ELSEVIER

Journal of Chromatography A, 917 (2001) 43–53

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Comparative studies on discrete and concatemeric DNA-Sepharose columns for purification of transcription factors

Himanshu Gadgil, William L. Taylor, Harry W. Jarrett*

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

Received 24 October 2000; received in revised form 13 February 2001; accepted 13 February 2001

Abstract

Concatemers, tandem copies of DNA elements ligated together, are widely used for the DNA affinity chromatography of transcription factors. Purification of different transcription factors using discrete, concatemeric and T_{18}:A_{18} tailed DNA affinity columns was studied. Columns having a discrete DNA sequence bound by cytidylic-adenylic-adenylic-thymidylic oligonucleotide (CAAT) enhancer binding protein (C/EBP) yields significantly more green fluorescent protein–C/EBP (GFP–C/EBP) fusion protein than a concatemeric DNA column made from five tandem repeats of the same DNA sequence. For lac repressor protein, the concatemeric and T_{18}:A_{18} tailed columns show greater retention times than a discrete, untailed DNA affinity column. It was observed that the T_{18}:A_{18} tailed column gives better resolution than either the discrete or concatemeric columns, of mixtures containing both lac repressor and GFP–C/EBP. Discrete concatemeric and T_{18}:A_{18} tail columns all bound the Sp1 transcription factor and showed similar retention. The T_{18}: A_{18} tailed column gives higher yield for Sp1 than the other columns. Our study shows concatemeric columns do not have any distinct advantage for the three different transcription factors we studied including Sp1, the original justification for the concatemeric approach.

Keywords: Transcription factors; Concatemers; Stationary phases, LC; DNA; Proteins

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 [1–3].

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 [1,2]. 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 [4,5]. Sp1 binds a GC-rich decanucleotide, the GCbox. The early promoter of the SV40 virus contains six

*Corresponding author. Tel.: +1-901-4487-078; fax: +1-901-4487-360.

E-mail address: hjarrett@utmem.edu (H.W. Jarrett).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)00647-1

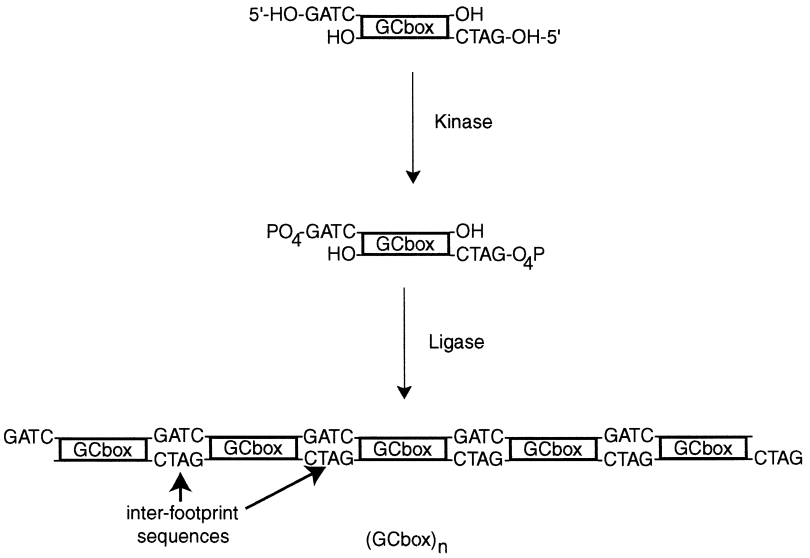

Fig. 1. Schematic representation of the strategy used for production of concatemeric DNA sequences.

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 [1]. Concatemers are usually made by ligating single copy DNA strands having a complementary overhang sequence. Catenating DNA makes it more complex[1] 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 Fig. 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 *Escherichia 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 [6]. 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 [7,8]. 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 [9] 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 DNA columns. Both fusion proteins have DNA binding properties

[1]DNA complexity and length are related but not identical concepts. A longer DNA sequence is of necessity more complex (i.e., complicated) than a shorter one. However, two sequences of the same length can differ in complexity if one is more repetitive than the other. Introducing homopolymeric sequence such as T_{18}:A_{18} results in increase in length of the DNA with little effect on its complexity.

comparable to the native transcription factors [10–12]. 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.

2. Methods

2.1. DNA Sepharose preparation

The oligonucleotides shown in Table 1 were used for coupling to Sepharose. All strands having Aminolink were coupled to CNBr-preactivated Sepharose 4B (Sigma). 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. 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 and Tjian [5]. A schematic description of concatemer formation is given in Fig. 1. A mixture of oligomers containing one to 20 copies of the original sequence was obtained as determined by their mobility on agarose gel electrophoresis. 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 nmol, 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 mass 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 $GCboxT_{18}$, respectively. All columns were stored at 4°C in TE0.1 buffer (10 m*M* Tris, pH 7.5, 1 m*M* EDTA, 0.1 *M* NaCl) containing 10 m*M* NaN_3 when not in use.

2.2. Production of proteins

Lac repressor-β-galactosidase fusion protein was produced as described earlier [13] by growing clone BMH-72-19-1, which was the generous gift of Dr. David Levens (Laboratory of Pathology, National Cancer Institute, Bethesda, MD, USA). 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 [12]. The crude bacterial extract (2.9 mg/ml, 1.4–2% of which is GFP–C/EBP) was used for the experiment in Table 1. The

Table 1
Oligonucleotides used for making DNA affinity columns

Name	Sequence of strand which was coupled	Sequence of complementary strand
Op1	5′-NH_2AATTGTTATCCGCTCACAATTCCAC**	5′GTGGAATTGTGAGCGGATAACAATT**
$Op1T_{18}$	5′-NH_2-$(T)_{18}$AATTGTTATCCGCTCACAATTCCAC	5′GTGGAATTGTGAGCGGATAACAATT$(A)_{18}$
$(Op1)_4$	5′-NH_2(AATTGTTATCCGCTCACAATTCCAC)$_4$	(5′GTGGAATTGTGAGCGGATAACAATT)$_4$
EP18	5′-NH_2GCAGATTGCGCAATCTGC	NA*
$(EP18)_5$	5′-NH_2-(GCAGATTGCGCAATCTGC)$_5$	NA*
GCbox	5′-NH_2GGGGCGGGGC	5′GCCCCGCCCC
$GCboxT_{18}$	5′-NH_2-$(T)_{18}$GGGGCGGGGC	5′GCCCCGCCCC$(A)_{18}$
$(GCbox)_n$	5′-GATCGGGGCGGGGC	5′-GATCGCCCCGCCCC

*NA stands 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 an aminoethyl group added on the last synthetic cycle with the Aminolink reagent (Applied Biosystems).

protein was also purified by Ni^{2+}-NTA-agarose (Qiagen) as described [12].

HeLa cell nuclear extract (1.8 mg/ml) used for studies on Sp1 was obtained by the procedure in Ref. [5].

2.3. Chromatography

All columns were 1 ml bed volume syringe columns initially equilibrated in TE0.1 (10 m*M* Tris, pH 7.5, 1 m*M* EDTA, 0.1 *M* NaCl) 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 m*M* Tris, pH 7.5, 1 m*M* EDTA, 1.2 *M* NaCl). The flow-rate was 0.5 ml/min and 1 ml fractions were collected.

2.4. Assay of fusion proteins

GFP–C/EBP was assayed by measuring fluorescence as described earlier [12]. LacIZ was assayed for β-galactosidase activity by mixing 150 μl of Buffer O (3 m*M* *o*-nitrophenyl-β-D-galactopyranosidase, 0.1 *M* NaH_2PO_4, 1 m*M* $MgCl_2$ and 45 m*M* β-mercaptoethanol, pH 7.2) with 50 μl of each fraction to be assayed. The reaction was carried out on microtiter plates and monitored continuously for absorption at 405 nm over time at 25°C.

2.5. Protein assay

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

2.6. Blotting experiments

A 0.6-ml volume of each column fraction was applied to 0.45-μm pore nitrocellulose filter paper in a Bio-Rad 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 m*M* 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 (1 mg/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 (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.

2.7. Polyacrylamide gel electrophoresis

All samples were concentrated using Centriplus centrifugal filter devices supplied by Millipore (Bedford, MA, USA). One fourth of each sample was applied to 4–15% Bio-Rad precast gradient sodium dodecyl sulfate polyacrylamide gels using the method of Laemmli [14] and stained after electrophoresis with silver using the Bio-Rad laboratory kit (Richmond, CA, USA).

3. Results and discussion

Fig. 2 shows elution of highly purified GFP–C/EBP from the EP18 and the $(EP18)_5$ columns, the latter containing five 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 and 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

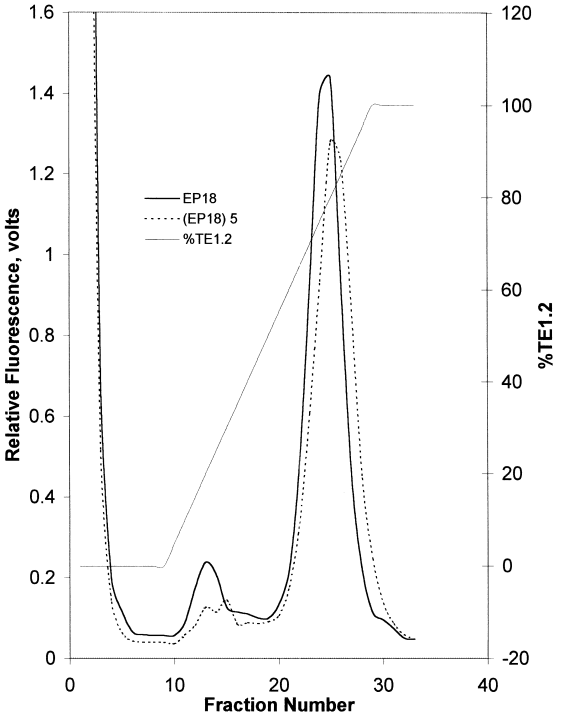

Fig. 2. Elution of GFP–C/EBP from EP18 and $(EP18)_5$ columns. A 100-μl volume of purified GFP–C/EBP was loaded on a 1 ml EP18-Sepharose (solid line) or $(EP18)_5$ (dashed line) column. Chromatography was performed as shown in Section 2.3.

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 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 $(EP18)_5$ column with $P=0.008$ (i.e., significantly different at the 1% 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 Fig. 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 foreshortened sequences as a consequence of less than 100% efficiency at each step of synthesis. These foreshortened sequences may have imperfect elements which function poorly or not at all. Since DNA complexity had little effect on retention time (Fig. 2), other ways of generating complex DNA, i.e., tailing, were not investigated.

As a consequence of characterizing columns prepared by enzymatic primer extension [9], 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 Fig. 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 detail elsewhere [9]. Interestingly, the more complex columns also appear to resolve more than one species of lac repressor fusion protein.

Table 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 (V/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 Section 2.3. Active fractions were pooled for assay.

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

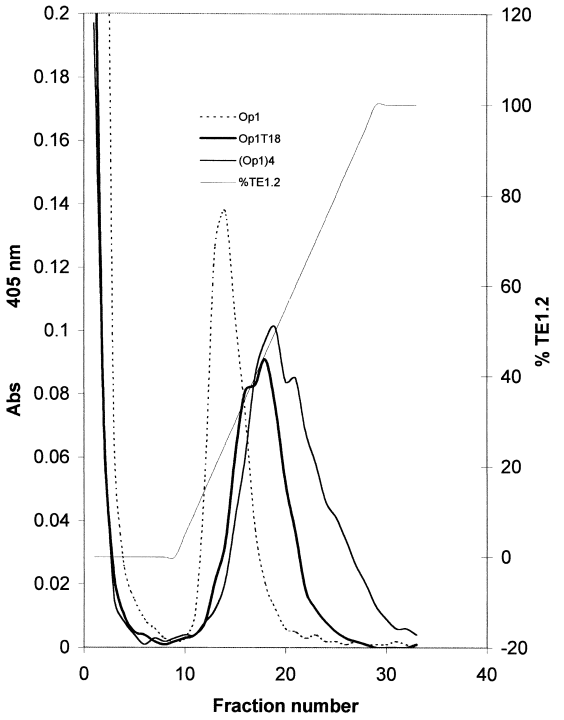

Fig. 3. Elution of lacIZ from the Op1, the $Op1T_{18}$ and the $(Op1)_4$ columns. A 100-μl volume of crude bacterial extract containing lacIZ was loaded onto a 1 ml Op1, (dashed line), $Op1T_{18}$ (bold line), or $(Op1)_4$-Sepharose (solid line) column.

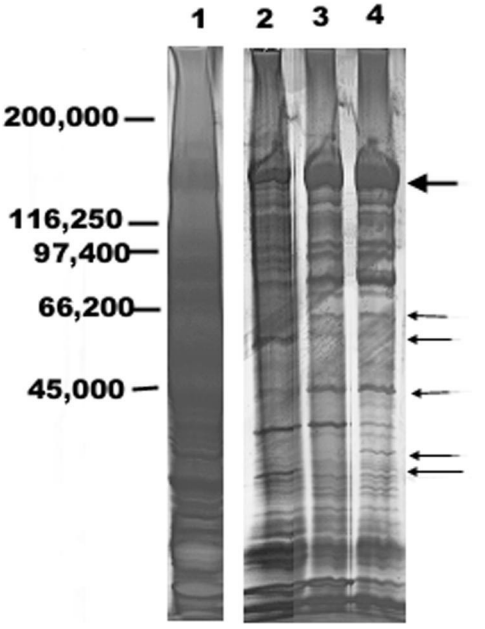

Fig. 4. Purity of lacIZ from the Op1, the $Op1T_{18}$ and the $(Op1)_4$ columns. A 500-μl volume of crude bacterial extract (as in Fig. 3) containing lacIZ was loaded onto a 1 ml Op1, $Op1T_{18}$ or $(Op1)_4$-Sepharose column. Lane 1 shows 100 times diluted crude extract while lanes 2, 3 and 4 represent 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 masses of standards run on the same gel.

Previously [9] 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 Fig. 4 that lacIZ (indicated by the larger, darker arrow near the top of the gel) obtained from $Op1T_{18}$ column (lane 3) 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.

Fig. 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 [13]. 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 Table 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 Fig. 5A that the Op1 column shows poor resolution of the two proteins with R_s=0.143. R_s, the resolution factor, is the ratio of the separation between the two peaks divided by the mean peak width. A value of R_s 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.

Fig. 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_s= 0.133, primarily because the peaks are broad on this concatemer column. Broad peaks could result from a heterogeneous stationary phase which is to be expected for this 100-mer DNA, heterogeneous because of sequence foreshortening as described above.

Of all the three columns $Op1T_{18}$ column gives the best resolution with R_s=0.4 (Fig. 5C). The better resolution is because of the greater 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, Fig. 5 reveals the basis of the selectivity differences observed in Fig. 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 coupling step. Concatemers produced by ligation would also be heterogeneous mixtures. This contributes heterogeneity to the stationary phase and this may account 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 [5]. To our surprise Sp1 binds to all three columns equally well (Fig. 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 Fig. 6 shows that the elution behavior was similar for all three columns, Fig. 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, twofold serial dilutions of each pool were made and blotted onto nitrocellulose paper and Sp1 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 $GCboxT_{18}$ column is about fourfold 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 fourfold 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 [9]. The $(GCbox)_n$ column, on the other hand, has several tandem repeats of the GCbox. Ligation introduces additional DNA sequences [5], as can be seen from the schematic in Fig. 1. The additional sequences are comprised of the inter-footprint sequences produced by ligation, which may be binding sites for other DNA binding proteins; and the binding of these proteins to the DNA could block the binding site for Sp1 and decrease yield. Alternatively, the protracted procedure for preparing concatem-

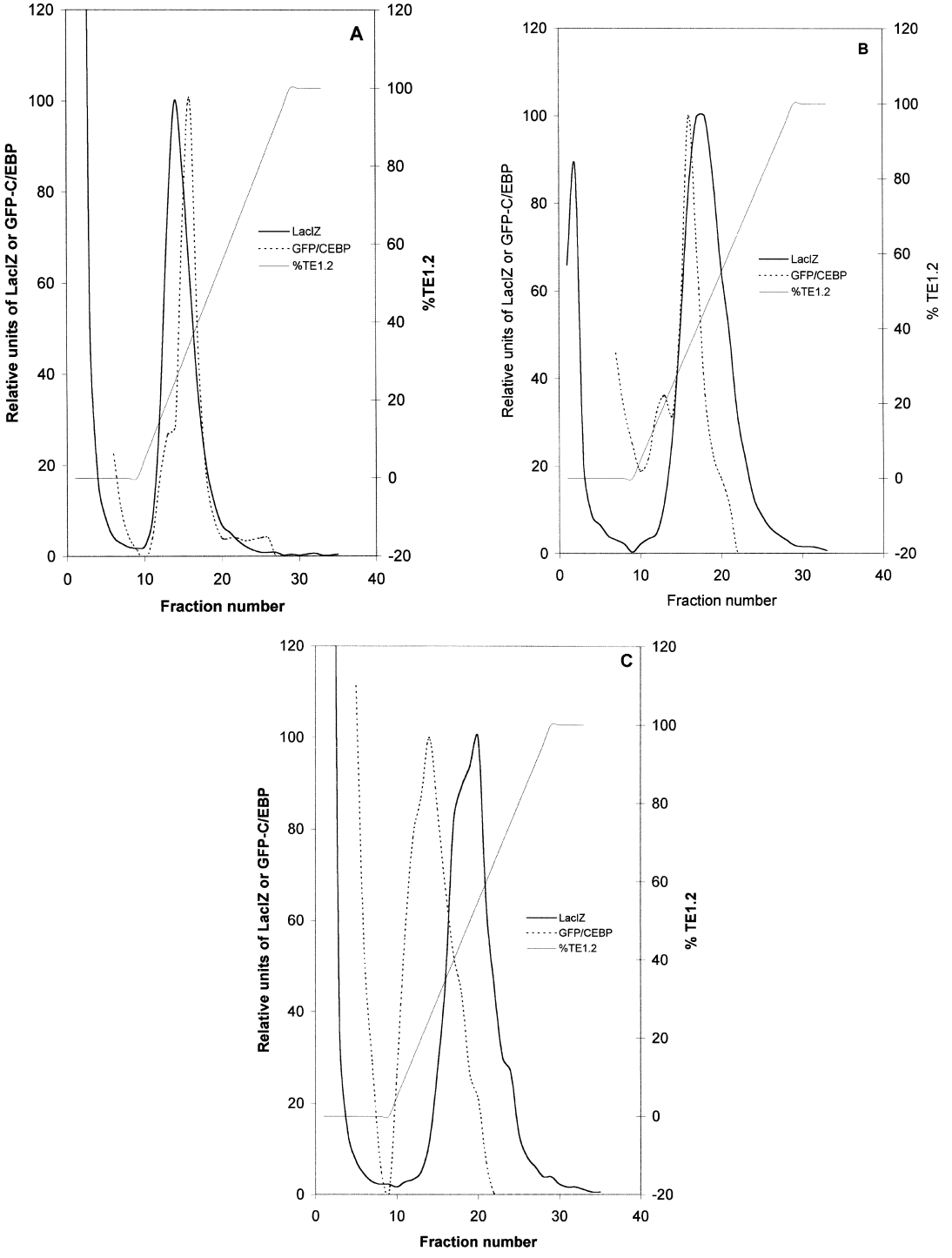

Fig. 5. Resolution of lacIZ and GFP–C/EBP on different DNA-Sepharose columns. A 100-μl volume of a crude preparation of lacIZ (as in Fig. 3) was mixed with 100 μl of purified GFP–C/EBP and loaded onto Op1 (A), the $(Op1)_4$, (B) or the $Op1T_{18}$ column (C). Fluorescence of GFP–C/EBP for early fractions is not shown because of interfering fluorescence of crude cell constituents flowing through the column.

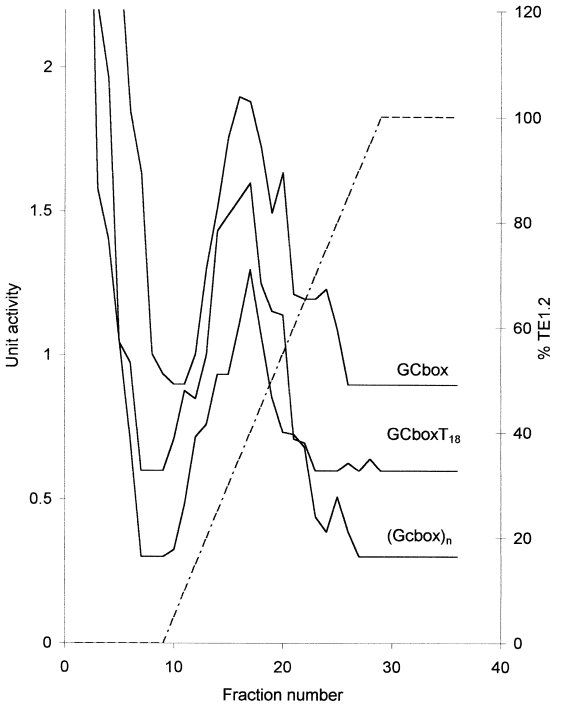

Fig. 6. Elution of Sp1 from different GCbox columns. A 500-μl volume of HeLa Cell nuclear extract was loaded onto a 1 ml GCbox, $(GCbox)_n$, or $GCboxT_{18}$-Sepharose column. A 600-μl volume of each fraction (1 ml) was blotted onto nitrocellulose paper using Bio-Rad slotblot apparatus and probed with 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.

Fig. 7. $GCboxT_{18}$-Sepharose yields more Sp1 than GCbox and$(GCbox)_n$-Sepharose. Peak fractions from each of the column showing Sp1 in Fig. 6 were pooled and twofold serial dilutions were made. A 600-μl volume of each dilution was blotted on nitrocellulose paper and probed with anti Sp1 (PEP 2) antibody.

ers 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 (Fig. 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 (Fig. 3). Of these two types, Sp1 falls in the first group since it elutes from simple and complex columns with quite similar retention times (Fig. 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.

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 sequences 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 (Figs. 2, 3 and 5). Chromatographically inactive DNA is probably the cause of the decreased yield found in Table 1 and Fig. 7. Thus, while DNA complexity can aid in purification, it has the harmful consequence of sometimes increasing peak width and decreasing column capacity and resolution. This was shown most dramatically by investigating two proteins which bound to the same column (Fig. 5). What happened was that on simple DNA columns these two proteins (C/EBP and lac repressor) essentially 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 in the normal course of purification: DNA columns bind multiple proteins that are resolved when their retention times or peak widths are altered favorably.

The origin of concatemer DNA-affinity chromatography is the purification of Sp1, where concatemers mimic the repetitive GCbox of the early promoter of SV40 [4]. However, Sp1 also binds to discrete elements in other promoters and thus tandem copies must not be a prerequisite of binding. 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 [15]. 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* homologue of Sp1 or a similar protein that binds specific Sp1 antiserum (unpublished data, W.L. Taylor and W.T. Penberthy). Since the identity of B2 was not 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 [16–19]. Such discrete element oligonucleotides are also effectively used in most electrophoretic mobility shift assays [20] (including Sp1 [4]) and in filter binding assays [21]. 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.

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 [1]. This is not only time consuming but also leads to loss of oligonucleotides during each step. Our study shows concatemeric columns do not 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.

Acknowledgements

We thank Ms. F. Darlene Robinson and Pallavi T. Patel for their excellent technical assistance and Ms. Shilpa Oak for helpful discussions and review of the manuscript. We also thank Dr. David Leven for the gift of lacIZ expressing *E. coli* and Drs. Steven McKnight and Jon Shuman for providing plasmids. This work was supported by National Institutes of Health, grant GM43609.

References

[1] J.T. Kadonaga, Methods Enzymol. 208 (1991) 10.
[2] H.W. Jarrett, J. Chromatogr. B 618 (1993) 315.
[3] C. DiRusso, R.P. Rogers, H.W. Jarrett, J. Chromatogr. A 677 (1994) 45.
[4] M.R. Briggs, J.T. Kadonaga, S.P. Bell, R. Tjian, Science 234 (1986) 47.
[5] J.T. Kadonaga, R. Tjian, Proc. Natl. Acad. Sci. USA 83 (1986) 5889.
[6] A.M. Khoury, H.S. Nick, P. Lu, J. Mol. Biol. 219 (1991) 623.
[7] Z. Cao, R.M. Umek, S.L. McKnight, Genes Dev. 5 (1991) 1538.
[8] W.H. Landschulz, P.F. Johnson, S.L. McKnight, Science 240 (1988) 1759.
[9] F.D. Robinson, H. Gadgil, H.W. Jarrett, J. Chromatogr. A 849 (1999) 403.
[10] A.J. Brake, A.V. Fowler, I. Zabin, J. Kania, B. Muller-Hill, Proc. Natl. Acad. Sci. USA 75 (1978) 4824.
[11] B. Muller-Hill, J. Kania, Nature 249 (1974) 561.
[12] H.W. Jarrett, W.L. Taylor, J. Chromatogr. A 803 (1998) 131.
[13] H. Gadgil, H.W. Jarrett, J. Chromatogr. A 848 (1999) 131.
[14] U.K. Laemmli, Nature 227 (1970) 680.
[15] S.L. Pfaff, W.L. Taylor, Dev. Biol. 151 (1992) 306.
[16] P.J. Rosenfeld, T.J. Kelly, J. Biol. Chem. 261 (1986) 1398.
[17] R. Blanks, L.W. McLaughlin, Nucleic Acids Res. 16 (1988) 10283.
[18] A. Hatamochi, P.T. Golumbek, E. Van Schaftingen, B. de Crombrugghe, J. Biol. Chem. 263 (1988) 5940.
[19] C.H. Duncan, S.L. Cavalier, Anal. Biochem. 169 (1988) 104.
[20] M. Fried, D.M. Crothers, Nucleic Acids Res. 9 (1981) 6505.
[21] A.D. Riggs, H. Suzuki, S. Bourgeois, J. Mol. Biol. 48 (1970) 67.