77) (74). Either nonspecific or specific DNA affinity columns are used for these purifications. Nonspecific columns were the first DNA affinity columns to be made and were made by coupling diverse DNA sequences, such as fragment salmon sperm genomic DNA to cellulose (67). Since then, there have been several new advances in the technique. Highly specific columns made by using the footprint region, the region of DNA that is protected from cleavage by DNAse1 upon binding of a specific DNA binding protein, have replaced nonspecific columns for transcription factor purifications for most applications. Various supports such as Sepharose, cellulose and silica are routinely used for coupling of DNA and several coupling chemistries are available for attaching DNA to these supports (76). The most commonly used method is coupling of the amino groups either inherently present in DNA or introduced during oligonucleotide synthesis to cyanogen bromide-activated Sepharose (95).

Chemical coupling of the oligonucleotide can cause modifications of nucleotides within the DNA sequence. Such modifications can potentially affect the specific DNA-protein interaction, which in turn may lead to decreased efficiency of the DNA affinity columns in protein purification. Modification of the nucleotide bases can also lead to decreased capacity of the column for its specific protein. There are only a few methods that allow the use of unmodified DNA sequences. The enzymatic synthesis method (81) which was developed

previously in our lab, involves synthesis of an unmodified DNA sequence on the column by using the Klenow fragment of DNA polymerase. In the other method, the highly specific biotin streptavidin interaction is used to trap a biotynilated DNA-protein complex (72) (31). In this method a biotinylated oligonucleotide is immobilized on a streptavidin containing support, which is then used for the affinity chromatography of DNA binding proteins. Streptavidin coated magnetic beads can be used as the support for coupling of biotinylated oligonucleotide as they can be easily separated from solution with a strong magnet. In some cases a biotinylated oligonucleotide is allowed to interact with proteins in solution and the protein-DNA complex is then trapped onto a streptavidin containing support (72) (118). In a third method, specific DNA affinity columns are made by immobilizing a footprint region containing a 3' polyA tail onto a polyT-agarose column (119). These columns are then used to purify DNA binding proteins. Other methods that allow coupling of unmodified DNA include oligonucleotide synthesis directly on a Teflon fiber support (103) and coupling of thiophosphorylated oligonucleotide to bromoacetyl agarose (120) but are not routinely used for making DNA affinity columns.

Most of the methods mentioned above have certain disadvantages. The enzymatic synthesis method can lead to tail length heterogeneity of the DNA. The heterogeneity of these columns could potentially lead to peak broadening and decreased purity. Comparative studies between enzymatically and chemically synthesized columns have shown that there is no distinct advantage in using enzymatic synthesis method for transcription factor purification (75). The

119

streptavidin supports have a high affinity for biotinylated oligonucleotides but such supports can also bind to several other proteins in the crude extract, especially protein containing the biocytin group. Many proteins can bind to streptavidin nonspecifically and elute during chromatography. These proteins can potentially coelute with the protein of interest and decrease the purity. Newer forms of streptavidin such as monomeric avidin and NeutrAvidin are supposed to show lower nonspecific interactions.

We have used the green fluorescent protein CAAT enhancer binding chimeric protein (GFP-C/EBP) to study a new method called the oligonucleotide trapping method. CAAT enhancer binding protein (C/EBP) regulates expression of several genes in mammals and certain viruses (121). C/EBP is one of the most studied transcription factors and the DNA sequence that is bound by C/EBP has been well-characterized (122). We have shown previously that GFP-C/EBP has DNA binding properties similar to that of C/EBP (98). We have also applied the trapping method to purify B3, which is a developmentally regulated transcription factor and regulates TFIIIA transcription in early Xenopus oocytes (99).

In the oligonucleotide trapping method described in this chapter, we have used the highly specific interaction between complementary strands of DNA to trap the protein-DNA complex. Similar techniques have been used to purify mRNA's from crude mixtures (123) and to make specific DNA affinity columns as discussed earlier (119). In our method a column attached, single stranded $(AC)_5$ oligonucleotide is used to trap a double stranded footprint region which has a

$(GT)_5$ tail on both strands from solution. The interaction between a specific protein and its footprint element is carried out in solution and the protein-DNA complex is passed over $(AC)_5$ –Sepharose. The latter is able to trap the DNA protein complex because of highly specific annealing of $(AC)_5$ and its complement $(GT)_5$ present at the ends of each DNA strand. The protein alone, can then be eluted by using a buffer containing high salt to weaken the protein-DNA interaction or the intact DNA-protein complex can be eluted by using moderate temperatures and low salt concentrations to melt the interaction between $(AC)_5$ and $(GT)_5$. This approach gives lesser contaminants than the biotin-streptavidin method.

In conventional DNA affinity chromatography the concentration of DNA that is coupled is very high. It is generally thought that the high DNA concentrations are necessary for obtaining high yields. Although this is probably true, having high DNA concentrations has a major disadvantage. Certain nonspecific proteins, which bind to DNA sequence with low affinity, are able to bind to the column because of these high DNA concentrations and can contaminate the protein of interest. We have shown that GFP-C/EBP obtained by using low concentrations of oligonucleotide and by using our trapping method is at least three fold purer than that obtained by conventional DNA affinity chromatography.

6.2. Methods

6.2.1 Coupling of DNA to Sepharose

EP24 (NH_2-GCTGCAGATTGCGCAATCTGCAGC),$(AC)_5$ (NH_2-ACACACACAC) and E3 (5'-NH_2-TGTGGTTACTAGGTTACAAATTACCCTAGCAACCATG) were coupled to CNBr-preactivated Sepharose 4B (Sigma St. Louis, MO USA). Coupling and end capping were carried out according to the protocol provided by the manufacturer. "5' NH_2" in oligonucleotide sequences represents an aminoethyl group added on the last synthesis cycle with the Aminolink reagent (Applied Biosystems). The E3 column was made double stranded by adding the corresponding complementary strand αE3 (5'-CATGGTTGCTAGGGTAATTTGTAACCTAGTAACCACA). The mixture was then heated to 95ºC and allowed to cool slowly to room temperature. EP24 is self-complementary and does not require the addition of a complementary strand. The amount of DNA coupled was determined by the difference in the UV absorption of DNA added and recovered after coupling. Approximately 20 nmoles of both EP24 and $(AC)_5$ oligonucleotides and 36 nmoles of E3 were coupled per gram of Sepharose.

6.2.2 Production of proteins

GFP-C/EBP was produced by growing *Escherichia coli* strain BL21 containing plasmid pJ22-GFP-C/EBP as described previously (98).

Xenopus laevis oocyte extract used for B3 purification was made from

stage I-II *Xenopus* oocytes as described in (99).

6.2.3 Chromatography

All supports were packed in 1 ml bed volume syringe columns initially equilibrated in TE0.4 buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 0.4 M NaCl). Details of elution and the gradient used are given in the figure legends.

For the oligonucleotide trapping method EP24(GT)$_5$ (GCTGCAGATTGCGCAATCTGCAGCGTGTGTGTGT), Bi-EP24 (Bi-GCTGCAGATTGCGCAATCTGCAGC) where "Bi" represents Biotin introduced during oligonucleotide synthesis or (GT)$_5$E3/αE3(GT)$_5$ (5'-NH2-TGTGGTTACTAGGTTACAAATTACCCTAGCAACCATGTGTGTGTGTG/5'-CATGGTTGCTAGGGTAATTTGTAACCTAGTAACCACATGTGTGTGTG) were incubated with either purified or crude GFP-C/EBP or Xenopus oocyte extract and passed over the appropriate column. Heparin, salmon sperm DNA (both obtained from Sigma, St. Louis, MO, USA), T$_{18}$(TTTTTTTTTTTTTTTTTT), and μE3 (TGTAACAACTAAACAACAAATTGTTCTAGCTGTTAATGCATTG/ACATTGTTGATTTGTTGTTAATCAAGATCCACAATACGTAAC) were used in some experiments as competitors. The details are given in figure legends. NeutrAvidin and monomeric avidin-agarose were obtained from Pierce (Rockford IL, USA) and packed in 1 ml syringe columns.

6.2.4 Protein assay

Protein concentrations were determined by the Bradford method (124).

6.2.5 Polyacrylamide gel electrophoresis

All the samples were concentrated using Centriplus centrifugal filter devices supplied by Milipore Corporation (Bedford, USA). One fourth of each sample was applied to a sodium dodecylsulfate 4-15% polyacrylamide Bio-Rad precast gradient gel using the method of Laemmli (87) and stained with silver using the Bio-Rad Laboratory kit (Richmond, CA, USA).

6.2.6 Western blot analysis

Electrophoresis was carried out as above, proteins were then transferred to nitrocellulose filters as described by Towbin et al. (117). A 1: 5000 dilution of rabbit polyserum generated against purified B3 (HTI Bioproducts) was used as a primary antibody for detection of B3. Immunoreactive proteins were visualized by using 1:3000 diluted rabbit secondary antibody with alkaline phosphatase or horseradish peroxidase conjugates and stained by using nitroblue tetrazolium or the electro-chemiluminascence (ECL) method, respectively (61).

6.3. Results

Figure 6.1 is a schematic of the oligonucleotide trapping method. In this method, a footprint region symbolized as NNNNNN bound specifically by the protein of interest is extended with a single stranded $(GT)_5$ sequence on each

Figure 6.1. *Schematic of the oligonucleotide trapping method.*

The strategy used in the oligonucleotide trapping method is shown diagrammatically. The circled S represents the chromatographic support, in this case, Sepharose. First 5'-aminoethyl$(AC)_5$ oligonucleotide is chemically coupled to Sepharose. A footprint region having a $(GT)_5$ extension on both strands is incubated with extract containing the protein of interest. This mixture is then passed over the $(AC)_5$-Sepharose column. The protein alone can then be eluted by using buffer containing high salt or the DNA protein complex can be eluted using high temperature.

NNNNNNTGTGTGTGTG

GTGTGTGTGT<u>NNNNNN</u>

Crude
protein
extract
containing
the protein
of interest

Protein of
interest

Specific DNA sequence
which binds to protein of
interest

NNNNNNTGTGTGTGTG

GTGTGTGTGT<u>NNNNNN</u>

Protein of
interest

Incubate on
ice

ACACACACAC **S**

((AC)$_5$-Sepharose**)**

ACACACACAC **S**

NNNNNTGTGTGTGTG

(GT) $_5$NNNNN

Protein of
interest

Elute with salt to
elute protein from
the column

Elution with temperature to
elute protein-DNA complex

strand. For our studies, with GFP-C/EBP we have used EP24 extended with $(GT)_5$ $(EP24(GT)_5)$. EP24 contains the consensus sequence for binding of C/EBP. This $EP24(GT)_5$ sequence is incubated with extracts containing GFP-C/EBP for 30 minutes and then passed over $(AC)_5$-Sepharose. $(AC)_5$- Sepharose is able to trap the DNA-protein complex because of the specific hybrid formation between the $(GT)_5$ region of $EP24(GT)_5$ and the $(AC)_5$ region bonded to the Sepharose. Elution can be achieved using high salt to disrupt the DNA-protein interaction or by using moderate temperatures and low salt to melt the hybrid between $(AC)_5$ and $(GT)_5$. When high salt is used GFP-C/EBP alone, free of DNA, is eluted. When temperature is used, GFP-C/EBP that is still bound to $EP24(GT)_5$ is eluted. A similar protocol was followed for the purification of B3 from Xenopus oocyte extract except $(GT)_5E3/\alpha E3(TG)_5$ was used in the place of $EP24(GT)_5$.

Two different variants of the oligonucleotide trapping experiment performed with GFP-C/EBP are shown in Figure 6.2. It can be seen from the figure that sharp peaks containing GFP-C/EBP are obtained upon elution with either salt or temperature. The peaks obtained with the two different elution schemes are similar in peak height and width. GFP-C/EBP alone, in the absence of $EP24(GT)_5$ was did not bind to $(AC)_5$- Sepharose (data not shown) and hence, both the peaks observed are due to the specific interaction between GFP-C/EBP and EP24.

We compared our oligonucleotide trapping method with the biotin streptavidin method. Figure 6.3 shows the purity of GFP-C/EBP obtained.
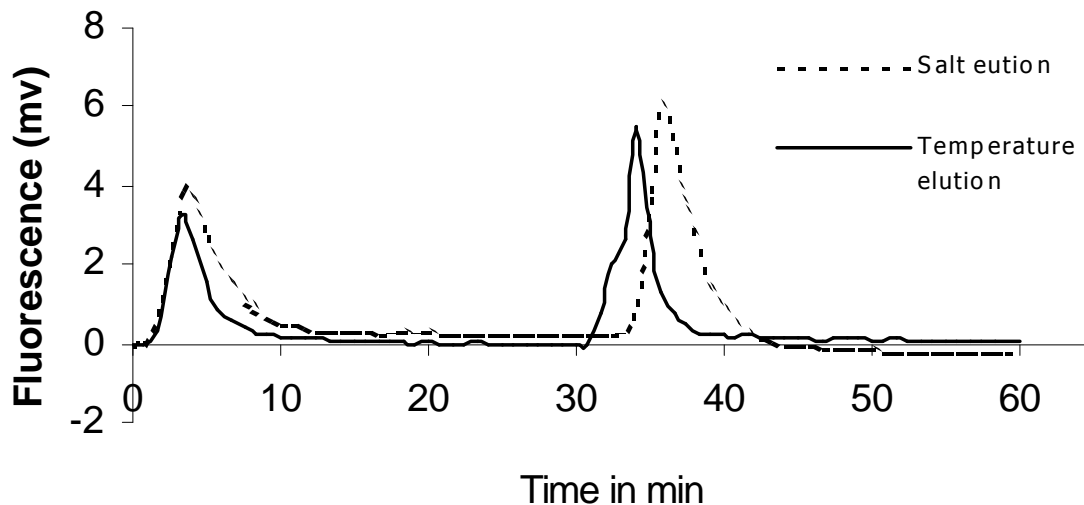
Figure 6.2. Oligonucleotide trapping method for GFP-C/EBP.

20 μl of purified GFP-C/EBP was incubated with 5 nmoles of EP24(GT)$_5$

on ice for 30 min in TE0.4 (10 mM Tris, pH 7.5, 1 mM EDTA, 0.4 M NaCl). This

mixture was then passed over a 1 ml (AC)$_5$-Sepharose column. The column was

washed with 15 ml of TE0.4 at 4°C. The protein was either eluted with TE1.2 (10

mM Tris, pH 7.5, 1 mM EDTA, 1.2 M NaCl) at 4°C for salt elution or with TE (10

mM Tris, pH 7.5, 1 mM EDTA) at 37°C for temperature elution. The flow rate was

maintained at 0.5 ml/min throughout the experiment.

Figure 6.3. *Comparison of the oligonucleotide trapping method and avidin-biotin trapping method.*

100 µl of a crude bacterial extract containing GFP-C/EBP was mixed with either BiEP24 (where Bi stands for Biotin group introduced at the 5' end during synthesis) or EP24(GT)$_5$ and incubated on ice for 30 min. The mixture containing BiEP24 was passed over a 500 µl NeutrAvidin column (lane 1) or 500 µl monomeric avidin column (lanes 2 and 3). The mixture containing EP24(GT)$_5$ was passed over 1 ml (AC)$_5$-Sepharose (lanes 4 and 5). All the columns were washed with PB0.4 (0.1 M KH$_2$PO$_4$, pH 7.5 and 0.4 M NaCl). The NeutrAvidin column (lane 1), monomeric avidin column (lane 3) and (AC)$_5$-Sepharose (lane 5) were eluted with PBS1.2 (0.1 M KH$_2$PO$_4$, pH 7.5 and 1.2 M NaCl). The monomeric avidin column (lane 2) was eluted with PBS0.4 containing 2 mM biotin, while (AC)$_5$-Sepharose (lane 4) was eluted with PB (0.1 M KH$_2$PO$_4$, pH 7.5 and 0.1 M NaCl) at 37°C. Lanes 6, 7, 8, 9 and 10 are western blots of proteins in lane 1, 2, 3, 4 and 5, respectively stained with silver stain.
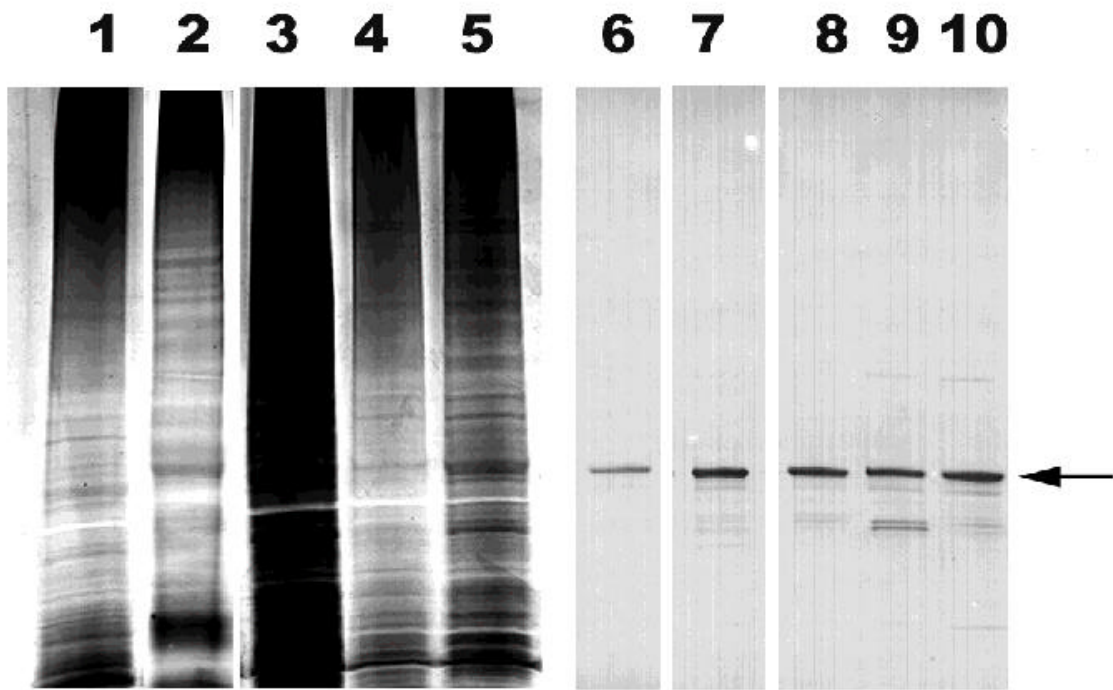
Lanes 1, 2 and 3 in Figure 6.3 represent proteins obtained from NeutrAvidin column with salt elution, monomeric avidin column with biotin elution and monomeric avidin column with salt elution respectively. NeutrAvidin and monomeric avidin are genetically modified forms of avidin that are thought to show minimal nonspecific interactions. Lanes 4 and 5 represent samples obtained with oligonucleotide trapping method using temperature and salt elution respectively. The GFP-C/EBP obtained from NeutrAvidin-agarose (lane 1) and monomeric avidin with biotin elution (lane 2) are significantly purer than that obtained from the monomeric avidin column with salt elution (lane 3). The oligonucleotide trapping method with temperature elution (lane 4) yields the purest GFP-C/EBP. From the silver stained gel it appears that more GFP-C/EBP is obtained with salt elution in the oligonucleotide trapping method (lane 5) than with temperature elution. However, western blotting of the same samples (lanes 9 and 10) shows that bands of similar intensities are observed. Hence the larger size of band for GFP-C/EBP seen in lane 5 is probably because of comigration of some contaminant protein along with GFP-C/EBP. The same is also true for the biotin-eluted sample in lane 2.

As discussed before chemical coupling of DNA to Sepharose can cause modification within the DNA, which could affect efficiency of the column. Our method on the other hand allows the use of unmodified DNA, which could lead to higher capacity for proteins. It can be seen from Figure 6.4 that the oligonucleotide trapping method has higher capacity for GFP-C/EBP and is able a bind to greater amount of GFP-C/EBP than a conventional DNA affinity column

131
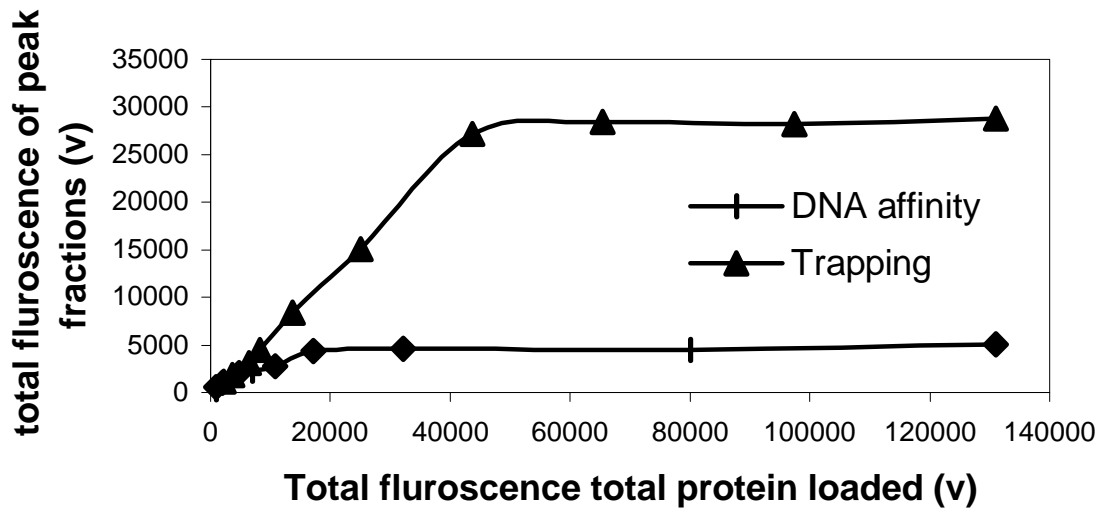
Figure 6.4. *Capacity of oligonucleotide trapping and DNA affinity chromatography.*

5 nmoles of EP24(GT)$_5$ was mixed with different amounts of purified GFP-C/EBP. The mixture was incubated on ice for 30 min and loaded onto a 1 ml (AC)$_5$-Sepharose column. The column was washed with 10 ml TE0.4 and eluted with TE1.2. Different amounts of GFP-C/EBP were also loaded onto a 250 μl (5 nmoles DNA) EP24-Sepharose columns, the column was then washed and eluted as described for the (AC)$_5$-Sepharose column. The fluorescence for GFP-C/EBP was monitored continuously by using a Shimadzu fluorescence spectromonitor RF-530 and the peak areas calculated by using Gilson Unipoint software. All the columns were run at 4°C and the flow rate for all the column runs was maintained at 0.5 ml/min.

having comparable amount of DNA to that used in the trapping experiments. It can be seen from the figure that the maximum amount of GFP- C/EBP bound in the trapping experiments is 28,000 fluorescence units, which is equivalent to 0.19 mg of GFP-C/EBP (data not shown). 5 nmoles of EP24 that is used for the trapping experiment can theoretically bind up to 0.2 mg of GFP-C/EBP assuming that C/EBP binds to its DNA element as a dimer. Hence around 95% of the theoretical capacity can be obtained in the oligonucleotide trapping method. The DNA affinity column also containing 5 nmoles of EP24, can maximally bind only 5,000 fluorescence units of GFP-C/EBP, which corresponds to around 0.03 mg of GFP-C/EBP, and hence only 15% of the theoretical capacity was achieved in conventional DNA affinity chromatography.

DNA affinity columns used routinely have a high concentration of DNA coupled, ranging from 15-200 μM. These high concentrations may be necessary to obtain high yields. This approach has a major drawback; the high concentration of DNA on the column can encourage nonspecific binding of proteins that have a low affinity for DNA. In a crude extract there are typically many DNA binding proteins and most of these proteins would bind to the DNA on the column. The lac repressor protein which is the most widely studied transcription factor has an affinity of Kd = 0.7-11 $\times$ 10^{-5} for nonspecific DNA sequence (125) and Kd = 3.5 $\times$ 10^{-10}-1.7 $\times$ 10^{-12} for a specific operator sequence (126). Because of this difference in specific and nonspecific affinities of DNA binding proteins, the specific protein alone would bind with a very high affinity to the DNA affinity column. However, the high DNA concentration on the column

drives the binding of even lower affinity binding proteins regardless of the DNA sequence. Separation of these nonspecific proteins from the protein of interest with suitable elution scheme would be challenging. This contamination could be minimized by keeping the concentration of the column DNA low. Although this could be achieved in conventional DNA affinity columns, it would require low DNA concentrations in large columns to allow reasonable capacity. The trapping method can be more efficiently utilized for this purpose.

Figure 6.5 illustrates how the binding equilibria would affect the binding of contaminant protein having a Kd of 4×10^{-5} M for the DNA footprint used in conventional DNA affinity chromatography and the oligonucleotide trapping method. Since the binding constants of C/EBP or B3 for non specific DNA are not known, we chose to use the constants for lac repressor protein to illustrate the point. In this model the concentration of DNA in DNA affinity chromatography is assumed to be 20 μM which is about the average of DNA concentration in most of our DNA affinity columns. The DNA concentration in the oligonucleotide trapping method is assumed to be 50 nM as this should be low enough to discourage binding of nonspecific proteins but sufficient to purify milligrams of the intended protein. The model also assumes that in either case, the DNA is in excess and all of the DNA binding proteins in the extract could be bound. Furthermore, it assumes binding stoichiometry is 1:1 and that equilibrium is obtained. It can be seen from the figure that conventional DNA affinity column retains at least three fold more contaminant protein than the oligonucleotide trapping method though this 3-fold outcome is highly dependant on the
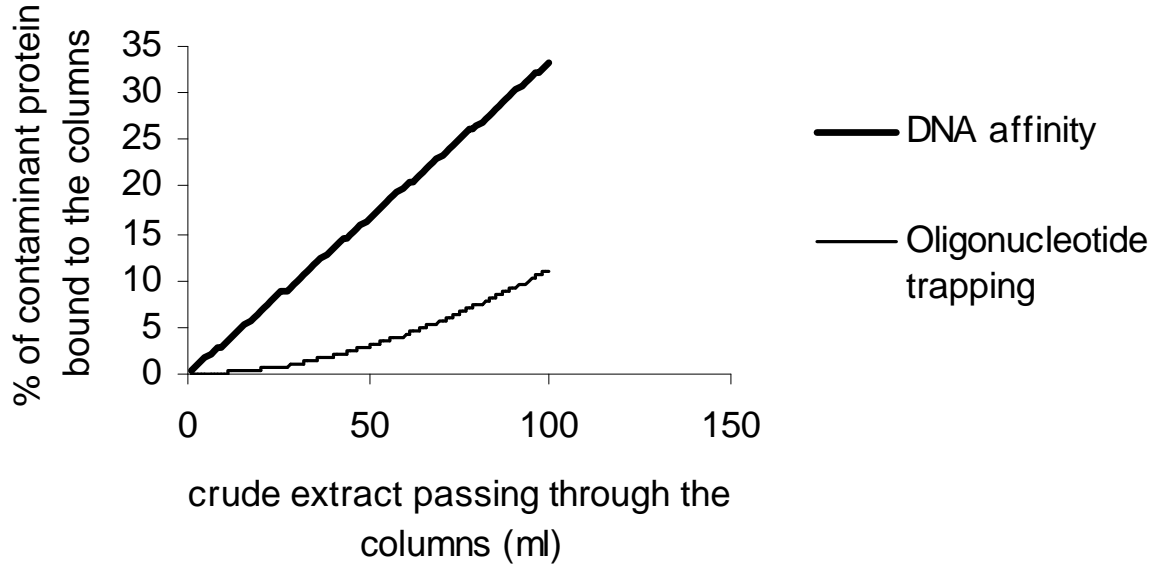
Figure 6.5. *Theoretical model for binding of nonspecific protein in the*

oligonucleotide trapping method and DNA affinity chromatography.

Binding of a nonspecific protein having an affinity of 4×10^{-5} M for the

footprint region in conventional DNA affinity (━) and oligonucleotide trapping (┅)

is shown. The concentration of DNA in DNA affinity chromatography and the

oligonucleotide trapping method is assumed to be 20 μM and 50 nM,

respectively. The binding constant of proteins for the small AC10 sequence on

the column is considered to be negligible. The model accounts for each ml of the

crude extract that passes through each column and was constructed in a

Microsoft Excel spreadsheet.

assumptions made in our model. However, different assumptions also result in higher purity by trapping method. According to the model both the methods were able to trap close to 100% of a protein of interest having an affinity of 4×10^{-12} for the DNA on the column (calculations not shown). The figure shows retention of a hypothetical contaminant protein as a crude extract passes through columns. The same fraction of the contaminant protein is retained throughout DNA affinity chromatograph and accumulates on the column because of its high DNA concentration. In the trapping method only a negligible fraction of contaminant protein is retained when the first 50 ml of crude extract passes through the column because little of the 50 nmole of DNA in the extract has yet bound to the column, increasing its DNA concentration. The major fraction of the total contaminant protein retained in the trapping method is retained only when the last few ml pass through the column, when column DNA concentrations has reached appreciable amounts. This is because, in the oligonucleotide trapping method, the initial concentration of the footprint element on the column is very low and builds up only towards the end. While in the conventional DNA affinity columns the concentration of DNA footprint remains constant and high through out.

We tried to apply this model for the actual purification of crude GFP-C/EBP. The amount of GFP-C/EBP was adjusted (by adding crude extracts from non-expressing bacteria) so that it was comparable to the low amounts present in the cells. The concentration of DNA used in the oligonucleotide trapping method and DNA affinity chromatography were similar to those in the model experiment.

It can be seen from Table 6.1 that GFP-C/EBP obtained with the oligonucleotide trapping method is purer than that obtained with conventional DNA affinity chromatography. Table 6.1 shows 3-fold higher purity; that is reasonable considering Figure 6.5. That both the model and the experiment gave 3-fold improvement is totally fortuitous but the improvement itself confirms the soundness of the basic model. It can also be seen that the yield obtained with the oligonucleotide trapping method is comparable to that obtained with DNA affinity chromatography although almost ten fold more DNA is used in the latter method.

An SDS polyacrylamide gel electrophoresis (PAGE) of the active fractions fromTable 6.1 is shown in Figure 6.6. It can be seen from the figure that GFP-C/EBP obtained (indicated by the arrow) from the oligonucleotide trapping method is significantly purer than that obtained with conventional DNA affinity chromatography. Several contaminant protein bands that are present in samples obtained from DNA affinity chromatography (lane A) are either completely absent or are highly reduced in samples obtained with oligonucleotide trapping method (lane T).

Purification of transcription factor B3 was carried out from a *Xenopus* oocyte extract by trapping and conventional DNA affinity chromatography. It can be seen from Figure 6.7 that B3 (indicated by an arrow) obtained with the trapping method (lane T) is purer than that obtained with DNA affinity chromatography (lane A) and several contaminant bands present in lane A are either absent or greatly reduced in B3 obtained from the trapping approach. Furthermore, in DNA affinity chromatography, B3 on the gel is the upper band of

Table 6.1. Balance sheet for purification of GFP-C/EBP with DNA affinity and oligonucleotide trapping method.

	Fold Purification	Yield (%)
trapping	3668 +/- 491	36 +/- 16
DNA affinity	1028 +/- 226	24 +/- 2

The results of three experiments were averaged (n=3) and averages are reported for both the columns. For DNA affinity chromatography 100 µl of purified GFP-C/EBP was mixed with 10 ml of crude bacterial extract and the volume was adjusted to 50 ml with TE0.4 and loaded onto 1 ml EP24-Sephorose column. For the oligonucleotide trapping experiment 2.5 nmoles of EP24(GT)$_5$ was added to the same extract, incubated on ice for 30 min and then loaded onto (AC)$_5$-Sepharose. Both columns were washed with 25 ml of TE0.4 and then eluted with TE1.2. Active fractions were pooled for assay. The flow rate was maintained at 0.5 ml/min and all the experiments were carried out at 4° C.

The probability (P) that the fold purification obtained by the two methods is not different is 0.00054. The probability that the yield is not different is 0.146.
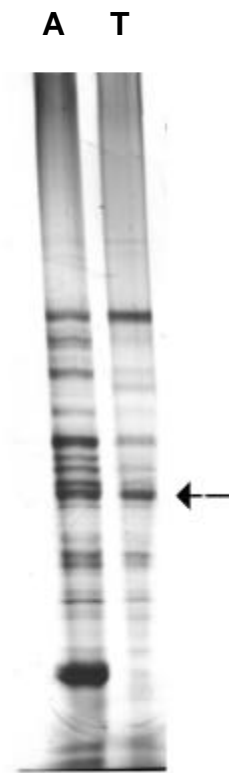
Figure 6.6. *Purity of GFP-C/EBP obtained with DNA affinity and the oligonucleotide trapping method.*

The active fractions from the balance sheet in Table 6.1 were concentrated and equal amounts of GFP-C/EBP were applied to the polyacrylamide gel and stained with silver. Lanes A and T shows proteins purified with DNA affinity chromatography and oligonucleotide trapping method, respectively.
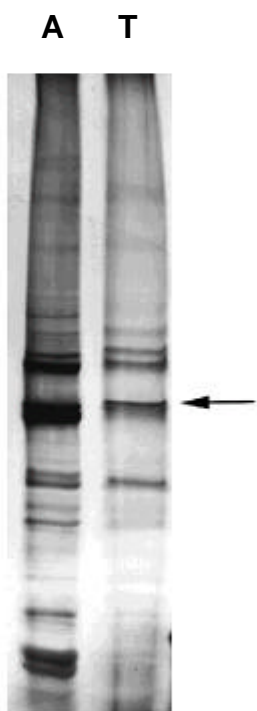
Figure 6.7. *Purification of B3 using the trapping approach.*

For DNA affinity chromatography (A) 2.5 ml of *Xenopus* oocyte extract was diluted to 50 ml with TE0.4 and loaded onto 1 ml E3-Sepharose column. For the oligonucleotide trapping experiment (T) 2.5 nmoles of E3TG$_5$/αE3TG$_5$ was added to the same extract, incubated on ice for 30 min and then loaded onto AC$_5$-Sepharose. Both columns were washed with 25 ml of TE0.4 and then eluted with TE1.2. Active fractions were concentrated and applied to SDS polyacrylamide gel and stained with silver. Lanes A and T shows proteins purified with DNA affinity chromatography and oligonucleotide trapping method, respectively. The flow rate was maintained at 0.5 ml/min and all the experiments were carried out at 4° C.

closely spaced doublet which would be difficult to remove. This contaminant is not retained by trapping.

While Figure 6.7 shows that trapping can be successfully used for obtaining high purity B3 Figure 6.8 shows that trapping approach can be further improved by using different competitors. It can be seen from the figure that heparin alone (lanes 2 and 7) and heparin along with T18 (lanes 3 and 8) could be successfully used for improving purity of B3 and the latter approach gives highly purified B3 as seen in lane 3. Specific competitor μE3 (lanes 4 and 9) and salmon sperm DNA (lanes 5 and 10) could not be used as competitors because they greatly affect the yield of B3 as can be seen in the western blot in lanes 9 and 10.

High purity B3 was also obtained with the Bi-column method (Figure 5.6). The major drawback of the bi-column method is the lower yields obtained. The Bi-column method could be scaled up to obtain higher amounts of B3 but this would involve synthesis of at least 10 ml resins of two different DNA affinity columns. Trapping on the other hand is easier to scale up. It can be seen from Figure 6.9 that highly pure B3 can be obtained (lane 2) after scaling up the trapping method used in Figure 6.8 by 10 fold. Around 5 μg B3 (data not shown) was obtained by this approach starting from 2 ml of crude oocyte extract and could be easily stained with Coomassie blue staining Figure 6.9.

Figure 6.10 shows that the B3 obtained with the trapping method is active. It can be seen from the figure that as little as 1μl of purified protein is able to shift the mobility of E3 oligonucleotide in the presence of specific competitor. 15 μl of
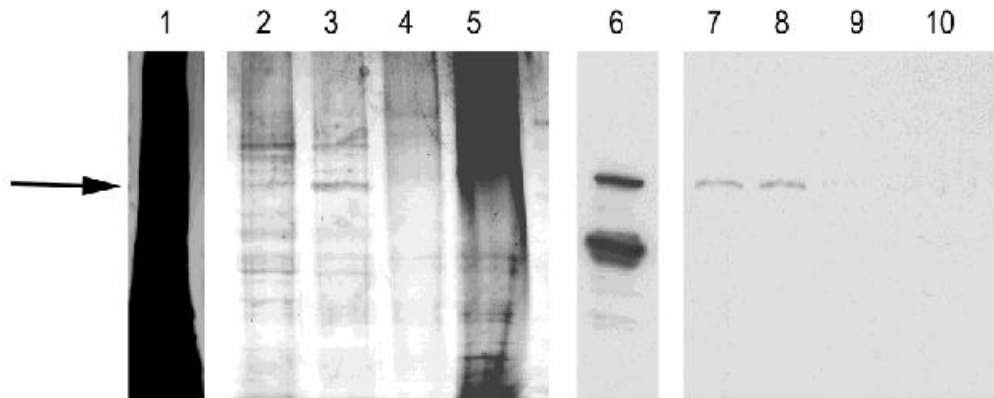
Figure 6.8. *Purification of B3 using the trapping approach in the presence of different competitors.*

200 μl of Xenopus oocyte extract was diluted to 4 ml using TE0.4 and incubated with 0.5μM of E3TG$_5$/αE3TG$_5$ along with competitors 4 mg/ml heparin (lanes 2 and 7), 4 mg/ml heparin and 20μM of T$_{18}$ (lanes 3 and 8), 5μM μE3 and 20μM T18 (lanes 4 and 9), and 1 mg/ml salmon sperm DNA (lanes 5 and 10). The mixtures were incubated on ice for 30 min and then loaded onto AC$_5$-Sepharose. Columns were washed with 25 ml of TE0.4 and then eluted with TE1.2. Active fractions were concentrated and run on a SDS 4-15 % polyacrylamide gel (lanes 1-5) and stained with silver, or subjected to western blotting (lanes 6-10) developed with NBT. Lanes 1 and 6 show 2μl crude extract.
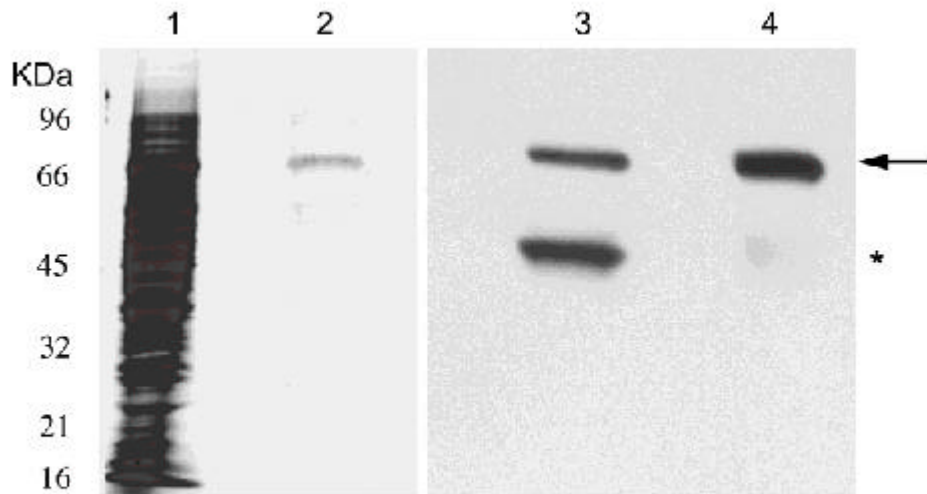
Figure 6.9. *Large scale purification of B3 using the trapping approach.*

2 ml of Xenopus oocyte extract was diluted to 40 ml using TE0.4 and incubated with 0.5μM of E3TG$_5$/αE3TG$_5$ along with competitors 4 mg/ml heparin and 20μM of T$_{18}$. The mixtures were incubated on ice for 30 min and then loaded onto a 5 ml AC$_5$-Sepharose. The column was washed with 40 ml of TE0.4 and then eluted with TE1.2. Active fractions were concentrated and applied to SDS polyacrylamide gel and stained with Coomassie blue lane 2 and 2. Or subjected to western blotting lane 4. Lanes 1 and 3 show 2μl crude.

Figure 6.10. *Gel mobility shift assay of the purified B3.*

Gel mobility shift was carried as described in (99) 20 picomoles of labeled E3TG$_5$/αE3TG$_5$ was incubated with different amounts (in µl and indicated top on of the well) of purified B3 in presence of the specific competitor. The reaction was resolved on a nondenaturing PAGE as described by Pfaff et al. (99).

purified B3 completely saturated 20 picomoles moles of E3 oligonucleotide used in the reaction.

6.4. Discussion

Coupling of DNA to solid supports in conventional DNA affinity chromatography has allowed purification of many transcription factors. But this technique suffers from two major drawbacks. As already discussed, coupling can lead to modification in the DNA sequence unless special coupling procedures are used. Base modification can affect the interaction with the protein of interest. Methods such as enzymatic synthesis or biotin avidin technologies can solve this problem to a certain extent but each of these methods has its own limitations. From Figure 6.2 it can be seen that a purer protein can be obtained with our oligonucleotide trapping method than the biotin-streptavidin method, even after using modified forms of streptavidin such as NeutrAvidin and monomeric avidin which have lower nonspecific interactions with proteins. A method similar to our method in which a footprint extended with polyA is used, has also been described (119). In that method, salt elution was used to obtain the protein and from our results it is clear that elution with high temperature is a better approach than salt elution in terms of purity of GFP-C/EBP obtained.

The second drawback of DNA affinity chromatography is that the DNA immobilized on the column is not in true solution and hence, the kinetic parameters of binding of proteins to DNA cannot be extrapolated to column chromatography. Methods such as magnetic bead purification and the previously

described polyA trapping method (119) could have been efficiently used to solve this problem, but most of the methods described so far fail to do so and employ biotin group or oligonucleotide extensions as other means of coupling a footprint region to a solid support and then perform conventional affinity chromatography. In other cases when the binding is carried out in solution the concentration of DNA used is very high. Such high concentration can encourage binding of nonspecific proteins.

In our method the binding between protein and DNA occurs in solution and only then is an $(AC)_5$ –Sepharose used to trap the DNA-protein complex. By carrying out the reaction in solution we can adjust the concentrations of DNA to levels which do not favor the binding of nonspecific proteins. This can be very rarely achieved by conventional column chromatography. The advantage of using low concentrations of DNA are clear from Table 6.1 and Figure 6.6 which show that GFP-C/EBP obtained from our approach is significantly purer than that obtained with conventional chromatography. It is important to note that we have used high salt to elute GFP-C/EBP in the oligonucleotide trapping method in Table 6.1 and Figure 6.6. High salt is not the best method for elution and better purity can be obtained with temperature elution (Figure 6.3). However since accurate measurements of fluorescence are required for the balance sheet in Table 6.1 and the DNA that is bound to protein after temperature elution can affect its fluorescence, salt elution was chosen.

Several competitors such as heparin, specific DNA (μE3), nonspecific DNA (salmon sperm DNA), and single stranded DNA (T_{18}) could be used

together with the trapping approach. It is noteworthy that the use of single stranded DNA along with heparin was the most efficient in purifying B3. T_{18} probably competes with AC5 for binding to single strand binding proteins and hence prevents contamination by these proteins.

From the overall results it can be seen that our oligonucleotide trapping method is a better alternative to conventional DNA affinity chromatography. The oligonucleotide trapping method can be further enhanced by using competitors to obtain highly purified DNA binding proteins.

Chapter: 7. Conclusions

Sequence specific DNA affinity chromatography has the highest selectivity and is the most powerful chromatographic technique known for the purification of transcription factors and other DNA binding proteins. As seen from Table 1.1 DNA affinity chromatography has been widely used for the purification of transcription factors. Pioneering work such as coupling of DNA to silica, the enzymatic synthesis of DNA affinity columns, and the use of competitor DNA has led to significant advances in the technique. However, despite of these efforts DNA affinity chromatography alone has rarely yielded homogenous protein. This is mainly because the technique in its original form had a restricted number of available elution strategies and other aspects, such as the appropriate length and concentration of DNA used for making the affinity columns, were also not fully characterized. Hence the main focus of our study was to optimize DNA affinity chromatography by developing new elution strategies and by defining the nature of DNA used for making sequence specific columns and to apply the knowledge obtained to develop new affinity methods for the purification of transcription factors.

We have shown that concatemeric DNA affinity columns suffer from low resolution, which affects the purity of proteins obtained (Chapter 3). Moreover the most efficient way to increase the complexity of DNA sequence used for coupling is to include homopolymeric sequences such as polydA:polydT at either 5' or 3' end of the footprint region (Chapters 2 and 3).

Heparin elution was developed as an alternative to salt elution (Chapter 4). In our experiments with lac repressor protein we got higher purity with heparin

elution than with salt elution. Other DNA binding proteins when eluted with heparin may or may not have higher purity than salt elution. Heparin elution offers a new variation in DNA affinity chromatography and allows repeated use of DNA affinity chromatography with different elution strategies. This is epitomized in the Bi-column method, which involves the use of two sequential sequence specific columns with different specificity and two different elution strategies, viz. Salt and heparin elution. This method was based on our prior knowledge that DNA affinity columns containing an untailed footprint region show shorter retention times than columns containing more complex, tailed footprints (Chapters 2 and 3). This method was successfully used for purification of three different transcription factors (Chapter 5) to near homogeneity in a single operation.

Conventional DNA affinity columns contain high concentration of DNA which can lead to retention of nonspecific proteins. Using lower concentration of DNA in should lead to increased purity but would be expected to adversely affect the yield. The oligonucleotide trapping method was developed to allow purification of milligram quantities of protein by using a tailed footprint region at concentration low enough to discourage binding of nonspecific proteins. By binding protein to DNA in solution, yield does not suffer table 6.1) and capacity is actually improved (Figure 6.4). This method was successfully used to obtain highly pure B3 in quantities high enough to allow further analysis.

This work has improved our understanding of DNA affinity chromatography and has led to development of new methods for the purification

of transcription factors. Some of the highlights of this work have been single step purification of B3 and up to 3000 fold purification of recombinant C/EBP which is the highest fold purification obtained using DNA affinity chromatography in our lab. The methods developed could also be used for the purification of other DNA binding proteins with only slight modifications and whether or not this work finds future application remains to be seen.

List of references

1. Liu, S., Shapiro, R. A., Nie, S., Zhu, D., Vodovotz, Y., and Billiar, T. R. (2000) *Gene* **250**(1-2)**,** 137-47

2. Heath, V. J., Gillespie, D. A., and Crouch, D. H. (2000) *Exp. Cell Res.* **254**(1)**,** 91-8

3. Izumia, M., Yokoib, M., Nishikawab, N. S., Miyazawab, H., Suginoc, A., Yamagishid, M., Yamaguchid, M., Matsukaged, A., Yatagaiab, F., and Hanaokabe, F. (2000) *Biochim. Biophys. Acta* **1492,** 341-52

4. Muller-Hill, B. (1975) *Prog. Biophys Mol. Biol.* **30,** 227-52

5. Ohshima, Y., Mizokoshi, T., and Horiuchi, T. (1974) *J. Mol. Biol.* **89,** 127-36

6. Ackers, G. K., Johnson, A. D., and Shea, M. A. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 1129-33

7. Latchman, D. S. (1996) *N. Engl. J. Med.* **334,** 28-33

8. Latchman, D. S. (1993) *Int. J. Exp. Pathol.* **74,** 417-22

9. Latchman, D. S. (1997) *Int. J. Biochem. Cell Biol.* **29**(12)**,** 1305-12

10. Johnson, P. F., and McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58,** 799-839

11. Mitchell, P. J., and Tjian, R. (1989) *Science* **245**(4916)**,** 371-8

12. Frankel, A. D., and Kim, P. S. (1991) *Cell* **65**(5)**,** 717-9

13. Gralla, J. D. (1992) in *Transcriptional Regulation*, pp. 629-42, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

14. Harrison, S. C. (1991) *Nature* **353**(6346)**,** 715-9

15.   Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053-95

16.   Draganescu, A., Levin, J. R., and Tullius, T. D. (1995) *J. Mol. Biol.* **250**(5)**,** 595-608

17.   Wintjens, R., and Rooman, M. (1996) *J. Mol. Biol.* **262**(2)**,** 294-313

18.   Harrison, S. C., and Aggarwal, A. K. (1990) *Annu. Rev. Biochem.* **59,** 933-69

19.   Alber, T. (1992) *Curr. Opin. Genet. Dev.* **2**(2)**,** 205-10

20.   Garrell, J., and Campuzano, S. (1991) *Bioessays* **13**(10)**,** 493-8

21.   Briggs, M. R., Kadonaga, J. T., Bell, S. P., and Tjian, R. (1986) *Science* **234,** 47-52

22.   Cao, Z., Umek, R. M., and McKnight, S. L. (1991) *Genes Dev.* **5,** 1538-52

23.   Perkins, K. K., Admon, A., Patel, N., and Tjian, R. (1990) *Genes Dev.* **4**(5)**,** 822-34

24.   Riggs, A. D., Bourgeois, S., Newby, R. F., and Cohn, M. (1968) *J. Mol. Biol.* **34**(2)**,** 365-8

25.   Oehler, S., Alex, R., and Barker, A. (1999) *Anal. Biochem.* **268**(2)**,** 330-6

26.   Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) *Biochemistry* **20,** 6929-48

27.   Winter, R. B., Berg, O. G., and von Hippel, P. H. (1981) *Biochemistry* **20**(24)**,** 6961-77

28.   Lane, D., Prentki, P., and Chandler, M. (1992) *Microbiol. Rev.* **56**(4)**,** 509-28

29.   Ludwig, L. B., Hughes, B. J., and Schwartz, S. A. (1995) *Nucleic Acids*

Res. **23**(18)**,** 3792-3

30.   Suske, G., Gross, B., and Beato, M. (1989) *Nucleic Acids Res.* **17**(11)**,**
4405

31.   Ozyhar, A., and Kiltz, H. H. (1991) *J. Chromatogr.* **587**(1)**,** 11-7

32.   Lim, W. A., Sauer, R. T., and Lander, A. D. (1991) *Methods Enzymol.* **208,**
196-210

33.   Stebbins, M. A., Hoyt, A. M., Sepaniak, M. J., and Hurlburt, B. K. (1996) *J*
Chromatogr. B Biomed. Appl. **683**(1)**,** 77-84

34.   Foulds, G. J., and Etzkorn, F. A. (1998) *Nucleic Acids Res.* **26**(18)**,** 4304-5

35.   Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A.,
and Tullius, T. D. (1991) *Methods Enzymol.* **208,** 380-413

36.   Hayes, J. J., Kam, L., and Tullius, T. D. (1990) *Methods Enzymol.* **186,**
545-9

37.   Heyduk, T., Ma, Y., Tang, H., and Ebright, R. H. (1996) *Methods Enzymol.*
274, 492-503

38.   Yang, W. P., Wu, H., and Barbas, C. F. (1995) *J. Immunol. Methods*
183(1)**,** 175-82

39.   Kovelman, R., and Roeder, R. G. (1992) *J Biol Chem* **267**(34)**,** 24446-56

40.   Malhotra, P., Manohar, C. F., Swaminathan, S., Toyama, R., Dhar, R.,
Reichel, R., and Thimmapaya, B. (1993) *J Biol Chem* **268**(27)**,** 20392-401

41.   Potter, D. A., Larson, C. J., Eckes, P., Schmid, R. M., Nabel, G. J.,
Verdine, G. L., and Sharp, P. A. (1993) *J Biol Chem* **268**(25)**,** 18882-90

42.   Wen, P., and Locker, J. (1994) *Mol Cell Biol* **14**(10)**,** 6616-26

43.   Moorefield, B., and Roeder, R. G. (1994) *J Biol Chem* **269**(33)**,** 20857-65

44.   Corthesy, B., and Kao, P. N. (1994) *J Biol Chem* **269**(32)**,** 20682-90

45.   Liu, F., and Bateman, E. (1994) *J Biol Chem* **269**(28)**,** 18541-8

46.   Bukenberger, M., Dingermann, T., Meissner, W., Seifart, K. H., and
       Winckler, T. (1994) *Eur J Biochem* **220**(3)**,** 839-46

47.   Eggert, M., Mows, C. C., Tripier, D., Arnold, R., Michel, J., Nickel, J.,
       Schmidt, S., Beato, M., and Renkawitz, R. (1995) *J Biol Chem* **270**(51)**,**
       30755-9

48.   Suzuki, H., Suzuki, S., Kumar, S., and Ozawa, T. (1995) *Biochem Biophys
       Res Commun* **213**(1)**,** 204-10

49.   Ripperger, J. A., Fritz, S., Richter, K., Hocke, G. M., Lottspeich, F., and
       Fey, G. H. (1995) *J. Biol. Chem.* **270,** 29998-30006

50.   Adams, A. D., Choate, D. M., and Thompson, M. A. (1995) *J Biol Chem*
       270(12)**,** 6975-83

51.   Powers, C., Krutzsch, H., and Gardner, K. (1996) *J Biol Chem* **271**(47)**,**
       30089-95

52.   Berthelsen, J., Vandekerkhove, J., and Blasi, F. (1996) *J Biol Chem*
       271(7)**,** 3822-30

53.   He, F., Narayan, S., and Wilson, S. H. (1996) *Biochemistry* **35,** 1775-82

54.   Ishisaki, A., Murayama, T., Ballagi, A. E., and Funa, K. (1997) *Eur J
       Biochem* **246**(1)**,** 142-6

55.   McPherson, L. A., Baichwal, V. R., and Weigel, R. J. (1997) *Proc Natl
       Acad Sci U S A* **94**(9)**,** 4342-7

56. Lam, L. T., and Bresnick, E. H. (1998) *J Biol Chem* **273**(37)**,** 24223-31

57. Bing, Z., Reddy, S. A., Ren, Y., Qin, J., and Liao, W. S. (1999) *J Biol Chem* **274**(35)**,** 24649-56

58. Wang, D., and Kudlow, J. E. (1999) *Biochim Biophys Acta* **1449**(1)**,** 50-62

59. Jarrett, H. W. (2000) *Anal. Biochem.* **279,** in press.

60. Gadgil, H., and Jarrett, H. W. (1999) *J. Chromatogr. A* **848,** 131-8

61. Gadgil, H., Taylor, W. L., and Jarrett, H. W. (2001) *J. Chromatogr. A* **, 905**, 133-9

62. Yeh, W. C., Hou, J., and McKnight, S. L. (1996) *Methods Enzymol.* **274,** 101-12

63. Nelson, T. J., and Kaufman, S. (1987) *Arch. Biochem. Biophys* **257**(1)**,** 69-84

64. Govindan, M. V., and Gronemeyer, H. (1984) *J. Biol. Chem.* **259**(20)**,** 12915-24

65. Abdullah, M., Siddiqui, A. A., Hill, J. A., and Davies, R. J. (1983) *Arch. Biochem. Biophys* **225**(1)**,** 306-12

66. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science* **240,** 1759-64

67. Alberts, B. M., Amodio, F. J., Jenkins, M., Gutman, E. D., and Ferris, F. J. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 289-305

68. Chockalingam, P. S., Jurado, L. A., Robinson, F. D., and Jarrett, H. W. (2000) *Methods in Molec. Biol.* **147,** 141-153

69. Goss, T. A., Bard, M., and Jarrett, H. W. (1990) *J. Chromatogr. A* **508,**

157

279-87

70. Wada, T., Watanabe, H., Kawaguchi, H., and Handa, H. (1995) *Methods Enzymol.* **254,** 595-604

71. Inomata, Y., Wada, T., Handa, H., Fujimoto, K., and Kawaguchi, H. (1994) *J. Biomater. Sci. Polym. Ed.* **5**(4)**,** 293-302

72. Leblond-Francillard, M., Dreyfus, M., and Rougeon, F. (1987) *Eur. J. Biochem.* **166,** 351-5

73. Kasher, M. S., Pintel, D., and Ward, D. C. (1986) *Mol. Cell Biol.* **6,** 3117-27

74. DiRusso, C., Rogers, R. P., and Jarrett, H. W. (1994) *J. Chromatogr. A* **677**(1)**,** 45-52

75. Robinson, F. D., Gadgil, H., and Jarrett, H. W. (1999) *J. Chromatogr. A* **849,** 403-12

76. Jarrett, H. W. (1993) *J. Chromatogr. B* **618,** 315-39

77. Kadonaga, J. T., and Tjian, R. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83,** 5889-93

78. Kadonaga, J. T. (1991) *Methods Enzymol.* **208,** 10-23

79. Whitson, P. A., Olson, J. S., and Mathews, K. S. (1986) *Biochemistry* **25,** 3852-8

80. Alberts, B., and Herrick, G. (1971) *Methods Enzymol.* **21,** 198-217

81. Solomon, L. R., Massom, L. R., and Jarrett, H. W. (1992) *Anal. Biochem.* **203,** 58-69

82. Jarrett, H. W. (1996) *J. Chromatogr. A* **742,** 87-94

83. Jarrett, H. W. (1995) *J. Chromatogr. A* **708**(1)**,** 13-8

84. Brake, A. J., Fowler, A. V., Zabin, I., Kania, J., and Muller-Hill, B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4824-27

85. Kania, J., and Brown, D. T. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3529-33

86. Kania, J., and Muller-Hill, B. (1977) *Eur. J. Biochem.* **79,** 381-6

87. Laemmli, U. K. (1970) *Nature* **227,** 680-5

88. Hsieh, M., and Brenowitz, M. (1997) *J. Biol. Chem.* **272,** 22092-96

89. Riggs, A. D., Bourgeois, S., and Cohn, M. (1970) *J. Mol. Biol.* **53,** 401-17

90. Richter, P. H., and Eigen, M. (1974) *Biophys. Chem.* **2,** 255-63

91. Ruusala, T., and Crothers, D. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4903-07

92. Whitson, P. A., and Mathews, K. S. (1986) *Biochemistry* **25,** 3845-52

93. Khoury, A. M., Lee, H. J., Lillis, M., and Lu, P. (1990) *Biochim. Biophys. Acta* **1087,** 55-60.

94. Lin, S. Y., and Riggs, A. D. (1970) *Nature* **228,** 1184-6.

95. Arndt-Jovin, D. J., Jovin, T. M., Bahr, W., Frischauf, A.-M., and Marquardt, M. (1975) *Eur. J. Biochem.* **54,** 411-8

96. Khoury, A. M., Nick, H. S., and Lu, P. (1991) *J. Mol. Biol.* **219,** 623-34

97. Muller-Hill, B., and Kania, J. (1974) *Nature* **249,** 561-3

98. Jarrett, H. W., and Taylor, W. L. (1998) *J. Chromatogr. A* **803,** 131-9

99. Pfaff, S. L., and Taylor, W. L. (1992) *Dev. Biol.* **151,** 306-16

100. Rosenfeld, P. J., and Kelly, T. J. (1986) *J. Biol. Chem.* **261,** 1398-408

101. Blanks, R., and McLaughlin, L. W. (1988) *Nucleic Acids Res.* **16,** 10283-

102. Hatamochi, A., Golumbek, P. T., Van Schaftingen, E., and de Crombrugghe, B. (1988) *J. Biol. Chem.* **263,** 5940-7

103. Duncan, C. H., and Cavalier, S. L. (1988) *Anal. Biochem.* **169,** 104-8

104. Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res.* **9,** 6505-25

105. Vlatakis, G., and Bouriotis, V. (1991) *Anal. Biochem.* **195,** 352-7

106. Fusi, P., Tedeschi, G., Aliverti, A., Ronchi, S., Tortora, P., and Guerritore, A. (1993) *Eur J Biochem* **211**(1-2)**,** 305-10

107. Soncin, F., Prevelige, R., and Calderwood, S. K. (1997) *Protein Expr Purif* **9**(1)**,** 27-32

108. Calaycay, J., Pande, H., Lee, T., Borsi, L., Siri, A., Shively, J. E., and Zardi, L. (1985) *J Biol Chem* **260**(22)**,** 12136-41

109. Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E., and Tulinsky, A. (1993) *J Biol Chem* **268**(24)**,** 17651-4

110. Cuatrecasas, P. (1970) *J. Biol. Chem.* **245,** 3059-65

111. Mulder, E., Foekens, J. A., Peters, M. J., and van der, M. (1979) *FEBS Lett* **97**(2)**,** 260-4

112. Dickson, R. C., Abelson, J. N., Barnes, W. M., and Reznikoff, W. S. (1975) *Science* **182,** 27-35

113. deHaseth, P. L., Gross, C. A., Burgess, R. R., and Record, M. T., Jr. (1977) *Biochemistry* **16**(22)**,** 4777-83

114. Hughes, M. J., Liang, H. M., Jiricny, J., and Jost, J. P. (1989) *Biochemistry* **28**(23)**,** 9137-42

115. Kaufman, P. D., Doll, R. F., and Rio, D. C. (1989) *Cell* **59**(2)**,** 359-71

116. Gorski, K., Carneiro, M., and Schibler, U. (1986) *Cell* **47**(5)**,** 767-76

117. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc Natl Acad Sci U S A* **76**(9)**,** 4350-4

118. Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986) *Molec. Cell Biol.* **6,** 4723-33

119. Rutberg, S. E., and Ronai, Z. (1992) *Nucleic Acids Res.* **20,** 1815

120. Kang, S., Xu, X., Heidenreich, O., Gryaznov, S., and Nerenberg, M. (1995) *Nucleic Acids Res.* **23,** 2344-5

121. Lekstrom-Himes, J., and Xanthopoulos, K. G. (1998) *J Biol Chem* **273**(44)**,** 28545-8

122. Aniskovitch, L. P., and Jacob, S. T. (1997) *Arch Biochem Biophys* **341**(2)**,** 337-46

123. Hornes, E., and Korsnes, L. (1990) *Genet Anal Tech Appl* **7**(6)**,** 145-50

124. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248-54

125. Revzin, A., and von Hippel, P. H. (1977) *Biochemistry* **16,** 4769-76

126. Hsieh, W. T., Whitson, P. A., Matthews, K. S., and Wells, R. D. (1987) *J Biol Chem* **262**(30)**,** 14583-91

Vita

Himanshu S. Gadgil, was born in Pune, India, on March 5, 1974. He received the Master's degree in Microbiology from the University of Pune. He joined the University of Tennessee, Memphis, in fall 1997.