Purification and characterization of a poly(dG).poly(dC)-binding protein from *Parenchinus angulosis*

Danielle Patterton

Thesis Presented for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry

University of Cape Town

September 1992

CONTENTS

ABSTRACT

Purification and characterization of a poly(dG).poly(dC)-binding protein from *Parenchinus angulosis*

Danielle Patterton

Department of Biochemistry
University of Cape Town
September 1992

A poly(dG).poly(dC)-binding protein (suGF1) had previously been identified in sea urchin (*Parenchinus angulosis*) embryonic nuclear extracts (J.P. Hapgood, personal communication). suGF1 may be involved in the regulation of early histone gene expression by interaction with a nucleosome which has been shown to be positioned *in vitro* over the H1-H4 intergenic region of the early histone gene battery of *Psammechinus miliaris* (189, H.-G. Patterton & J.P. Hapgood, unpublished).

In this investigation suGF1 was purified to near homogeneity by DNA-affinity chromatography. The purification procedure involved a cation exchange step, followed by poly(dG).poly(dC)-affinity chromatography, and finally by affinity chromatography utilizing multimerized specific DNA-binding sites of suGF1. The 59,5 kDa purified protein was identified as suGF1 by renaturation of sequence-specific DNA-binding activity from a SDS-PAGE gel slice, by Southwestern blotting and by DNase I footprinting. Ultraviolet crosslinking of the nuclear extract and purified suGF1 revealed the presence of the same specific bands on SDS-PAGE. Optimal suGF1 DNA-binding was shown to occur at relatively high ionic strength (175 mM). suGF1 DNA-binding was sensitive to EDTA, implying a requirement for a divalent cation for DNA-binding.

The suGF1 DNA-binding interaction was investigated by methylation interference, and DNase I and hydroxyl radical footprinting. The footprinting data was analyzed in difference

probability plots, from which a model for the suGF1-DNA complex was inferred. In the model suGF1 approaches the DNA helix mainly from one side, and interacts with guanine residues in the major groove. Contacts are made to one of the DNA sugar phosphate backbones abutting this major groove. The data is consistent with the DNA in the complex being curved, and/or exhibiting a sharp bend at the site of suGF1 contact in the major groove. suGF1 does not seem to bind to DNA as a rotationally symmetrical dimer.

The results of this investigation are discussed in terms of the literature. suGF1 may be related to the chicken erythrocyte-specific factor BGP1, which has been shown to bind to 16 contiguous guanines in the β^A-globin promoter, both in the naked DNA molecule and wrapped around a histone octamer (37, 139).

ACKNOWLEDGMENTS

The following people are gratefully acknowledged for their contribution to this research project :

My supervisor, Janet Hapgood, for all her encouragement and advice, and for proofreading of the manuscript,

Hugh Patterton for constant support and encouragement, for the plasmid pHP2, and for computer programs for the manipulation of densitometric files, the construction of difference probability plots and the generation of the schematic DNA helix used in Figure 7.1,

C. von Holt for support,

B. T. Sewell and A. Roseman for computer software for densitometry,

Madeleine Vermaak for printing of the manuscript,

my family, friends and colleagues for support, discussion and encouragment,

and AECI Ltd. for a Postgraduate Research Fellowship.

LIST OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

A/X fragment	Double-stranded DNA restriction fragment obtained from *Asp* 718/*Xba* I digestion of plasmid pHP2 (see Figure 2.1)
bp	DNA base-pair
BrdU	Bromodeoxyuridine
BSA	Bovine Serum Albumin
Buffer C	Column buffer (0,X buffer C denotes buffer C containing 0,X M KCl
BS-affinity chromatography	Binding Site-affinity chromatography; the matrix consists of multimerized suGF1 binding sites attached to sepharose CL-4B
CRB	Column regeneration buffer
CSB	Column storage buffer
DBD	DNA-binding domain
deg	Degrees Celsius (°C)
DMS	Dimethylsulfate
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
EGTA	Ethyleneglycol-bis-(2-amino-ethyl ether) N,N'-tetra-acetic acid
E/H fragment	Double-stranded DNA restriction fragment obtained from *Eco* RI/*Hind* III digestion of plasmid pHP2 (see Figure 2.1)
EMSA	Electrophoretic mobility shift assay
G	Guanine residue of double-stranded DNA
G-string	Oligo(dG).oligo(dC) region
Hepes	4-(2-Hydroxyethyl)-1-piperazineethane sulphonic acid
kDa	kiloDalton
MW	Molecular weight
NE	Nuclear extract
NP-40	Nonidet-P40
oligo(dG).oligo(dC)	Oligodeoxyguanidylic-oligodeoxycytidylic acid
oligo-NS	Nonspecific oligodeoxyribonucleotide (see Figure 2.1)
oligo-S	Specific oligodeoxyribonucleotide (see Figure 2.1)
oligonucleotide	Oligodeoxyribonucleotide
OP	*o*-Phenanthroline (1,10-Phenanthroline)
PCI	phenol : chloroform : isoamylalcohol (25 : 24 : 1)
pdAdT	Polydeoxyadenylic-deoxythymidylic acid
pdIdC	Polydeoxyinosinic-deoxycytidylic acid
PMSF	Phenylmethylsulphonylfluoride
poly(dG).poly(dC)	Polydeoxyguanidylic-polydeoxycytidilic acid
PP	*p*-Phenanthroline (4,7-Phenanthroline)
RNA	Ribonucleic acid
RNaseA	Ribonuclease A
SDS	Sodium dodecyl sulfate
SDS PAGE	SDS polyacrylamide gel electrophoresis
suGF1	sea urchin G-binding Factor 1

TBE	Tris-borate-EDTA (see text for concentrations
TCA	Trichloroacetic acid
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
U	Units of suGF1 activity (see text)
UV	Ultraviolet
WGA	Wheat Germ Agglutinin

CHAPTER 1

INTRODUCTION

1.1 Aim of this investigation

The aim of this investigation was two-fold. Firstly, to purify a poly(dG).poly(dC)[1]-binding protein (suGF1) from sea urchin (*Parenchinus angulosis*), and secondly, to characterize the protein DNA-binding interaction of purified suGF1.

Chapter 3 serves as a brief introduction to the purification of DNA-binding proteins, and contains a discussion of specific methods used in setting up a strategy for the purification of suGF1. The purification and identification of the purified protein as suGF1 is dealt with in chapter 4. The dependence of suGF1 DNA-binding on mono- and divalent cations is presented in chapter 5. Chapter 6 contains the results of methylation interference, as well as DNase I and hydroxyl radical footprinting of the suGF1-DNA complex. A model for suGF1 bound to an oligo(dG).oligo(dC)-containing sequence in the H1-H4 intergenic region was deduced from the data presented in chapter 6. This model is presented in chapter 7, where the results of the investigation are finally discussed and related to other work.

In view of the aims presented above, this chapter is divided into three sections. The first section deals with the identification and initial characterization of a poly(dG).poly(dC)-binding protein, suGF1, in sea urchin nuclear extracts. The possible involvement of suGF1 in regulation of histone gene expression in sea urchins is discussed. Since the DNA structure of the cognate sequence was thought to play a role in the binding of suGF1 to DNA, the structural properties of poly(dG).poly(dC) are briefly reviewed in the second section. In view of the low resolution model of the suGF1-DNA complex proposed in chapter 7, the third section of this review deals with the structures of DNA-binding domains of eukaryotic proteins. (Other proteins which may be similar to suGF1 due to their oligo(dG).oligo(dC)-rich recognition sequences, are compared to suGF1 in chapter 7, in terms of the properties of

[1] Poly(dG).poly(dC) refers to the homopolymer, i.e. one strand of the entire DNA molecule containing only guanine residues (G's) and the other only cytosine residues (C's). Oligo(dG).oligo(dC) is used when referring to short regions of contiguous G's which may be flanked by other sequences. suGF1 is referred to as a poly(dG).poly(dC)-binding protein, but can also bind to certain oligo(dG).oligo(dC)-containing sequences (see text).

suGF1 investigated in this study.)

1.2 Identification of a sea urchin poly(dG).poly(dC)-binding protein (suGF1) which may be involved in regulation of histone gene expression

A factor(s) contained in *P. angulosis* embryonic nuclear extracts has been shown to bind sequence-specifically *in vitro* to a site present in the H1-H4 intergenic region of the early histone gene battery of *Psammechinus miliaris* (200), by electrophoretic mobility shift assays (EMSAs; see chapters 2 and 3) and DNase I footprinting (J.P. Hapgood, unpublished). It was noted that the binding site contained 11 contiguous guanines (G's). Furthermore, the oligo(dG).oligo(dC) region (G-string) was flanked on the 5' side by an unusual $(GA)_{16}$ repeat. This $(GA)_{16}G_{11}$ sequence has been shown to adopt an unusual DNA-structure (H-DNA) with enhanced susceptibility to S1-nuclease at low pH in linear and supercoiled plasmids, and at neutral pH in supercoiled plasmids (91, 146, 214, H.-G. Patterton and C. von Holt, submitted).

Poly(dG).poly(dC), but not poly(dA-dG).poly(dC-dT), competed for binding of the factor to a radiolabeled DNA fragment containing the H1-H4 intergenic region (E/H fragment) in an EMSA (J.P. Hapgood, unpublished). Additional evidence for specificity of the factor for the G-string rather than the $(GA)_{16}$-repeat, was provided by the DNase I footprint which extended across the entire G-string, but only over two (GA) repeats flanking the G-string. The binding specificity was further investigated in competitive EMSAs and DNase I footprinting experiments with different DNA restriction fragments. A fragment containing nine contiguous G's from a chicken *c-myc* promoter was found to compete to a lesser extent for binding of the sea urchin factor, than the fragment containing the 11 contiguous G's from the H1-H4 intergenic region (D. Patterton, J.P. Hapgood, unpublished). A fragment containing 18 contiguous G's known to bind a chicken factor BGP1 (139) and located upstream of the chicken β^A-globin gene promoter, competed for binding of the sea urchin factor to approximately the same extent as the H1-H4 intergenic region fragment. DNase I footprints over the oligo(dG).oligo(dC) regions were obtained on the Watson and Crick strands of the *c-myc* and β-globin fragments with *P. angulosis* nuclear extracts (D. Patterton, J.P. Hapgood, unpublished). A variety of fragments not containing oligo(dG).oligo(dC) regions, did not compete for factor binding in an EMSA. Thus the factor was defined as a sequence-specific DNA-binding factor specific for poly(dG).poly(dC) and oligo(dG).oligo(dC)-containing sequences, and was termed suGF1 (sea urchin G-binding Factor 1).

Two bands (factor-DNA complexes B1 and B2) were always present in EMSAs with *P. angulosis* embryonic nuclear extracts and the radiolabeled specific DNA fragment. The two complexes showed the same sequence-specificity and relative affinity for DNA competitors in

EMSAs (J.P. Hapgood, D.Patterton, unpublished). Methylation of the same seven G's located at the 5' end of the 11 bp G-string interfered with factor-DNA binding in complexes B1 and B2 in nuclear extracts (J.P. Hapgood, unpublished). This result strongly suggested that the same factor contained in the nuclear extracts, was involved in the two complexes. The two bands were always present in the same relative ratio in nuclear extracts, and no evidence of proteolysis could be found (J.P. Hapgood, unpublished). The difference in electrophoretic mobility of B1 and B2 did not seem to result from differential phosphorylation of suGF1, as shown by treatment with Calf Intestinal Phosphatase (J.P. Hapgood, unpublished). The factors(s) (suGF1) involved in complexes B1 and B2 were shown to be a protein (or proteins) by virtue of their sensitivity to proteases (J.P. Hapgood, unpublished). No evidence was found for RNA in the complexes B1 and B2, as shown by the insensitivity of the amount or electrophoretic mobility of complexes on EMSA after treatment of nuclear extract with RNases (J.P. Hapgood, unpublished). Ultraviolet protein-DNA crosslinking revealed the presence of two species on SDS-PAGE (apparent molecular weight approximately 70 and 95 kDa), binding sequence-specifically to the region containing 11 contiguous G's (J.P. Hapgood, unpublished). These two crosslinked species were present in both complexes B1 and B2.

Functional significance of binding of suGF1 to the H1-H4 intergenic region of the early histone gene battery of *P. miliaris* has not been determined. Sequencing of the early histone gene battery of *P. angulosis* is currently underway (J.S. Rees and C. von Holt, personal communication). Although it is not known if the suGF1 binding site is conserved in the H1-H4 intergenic region of *P. angulosis*, oligo(dG).oligo(dC) regions have been detected in other intergenic regions of this early histone gene battery (J.S. Rees and C. von Holt, personal communication). The complete intergenic sequences of early histone gene batteries of other sea urchin species are not available, but oligo(dG).oligo(dC) sequences are also found in the intergenic regions of the battery of *Lytechinus pictus* (J.P. Hapgood, personal communication). It is possible that suGF1 is involved in regulation of expression of the sea urchin early histone genes.

The sea urchin early histone genes are organized in tandem repeats with each repeat unit consisting of the H4, H2B, H3, H2A and H1 genes separated by intergenic (noncoding) regions (199, see references 92 and 152 for reviews). Early histone genes are activated upon meiotic maturation of the egg (119) and reach maximum levels of transcription at about 64 to 128 cell stage (see references 92 and 152 for reviews). The early histone genes are already repressed, never to be transcribed again, in the hatching blastula embryo, when late histone gene transcripts begin to accumulate rapidly (27, 92, 112, 119, 152, 153).

The chromatin structure of the actively transcribed early histone gene battery differs

drastically from that of the inactive state (4, 22, 68, 211, 235). MNase digestions of nuclei of *Strongylocentrotus purpuratus* revealed the absence of a regular nucleosomal repeat pattern over gene and spacer sequences of the whole battery in the expressed stage (68). The nucleosomal pattern seemed to be partly restored just after switch-off of transcription, and was completely restored in nuclei several cell divisions later. The promoter regions of early histone genes are sensitive to MNase and DNase I when being transcribed (4, 22, 68, 211, 235) and the disappearance of these hypersensitive sites either coincides with the shut-down of transcription (MNase sites) or follows a few cell divisions later (DNase I sites) (22, 68, 211, 235). Such DNase I hypersensitive sites have been mapped precisely in the promoters of all five early histone genes of *S.purpuratus*, and are postulated to result from an alteration in DNA conformation caused by the binding of transcription factors (50,68).

Elucidation of the regulatory mechanisms underlying the coordinate temporal expression pattern of the sea urchin early histone genes has been a major research goal for some time. As regulation seems to be predominantly at the transcriptional level (see for example reference 119), much attention has been given to the identification of possible transcription factors and DNA *cis*-elements necessary for transcription. Sequence analysis has revealed the existence of several homologous sequences 5' and 3' of sea urchin histone genes, some of which are shared with other histone genes investigated, or almost all genes transcribed by RNA polymerase II (92, 152). These sequences include a TATA box, GATCC sequence, CAP box, CCAAT box and a sequence (CAPyNATG) containing the first codon (92, 152). Some homology exists between promoter regions of the same histone gene type for different species, suggesting a possible regulatory role for these regions (see for example reference 130). However, contrary to initial expectations, no obvious *cis*-element governing the coordinated expression of early histone genes has been detected by sequence analysis and/or experimental procedure (226). Surprisingly, it seems that early or late sea urchin genes of the same histone class contain the most striking homologies in their flanking regions (130, 225).

Faithful expression of sea urchin histone genes after microinjection of these genes and their flanking sequences into sea urchin eggs revealed that species and/or stage specific factors are involved in the regulation. The major control elements required for the temporal expression pattern occur within the 5' flanking sequences of the individual early or late genes (38, 50, 228). Organization of early histone genes into a battery was found not to be a prerequisite for correct temporal expression. If alterations in chromatin structure are involved in early histone gene regulation, the effects are likely to be very localized (38, 50). When early and late H2B genes and respective flanking sequences were linked on the same plasmid and microinjected into sea urchin eggs, the temporal expression pattern was similar to that of the endogenous genes, thereby indicating that differential timing of DNA replication during the cell cycle is

probably not necessary for the developmental regulation of the early and late genes (38). It should however be noted that these microinjection experiments should be interpreted with caution. Large amounts of DNA are injected into the oocyte and the measured expression patterns of the injected genes are not identical to the *in vivo* situation.

The general picture for regulation of sea urchin histone gene transcription emerging from *in vitro* protein-DNA binding studies (see for example references 9, 10, 50, 226) linked to functional assays such as microinjection (see for example references 38, 50, 130, 175, 228) and *in vitro* transcription with developmental stage- or tissue-specific nuclear extracts (see for example references 10, 225, 226) , is one of a complex interplay of DNA *cis*-elements and DNA-binding factors (see for example references 50, 130, 225). Differential transcriptional expression could be brought about by a combination of histone-specific- and ubiquitous transcription factors, the activities of which could be regulated in a developmental stage-specific manner (9, 10, 50, 130, 175, 225, 226).

It is envisaged that disruption of positioned nucleosomes over promoters of the early histone gene battery could be accomplished by certain key factors, enabling binding of other factors constituting a functional promoter. Such a mechanism has been investigated for other genes, for example the yeast PHO5 promoter and the MMTV promoter (see reference 207 for a review). A nucleosome has been shown to be rotationally and translationally positioned in the H1-H4 intergenic region of the early histone gene battery of *P. miliaris in vitro* (189, H.-G. Patterton and C. von Holt, submitted). Preliminary investigations indicate that suGF1 seems to be able to recognize and bind to its recognition sequence when contained in this positioned nucleosome, suggesting a regulatory role for suGF1 (H.-G. Patterton, J.P. Hapgood, unpublished). A key factor involved in the removal or destabilization of a nucleosome would be expected to be present in an active form, correlating with transcriptional activity of the regulated early histone gene. suGF1 DNA-binding activity in EMSAs has been detected at different developmental stages of the sea urchin embryo, but has not been quantitated. Comparative quantitation of protein at different developmental stages is however known to be difficult, since accurate results can only be obtained by using cDNA or antibody probes to the protein.

1.3 The structure of poly(dG).poly(dC)

A model for the structure of poly(dG).poly(dC) has been proposed, based on the X-ray crystallographic structure of d(G-G-G-G-C-C-C-C) (154). The overall features of the structure proposed in the model are similar to those of A-type DNA. The minor groove is shallow and wide, whereas a deep, narrow major groove is found. These features seem to be imposed by

the preferred stacking of the five-membered rings on the six-membered neighbouring rings in each GpG step. No (or very little) same-strand overlapping of bases is however seen in the CpC steps of the other strand. In the model for poly(dG).poly(dC), GpG steps exhibit a low roll (mean value of 5°), preferring a parallel arrangement of the bases to optimize stacking. The base-pairs show a large lateral displacement from the helical axis (high slide; mean value of 1,9 Å), and mean twist of 32,1° (11,2 bp per turn). A mean tilt of 12° is proposed. The poly(dG).poly(dC) helix may be underwound in the proposed model, since the helical repeat of poly(dG).poly(dC) measured in solution is 10,7 bp (182).

Oligo(dG).oligo(dC) regions are found in the noncoding regions of many genes transcribed by RNA-polymerase-II (see reference 37 for a list, and chapter 7 for examples of proteins binding to such sequences). These sequences may play a role in gene regulation, possibly *via* families of oligo(dG).oligo(dC)-binding proteins with related functions (37). In addition, oligo(dG).oligo(dC) regions have been shown to adopt unusual structures such as H-DNA at low pH in supercoiled plasmids (120, 122, 123, 124). Kohwi and Kohwi-Shigematsu (121) have investigated the possible functional role in transcriptional regulation of an oligo(dG).oligo(dC) region capable of forming H-DNA. Their results suggest that oligo(dG).oligo(dC) regions of 35 bp and longer can function as negative regulators of transcription *in vivo*, by adopting intramolecular triple helices under localized superhelical stress, and thus preventing the binding of transcription factors (121). Most oligo(dG).oligo(dC) regions known to occur upstream of genes are, however, shorter than 35 bp.

A region containing 16 (or in certain strains 18) contiguous G's in the chicken β^A-globin promoter has been shown to exhibit enhanced susceptibility to S1 nuclease cleavage in supercoiled plasmids and at low pH (133, 164, 202). S1 hypersensitivity has also been mapped to the region of the G-string (approximately 200 bp upstream of the β^A-globin gene) in nuclei when the gene is expressed (133). The enhanced susceptibility to cleavage by S1 nuclease *in vitro* is probably indicative of an unusual DNA structure containing single-stranded regions, such as H-DNA (37, 133, 164, 202). It has also been suggested that the region could be preferentially cut by S1 nuclease even when the base-pairs remain intact, due to the variable swiveling of the cytosine-containing strand around the stacked guanine-containing strand (154).

A chicken erythrocyte-specific factor (BGP1) has been shown to bind to the G-string *in vitro* (60). A faint DNase I footprint was obtained over the G-string *in vivo*, when the β^A-globin gene is expressed (99). The binding of BGP1 correlates with enhanced susceptibility to nucleases and restriction enzymes of the entire region of approximately 200 bp upstream of

the gene (59, 60, 133, 155, 212). A nucleosome has been shown to be positioned over this region *in vitro*, with the 5' border of the nucleosome situated at the G-string (the binding site of BGP1) (113). Deletion of the G-string has no effect on transcription *in vitro* (in the absence of nucleosomes) (61). BGP1 is however proposed to be involved in transcriptional regulation *via* removal or destabilization of a positioned nucleosome *in vivo* (37, 113, 139). This could allow access of other transcription factors to DNA *cis*-elements in the promoter (60). As mentioned in section 1.2, suGF1 may play a similar role in the regulation of transcription of a histone gene through an alteration in chromatin structure (H.-G. Patterton and J.P. Hapgood, unpublished).

1.4 Structural aspects of the binding of regulatory proteins to DNA

1.4.1 Introduction

The mechanism of activation of transcriptional initiation in eukaryotes is currently not understood. It is known that DNA-binding proteins play a central role in the regulation of transcription by interacting with proximal and distal DNA *cis*-elements, and with each other (see references 87, 149, 159, 160 and 195 for reviews). This sometimes involves looping of intervening DNA, and results in the formation of large protein-DNA complexes at the start sites of transcription (see references 64, 149 and 195 for reviews). Certain DNA-binding proteins have been shown to recognize and bind to their cognate sequences when the DNA is wrapped around a nucleosome (see references 126 and 207 for reviews). These factors may be involved in displacement of positioned nucleosomes, enabling binding of other transcription factors to constitute a functional promoter. A large number of different proteins interacting with varying degrees of specificity and affinity with DNA *cis*-elements and each other, contribute to differential regulation of transcription. Changes in the concentration and/or activity of these proteins may be effected by, for example, tissue-, cell-cycle- or developmental stage-specific gene expression or posttranslational modification (see for example references 16, 100 and 156).

The study of the properties, structures and modes of DNA-binding of an ever-increasing number of DNA-binding regulatory proteins has lead to the conclusion that these proteins usually consist of modules or domains with separable functions, such as DNA-binding, transcriptional activation, or dimerization (see references 64 and 104 for reviews). Surprisingly, it seems that most transcription factors characterized up to date belong to one of a few main classes (families) of DNA-binding domains. Within one such family, the general structural features of the DNA-binding domain, for example two α-helices connected by a turn, are conserved.

Hydrogen-bonding to bases in the major or minor groove has emerged as a major factor in sequence-specific protein DNA-binding (see reference 15 for a review). The sequence-dependent DNA structure could however also play a role in determining the binding affinity and specificity (see references 220 and 221 for reviews). The "steric fit" of the protein, or the contribution of direct contacts to the phosphate backbone of the DNA, would for example depend on the local geometry and/or flexibility of the DNA molecule in the region of protein binding. It is important to realize that both the protein and DNA can undergo structural changes upon binding (see 88 and 220 for reviews). The DNA can for example be bent in protein-DNA complexes (41, 114, 220, 221).

Protein DNA-binding domains (DBDs) usually contain extended structures such as α-helices which can contact DNA bases in the major groove. Such contacts may be established by direct- or water-mediated hydrogen bonding or non-polar van der Waals interactions (see references 15 and 87 for reviews). The extended structures are often positioned for interactions with specific bases *via* interactions with the sugar phosphate backbone (87). Dimerization has emerged as an extremely important way of increasing the DNA-contact area (and affinity constant) of relatively simple and small binding domains with recognition sequences which often contain two-fold rotational symmetry (131, 227). Moreover, formation of homo- or heterodimers resulting in changed DNA-binding affinity or specificity has enabled complex regulatory pathways to be set up with limited and simple protein components (see references 87 and 104 for reviews).

In recent years, structural details of several different DBDs, either in isolation or complexed to DNA recognition sequences, have been derived from NMR and X-ray crystallographic studies (see reference 87 for a recent list of published structural data). Although such studies are the only way of obtaining absolute structural data, information on DNA-protein contacts can be obtained by nuclease digestion and chemical interference- and protection experiments. Such investigations can be performed more easily and rapidly in comparison to X-ray crystallography.

In this section I aim to briefly review the main families of eukaryotic DNA-binding domains identified up to date in terms of general structure of the domain itself, as well as structural details of at least one example of a complex with DNA inferred from X-ray crystallography. These results will be compared to those obtained by nuclease and chemical protection and interference experiments.

1.4.2 Zn-binding domains

Three classes of DNA-binding domains requiring Zn(II) as structural element have been identified (see references 87, 14 and 143 for reviews).

1.4.2.1 Class 1 Zn-binding domains (Zn Fingers)

Zn fingers (class 1), which contain modules of approximately 30 amino acid residues, are usually present as tandem repeats. Each repeat (finger) can be defined by the following consensus sequence :

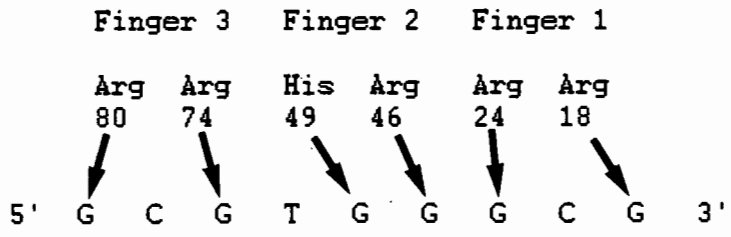
$OXCX_{2-4}CX_3OX_5X_2HX_{3-4}HX_{2-6}$

with O being a hydrophobic amino acid, C and H being cysteine and histidine residues respectively, and X being variable (21, 158).

The structures of single Zn fingers in solution have been solved by NMR (103, 136, 172). In addition, the structure of the three Zn finger peptide from Zif268 (also called Krox-24, EGR-1 or NGFI-A) bound to its DNA recognition sequence has been determined by X-ray crystallography (181). A zinc finger consists of a N-terminal β-hairpin packed against a 12 residue C-terminal α-helix (13, 136). The Zn ion is tetrahedrally coordinated between the two histidine residues on the inward-facing side of the α-helix and the two cysteines flanking the turn in the β-hairpin.

In the X-ray structure (181) the three Zn fingers follow the major groove of the DNA in a semicircular (C-shape) structure. The peptide is oriented in a C- to N-terminal direction ("antiparallel") on the 5' to 3' G-rich DNA strand to which hydrogen bonds are made (primary strand). The α-helix of each finger fits into the major groove, contacting a 3 bp subsite with no gaps in the DNA sequence between these subsites. In each finger, the side chain of the amino acid residue immediately preceding the α-helix (Arg) forms hydrogen bonds to the third base on the primary strand of the 3 bp subsite contacted by the relevant finger (5'--G) (181; see Figure 1.1). The third (His) and sixth (Arg) residues of the α-helix form hydrogen bonds to the second (5'-G-) and third bases (5'G--) of the subsite, respectively (181; see Figure 1.1). In the cocrystal, the α-helix is positioned and held in place by contacts to the sugar phosphate backbone, located mostly along the G-rich DNA strand (181). The involvement of one of the metal-liganding His residues in contacts to the phosphate backbone is thought to be especially important, since this residue is conserved between all Zn finger DBDs and is thus probably

(A)



(B)

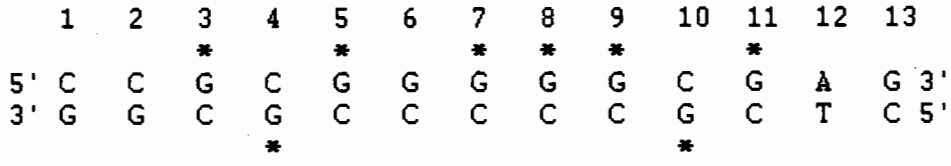


Figure 1.1 Schematic representation of hydrogen bonds formed by Zif268 Zn fingers (181) and results from DMS methylation interference (35)

(A) Numbers underneath three letter amino acid abbreviations indicate protein sequence positions as given in reference 181. Arrows denote hydrogen bonds to bases in the G-rich strand.

(B) A double-stranded DNA binding site for Zif268 is shown. G's interfering with Zif268 binding when methylated are indicated by asterisks (35).

generally used to orient fingers for sequence-specific recognition (87, 181).

The structure of the DNA in the complex with Zif268 was found to be essentially B-type, although the average helical twist of 32,0° (11,3 bp per turn) was somewhat smaller than expected (34,3°, 10,5 bp per turn) possibly due to crystal packing forces (181). The helical twist also varied considerably between the 2nd and 3rd bp (24,3°) and 3rd and 4th bp (39,2°) of the subsite of finger 3, with an analogous situation found at the first subsite (181).

No detailed analyses of chemical and nuclease probing results by difference probability plots have been published for the Zif268-DNA interaction. The region containing the recognition sequence is however protected from DNase I digestion over 18 to 20 bp (35). Limited dimethylsulfate (DMS) methylation of G's in the sequence CCGCGGGGGCGAG was shown to interfere with Zif268 binding at the G's indicated by asterisks in Figure 1.1 (35, 98). The methylation interference result thus agrees very well with the X-ray structure of the complex, as only the G's within the 9 bp contact region interfere with binding *via* the major groove. Noticeably methylation of the G at position 6 in Figure 1.1 does not interfere with binding, indicating that no hydrogen bonding takes place to the N-7 position of this G. This finding is in agreement with the X-ray structure where contacts are made to G's at positions 7 and 8, and a T was used in position 6 in the crystallized complex.

Although no NMR or X-ray structure is available, transcription factor Sp1 most probably interacts with DNA through three Zn fingers (109, 163). The binding site for Sp1 was originally identified as a GC box (GGGGCGGGGC) (57, 110), but Sp1 seems to be able to bind sequence-specifically to many variations of this sequence (20). The consensus sequence for Sp1 binding is $^{G}/_{T}{}^{G}/_{T}$GGCG$^{G}/_{T}{}^{G}/_{A}{}^{G}/_{A}{}^{C}/_{T}$ (20, 110), but all possible binding sites do not seem to be included in this sequence. Sp1 protects approximately 21 bp, including the GC-box recognition sequence, from cleavage by DNase I (56, 57, 74, 138). DMS methylation protection experiments revealed protection of the last five G's of the sequence GGGCGGGG by Sp1, indicative of major groove interactions consistent with the general features of the Zif268 X-ray structure (74, 181). Methylation of the N-7 position of the one G located in the center of the recognition sequence on the opposite strand, did not interfere with Sp1 binding (74). However, it is possible that Sp1 contacts this central base-pair by hydrogen bonding which does not involve the N-7 position of the G, where the methylation takes place (138). In an independent study, three G's located 5' of the central C on the G-rich strand (sequence GGGCGGG) interfered with Sp1 binding (105).

The reduced stringency in the sequence recognized by Sp1 might be explained by the presence of a large number of hydrogen bonds in the Sp1-DNA complex involving the optimal binding

site, with binding still taking place if only subsets of these hydrogen bonds can be made. Specific DNA-binding will then still take place even though, for example, only two of the three Zn fingers are involved in favourable interactions. Letovsky and Dynan (138) determined the equilibrium constant for Sp1 binding to GC-boxes where the central C/G bp had been mutated. Replacement of the central C in the G-string by an A or G would be expected to have a similar effect on the local DNA structure. Purine stacking should alleviate the disruption of the local structure of the G-string by a central pyrimidine (C or T). Mutation to a G/C bp, but not to an A/T or T/A bp, was however found to severely disrupt Sp1 binding. This result could be explained by hydrogen bonding of Sp1 to the 4-amino group of C and the 6-carbonyl group of G at the central position in the wild type GC box (138). Analogous hydrogen bonding contacts could be provided by mutation to an A/T or T/A bp, but not to a G/C (138). This experiment thus seems to support the importance of hydrogen bonding in DNA-binding of Zn fingers, as opposed to recognition of a distinct local DNA structure.

In the Zif268-DNA cocrystal, hydrogen bonds by the amino acids in the second, third and sixth positions of the α-helix and the amino acid immediately preceding the α-helix, are the main determinants of sequence-specific DNA-recognition. This seems to reflect a general pattern whereby all Zn fingers binding to G-rich sequences, or possibly all Zn fingers, recognize their DNA targets (118, 163). Nardelli et.al. (163) have shown that the specificity of individual fingers of Sp1 and Krox-20 can be interchanged between recognizing a GGG and GCG subsite, by mutations in amino acids at positions three and six in the α-helices. This would explain the observed binding specificity of the wild type proteins for GGGGCGGGG and GCGGGGGCG recognition sequences respectively. Such a simple scheme of recognition seems to hold true for the GSG-element (GCGGGGGCG)-binding family of Zn finger domains (163). Although such experiments support the likelihood of a very similar manner of protein-DNA interaction for all Zn finger DBDs, the general applicability of this simple scheme for recognition remains to be established (118).

Because of the highly conserved structure of Zn fingers (13, 136, 172, 181), it could reasonably be expected that Zn fingers interacting with less G-rich sequences would bind in a similar way. Consecutive fingers would be expected to contact 3 bp subsites by major groove hydrogen bonds and positioning contacts to the phosphate backbone, following the major groove without crossing over the minor groove. The DNA-interaction of transcription factor IIIA (TFIIIA), which involves nine tenderly repeated Zn fingers, presents a major challenge to this model.

The TFIIIA-DNA interaction represents the most extensive chemical and nuclease probing analysis of a Zn finger-DNA interaction. Methylation and ethylation interference (198),

DNase I (190), DNase II (190), MNase (62), DMS (62) and hydroxyl radical (26, 230) footprinting, combined with extensive analysis by difference probability plots (36, 62, 190), is consistent with TFIIIA binding predominantly to one face of the helix over a specific region of approximately 50 bp. Interactions seem to be made mainly with the nontemplate strand. Patches of major groove contacts extend throughout the region of interaction, which was noted to contain a 5 bp sequence periodicity, in terms of the distribution of G's (194). Two models based on these results (36, 62) proposed interaction of successive Zn fingers from one side of the helix in consecutive major grooves. These fingers would be connected alternately in the major groove or by crossing of a linker over the minor groove. Crossing of linkers over the minor groove seemed to be easier to reconcile with the results of the nuclease and chemical interference studies (62). Based on structural considerations, Berg (13) however favoured the model where the Zn fingers followed the major groove with a 5 bp repeat. The idealized 5 bp spacing of consecutive Zn fingers of both models (36, 62) seems to be contradicted by binding studies involving TFIIIA with consecutively deleted Zn fingers (230). These experiments point to interaction by clusters of fingers to three main regions within the 50 bp recognition site (230). It is formally possible that the 10 to 11 bp periodicity detected by hydroxyl radical footprinting could be due to changes in the local geometry of the DNA in the complex (36), imposed by the binding of three clusters of Zn fingers, with each finger interacting with a 3 bp subsite. Although interaction with 5 bp subsites seem unlikely in view of the Zif268 X-ray structure where consecutive fingers follow the major groove and interact with 3 bp subsites, it is formally possible and cannot be ruled out in the absence of rigorous structural data.

Huber *et al.* (96) have proposed that widening of the major groove in the free DNA at local junctions between A- and B-type DNA structures correlates with the regions where individual fingers of TFIIIA seem to be binding. Although the above-mentioned simple scheme of Zn finger-DNA recognition based on hydrogen bonding favours a relatively unimportant role for DNA structure in the recognition process, it remains to be determined if DNA structure or flexibility could play an important role, especially in the case of TFIIIA.

In conclusion it can be said that the general features of the Zn finger-DNA X-ray crystallographic structure predict a recognizable pattern of Zn finger-DNA interaction, at least for G-rich sites. Chemical and nuclease interference and protection experiments are characterized by a *bias* towards protection on the G-rich strand as well as strong methylation interference or protection of at least half of the G's in a 3 bp subsite unit, varying with the observed sequence stringency of the consensus recognition site.

1.4.2.2 Class 2 Zn-binding domains (Double loop-Zn-helix)

The 70 to 80 residue class 2 Zn-binding domains, also referred to as double loop-zinc-helix (LZnH) domains (87), are found in receptors for steroids and related hormone-like molecules which bind to similar receptor DNA elements with two-fold symmetry (see references 87 and 205 for reviews). The structures of such DNA-binding domains from the glucocorticoid receptor (GR) and oestrogen receptor (ER) have been determined by NMR, and the structure of the GR bound to DNA has been solved by X-ray crystallography (86, 144, 204).

Each binding domain consists of one polypeptide comprising two LZnH modules (86, 144, 204). Each of these modules contains a N-terminal loop followed by an α-helix with a Zn(II) ion liganded by two Cys residues at the beginning of the loop and two Cys residues at the N-terminus of the α-helix (86, 144, 204). The α-helices of the two LZnH modules of one domain are packed against each other at approximately 90° to form a compact, globular folded structure (the "Double LZnH domain"). The proteins bind as dimers with two-fold rotational symmetry to their two-fold symmetrical recognition sequences (reviewed in 222). This is accomplished by dimerization of two double LZnH domains *via* amino acid residues of the C-terminal loops of each double LZnH domain (86, 144, 204).

In the X-ray structure (144) the N-terminal α-helices of the two double LZnH domains fit into successive major grooves on one side of the DNA double helix, and side chains from the α-helix contact bases in the major groove by hydrogen bonding. These N-terminal helices are positioned by interactions with the C-terminal helices as mentioned above, and by phosphate contacts of the domain to the backbone on either side of the major groove. There are no protein-DNA contacts to the minor groove between the two successive major grooves contacted by the α-helices in the DNA-GR-structure (144), but such contacts have been proposed for the DNA-ER interaction (204), for which a structure is not yet available. The distance between the two successive major groove interaction sites seems to be a major determinant of binding-specificity (see reference 205 for a review). The DNA in the GR-DNA X-ray structure was shown to have a B-type structure with no significant bending or distortions (144).

The X-ray crystallographic structure for the GR-DNA complex is entirely consistent with methylation interference and protection experiments carried out with progesterone, estrogen and glucocorticoid receptors, which reveal interactions in consecutive major grooves on one face of the DNA helix (117, 144, 201, 222).

1.4.2.3 Class 3 Zn-binding domains

Class 3 Zn-binding domains have been found in many transcriptional activators of fungi, and contain 2 Zn(II) ions coordinated by 6 Cys residues (see references 82, 148 and 177 for reviews). These proteins bind as dimers to 17 bp sequences (Upstream Activation Sequences or UAS's) which display pseudo two-fold rotational symmetry. Preliminary NMR studies indicated that each Zn(II) ion was tetrahedrally coordinated between 4 Cys residues (71, 176, 177). Two Cys residues were involved in contacts to both Zn ions (bridging ligands). The structure of the class 3 Zn-binding domain of GAL4 has recently been determined by NMR, and the structure of the complex with DNA has been solved by X-ray crystallography (8, 127, 148).

In the cocrystal each monomer consists of a small, metal-binding module, an extended linker, and an α-helical dimerization motif (148). The metal-binding module contains two short α-helical segments, each followed by an extended strand, and is compactly folded with the coordinated Zn(II) ions located in the center. Each metal-binding module of the dimer contacts three base-pairs (CCG) in the major groove, with positioning contacts being made to the phosphate backbone. The two CCG major groove contact sites of the dimer are separated by approximately one-and-a-half turns, thus placing them on opposite sides of the helix. The coiled-coil dimerization element extends at right angles to the imaginary DNA helical axis and contacts the minor groove at the dyad between the two pseudo-symmetrical binding sites. Protein-DNA contacts (linkers) extend from the base of the coiled-coil dimerization element in opposite directions along the phosphate backbone strands (following the minor groove) towards the metal-binding modules. The major groove opposite the dimerization element (at the dyad) is available for potential additional interactions with other DNA-binding proteins or other domains of GAL4. The DNA is very slightly bent (about 7°) towards the dimerization element, and shows only small deviations from standard B-type DNA.

The GAL4-DNA X-ray crystallographic structure shows remarkable similarity to a model predicted by nuclease and chemical probing experiments for the interaction of two class 3 Zn-binding domains (GAL4 and LAC9) with their recognition sites (82, 33). Methylated G's near the ends of the UAS interfered with GAL4 binding (75). Protection from hydroxyl radical cleavage, and interference of ethylated phosphates with GAL4 binding was evident. These contacts extended approximately 5 bp outward from the center of symmetry of the UAS in the 3' direction on each strand (33).

1.4.3 Basic leucine zipper coiled-coil domain

Proteins in the basic leucine zipper coiled-coil (bCC or bZIP) class were first identified by virtue of a dimerization motif termed a leucine zipper. This motif comprises an α-helical segment (in each protein monomer), with four to five heptad repeats containing predominantly non-polar residues at the first and fourth positions, and almost exclusively leucine residues at the fourth position (see 26, 87, 132, 203, 227 for reviews). Dimerization occurs by the formation of a coiled-coil complex between the two α-helices (168, 170). Upon dimerization, the positively charged basic regions found immediately N-terminal to the leucine zipper region in each monomer are brought in close proximity to each other to form the DNA-binding domain (87).

Although no X-ray crystallographic structure has been published for this type of protein-DNA complex, it is known that the basic regions which appear unstructured in the absence of DNA, become almost entirely α-helical when bound to the 9 to 10 bp two-fold symmetrical DNA recognition sequences (93, 203; see reference 87 for additional references). Since contact is made in the major groove over at least 10 bp (73, 165, 169, 227), the basic regions must wrap around the DNA by at least one half-turn to either side of the center of the binding site. Based on computer modeling, Vinson *et al.* (227) proposed a "Y-shaped Scissors grip" model wherein the α-helical basic regions diverge from the coiled-coil dimerization domain to fit into the major groove. A similar induced-helical fork model was proposed by O'Neil *et al.* (167). A kink in each α-helix extending outwards from the center of the binding site was predicted, to enable contact to bases located at the most distant positions from the dyad. There might also be a possible sharp kink at the junction between the coiled-coil and basic regions.

Chemical protection and interference studies support this Y-shaped Scissors grip model. Methylation of G's in binding sites for the bCC protein C/EBP indicate major groove contacts extending from the dyad over 5 bp or more in each direction (165, 227). Ethylation interference and weak hydroxyl radical protection occur on both strands and show a 5' stagger from the dyad center, consistent with extensive major groove docking on each half-site, leaving the minor groove exposed (165, 227). Fe(EDTA) affinity cleaving experiments were consistent with the location of N-terminal amino acids in the major groove, symmetrically displaced from the dyad and separated by 9 to 10 bp, as predicted by the Scissors grip model (169). Purine methylation interference, missing contact footprinting and ethylation interference of a GCN4-DNA complex indicated multiple sites of DNA-protein contact over nearly 1,5 turns, with the minor groove accessible to hydroxyl radical cleavage at all positions (73).

Although GCN4-induced DNA-bending could not be detected (73), Kerppola *et al.* (114) showed that Fos and Jun (two bCC monomers) induced flexure at an AP-1 DNA-binding site. Moreover, Fos-Jun heterodimers bend DNA toward the major groove, whereas Jun homodimers bend DNA toward the minor groove. This finding suggested considerable differences in the geometry's of the two complexes and seems difficult to reconcile with the predicted general models for the bCC-DNA interaction (114, 167, 227).

1.4.4 Helix-turn-helix domains

Most prokaryotic transcriptional regulatory proteins bind as dimers to DNA *via* helix-turn-helix (HTH) DBDs. The HTH motif has also been identified in eukaryotic homeodomains, which bind as monomers (see references 87, 88 and 174 for reviews). X-ray crystallographic structures of the *engrailed* homeodomain bound to its TAAT-core recognition sequence, and of numerous prokaryotic HTH protein-DNA complexes have been solved (116; see references 87 and 88 for recent lists).

The HTH motif can be defined as a segment of 20 residues or longer, containing two α-helices positioned at an angle of 120° to each other. Additional elements of secondary structure function to stabilize the HTH-motif by forming a hydrophobic core (see reference 88 for a review). The C-terminal (second) helix of the HTH motif is termed the "recognition helix" and invariably contacts DNA bases in the major groove. Dimerization results in a rotationally symmetric molecule which can contact symmetrical half-sites located one helical turn apart on the same face of the DNA. Some of the secondary elements can contribute additional direct base-pair contacts or positioning and stabilizing contacts to the phosphate backbone. In two cases the helices from the subunits of a dimer interdigitate to form a complete DNA-binding structure. The prokaryotic protein-DNA interactions inferred from X-ray crystallography of many different HTH-containing complexes will not be discussed in detail here, as attention is focused on the eukaryotic homeodomains which also contain HTH-motifs for DNA-binding.

In the X-ray crystallographic structure of the *engrailed* homeodomain-DNA complex, the homeodomain consists of three α-helices and an extended N-terminal arm (116). Helix 2 and 3 represent the HTH-motif. Helix 3, which is much longer than the recognition helix of prokaryotic HTH-motifs, lies in the major groove and contacts bases in this groove as well as the sugar phosphate backbone through its hydrophilic face. The hydrophobic face of helix 3 packs against helix 1 and helix 2 which span the major groove wherein helix 3 lies. An unusual interaction is found in the N-terminal arm as it fits into the minor groove to form direct hydrogen bonds to bases. The helix 3 major groove contacts map to the 3' end of the

TAAT subsite which is found in most homeodomain binding sites (see reference 206 for a review) and the minor groove contacts of the N-terminal arm are made near the 5' end of the TAAT subsite. In the overall three-dimensional structure most of the backbone contacts are clustered into two regions flanking the major groove contacts of helix 3. The DNA in the complex is relatively straight and on average B-type, although the major groove where helix 3 interacts is wider than normal. The X-ray structure essentially agrees with a model derived from genetic analysis of HTH-domains (84) and with the NMR structure of the *antennapedia* homeodomain (173, 187).

Ethylation interference experiments with the *antennapedia*-DNA complex show that the protein contacts at the binding site are in agreement with the phosphate contacts seen in the *engrailed*-DNA cocrystal, and are consistent with phosphate contacts on one face of the helix (1, 116). Methylation of the N-7 of guanines (projecting into the major groove) or N-3 of adenines (projecting into the minor groove) were found to interfere with binding of the *antennapedia* protein over a region covering the major and minor groove interactions seen in the cocrystal (1). Supporting results were obtained in chemical interference and protection experiments with the wild type and mutant *ftc* homeodomain-DNA complexes (183).

Because of the relatively small DNA-contact region of the homeodomain, additional sequence-specific contacts are probably needed to obtain levels of specificity required for differential gene expression (116). The effective specific activity could be increased by cooperative binding of homeodomains to neighbouring sites through interactions with other proteins by homeodomain dimerization in certain systems, or by other DNA-binding domains located in the homeodomain-containing protein. Conserved clusters of residues termed POU domains have for example been found in a number of homeodomain proteins. These domains may recognize neighbouring subsites on the DNA, increasing both the affinity and sequence-specificity of the binding (90, 97, 116).

1.4.5 Other domains

Distinct from the helix-turn-helix domain, a helix-loop-helix DNA-binding domain has been defined by virtue of sequence comparisons. This domain seems to involve dimerization of two monomers, each containing a N-terminal basic region, and two regions thought to represent amphipathic α-helices separated by a loop (87). A novel cysteine-rich sequence motif, common to certain proteins interacting with DNA, hints at the existence of uncharacterized DBDs stabilized by Zn in eukaryotes (66). Two classes of prokaryotic proteins have been shown to interact with DNA through β-ribbon recognition elements and are reviewed in reference 87. Nuclease and chemical interference and protection studies of

many DNA-binding proteins cannot be reconciled with the structures of known DBDs. These proteins appear to contain novel DBDs (see references 66, 87 and 166 for examples). Much remains to be discovered in the field of protein-DNA binding.

1.4.6 Conclusion

From the vast number of DNA-binding proteins that have been identified, the structures of the binding domains of only a handful of these proteins have been determined by NMR or X-ray crystallography. Even fewer structures of binding domain-DNA complexes have been solved. Sequence comparisons, mutational studies, and nuclease or chemical protection or interference analyses have however been immensely important in revealing the existence of families or classes of DNA-binding domains to which a large proportion of DNA-binding proteins seem to belong. In the above sections I have briefly described the structures and general models proposed for DNA-interaction of the most important eukaryotic DBDs. Examples of nuclease and/or chemical protection or interference experiments have also been discussed and the consistency of results of such experiments with known structural details of protein-DNA interactions have been noted.

Noticeably, DNA recognition by Zn-binding DBDs classes 2 and 3, bCC domains and homeodomains all seem to involve protein interactions with specific, recognizable DNA consensus sequences. The DNA recognition sequences of Zn fingers seem to be more variable and flexible, although many examples of GGG and GCG recognition subsites occur. Furthermore, the different DBDs are characterized by recognizable, and often unique molecular mechanisms. Examples include minor groove contacts flanked by major groove contacts (homeodomain-DNA interactions), major groove interactions extending over 10 bp (bCC domains), and binding in consecutive major grooves on the same side of the helix, but "snaking" along the minor groove (Zn class 3 domains). When studying a novel protein, the combination of the results of several different chemical and nuclease probes can usually be used to map protein-DNA contacts to the major and/or minor grooves, and the side of approach of the protein to the DNA helix can often be inferred. Models based on such experiments could be compared to the interactions of known DBDs, aiding in characterization of the protein.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

All chemical reagents and solvents were analytical grade unless otherwise stated. Where not stated the source was found not to be important. All H_2O was distilled twice. All solutions, glassware and plastics were sterilized by autoclaving or sterile filtering.

2.2 Plasmid propagation and isolation

The plasmid pHP2 contains part of the H1-H4 intergenic region of the major early histone gene battery of *P. miliaris* (h22) (200) as a 201 bp *Hind* III/*Afl* III insert (H.-G. Patterton & C. von Holt, submitted). The plasmid was propagated in *E.coli* strain HB101, grown in Luria-Bertani growth medium containing 25 µg/ml chloramphenicol, and isolated by the triton lysis method (6). Briefly, 2 litres of bacterial culture was centrifuged for 30 min at 4 200 rpm (JA14 rotor, Beckman) at 4°C, and the pellet resuspended in approximately 15 ml 50 mM Tris.HCl (pH 7,5). The suspension was incubated for 30 min with 1 ml lysozyme (10 mg/ml), 10 min with 1 ml EDTA (0,5 M), 20 min with 100 µl RNaseA (20 mg/ml), 200 µl Triton (10% (v/v)) and centrifuged for 45 min at 20 000 rpm (JA20 rotor). The supernatant was extracted three times with an equal volume of neutralized phenol (redistilled from Holpro), and twice with an equal volume of chloroform. The sample was adjusted to 300 mM sodium acetate. The DNA was precipitated with 2,5 volumes absolute ethanol for 30 min at -70°C, and was recovered by centrifugation at 20 000 rpm (JA20) for 20 min.

The pellet was washed with 75% (v/v) ethanol, dried, and the plasmid banded in a cesium chloride/ethidium bromide gradient (Vti65 rotor, 55 krpm, 16 hours; cesium chloride from Sigma). This step was performed twice, whereafter the supercoiled plasmid was recovered and ethidium bromide removed by repeated extractions with isoamylalcohol (6). The supercoiled plasmid was dialyzed against TE (pH 7,5), precipitated as above, and stored in aliquots in TE (pH 7,5) at -20°C. An aliquot was electrophoresed on a 1% (w/v) agarose gel (6) to ensure that the level of nicking was 5% or less.

2.3 Synthesis and annealing of oligodeoxyribonucleotides

Oligodeoxyribonucleotides (oligonucleotides) were synthesized on a Beckman Systems 1+ DNA Synthesizer and purified by established procedures (6). Concentrations were determined spectrophotometrically. The molar extinction coefficient for each oligonucleotide was estimated from the extinction coefficients of the individual bases (6). Complementary strands were annealed at a molar ratio of 1:1, by incubating at 88°C for 2 min, 65°C for 10 min, 37°C for 10 min, 25°C for 5 min, and finally placing the sample on ice. Oligo-S (specific oligodeoxyribonucleotide) contained a sequence from the H1-H4 intergenic region to which suGF1 bound specifically (see Figure 2.1 and section 1.2). Oligo-NS (nonspecific oligodeoxyribonucleotide) contained a random sequence to which suGF1 did not bind specifically. A 9 bp primer (sequence AATTCTCCC) contained sequences complementary to the Watson strand of oligo-S (see section 2.15).

2.4 Isolation and radioactive labeling of fragments

A 335 bp *Eco* RI/*Hind* III (E/H) and a 216 bp *Asp* 718/*Xba* I (A/X) fragment (See Figure 2.1) were prepared from pHP2 (6). Both fragments contain the binding site of suGF1, including 11 contiguous G's (Watson strand). Fragments obtained by restriction enzyme digestion were resolved on 4% (for E/H digestion) or 10% (for A/X digestion) polyacrylamide gels (acrylamide from Merck, bisacrylamide from Sigma) in TBE (0,089 M Tris-borate, 0,089 M boric acid, 0,002 M EDTA). The relevant bands were excised from the gel after visualization by UV shadowing (6). DNA was eluted from the chopped-up gel slices by shaking overnight at 37°C in approximately three to four volumes of elution buffer (0,5 M ammonium acetate, 1 mM EDTA (pH 8,0)). The solution was filtered through siliconized glass wool. The DNA was precipitated with two volumes absolute ethanol (10 min, -70°C), and was recovered by centrifugation (13 000 rpm, 10 min in a microfuge). The DNA was reprecipitated from 100 µl TE (pH 7,5), washed with 75% (v/v) ethanol, dried and stored in aliquots in TE (pH 7,5) at -20°C. The DNA concentration was determined spectrophotometrically or by ethidium bromide spotting (6).

Restriction fragments and oligonucleotides were 3' end-labeled by a Klenow fill-in reaction (6). The Watson strand of the E/H fragment could be labeled selectively by filling in the *Hind* III site using [α-^{32}P]dCTP (Amersham) as radioactive nucleotide (see Figure 2.1). The Watson strand of the A/X fragment was labeled at the *Xba* I site by using [α-^{32}P]dCTP, while the Crick strand could be labeled at the *Asp* 718 site with [α-^{32}P]dGTP (see Figure 2.1). Specific activity of fragments were typically 25 000 to 65 000 dpm/ng.

Sequence of restriction fragments of plasmid pHP2 :

```
  gaattctc atgtttgaca gcttatcatc gccctgactg agtcgagccc
  cttaagag tacaaactgt cgaatagtag cgggactgac tcagctcggg
    Eco RI
                       -440        -430        -420
aattcgagct cggtacccCA CGTAGAGGAA AAGAGAGTTA TACCACTCCT
ttaagctcga gccatgggGT GCATCTCCTT TTCTCTCAAT ATGGTGAGGA
             Asp 718
-410        -400        -390        -380        -370
GACATGAAAC ACACTCAATT CAACATATTT AGAGGAAGGG AGAGAGAGAG
CTGTACTTTG TGTGAGTTAA GTTGTATAAA TCTCCTTCCC TCTCTCTCTC

-360        -350        -340        -330        -320
AGAGAGAGAG AGAGAGAGAG AGGGGGGGGG GGAGGGAGAA TTGCCCAAAA
TCTCTCTCTC TCTCTCTCTC TCCCCCCCCC CCTCCCTCTT AACGGGTTTT

-310        -300        -290        -280        -270
CACTGTAAAT GTAGCGTTAA TGAACTTTTC ATCTCATCGA CTGCGCGTGT
GTGACATTTA CATCGCAATT ACTTGAAAAG TAGAGTAGCT GACGCGCACA

-260        -250
ATAAGGATGA TTATAAGCTg gggatcctct agagtcgacc tgcaggcatg
TATTCCTACT AATATTCGAc ccctaggaga tctcagctgg acgtccgtac
                              Xba I

caagctgggc tcgacttagt cagggtcacc gataagctt     Watson
gttcgacccg agctgaatca gtcccagtgg ctattcgaa     Crick
                                   Hind III
```

Synthetic Oligonucleotides :

```
                -344                      -319
Oligo-S     gatcAGAGAGGGGGGGGGGGAGGGAGAATT        Watson
                TCTCTCCCCCCCCCCCTCCCTCTTAActag    Crick

Oligo-NS    gatcTTCTGCACTCTCACCGGTACTGGACT        Watson
                AAGACGTGAGAGTGGCCATGACCTGActag    Crick
```

Figure 2.1 Sequences of DNA restriction fragments and of synthetic oligodeoxyribonucleotides

Part of the sequence of plasmid pHP2 (small letters) is shown with the 201 bp insert from the H1-H4 intergenic region of the P. miliaris early histone gene battery (h22) (200) in capital letters. Numbering is with respect to the major cap site of the mRNA of H4 denoted +1 (92). A 335 bp *Eco* RI/*Hind* III fragment (E/H fragment) and a 216 bp *Asp* 718/*Xba* I fragment (A/X fragment) were prepared from pHP2 and radiolabeled on one strand as described in the text. The sequences of two synthetic double-stranded 30 bp oligonucleotides, oligo-S and oligo-NS, are shown. Oligo-S contains 26 bp of the H1-H4 intergenic region of the early histone gene battery of P. miliaris (h22) (capital letters), and 4 base single-stranded 5' overhangs (small letters) to enable multimerization. Upper and lower strands are denoted Watson and Crick respectively. Oligo-NS contains a random DNA sequence.

2.5 Growth of sea urchin embryos

Growth of sea urchin embryos was based on several methods (J.P. Hapgood, personal communication, 9, 47, 78, 161, 229). Sea urchins (*P. angulosis*) were collected manually in rock pools on the West Coast of the Cape Peninsula (Melkbos Beach). Sea urchins were induced to spawn by injection with 2 ml 0,5 M KCl. Eggs were collected, filtered through two layers of cheesecloth and washed three to four times with filtered sea water. All sea water was filtered through Whatman 3MM paper. 50 ml sperm (1/500 dilution) was added per liter of egg suspension (4% (v/v)) which contained 100 mg/l penicillin and 50 mg/l streptomycin in sea water. Where embryos of less than nine hours were grown, 1 mM 3-amino-1,2,4-triazole (ATA) was added prior to fertilization. The nine- or fourteen-hour cultures were shaken at 21°C, 180 rpm for the appropriate period. Four-hour cultures were shaken for 15 min followed by a 30 min shaking period in the presence of pronase (50 µg/ml). Four- and nine-hour embryos were at the 128 cell-stage and late blastula stage respectively (47). Fourteen-hour embryos were at a stage between the late blastula (nine-hour) and early gastrula (eighteen-hour). The cultures were allowed to settle and the Pronase- and ATA-containing sea water replaced with filtered sea water containing antibiotics (as above). The cultures were then shaken for the required period. All subsequent steps were performed at 4°C.

The cultures were allowed to settle, centrifuged (5 000 rpm, 1 min, Beckman JA14 rotor), washed twice with 0,5 M KCl and once with Buffer A containing 1 M hexylene glycol (Hex-A). Buffer A was 15 mM Tris.HCl (pH 8), 65 mM KCl, 15 mM NaCl, 0,15 mM spermine, 0,5 mM spermidine, 0,2 mM EDTA, 0,2 mM EGTA, 10 mM β-mercaptoethanol and 0,1 mM PMSF. Embryos were either frozen at -70°C in 2 to 3 volumes nuclear storage buffer (Buffer A containing 25% (v/v) glycerol), or processed immediately.

2.6 Preparation of nuclei

A method based on several procedures was followed (J.P. Hapgood, personal communication, 9, 49, 162, 178). All steps were carried out at 4°C. Embryos were washed once with Hex-A and resuspended in the same buffer. The suspension was rolled for two hours to allow swelling of the cell membranes, enabling the use of gentle methods for cell disruption. Four-hour embryos were pushed twice through two layers of 50 µm nylon mesh, whereas nine or fourteen-hour embryos were homogenized with 15 strokes of a tight dounce homogenizer. Dissociated embryos with broken cell walls but intact nuclei (monitored by fluorescent microscopy with a Nikon DM400 filter after staining with Hoechst 33258 dye (78)) were centrifuged (5 000 rpm, 1 min) and washed once with Hex-A. The pellet was resuspended in a minimum volume Hex-A and made up to 1,8 M sucrose by adding the required volume of

2,3 M sucrose in Buffer A. The homogeneous suspension was centrifuged at 15 000 rpm (SW28 rotor, Beckman) for 45 min. The pellet (nuclei) was washed once with Hex-A and processed immediately. Purity was assessed by fluorescent microscopy as above. Nuclei preparations were devoid of contaminating cellular membranes.

2.7 Preparation of embryonic sea urchin nuclear extracts

A method based on various procedures was followed (J.P. Hapgood, personal communication, 9, 49, 162, 178). All steps were carried out at 4°C. Nuclei from 1 liter of 4% (v/v) fourteen-hour culture were resuspended in 32 ml lysis buffer (15 mM Tris.HCl (pH 8), 100 mM KCl, 3 mM $MgCl_2$, 0,1 mM EDTA, 1 mM DTT and 0,1 mM PMSF). Ammonium sulfate (4 M) was added dropwise and with immediate mixing to a concentration of 0,4 M over a period of 10 min. The solution was rolled for 30 min and centrifuged at 26 000 rpm (SW28 rotor, Beckman) for 45 min. An additional 0,25 g/ml solid $(NH_4)_2SO_4$ was added to the supernatant. The suspension was rolled for 45 min and centrifuged at 26 000 rpm (SW28 rotor) for 15 min. The pellet was resuspended in 3,25 ml Column Buffer (20 mM Tris.HCl (pH 8), 2 mM $MgCl_2$, 0,2 mM EDTA, 20% (v/v) glycerol, 1 mM DTT and 0,5 mM PMSF) containing 100 mM KCl, for every litre of original 4% (v/v) culture. The extract was dialyzed overnight against approximately 400 volumes of the same buffer, centrifuged for 20 min in a microfuge and the supernatant ("nuclear extract") stored in aliquots at -70°C. The protein concentration was typically between 5 and 15 mg/ml.

2.8 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSAs) were carried out essentially as described (67,72). In the standard EMSA, 1 ng of end-labeled DNA restriction fragment was incubated with variable amounts of protein for 30 min at 4°C in EMSA incubation buffer (16 mM Hepes or Tris.HCl (pH 8), 150 mM to 250 mM KCl as stated, 16% (v/v) glycerol, 1,6 mM $MgCl_2$, 0,8 mM DTT, 0,4 mM PMSF, 1 mM EDTA, 0,5 µg pdIdC (Boehringer) and 1 to 1,5 µg of BSA (Molecular Biology Grade, Boehringer)) in a total volume of 25 µl. The incubation conditions differed from these standard conditions for certain applications, for example where the reaction was scaled up for preparative electrophoretic mobility shifts (see individual experiments for conditions).

Nondenaturing polyacrylamide gels (4%) (acrylamide Merck, bisacrylamide Sigma) (22 cm x 18,5 cm x 0,15 cm) were pre-electrophoresed at 30 mA for 2 hours. The electrophoresis buffer was changed and the EMSA incubation mixtures were loaded directly onto the gels. Electrophoresis was for four to eight hours at 30 mA per gel, or overnight at

20 mA per gel, at 4°C. A buffer system consisting of TGE (50 mM Tris.HCl (pH 8,4), 380 mM glycine (Merck), 2 mM EDTA) was employed. Gels were dried and exposed to preflashed X-ray film with an intensifying screen at -70°C.

suGF1 activity determined during the purification, was a measure of the sequence-specific DNA-binding of suGF1 to the E/H fragment by EMSA. The incubations for these EMSAs were in standard EMSA buffer (250 mM KCl), supplemented with 0,1% (v/v) NP-40. Free DNA and DNA-protein complexes were separated on nondenaturing polyacrylamide gels as above and visualized by "wet" autoradiography (see section 2.22). Gel slices corresponding to appropriate bands or background on the autoradiograph were excised from the gel, placed in liquid scintillation fluid (30% (w/v) Triton X-100, 0,5% (w/v) 2,5-diphenyloxazole, 0,5% (w/v) SDS and 93% (v/v) toluene (technical grade)) and counted for 1 min in a Beckman LS 5 000TD liquid scintillation counter. All incubations were carried out in duplicate and activity in two different preparations were compared by excision from the same gel and at approximately the same percentage shift (between 5 and 12% of the labeled probe shifted into suGF1-probe complexes; see section 3.3). suGF1 activity was measured in units, with one unit defined as the amount of suGF1 activity needed to shift 0,01 ng of the E/H fragment in an EMSA in the presence of nonspecific competitor DNA.

2.9 Protein determination

Protein concentration was determined either by the micro modified Lowry procedure (11) or by the micro Bicinchoninic acid (BCA) method essentially as described by Smith *et al.* (209) and modified by Zhang et. al. (238). The advantage of the latter method is that the detergent NP-40 does not interfere with the determination. BSA was used as a standard.

2.9.1 Micro BCA protein determination

Solution MA consisted of 0,92 M Na_2CO_3, 0,16 M $NaHCO_3$, and 33 mM disodium tartrate. The pH of this solution was not adjusted (238). Solution MB was 4% (w/v) Bicinchoninic acid (4,4'-dicarboxy-2,2'-biquinoline, disodium salt from Sigma). Solution MC was obtained by adding 4 volumes of 4% (w/v) $CuSO_4.5H_2O$ to 100 volumes MB. Micro working reagent (M-WR) consisted of an equal volume of solutions MC and MA. Care was taken to avoid metal ion contamination. Solution M-WR was stored at room temperature in the dark for no longer than 1 week, and was not used if an aliquot changed colour from green to purple when heated at 60°C for approximately 10 min.

Protein solutions were adjusted to 1 ml in Eppendorf vials with water. After addition of 10 µl

1,76% (w/v) sodium deoxycholate and mixing, the solutions were left standing for 15 min. This was followed by the addition of 333 µl 24% (w/v) trichloroacetic acid (TCA), thorough mixing, and centrifugation at 18 000 rpm (Beckman JA20.1 rotor) for 50 min at 4°C. The supernatants were carefully removed, and the pellets redissolved in 200 to 400 µl H_2O. An equal volume of M-WR was added and the vials incubated at 60°C for 1 hour. The solutions were allowed to cool down to room temperature prior to reading the optical density at 562 nm in a Beckman DU 68 spectrophotometer against a reagent blank. The protein concentration was determined from a standard curve (0 to 5 µg protein).

2.9.2 Micro modified Lowry protein determination

Proteins were precipitated as above, but the pellets were dissolved in 1 ml of Lowry reagent C (100 volumes 2% (w/v) Na_2CO_3 in 0,1 N NaOH, 1 volume 1% (w/v) $CuSO_4$, and 1 volume 2% (w/v) disodium-tartrate), followed by addition of 100 µl Folin-Ciocalteu's phenol reagent freshly diluted 1:1 with water. The optical densities were determined at 660 nm after incubation for 75 min at room temperature in the dark. The protein concentration was determined from a standard curve (0 to 10 µg protein).

2.10 Synthesis of affinity matrixes

Two affinity matrixes were synthesized by cyanogen bromide (CNBr) coupling of poly(dG).poly(dC) and concatemers of an oligonucleotide containing the suGF1 binding site respectively to Sepharose CL-4B resin, essentially as described by Kadonaga *et al.* (107, 111).

2.10.1 Synthesis of suGF1 Binding Site-affinity matrix

2.10.1.1 Preparation of ligated oligonucleotides

220 µg of each of the Watson and Crick strands of oligo-S (see Figure 2.1) in TE (pH 7,5) was added to 10µl 10 x T4 Polynucleotide Kinase buffer (500 mM Tris.HCl (pH 7,6), 100 mM $MgCl_2$, 1 mM spermidine, 1 mM EDTA) to give a final volume of 85 µl. The strands were annealed as described in section 2.3. The completeness of the annealing was monitored by electrophoresis of an aliquot on a 2% (w/v) agarose gel (6).

The 5' ends of the double stranded oligonucleotide were phosphorylated by addition of 3 µl 100 mM ATP, 0,5 µl DTT, 5 µl H_2O, 20 µCi [γ-^{32}P]ATP and 10 µl (10 U/µl) T4 Polynucleotide Kinase (Amersham) and incubating at 37°C for 2,5 hours. After addition of 70 µl 7,5 M ammonium acetate and 83 µl H2O, the reaction mixture was heated to 65°C for 15 min to

inactivate the kinase, and cooled to room temperature. The DNA was precipitated after addition of 750 µl ethanol for 10 min at -70°C, and pelleted for 15 min in a microfuge. The pellet was dissolved in 225 µl TE (pH 7,5) and extracted once with phenol-chloroform (1:1), and once with chloroform-isoamylalcohol (24:1). The DNA was reprecipitated by addition of 25 µl sodium acetate and 750 µl ethanol as above, and was dried.

The pellet was dissolved in 65 µl H_2O and 10 µl 10 x Linker-kinase buffer (660 mM Tris.HCl (pH 7,6), 100 mM $MgCl_2$, 150 mM DTT, 10 mM spermidine) by vortexing, followed by the addition of 20 µl 20 mM ATP and 5 µl (30 Weiss units) T4 DNA Ligase (Amersham). The ligation was allowed to proceed at 15°C and monitored by electrophoresis of 0,5 µl aliquots on agarose gels. Overnight incubation yielded an average length of 10-mers (approximately 300 bp). The ligated oligonucleotides were prepared for coupling by extracting once with an equal volume of phenol and once with chloroform-isoamylalcohol (24:1). The DNA was precipitated with 33 µl 10 M ammonium acetate and 133 µl isopropanol, reprecipitated from 225 µl TE (pH 7,5) with 25 µl 3 M sodium acetate and 750 µl ethanol, washed twice with 75% (w/v) ethanol, dried, and finally dissolved in 50 µl distilled H_2O.

2.10.1.2 Coupling of ligated oligonucleotides to resin

Sepharose CL-4B (5 ml, settled bed volume) (Pharmacia) was washed with 500 ml H_2O in a 60 ml sintered glass funnel, transferred to a 25 ml cylinder and adjusted to 10 ml with H_2O. The slurry was transferred to a 150 ml glass beaker in a 15°C waterbath over a magnetic stirrer in a fume cupboard. CNBr (0,56 g; Riedel-de Häen) dissolved in 1 ml N,N-dimethylformamide was added dropwise over 1 min to the stirring slurry. Sodium hydroxide (15 µl, 5 M) was immediately added, followed by another 15 µl addition every 10 seconds for 10 min, to a final volume of 900 µl. Ice-cold H_2O (50 ml) was immediately added and the mixture poured into a 60 ml sintered glass funnel under suction. Care was taken not to suck the resin to a dry cake. The resin was washed with 4 x 75 ml ice-cold H_2O and 2 x 75 ml ice-cold 10 mM potassium phosphate (pH 8,0), and the thick slurry immediately transferred to a siliconized SS34 tube (Sorvall). The DNA (50 µl) was immediately added, and the slurry rolled for 16 hours at room temperature.

The resin was washed with 2 x 75 ml H_2O, 1 x 75 ml 1 M ethanolamine-HCl (pH 8,0), and rolled in this ethanolamine solution for 6 hours at room temperature to inactivate unreacted CNBr-activated Sepharose. The final washes were with 75 ml each of 10 mM potassium phosphate (pH 8,0), 1 M potassium phosphate (pH 8,0), 1 M KCl, H_2O and column storage buffer (10 mM Tris.HCl (pH 7,8), 1 mM EDTA, 0,3 M NaCl, 0,04% (w/v) sodium azide). The resin was stored at 4°C in column storage buffer.

The coupling efficiency was estimated to be approximately 50% by comparing the level of radioactivity in the first few millilitres of the wash after the overnight coupling step with that of the washed resin.

2.10.2 Synthesis of poly(dG).poly(dC)-affinity matrix

A trace of poly(dG).poly(dC) (Boehringer) was labeled with [γ-^{32}P]ATP and T4 Polynucleotide Kinase after removal of 5' phosphates with Calf Intestinal Phosphatase (6), and mixed with unlabeled poly(dG).poly(dC). Approximately 900 µg of this homopolymer preparation was coupled to approximately 9 ml Sepharose CL-4B by the cyanogen bromide method as detailed above. Coupling efficiency was estimated to be approximately 60%. The resin was stored in column storage buffer at 4°C.

2.11 Purification of suGF1

A summary of steps used in the purification of suGF1 is shown in Figure 2.2. Column buffers containing different concentrations of potassium chloride are referred to as "0,X buffer C", where 0,X buffer C is 20 mM Hepes (pH 8,0), 2 mM $MgCl_2$, 0,2 mM EDTA, 20% (v/v) glycerol, 0,5 mM PMSF and 1 mM DTT, containing 0,X M KCl. Hepes was always used for the affinity chromatography steps, but substituted with 20 mM Tris.HCl (pH 8,0) for the P11 phosphocellulose chromatography step, with no detectable difference. All chromatographic steps were performed at 4°C and all fractions were stored at -70°C between manipulations. P11 phosphocellulose- and affinity chromatography were based on methods described in references 44 and 107 respectively.

2.11.1 P11 phosphocellulose chromatography

P11 phosphocellulose (Whatman) which had been swollen and washed according to reference 24 was packed in a column with radius 2,2 cm and bed volume approximately 180 ml (one "column volume"). The flow rate was 1 ml/min at all times. The column was equilibrated overnight in 0,1 buffer C containing no $MgCl_2$. Nuclear extract (40 ml, protein concentration between 5 and 15 mg/ml) from approximately 18 litres of fourteen-hour *P. angulosis* embryos, was loaded onto the column and washed with 2,5 column volumes of 0,1 buffer C lacking $MgCl_2$. Fractions of 15 ml each were collected. Bound protein was eluted stepwise with 3 column volumes 0,3 buffer C lacking $MgCl_2$, 3 column volumes 0,5 buffer C and 1 column volume of 0,8 buffer C. Protein elution was monitored spectrophotometrically at 280 nm against a reagent blank. Aliquots (1 µl) of the fractions were monitored for suGF1 activity by

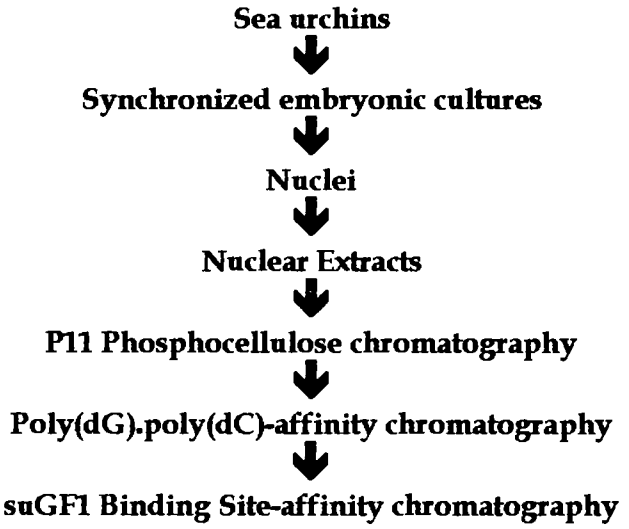


Figure 2.2 Summary of steps used in the purification of suGF1

EMSA (see section 2.8) in 250 mM KCl. The column was regenerated by washing with at least 10 column volumes of 2,5 M KCl until the resin was white. When not in use, the resin was stored in 50 mM Tris.HCl (pH 8), 100 mM KCl and 0,04% (w/v) sodium azide (added freshly).

2.11.2 Poly(dG).poly(dC)-affinity chromatography

A column with radius 4,75 mm and bed volume of approximately 9 ml was used. The flow rate was 0,5 ml/min and 9 ml fractions were collected. All buffers were supplemented with 0,01% (v/v) NP-40. The column was equilibrated in 0,35 buffer C. Fractions from the P11 column exhibiting suGF1 activity in an EMSA were pooled, diluted to 0,35 buffer C by addition of 0,0 buffer C (and NP-40), incubated with 800 µg pdIdC (1 µg/µl) for 10 min and loaded onto the column overnight. The flow-through was collected in a single fraction, whereafter the column was washed with 5 column volumes (5 x 9 ml) 0,35 buffer C. Bound proteins were eluted stepwise with 8 column volumes 0,55 buffer C, 5 column volumes 0,7 buffer C, and 3 column volumes 1,0 buffer C. Aliquots (1 µl) were monitored for suGF1 activity in EMSAs. The column was regenerated at room temperature by washing with 300 ml column regeneration buffer (CRB) (10 mM Tris.HCl (pH 8,0), 1 mM EDTA, 2,5 M NaCl, 1% (v/v) NP-40 (107)) followed by 300 ml column storage buffer (CSB) (As in 2.10.1.2). The matrix was stored at 4°C.

2.11.3 suGF1 Binding Site-affinity chromatography

suGF1 Binding Site-affinity chromatography (BS-affinity chromatography) columns (1 ml) were poured in Econo-Pac 10DG columns (Bio-Rad). The flow-rate was approximately 0,33 ml/min (gravity flow; 107) and fractions of 1 ml were collected. All buffers contained 0,01% and 0,02% (v/v) NP-40 for the first and second pass over the BS-affinity column respectively. The columns were equilibrated in 0,35 buffer C. For the first pass, fractions from the poly(dG).poly(dC)-affinity column containing suGF1 activity were pooled, diluted to 0,35 buffer C with 0,0 buffer C and incubated with 800 µg pdIdC for 10 min. This solution was divided into 8 aliquots and loaded onto separate 1 ml BS-affinity columns in order to decrease the time of exposure of highly purified suGF1 to a temperature of 4°C. The flow-through was collected whereafter the columns were washed with 6 ml (6 column volumes) 0,35 buffer C and the bound proteins eluted with 8 ml 0,5 buffer C, 6 ml 0,65 buffer C and 4 ml 1,0 buffer C. Aliquots (1 µl) were monitored for suGF1 activity in EMSAs (250 mM KCl). The active factions were pooled, diluted to 0,35 buffer C, incubated with 400 µg pdIdC for 10 min, and divided into four fractions, each of which was passed over a BS-affinity column exactly as for the first pass. Once again active fractions were identified from 1 µl aliquots monitored in an EMSA. The columns were regenerated by washing with at least 30 column volumes CRB at

room temperature, and stored at 4°C in CSB (107). Preparations containing purified suGF1 in 0,65 buffer C were concentrated at least 10-fold by ultrafiltration in P-10 centricon devices (Amicon) according to the manufacturers recommendations. suGF1 was stored in aliquots at -70°C, either in 0,65 buffer C or after dilution into 0,35 buffer C.

2.12 SDS polyacrylamide gel electrophoresis and silver staining

Samples were boiled in sodium dodecyl sulfate (SDS) sample application buffer (0,0625 M Tris.HCl (pH 6,8), 2% (w/v) SDS (Sigma), 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0,001% (w/v) bromophenolblue) and loaded directly onto 7% or 10 % (30 : 0,5 acrylamide (Merck) : bisacrylamide (Sigma)) SDS-PAGE gels (22 cm x 18,5 cm x 0,15 cm) as described (6, 128). Alternatively (see individual experiments), samples were TCA precipitated, resuspended in SDS sample application buffer, neutralized with NaOH, boiled, and subsequently loaded onto SDS gels as above. Electrophoresis was at constant current (60 V) until the ion front had run off, followed by fixing in 50% (v/v) methanol, 10% (v/v) acetic acid for at least 45 min. Silver staining was performed in plastic containers and all solutions were made up freshly. One of two different methods was used.

For silver staining method 1, gels were gently shaken in 50% (v/v) methanol, 12% (w/v) TCA, 2% (w/v) $CuCl_2$ for 30 min, and then for 10 min in each of Solution A (10% (v/v) ethanol, 5% (v/v) acetic acid), 0,01% (w/v) $KMnO_4$, Solution A, 10% (v/v) ethanol, water, and finally 0,1% (w/v) $AgNO_3$ (May&Baker) (J.P. Hapgood, personal communication). Gels were then immersed in H_2O for 20 sec followed by 10% (w/v) K_2CO_3 for 1 min, developed in 0,01% (w/v) formaldehyde, 2% (w/v) K_2CO_3 , stopped in Solution A, and photographed.

For silver staining method 2, gels were washed three times for 1 min each in 10% (v/v) ethanol, 5% (v/v) acetic acid and soaked in 32,4 µM DTT for 30 min, followed by a further 30 min in $AgNO_3$ (1 mg/ml) (6). Enough developer (0,5 ml 37% (v/v) formaldehyde per litre, 3% (w/v) Na_2CO_3) was added to just cover the gel which was photographed as bands appeared. High molecular weight standards (Sigma SDS-6H) were Carbonic Anhydrase from bovine erythrocytes (29 kDa), Egg Albumin (45 kDa), Bovine Albumin (66 kDa), Phosphorylase b from rabbit muscle (97,4 kDa), β-Galactosidase from *E.coli* (116 kDa) and Myosin from rabbit muscle (205 kDa).

2.13 Recovery and renaturation of suGF1 from SDS-PAGE gels

Approximately 1 µg suGF1 was TCA precipitated and electrophoresed on a 10% SDS-PAGE gel as described in section 2.12. The major protein band was identified by silver staining of a

parallel lane from the SDS gel, and excised from the unstained section of the gel. Recovery and renaturation were essentially as described by Calzone *et al.* (29). Briefly, the gel slice was cut into small pieces and rolled for 6 hours at room temperature in an Eppendorf vial with 1 ml elution buffer (50 mM Tris.HCl (pH 7,6), 0,1 mM EDTA, 0,1% (w/v) SDS, 5 mM DTT, 150 mM NaCl and 0,1 mM PMSF). The supernatant was recovered by centrifugation, and the gel pieces washed twice with 400 µl H_2O. The combined supernatants were precipitated with 5 volumes of acetone at -20°C overnight, centrifuged, washed once with acetone : H_2O (5 : 1 (v/v)) and dried briefly.

The pellet was resuspended in 100 µl denaturation buffer (20 mM Hepes (pH 8), 0,1 mM EDTA, 1 mM DTT, 6 M guanidinium-HCl, 1 mM $MgCl_2$) and the protein renatured by passing the sample over a 1 ml Bio-Gel P-6 (Bio-Rad) gel filtration column poured in a pasteur pipette and equilibrated in buffer F (10 mM Hepes (pH 8), 0,1% (v/v) NP-40, 1 mM DTT, 100 mM KCl, 10% (v/v) glycerol, 60 µg/ml BSA and 5 µM $ZnCl_2$). Fractions were assayed for suGF1 activity by EMSA. Pooled active fractions were tested for specific binding to the suGF1 recognition site in the H1-H4 intergenic region of the early histone gene battery of *P. miliaris* by competition EMSAs with oligo-S and oligo-NS as specific and nonspecific competitors at 10- and 50-fold molar excess over labeled E/H fragment.

2.14 Southwestern blotting

Southwestern blotting was carried out essentially as described (6, 208, 219). SDS-PAGE was at room temperature as described above, and approximately 750 units of suGF1 activity (as defined in section 2.8) was loaded per lane. The transfer and all other steps (including probing of the blot) were carried out at 4°C. Blotting was overnight (>16 hours, 220 mA) onto nitrocellulose (Hybond-C, Amersham) in pre-chilled transfer buffer (20% (v/v) methanol, 192 mM glycine (Merck), 25 mM Tris.HCl (pH 8), 0,5 mM β-mercaptoethanol). The gel was stained in Coomassie Brilliant Blue (6) to monitor transfer. The blot of the standard lane was cut off, stained in Amido Black stain solution (0,1% (w/v) Amido Black in 40% (v/v) acetic acid, 10% (v/v) methanol) for 15 min, and destained in 40% (v/v) acetic acid, 10% (v/v) methanol. The distances migrated by the standards were measured.

The blot containing the protein samples was blocked for 1 hour in "Blotto" (5% (w/v) non-fat dry milk (Elite), 50 mM Tris.HCl (pH 7,6), 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 0,1% (v/v) NP-40, 0,5 mM PMSF), and rinsed twice for 5 min in Southwestern wash buffer (25 mM Tris.HCl (pH 8), 50 mM NaCl, 5% (v/v) glycerol, 0,1 mM EDTA, 1 mM DTT, 2 mM $MgCl_2$, 0,1% (v/v) NP-40 and 0,2 mg/ml BSA (Fraction V, Boehringer)). The blot was probed for 1 hour in a sealed plastic bag with 5 to 10 ml Southwestern probe solution (25 mM Tris.HCl

(pH 8), 5% (v/v) glycerol, 0,1 mM EDTA, 1 mM DTT, 2 mM $MgCl_2$, 0,1% (v/v) NP-40, 0,2 mg/ml BSA, 175 mM KCl, 43 µg/ml double-stranded sonicated calf thymus DNA, 14,3 µg/ml single-stranded sonicated calf thymus DNA and 10^6 dpm/ml oligo-S or oligo-NS (approximately 25 ng)) per 100 cm^2 membrane. Where required, specific (oligo-S) or nonspecific (oligo-NS) competitor oligonucleotides were present at a 200-fold molar excess over the suGF1 binding site oligonucleotide (oligo-S).

The blot was washed four times for 6 min periods in Southwestern wash buffer, blotted dry on blotting paper, covered with plastic film and exposed to preflashed X-ray film with an intensifying screen at -70°C.

2.15 Protein-DNA Ultraviolet crosslinking

UV crosslinking was carried out as described (6) and modified by J.P.Hapgood (unpublished). Bromodeoxyuridine (BrdU) (Boehringer) and the BrdU-probe was kept in the dark as far as possible. The Watson strand of oligo-S (5 µg; See Figure 2.1) was annealed to a 9 bp complementary primer (3 µg, see section 2.3) in a total volume of 25 µl by heating at 90°C for 5 min and cooling to room temperature over a period of two hours in buffer M (50 mM NaCl, 10 mM Tris.HCl (pH 7,5), 10 mM $MgCl_2$ and 1 mM DTT). The noncoding strand was filled in with the Klenow fragment of DNA polymerase I (16 U), 250 µCi [α-^{32}P]dCTP, and deoxyribonucleotide mix (dGTP, dATP and BrdU, 63 µM each) in buffer M supplemented with 12,5 mM DTT (final volume 80 µl) for 4 hours at room temperature.

In order to isolate the probe where the complementary strand had correctly been extended to the full length of the template, the reaction mixture was passed over a Sephadex G-25 spincolumn (6) in TE (pH 7,5), extracted once with PCI (phenol : chloroform : isoamylalcohol, 25 : 24 : 1) and electrophoresed overnight at 150 V on a 20% polyacrylamide gel in TBE. The "wet" gel was autoradiographed (see section 2.22) and the appropriate band (double-stranded oligo-S ("BrdU probe")) excised. The BrdU probe ($5x10^6$ total dpm) was recovered from the gel slice as detailed in section 2.4. Optimum suGF1 DNA-binding conditions (small volume, high concentration of BrdU probe, and maximum shift with NE or suGF1 still yielding resolvable bands) were determined by trial EMSAs (100 000 dpm/incubation). In the actual crosslinking experiment, the probe was incubated in crosslink buffer (16 mM Hepes (pH 8,0), 1,6 mM $MgCl_2$, 180 mM KCl, 1,2 mM EDTA, 16% (v/v) glycerol, 60 µg/ml Egg Albumin (Sigma) and 100 µg/ml pdIdC) for 30 min at 4°C in the absence or presence of appropriate amounts of protein, and with or without specific (oligo-S) or nonspecific (oligo-NS; see Figure 2.1) competitors. The reaction mixtures (final volume 50 µl) were then irradiated with 305 nm ultraviolet light at a distance of 5 cm from the UV transilluminator lamp (intensity

7 000 μW/cm^2) for 1 hour at 4°C. Control reactions were shielded from the UV source by foil. After crosslinking, 45 μl of each sample was loaded onto a 4% nondenaturing polyacrylamide gel (preparative EMSA) while 5 μl of each sample was loaded onto another EMSA gel for analytical purposes.

Electrophoresis was carried out as described in section 2.8 and DNA-protein complexes detected in the nondenaturing polyacrylamide gel by wet autoradiography. The excised gel slices containing protein-DNA complexes were soaked in SDS sample application buffer (see section 2.12) for 30 min and fitted into the wells of the stacking gel of a 10% SDS-PAGE gel. Electrophoresis was as described in section 2.12. The SDS gel was stained in Coomassie Blue (6), destained until the standard lanes (SDS high molecular weight markers, section 2.12) were visible, dried and autoradiographed (section 2.22).

2.16 Metal chelation and DNA-binding reconstitution

2.16.1 Incubation of suGF1 with chelators without subsequent removal of possible metal-chelator complexes

EMSAs to test the effect on DNA-binding of incubation of suGF1 with 1,10-phenanthroline (OP) were carried out as described in section 2.8, with the following modifications : proteins were preincubated in 1,25 x EMSA incubation buffer (219 mM KCl) with an appropriate amount of OP as indicated in the text in a volume of 20 μl for 5 min at room temperature and 5 min on ice. Since the OP stock solution was prepared in ethanol, control experiments containing ethanol but no OP were performed. The reaction was started by the addition of end-labeled E/H fragment and pdIdC to a final volume of 25 μl. The OP concentrations cited in chapter 5 refer to final OP concentrations in the 25 μl binding incubations.

2.16.2 Incubation of suGF1 with chelators with removal of possible metal-chelator complexes by dialysis against EDTA

An aliquot (440 μl) of dilute purified suGF1 in 0,65 buffer C (see section 2.11) was made up to 1 ml with 0,1 buffer C, supplemented with EDTA, NP-40 and BSA to final concentrations of 10 mM, 0,1% (v/v) and 1 mg/ml respectively, and extensively dialyzed against two changes of 0,1 buffer C containing 50 mM EDTA, no $MgCl_2$ and 0,1% (v/v) NP-40 (suGF1+EDTA). A control sample (suGF1-EDTA) was dialyzed against 0,1 buffer C containing 0,1% (v/v) NP-40. Aliquots of the two dialyzed preparations (suGF1+EDTA and suGF1-EDTA) were stored at -70°C. The (suGF1+EDTA) sample was then dialyzed against 0,1 buffer C containing no $MgCl_2$, 0,2 mM EDTA and 0,1% (v/v) NP-40, while the (suGF1-OP) sample was dialyzed

against the same buffer containing 2 mM $MgCl_2$. Aliquots (5 µl) of the samples were tested for DNA-binding activity in an EMSA (see section 2.8; 175 mM KCl) with supplementation of buffers with $ZnCl_2$ or $MgCl_2$ as described in chapter 5.

2.17 Maxam-Gilbert G-sequencing reaction

Sequencing was performed with reagents from and essentially according to instructions provided with a NEK-010 DNA Sequencing System kit (Maxam-Gilbert Procedure (151), Du Pont NEN products).

Approximately 210 µl G-reaction buffer (50 mM sodium cacodylate (pH 8), 1 mM EDTA) containing 1 to 20 ng end-labeled DNA fragment was incubated with 1 µl Dimethylsulfate (DMS) for 5 min at 20°C. The methylation reaction was quenched by the addition of 50 µl 5 M ammonium acetate and the DNA precipitated with 750 µl ethanol and 1 µg tRNA as carrier. The pellet was resuspended in 50 µl 0,5 M ammonium acetate by vortexing, precipitated with 250 µl ethanol, washed with 75% (v/v) ethanol and dried.

The pellet was redissolved in 100 µl 1 M piperidine, incubated at 90°C for 30 min and lyophilized overnight. DNA was dissolved in 50 µl H_2O and lyophilized for two to three hours. This lyophilization step was repeated twice whereafter the DNA was dissolved in sequencing gel loading solution (see section 2.21). It was found that the radiolabeled DNA standard could be stored for up to two weeks at -20°C without significant loss of resolution on the sequencing gels.

2.18 DNase I footprinting

2.18.1 DNase I footprinting without subsequent separation of protein-bound and free DNA

An appropriate end-labeled DNA fragment (see section 2.4) was incubated with protein (nuclear extract or purified suGF1) in standard EMSA buffer (see section 2.8; final KCl concentration 175 mM) for 30 min at 4°C or 37°C in a total volume of 50 µl. The sample was adjusted to 15 mM $MgCl_2$, 15 mM $CaCl_2$, and 3 µl DNase I (Grade I Boehringer) was added (J.P. Hapgood, personal communication). DNase I was stored as a 1,0 mg/ml stock solution in DNase I dilution buffer (50 mM Tris.HCl (pH 7,5), 0,1 mM $MgCl_2$ and 0,01 mM $CaCl_2$) containing 50% (v/v) glycerol at -20°C, and diluted in the range of 1/10 to 1/80 with DNase I dilution buffer just before use, depending on the amount of DNA, length of DNA fragment and amount of protein present. The reaction was allowed to proceed at 4°C for exactly one

minute, and was then terminated by the addition of 8,4 µl stop solution A (12 µl 20 mg/ml proteinase K, 80 µl 0,5 M EDTA, 32 µl 10% (w/v) SDS, 210 µl H_2O), mixed well, and incubated at 37°C for 45 to 60 min. After addition of 1 µg tRNA and 6 µl 3M sodium acetate (pH 7), the solution was extracted once with 70 µl PCI. Nicked DNA was precipitated with 170 µl ethanol at -70°C for 30 min, pelleted in a microfuge for 30 min at 4°C, washed with 250 µl 75% (v/v) ethanol, dried briefly, dissolved in sequencing gel loading solution and resolved on a sequencing gel (see section 2.21).

2.18.2 DNase I footprinting with subsequent isolation of protein-bound and free DNA by electrophoretic mobility shift assay

A method based on reference 218 and that given in section 2.18.1 was used. Analytical EMSAs were carried out to determine conditions required for shifting approximately 50% of 1,2 ng of the 335 bp E/H fragment. The incubations giving the required shift were scaled up 5 times. 6 ng of end-labeled E/H fragment was incubated with an appropriate volume of protein and 6 µg pdIdC in a final volume of 100 µl EMSA incubation buffer (see section 2.8; 153 mM KCl and 72 µg/ml BSA) for 30 min at 4°C or 37°C. Reaction mixtures were placed on ice for 1 min, adjusted to 15 mM $MgCl_2$ and 15 mM $CaCl_2$, and 6 µl DNase I (diluted as in section 2.18.1) was added. The digestions were stopped after exactly 1 min by the addition of 16,8 µl stop solution B (80 µl 0,5 M EDTA, 254 µl 0,1 buffer C (see section 2.11)) and immediately loaded onto an electrophoresing EMSA gel at 4°C (see section 2.8).

After electrophoresis, the gel was exposed to X-ray film (see section 2.22). Gel slices corresponding to bands representing free DNA and suGF1-DNA complexes were excised. DNA was isolated essentially as described (6). Briefly, the polyacrylamide gel slices were embedded in 1% (w/v) agarose gels (Sigma) (25 cm x 19 cm x 1 cm) containing 0,05% (w/v) SDS (Sigma) in TBE and the DNA electrophoresed onto Na45 DEAE membranes (Schleicher & Schuell) for 45 min at 150 V in TBE. The DNase I digestion products were removed from the membranes by incubation in 400 µl DEAE elution buffer (20 mM Tris.HCl (pH 7,5), 1,5 M NaCl and 1 mM EDTA). The membranes were washed with 100 µl DEAE elution buffer and the combined supernatants centrifuged for 5 min in a microfuge at room temperature to pellet debris. The recovered supernatants were extracted once with PCI. The DNA was precipitated for 30 min at -70°C by addition of 1 ml 96% (v/v) ethanol and 1 µg tRNA, followed by centrifugation for 20 min at 4°C. The pellets were dissolved in 90 µl H_2O by vortexing for 20 sec, centrifuged for 10 seconds and combined with 10 µl 5 M ammonium acetate and 500 µl ethanol in new Eppendorf vials. DNA was precipitated as above, washed once with 75% (v/v) ethanol, dried and either stored as dry pellets at -70°C or dissolved in sequencing gel loading solution and immediately electrophoresed on sequencing gels (see section 2.21).

2.19 Methylation interference

Methylation interference was essentially as described (6). An appropriate amount of end-labeled 335 bp E/H fragment (typically $1{,}5x10^7$ dpm, specific activity approximately 45 000 dpm/ng) was methylated, precipitated, reprecipitated from ammonium acetate, washed and dried exactly as for the preparation of Maxam-Gilbert G-sequencing standards (151; see section 2.17). The pellet was resuspended in 50 µl H_2O by vortexing and was removed into a new Eppendorf vial. An aliquot (1 µl) was counted (typically 250 000 dpm/µl) before and after centrifugation (10 seconds in a microfuge) to ascertain that the DNA was not aggregated.

The binding conditions and volumes of suGF1-preparations needed to shift required amounts of the methylated fragment were determined in analytical EMSAs (see section 2.8), whereafter a preparative EMSA was carried out. Typically 12 ng of methylated end-labeled E/H fragment (approximately 500 000 dpm) was incubated with an appropriate volume of suGF1-preparation and 9 µg pdIdC in a final volume of 120 µl in EMSA incubation buffer (see section 2.8; 175 mM KCl, 60 µg/ml BSA). After electrophoresis, the gel was exposed to X-ray film and gel slices corresponding to bands representing free DNA and DNA-suGF1 complexes were excised (see section 2.22). The end-labeled, methylated DNA was isolated from the polyacrylamide gel slices, removed from the membranes, precipitated from DEAE elution buffer, reprecipitated from 0,5 M ammonium acetate, washed and dried exactly as described in section 2.18.2.

The dried DNA pellets representing populations of end-labeled fragments either enriched (isolated from "free DNA" bands of EMSA) or depleted (isolated from suGF1-DNA complexes obtained in the EMSA) in fragments methylated at G's which interfere with suGF1 binding, were cleaved with piperidine at all positions of methylated G's, lyophilized and electrophoresed on sequencing gels (approximately 25 000 dpm/lane) as described for the G-sequencing standards in section 2.17 (151).

2.20 Hydroxyl radical footprinting

Hydroxyl radical footprinting was based on published methods (51, 223, 224). An iron(II)EDTA stock solution containing 13 mM ferrous ammonium sulfate $[(NH_4)_2Fe(SO_4)_2.6H_2O)]$ (Aldrich), and 26 mM EDTA was stored in aliquots under nitrogen at -70°C in the dark. H_2O_2 (30% (v/v); BDH) was stored at 4°C and a 130 mM stock solution of ascorbic acid (pH 7; Merck) at -20°C. The possible adverse effects of these reagents on

suGF1 DNA-binding were assessed by DNase I footprinting and EMSAs (see chapter 6). Due to a high background of free DNA in the initial trial experiments, subsequent experiments were performed by isolation of free and bound DNA populations by preparative EMSA after hydroxyl radical cleavage.

An aliquot (2 ng) of end-labeled 216 bp A/X fragment (see Figure 2.1; specific activity typically 40 000 dpm/ng) was incubated with an appropriate volume of suGF1 (as determined in an analytical EMSA), and 2 µg pdIdC in a final volume of 76,9 µl for 30 min at 4°C or 37°C in hydroxyl radical incubation buffer (20,8 mM Hepes, 225 mM KCl, 1,3 mM $MgCl_2$, 1,1 mM DTT, 0,52 mM PMSF and 104 µg/ml BSA). BSA was prepared from a solid (Boehringer) to avoid the presence of glycerol which would quench the reaction. The final concentration of glycerol (contributed by the suGF1 preparation) was 0,2% (v/v).

Hydroxyl radical cleavage was carried out by briefly mixing 7,7 µl iron(II)EDTA stock solution with 7,7 µl ascorbate (1/10 dilution of 130 mM stock) and 7,7 µl H_2O_2 on the side of the Eppendorf vial, and initiating the reaction by allowing the droplet to slip into the solution. The final concentrations in the reaction mixture of iron(II), EDTA, H_2O_2 and ascorbate were 1 mM, 2 mM, 2,3% (v/v) and 1 mM respectively. The reaction was allowed to proceed for 2 min at the appropriate temperature and stopped by the addition of 10 µl stop solution (750 µl 80% (v/v) glycerol, 87,5 µl 2 M KCl, 162,5 µl H_2O; final glycerol concentration approximately 5,7% (v/v)).

Reaction mixtures were immediately electrophoresed on EMSA gels at 4°C (see section 2.8). Free A/X fragment and suGF1-DNA complexes were identified by wet autoradiography and excised from the gel (see section 2.22). DNA was recovered from gel slices exactly as described in section 2.18.2, dissolved in sequencing gel loading solution and electrophoresed on sequencing gels (see section 2.21) with appropriate standards.

2.21 Sequencing gels

Samples were dissolved in sequencing gel loading solution (98% (v/v) formamide, 1 mM NaOH, 1 mM EDTA, 0,1% (w/v) Xylene Cyanol, 0,1% (w/v) Bromophenol Blue) (6), heated at 90°C for 3 min, immediately chilled on ice, and loaded onto 6% polyacrylamide (19 : 1, acrylamide : bisacrylamide), 7 M urea (Merck) sequencing gels. Acrylamide was from Merck (Electran) and bisacrylamide from Bio-Rad (Electrophoresis purity reagent). Sequencing gels were electrophoresed at 1,5 kV for 1 to 4 hours at constant temperature (45° to 50°C), fixed in 10% (v/v) methanol, 10% (v/v) acetic acid for approximately 10 min, washed in H_20, dried and exposed to X-ray film (see section 2.22).

2.22 Autoradiography

Dry gels or nitrocellulose blots wrapped in plastic film were exposed to X-ray film (Cronex) (preflashed twice on both sides), with one or sometimes two intensifying screens (Kodak) at -70°C. Wet gels were wrapped in plastic film after removal of one of the glass plates, and two or more corners of the gel were marked with radioactive dye (6). The gel and markers were exposed to preflashed X-ray film (Cronex; with or without an intensifying screen), clamped between two glass plates in the dark at 4°C. Autoradiographs were aligned with radioactive dye markers to identify the positions of radioactive species in the gel.

2.23 Densitometry

A file of data points of a one-dimensional densitometric trace was obtained by scanning autoradiographs on a home-built densitometer (computer software by T.S. Sewell and A. Roseman). The computer program DENS_TOOLS (H.-G. Patterton, unpublished) was used to clip scans to a fixed total amount of data points from an internal reference point (Figure 5.2). Percentage shifted probe on autoradiographs from EMSAs was roughly estimated by cutting out peaks representing suGF1-DNA complexes and other radioactive species. The weight of the peaks representing suGF1-DNA complexes was expressed as a percentage of the weight of the peaks representing the total radioactive species.

2.24 Difference probability plots

Difference probability plots were constructed firstly in order to correct for secondary cleavage in one molecule between the primary site of cleavage and the radiolabeled end, and secondly to enable easy analysis of effects solely due to the presence of protein, above the high background of sequence- or structural specificties of enzymatic or chemical probes for free DNA.

Difference probabilities were calculated and plots constructed with the computer program DENS_TOOLS (H.-G Patterton, Unpublished). The peaks in the densitometric trace of each lane were identified using the known sequence, by comparison with Maxam-Gilbert G-sequencing standards. The area of each peak approximated as a gaussian curve, was determined by numeric integration. The extent of cleavage at each position was calculated by dividing the area of the relevant peak by the summed area of all the peaks above and including the relevant peak, according to Lutter (145). The difference in probability of cleavage due to the presence of protein was calculated by subtracting the natural logarithm of

the extent of cleavage in the absence of protein, from that obtained in the presence of protein, at each position in the sequence. A three-point running average was calculated at every datum point (except the first and last), as the average ln(difference probability) value of the specific datum point and the data points on either side. This three point running average was plotted against the sequence position to yield a difference probability plot (55, 145, 190, 191).

2.25 Fitting of polynomials

The computer program FIT_POLYN (written by H.-G. Patterton) was used to fit first to sixth order polynomials to distances migrated by SDS-PAGE high molecular weight standards, to obtain a standard curve by the least squares method. The polynomial with the smallest root mean square error generating a simple curve through the data points, was used to calculate unknown molecular weights of species with known distances of migration.

CHAPTER 3

SETTING UP A STRATEGY FOR PURIFICATION OF suGF1

3.1 Introduction to the purification of DNA-binding proteins

The key step in the purification of a DNA-binding protein has to be highly selective for the protein, as well as having a yield as high as possible. Calzone *et al.* (30) estimated the minimum prevalency of factors binding to the 5' domain of the CyIIIA cytoskeletal gene in late cleavage-stage sea urchin embryos, as a few hundred to a few thousand molecules per nucleus. Subsequent purification of one such low abundance factor (P3A2) from mid-blastula stage sea urchin embryos, revealed that there were only about 1200 molecules per nucleus, therefore requiring an enrichment factor of several thousand-fold for purification to homogeneity (29). Enormous amounts of biological starting material are needed to obtain sufficient purified protein for applications such as microsequencing. The solution to the required high enrichment depends on finding a suitable affinity chromatography step, and DNA-affinity chromatography is an obvious choice (20, 107, 108, 111).

DNA-affinity chromatography was originally carried out with nonspecific DNA (108), for example calf thymus DNA attached to cellulose (2) or agarose (5) supports. Many DNA-binding proteins bind to their DNA recognition sequences with high specificity. Sequence-specific DNA-affinity chromatography, where DNA containing the specific recognition sequence of the protein being purified is attached to a solid support, therefore became the key purification step in the majority of cases (108).

Sequence-specific DNA-affinity chromatography has recently been reviewed by Kadonaga (108). The DNA employed in this method can be divided into two categories, the first being plasmids or fragments containing the recognition sequence of the protein in addition to other DNA sequences being present. Plasmids containing identical multiple protein binding sites and adsorbed to cellulose (196), and biotinylated DNA fragments attached to a variety of solid supports by biotin-avidin or biotin-streptavidin interactions (69, 108, 134) have been used. The second category makes use of synthetic oligonucleotides containing a high affinity binding site which is specific for the required protein. This leads to increased specificity due to the absence of large proportions of nonspecific DNA (or DNA specific to other factors in the

nuclear extracts) as contained in plasmids and restriction fragments. Teflon-based supports on which oligonucleotides were synthesized have, for example, been used directly as matrix (108). Synthetic oligonucleotides have also been covalently attached to agarose supports as monomers (17, 234), or as concatenated multimers (111). The latter procedure has been employed with success in the purification of the majority of DNA-binding proteins to date (see reference 108 for examples). The DNA homopolymer poly(dG).poly(dC) covalently attached to agarose, has been used successfully in the purification of a chicken factor (BGP1) which binds specifically to this sequence (37).

Purification *via* multimerized oligonucleotides involves incorporation of a selected high affinity binding site for the protein in complementary synthetic oligonucleotides. These oligonucleotides are then annealed, 5' phosphorylated, ligated to an average length of 10-mers, and covalently coupled to cyanogen bromide-activated Sepharose CL-2B or CL-4B to yield the affinity resin (107, 108, 210). A crude or partially purified protein preparation is incubated with DNA to which the desired protein has very low affinity, and subsequently passed over the affinity column at relatively low ionic strength. Proteins having little or no affinity or specificity for the sequence of the multimerized oligonucleotides flow through the column while proteins binding specifically are retained. These proteins can be eluted with a stepwise increase in ionic strength (107, 108). A 500- to 1000-fold enrichment and 30% yield can typically be expected for two sequential affinity chromatography steps (108).

The sequence-specific DNA-affinity resins have a relatively low capacity and might be damaged by nucleases and other contaminants contained in nuclear extracts. The general strategy for purification of a DNA-binding protein thus usually consists of a first step by which nuclear extract can be fractionated in bulk amounts, for example gel filtration- or ion-exchange chromatography (see reference 108 and references therein). This preparation can then usually be passed directly over an affinity column. Sorger *et al.* (210) have suggested a general purification scheme where the first step (heparin-sepharose-affinity chromatography) selects for and fractionates polyanion-binding proteins. The second step (calf thymus cellulose chromatography) retains proteins binding non-specifically but tightly to DNA, and the last (sequence-specific DNA-affinity chromatography) selects for the protein(s) binding with high specificity to the multimerized synthetic oligonucleotide. One or more passes over the same DNA-affinity column, or one pass over the first and another over a different DNA-affinity column (containing for example the same recognition sequence but different flanking sequences) is often required to obtain a sufficient degree of purity for most applications (107, 210). The protein is sometimes purified to homogeneity by isolation from a SDS-PAGE- or nondenaturing polyacrylamide gel (29, 70).

Many transcription factors have been shown to contain terminal N-acetylglucosamine residues (101, 141). Human Sp1 is, for example, glycosylated with O-linked N-acetylglucosamine monosaccharides on multiple serine and/or threonine residues (101). A column containing covalently bound wheat-germ agglutinin (WGA), which binds terminal N-acetylglucosamine with high affinity (16), has been shown to bind more than 95% of the Sp1 in a crude nuclear extract from HeLa cells (102). Sp1 purified by this WGA chromatography step followed by sequence-specific DNA-affinity chromatography was shown to be at least 95% pure with an overall recovery of 80%. A recovery of only about 15% was obtained by the original purification method, consisting of ammonium sulfate precipitation of the crude nuclear extract, S300 gel filtration chromatography and two passes over a sequence-specific DNA-affinity column. The combination of WGA- and DNA-affinity chromatography has also been used very successfully in the purification of the liver-specific transcription factor HNF1 (141) and a glycosylated subset of the multiple polypeptide species of transcription factor CTF/NF-I (102). Purification by an antibody-column can usually only be set up after the protein has been purified or enriched substantially by other methods.

3.2 Starting material : Nuclear extracts

Sea urchins lend themselves especially well to the purification of proteins expressed during early development, as large amounts of synchronized embryos can be grown up with relative ease (e.g. 25 litres of 4% culture; see section 2.5). suGF1 activity (defined in section 2.8) had previously been detected in nuclear extracts from four-, nine- and fourteen-hour embryos by electrophoretic mobility shift assays (EMSAs) (J.P. Hapgood, unpublished). Accurate quantitation of suGF1 abundance at different embryonic stages is very difficult, and would have to await the availability of antibodies or cDNA probes (J.P. Hapgood, D. Patterton, unpublished; see section 1.2). Fourteen-hour embryonic nuclear extracts were chosen as source for the purification of suGF1, since the yield of suGF1 in four-hour extracts was most variable (possibly due to general protein degradation). In addition, nuclear extracts with less cytoplasmic contamination could routinely be obtained from fourteen-hour embryos (225) which could be conveniently grown overnight.

Nuclear extracts were prepared by two ammonium sulfate fractionation steps (see section 2.7). The first step involved lysis of nuclei and extraction of soluble nuclear proteins at 0,4 M ammonium sulfate from chromatin. The second step selected for a subset of these nuclear proteins including suGF1 by differential precipitation (J.P. Hapgood, unpublished). Upon resuspension and overnight dialysis (see section 2.5), a fraction of the protein was found to precipitate out again and was subsequently removed from the soluble proteins (referred to as the "nuclear extract") by centrifugation. It is thus clear that the nuclear extract contained a

relatively small percentage of protein originally present in the nuclei. Protein determinations (see section 2.9.2) of the various pellets and supernatants for one such nuclear extract preparation (results not shown) were in agreement with Calzone *et al.* (29) who reported the final yield of protein in their nuclear extracts prepared from hatching mesenchyme blastula embryos to be about 10% of the starting nuclear protein. This point will be raised in a later discussion of enrichment of suGF1 (see chapter 7).

A few nuclear extract preparations were found to contain nuclease activity detectable in electrophoretic mobility shift- and/or DNase I footprinting assays. These extracts were not passed directly over the DNA-affinity columns (see section 3.8).

3.3 Monitoring purification

suGF1 was originally defined as the protein(s) from *P. angulosis* embryonic nuclear extracts, which bound sequence-specifically to the oligo(dG).oligo(dC)-containing site in the E/H fragment (see section 1.2). Throughout the purification, suGF1 activity is defined as a measure of the amount of sequence-specific DNA-binding of suGF1 to the E/H fragment. This was measured by EMSAs, as described in section 2.8.

EMSAs are particularly suitable for monitoring purification, as they present a rapid, sensitive analytical technique to assay many samples at once, for sequence-specific DNA-binding. In an EMSA a small amount (about 5 fmol) of labeled DNA fragment or oligonucleotide is incubated with a protein preparation in the presence of nonspecific competitor DNA and electrophoresed on an agarose or nondenaturing polyacrylamide gel (see section 2.8; 7, 32, 67, 72, 191). Although protein-DNA complexes generally have life-times shorter than the period of electrophoresis, they are stabilized in the gel due to the so-called cage-effect which can be explained as an increased local concentration of the components of the complex due to slow diffusion in the gel, thus favouring reassociation (32). Factor-DNA complexes can be detected as they migrate shorter distances in the gel than the free DNA probe (7, 32, 67, 72).

In agreement with Sorger *et al.* (210), EMSAs were however found to be non-linear and far from an ideal method for accurate quantitation of the relative activity present at each step. It seemed that the nature and concentration of contaminants relative to suGF1 could affect the amount of sequence-specific binding. The function describing percentage suGF1-specific shift *vs.* concentration protein, was found to particularly deviate from linearity in EMSAs of impure preparations of suGF1. Quantitation of relative activity during the purification could be complicated by several factors : purified proteins are usually more sensitive to chemical and oxidative damage than proteins in crude extracts (191), and often adhere strongly to surfaces

even in the presence of considerable amounts of carrier protein and/or NP-40 (210). It is also possible that optimal conditions for binding, e.g. nonspecific competitor concentration or optimal ionic strength, can change considerably as the purification proceeds (210). The relative activity in different preparations was thus difficult to compare. The relationship between percentage suGF1-specific shift and concentration of protein was found not to deviate significantly from linearity at low percentage shifts (between 5 and 12% of the probe found in suGF1-DNA complexes). In addition to assaying at low percentage binding, EMSA incubations were carried out in 250 mM KCl, in an attempt to decrease competition for labeled fragment by abundant nonspecific or low-specificity DNA-binding proteins. (The optimal KCl concentration as tested in EMSAs was found to be identical for crude and purified suGF1 (see chapter 5)).

The activity of suGF1 measured in EMSAs (see section 2.8), was expressed in binding units. One binding unit (U) was defined as the amount of suGF1 activity needed to shift 0,01 ng of E/H fragment into specific suGF1-DNA complexes. The activity was always measured in an EMSA where the total percentage shift was approximately 10%[2]. The specific activity (U/µg) of suGF1 during the purification, was obtained by dividing the activity per unit volume (U/µl), by the amount of protein per unit volume (i.e. the protein concentration, in µg/µl).

Initially the protein concentration was determined by the micro modified Lowry method (11; see section 2.9.2). The only substance which could interfere substantially with the assay was DTT, which was present at a concentration of 1 mM in all buffers used in the purification. Controls with DTT however showed that the protein concentration was proportional to the absorbance in the range of 1 to 10 µg. A problem was encountered in highly enriched protein fractions which contained low amounts (0,01% to 0,1 % (v/v)) of the nonionic detergent NP-40, added to prevent aggregation and nonspecific binding to the Eppendorf vials. NP-40 was not removed by TCA precipitation and interfered severely with the Lowry method of protein determination by forming a cloudy suspension. This problem could either be solved by using a detergent which would not be precipitated by TCA, for example octyl-[β]-D-glucopyranoside (210), or an assay in which NP-40 would not interfere. The latter solution was employed by using the BCA protein determination method (see section 2.9.1) which, like the Lowry method, is based on monitoring Cu^{1+} produced in the reaction of protein with alkaline Cu^{2+} (biuret reaction). The BCA method relies on a different detection reagent (BCA), which forms a highly-specific, stable purple complex with Cu^{1+} (209).

[2] The activity was thus precisely quantitated in an EMSA using approximately 10 U of suGF1 activity (0,1 ng (10%) of the total of 1 ng E/H fragment shifted).

3.4 Selection of a first step in the purification of suGF1

A variety of columns were initially tested on a small scale to select a first step in the purification scheme. suGF1 eluted in the void volume from a Sephacryl S300 gel filtration column, and at low and high ionic strength respectively from anion- (DEAE sepharose, Pharmacia) and cation (P11 phosphocellulose, Whatman) exchange columns (results not shown). Two polyanion columns were tried, namely heparin-sepharose (prepared according to reference 46) and calf thymus DNA-sepharose (according to reference 210). Interestingly, it was noted that suGF1 in four-hour nuclear extracts eluted at 200 mM KCl from the DEAE column, while suGF1 in nine- and fourteen-hour extracts eluted at 300 mM KCl. In each case the extract had been loaded at 100 mM KCl in column buffer (see section 2.11) and eluted by a 100 mM stepwise increase in KCl. This observation might be explained by developmental stage-specific posttranslational modifications which alter the net charge on the protein, but was not pursued.

The cation exchanger P11 phosphocellulose was chosen as the first step in the purification because of superior yield and enrichment in comparison with the other columns tried. Using this matrix, suGF1 activity could be separated from several contaminating activities which bind to the 335 bp E/H fragment (see section 2.4). These contaminants form DNA-complexes migrating at different and distinct positions on a nondenaturing polyacrylamide gel compared to suGF1-DNA complexes (see section 4.2.1).

3.5 Selection and synthesis of sequence-specific DNA-affinity chromatography matrices

DNA-affinity chromatography has been used successfully in most of the examples of purification of DNA-binding proteins (see section 3.1). Since it was known that suGF1 was a sequence-specific DNA-binding protein (see section 1.2), it was therefore decided to use DNA as the suGF1-specific ligand. The investigation of other affinity matrixes such as WGA-affinity columns in combination with DNA-affinity chromatography, proved to be unnecessary when a sufficient degree of purification was obtained by a combination of cation exchange and DNA-affinity chromatography. In addition, no antibody against suGF1 was available.

In selecting the sequence of the DNA to be used as ligand on the affinity chromatography matrix, it is important to consider the selectivity and affinity of the DNA sequence to the specific protein (107, 108, 210). Competitive DNase I footprinting or EMSAs are generally

used to determine the specificity of the DNA-protein complex. The presence of unlabeled DNA containing the binding site of the protein at a 10-fold or greater molar excess over the same labeled DNA, should lead to the removal of label from sequence-specific complexes in an EMSA. Unaffected (nonspecific) complexes could be ascribed to low-affinity binding of abundant proteins present in excess over the DNA containing the binding site (210).

It had previously been established that suGF1 bound sequence-specifically to a region containing 11 contiguous guanines residues (G's), in the H1-H4 intergenic region of the early histone gene battery of *P.miliaris* (h22; 200) (J.P. Hapgood, unpublished; see section 1.2). This sequence was considered suitable for the affinity chromatography, since competition experiments had established it as a high affinity binding site (see section 1.2). It was decided to incorporate the above-mentioned sequence in a synthetic oligonucleotide. This oligonucleotide could then be multimerized and covalently coupled to a solid support to yield the affinity matrix. It was thought that such a matrix would exhibit a higher specificity for suGF1 compared to fragments or plasmids containing one or multiple copies of the binding site and additional interfering DNA sequences (see section 3.1; 108, 210). Moreover, the use of plasmid-columns have proven inefficient in the purification of Sp1 (111). Sepharose CL-4B was chosen as matrix, since agarose affinity columns were known to adsorb more nonspecific contaminating proteins than sepharose (210).

The length of the synthetic oligonucleotide (oligo-S; see Figure 2.1) was 30 bp, incorporating the region of the observed DNase I footprint (see section 1.2), and unable to circularize in the event of the oligonucleotide being naturally curved in solution (107, 108). Four base GATC single-stranded 5' overhangs were used to enable sticky-end ligation for oligomerization prior to coupling (see section 2.10.1.1). Furthermore, it is thought that a four base overhang is advantageous for coupling to the CNBr-activated sepharose which seems to take place mainly *via* primary amino groups on unpaired bases (107, 233), to form isourea derivatives, N-substituted imidocarbonates or N-substituted carbamates (233).

The Watson and Crick strands of oligo-S (see Figure 2.1) were annealed, 5' phosphorylated, and enzymatically ligated to an average of 10-mers (about 300 bp) (see Materials and Methods for details). These oligomers consisting of tandem arrays of the suGF1 recognition sequence, were then covalently coupled to Sepharose CL-4B which had been activated by CNBr under basic conditions. Unreacted CNBr-activated functional groups were quenched by incubation with ethanolamine, and the resin was washed and stored at 4°C as described in section 2.10.1.2. Coupling efficiency of oligo-S to the matrix was monitored by radiolabeling a trace amount of oligo-S during the phosphorylation step, prior to coupling. The levels of radioactivity in the resin could thus be compared with that in the first few millilitres of the

filtrate after the coupling step. Coupling efficiency was estimated to be approximately 50%, translating to approximately 44 µg DNA per ml resin, and a capacity of about 2,26 nmoles of suGF1 DNA-binding sites per ml ($1{,}36 \times 10^{15}$ 30 bp oligonucleotide-monomers per ml). This result is in agreement with Sorger *et al.* (210) who predict a coupling efficiency of 60% or lower (20 to 50 µg DNA/ml resin), as opposed to greater than 90% reported by Kadonaga (107).

As it became clear that suGF1 bound specifically to oligo(dG).oligo(dC) regions analogous to the chicken erythrocyte factor BGP1 (J.P. Hapgood, D. Patterton, unpublished) (37, 139), it was decided to synthesize a second sequence-specific DNA-affinity resin with poly(dG).poly(dC) covalently attached to sepharose CL-4B. Such an affinity column had been used successfully in conjunction with a nonspecific DNA column to purify BGP1 (37, 139). It could reasonably be expected that several factors in the sea urchin nuclei would bind with high affinity to poly(dG).poly(dC). Furthermore, the suGF1 DNase I footprint extended over a substantial amount of DNA-sequence flanking the 11 contiguous G's. It was therefore thought that a higher degree of purity could be obtained by combining a poly(dG).poly(dC)-affinity chromatography step (selecting only for oligo(dG).oligo(dC)-binding proteins) with the Binding Site-affinity (BS-affinity) chromatography step. The latter step might be expected to fractionate the oligo(dG).oligo(dC)-binding proteins on the basis of differential salt stability when bound to the oligo-S sequence.

Poly(dG).poly(dC) was coupled to sepharose CL-4B in exactly the same way as the multimerized binding site oligonucleotides. The coupling efficiency was estimated to be approximately 60% (53 µg DNA/ml resin). The maximum capacity of the resin was estimated as approximately 2,7 nmoles of suGF1 DNA-binding sites per ml. This was calculated on the basis that the minimum distance between the centers of adjacent suGF1 molecules is estimated at 30 bp from DNase I footprinting results, excluding the effects of possible stable protein-protein multimerization of suGF1 (see section 4.2.7) and ignoring binding in different frames within the G-string.

3.6 Use of nonspecific competitor DNA

Because of the relatively low capacity of sequence-specific affinity columns (210), protein solutions are normally preincubated with large amounts of nonspecific competitors for DNA binding such as pdIdC, pdAdT or sonicated genomic DNA (107, 108, 210). Nonspecific DNA-binding proteins and proteins which preferentially bind to DNA with specific sequences or structures differing greatly from that coupled to the resin, thus do not saturate the DNA binding sites on the affinity column purely by virtue of their capacity to bind DNA. pdIdC had previously been found to be an excellent nonspecific competitor for suGF1 binding. (No

competition by at least 4 µg pdIdC with 1 ng fragment present; D. Patterton, J.P. Hapgood, unpublished.) Sonicated calf thymus DNA competed for binding in EMSAs when present at relatively low amounts (0,5 to 1 µg). Since the amounts of pdIdC required were however very expensive, an alternative way of increasing the capacity of the sequence-specific affinity columns was investigated, namely loading the sample at high ionic strength (210). Because it was known that suGF1 could bind to its recognition sequence in the E/H fragment in EMSAs at relatively high ionic strengths (dealt with in chapters 4 and 5), loading at 350 mM KCl (in buffer C, see section 2.11) was tried, and met with success. No suGF1 activity could be detected in the flow-through, while nonspecific binding should be greatly impeded at this ionic strength (210).

3.7 Requirement for carrier protein but not Zn^{2+} in electrophoretic mobility shift assays of highly enriched suGF1 preparations

In initial trial runs of unfractionated nuclear extract over the affinity columns (see section 3.8), it was found necessary to modify the binding conditions for EMSAs with highly enriched protein preparations (191, 210). The results of these experiments are depicted in Figure 3.1 which shows an autoradiograph of an EMSA polyacrylamide gel (see section 2.8). Note the presence of two species (complexes B1 and B2) migrating slower than the free DNA probe (F). Both species are specific suGF1-DNA complexes (see section 1.2; discussed in chapter 7).

No activity could be detected in the absence of carrier protein (BSA). Figure 3.1 (A) clearly shows that BSA at a concentration of at least 40 µg/ml (lanes 9 and 10) was essential for maintenance of DNA-binding of suGF1 in highly enriched preparations, in an EMSA. No binding was detected at 10 µg/m1 BSA (lanes 5 and 6), and only a smear at 20 µg/ml (lanes 7 and 8), while BSA at 200 µg/ml (lanes 13 and 14) did not compete for suGF1 binding to the probe. BSA presumably prevents suGF1 from adhering to the walls of reaction vessels (191, 210), and is thought to stabilize DNA-binding proteins by lowering the chemical activity of the binding buffer (210).

The presence of endogenously added Zn^{2+} did not affect the DNA-binding ability of highly enriched suGF1 in an EMSA (Figure 3.1 (B)). Indeed, the amounts of suGF1-DNA complexes B1 and B2 are decreased with increasing concentrations of $ZnCl_2$, with no detectable shifts being present at concentrations higher than 50 to 80 µM. This result is in stark contrast to the greatly increased DNA-binding of factor H4TF-1 when buffers are supplemented with low concentrations of Zn^{2+}. Like suGF1, H4TF-1 binds to a very G-rich DNA sequence (42, 44,). The possible effect of Zn^{2+} on the suGF1-DNA interaction is dealt with in chapter 5.

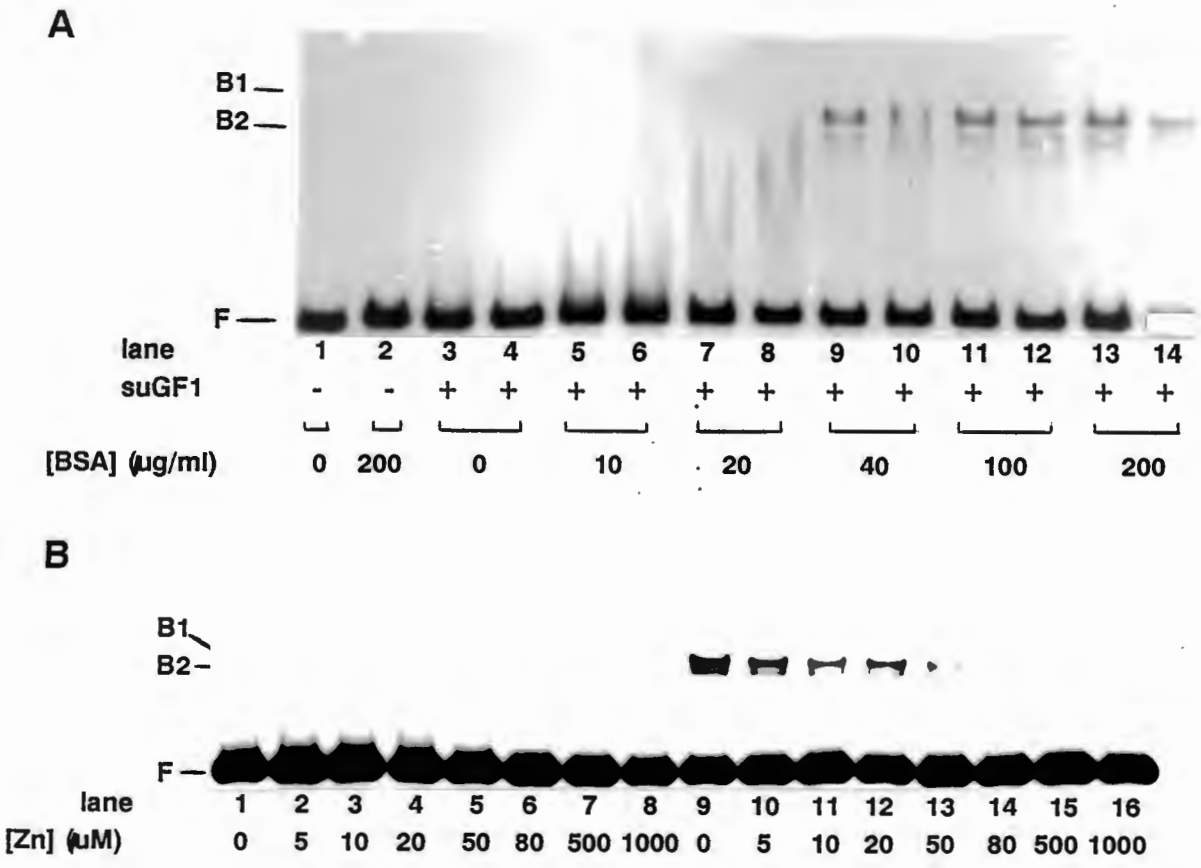


Figure 3.1 Affinity chromatography-enriched suGF1 requires at least 40 µg/ml BSA but no additional Zn^{2+} for DNA-binding

(A) End-labeled 335 bp E/H fragment (1 ng) was incubated with (lanes 3 to 14) or without (lanes 1 and 2) approximately 6 ng protein highly enriched in suGF1 by 1 pass over a BS-affinity column in standard EMSA buffer (see section 2.8; 250 mM KCl), in the presence or absence of BSA at the indicated concentrations (µg/ml) ([BSA]). Complexes were resolved and detected in 4% polyacrylamide gels as described in Materials and Methods. F is free end-labeled E/H fragment. B1 (bound 1) and B2 (bound 2) are specific suGF1-DNA complexes.

(B) End-labeled 335 bp E/H fragment (1 ng) was incubated in standard EMSA incubation buffer (see section 2.8; 250mM KCl) in the presence (lanes 9 to 16) or absence (lanes 1 to 8) of BSA (60 µg/ml) and in the presence of $ZnCl_2$ as indicated ([Zn]). Electrophoresis and autoradiography were as described in Materials and Methods. F, B1 and B2 refer to the same complexes as in (A).

3.8 Testing of sequence-specific DNA-affinity chromatography columns

It was necessary to establish the optimal ionic strength for binding of suGF1 in nuclear extracts and elution of suGF1 during affinity chromatography. suGF1 was found to bind to both affinity matrices at 350 mM KCl (see section 3.6). Preincubation with pdIdC increased the binding capacity of the columns for suGF1. suGF1 activity eluted from the BS- and poly(dG).poly(dC)-affinity columns at approximately 600 and 700 mM KCl in buffer C (see section 2.11) with a 100 mM stepwise increase in KCl, respectively. The result of an EMSA where aliquots of fractions loaded on and eluting from the poly(dG).poly(dC)-affinity column were tested for activity is shown in Figure 3.2. Approximately 270 µl *P. angulosis* fourteen-hour embryonic nuclear extract (340 µg total protein) was loaded onto the 1 ml affinity column which was subsequently washed with 8 ml 0,35 buffer C. (0,35 buffer C refers to 0,35 M KCl in buffer C as explained in section 2.11.) Fractions of 1 ml were collected. Bound proteins were eluted stepwise with 8 ml 0,55 buffer C, 7 ml 0,7 buffer C and 2 ml 0,8 buffer C.

In Figure 3.2 the suGF1-DNA complexes B1 and B2 are clearly visible in lanes 2 and 3 where nuclear extract was incubated with the labeled probe. The flow-through (FT) contained no or very little suGF1 activity, but seemed to contain other DNA-binding activities as evidenced by the smears migrating slower than the labeled probe in lanes 4 to 7. All suGF1 activity was eluted from the column with the 0,7 buffer C step in approximately 2 column volumes (lanes 21 and 22). A third, faster migrating complex (in lanes 21 and 22), which is also present in the nuclear extract incubations, indicated that a contaminant with high affinity for poly(dG).poly(dC) and/or high abundance was present in the nuclear extract and co-eluted with the suGF1 activity. In the initial pilot experiments with nuclear extract, an enrichment factor of at least 100-fold and yield of at least 67% was obtained for each of the affinity columns.

Upon TCA precipitation, SDS-PAGE and silver-staining of these fractions (Figure 3.3; see section 2.12, silver-staining method 1), it was evident that a very large proportion of the nuclear extract proteins did not bind to the poly(dG).poly(dC)-affinity column under the loading conditions, and were found in the flow-through (lanes 3 and 4). Several proteins (lanes 5 and 6) which seemed to correspond to some of the high abundance factors in the nuclear extract (lane 2), were subsequently washed from the column with 0,35 buffer C. The 0,55 buffer C step seemed to remove more of these species, as well as higher molecular weight proteins (lane 8). The proteins in the 0,35 and 0,55 buffer C steps probably represent DNA-binding proteins with some degree of specificity for poly(dG).poly(dC). The 0,7 buffer C step (lane 15) which corresponds to the suGF1 activity as determined in the EMSA (Figure 3.2) contained several proteins. Artifactual bands of apparent molecular weights 54 kDa and a

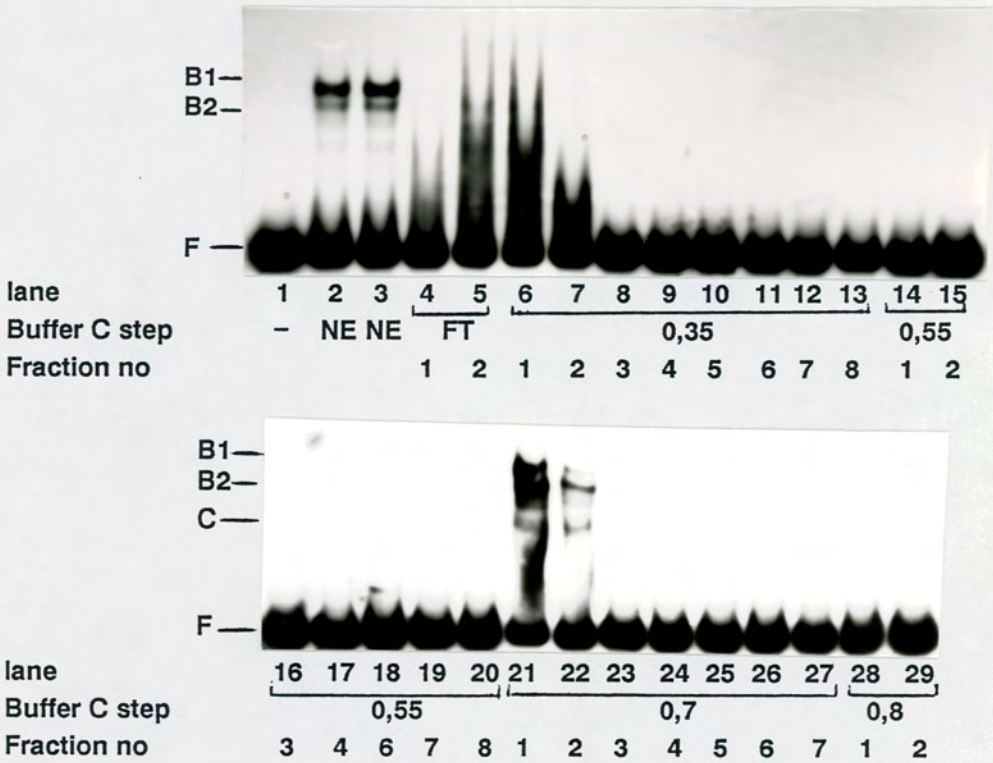


Figure 3.2 suGF1 elutes at 0,7 M KCl from a poly(dG).poly(dC)-affinity column

P. angulosis fourteen-hour embryonic nuclear extract (approximately 340 µg protein) was loaded onto a 1 ml poly(dG).poly(dC)-affinity column pre-equilibrated with 0,35 buffer C (see section 2.11.2). The flow-through was collected, followed by stepwise elution of bound proteins with 8 ml 0,35 buffer C, 8 ml 0,55 buffer C, 7 ml 0,7 buffer C and 2 ml 0,8 buffer C. Fractions of 1 ml were collected. 6 µl buffer C (lane 1), an aliquot of the nuclear extract (0,26 µl NE; lanes 2 and 3) or 6 µl of the collected fractions (lanes 4 to 29) were incubated with 1 ng end-labeled E/H fragment in a standard EMSA (see section 2.8; 250 mM KCl). Electrophoresis and autoradiography were as described in Materials and Methods. F is free end-labeled E/H fragment. B1 and B2 are specific suGF1-DNA complexes. C is an additional factor-DNA complex.

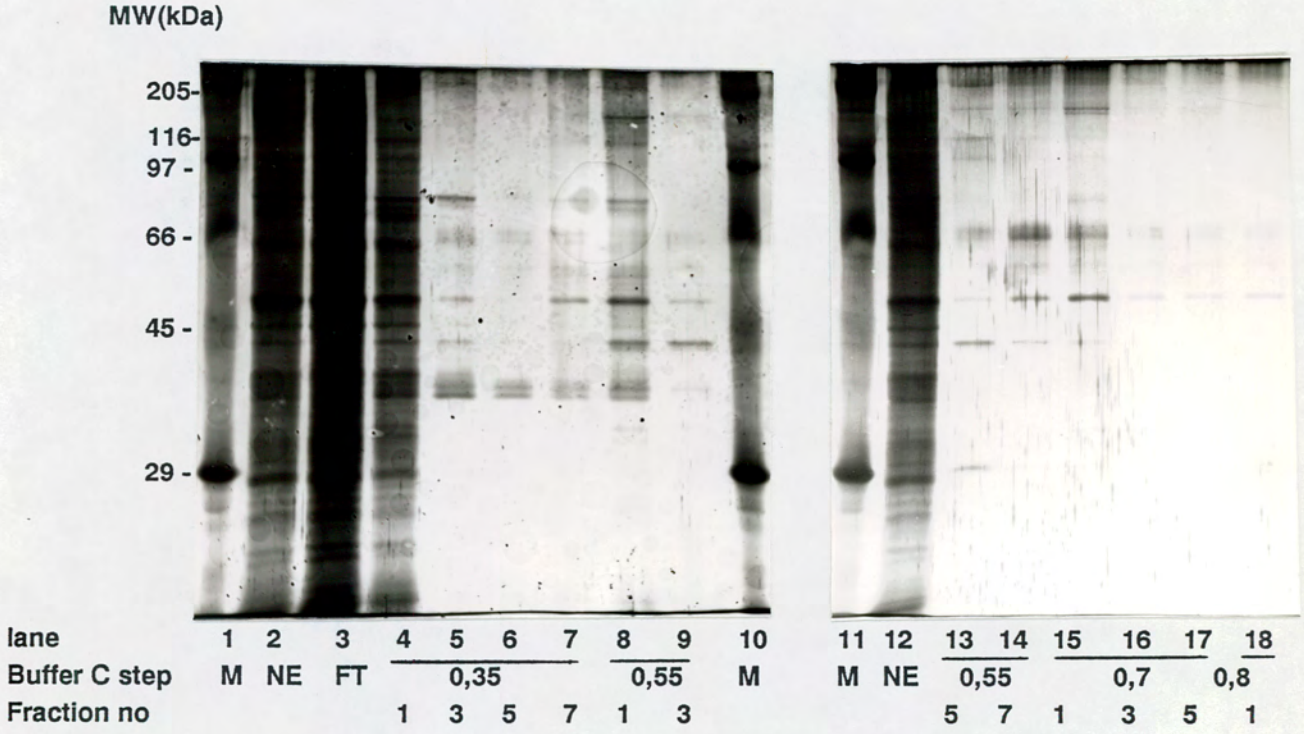


Figure 3.3 A large proportion of the nuclear extract proteins do not bind to a poly(dG).poly(dC)-affinity column

An aliquot (3,5 µl) of nuclear extract (NE; lanes 2 and 12), 800 µl flow-through (FT; lane 3) or a combination of 400 µl of each two consecutive fractions eluting from a 1 ml poly(dG).poly(dC)-affinity column (lanes 4 to 9 and 13 to 18; fraction numbers indicate the first of the two consecutive fractions of which 400 µl of each was combined) were TCA precipitated (see section 2.9.1), dissolved in SDS loading dye, neutralized with NaOH and electrophoresed on a 10% SDS gel which was silver-stained (see section 2.12; silver-staining method 1). Sizes of molecular weight markers are indicated in the left margin (MW (kDa)). The two panels represent two different gels which were prepared, electrophoresed, stained, photographed and printed in an identical manner.

doublet at 68 kDa were present. These artifactual bands are often seen upon silver-staining of SDS gels, and are dependent on the presence of β-mercaptoethanol or DTT (217). All subsequent silver-stained gels were stained by a different method (see section 2.12; method 2) which was found to generally be less sensitive to artifactual bands, more sensitive to proteins and less labour-intensive. SDS gels of fractions eluting from the BS-affinity column in trial runs were similar to those of the poly(dG).poly(dC) column (results not shown).

In summary, the collection of bands obtained in silver-stained gels of the active fractions after one pass of nuclear extract over one of the affinity columns, emphasized the need for a preliminary purification step and/or more than one pass over one or both of the affinity columns. The P11 phosphocellulose step had the advantage of removing a large proportion of contaminants binding to the E/H fragment in EMSAs, and was furthermore needed to remove nucleases from the extracts and increase the effective capacity of the affinity columns. It was therefore decided that the purification strategy following the P11 cation exchange step was to be one pass over the poly(dG).poly(dC)-affinity column followed by one or more passes over the BS-affinity column.

CHAPTER 4

PURIFICATION AND IDENTIFICATION OF suGF1

4.1 Introduction

In this chapter purification of suGF1 by P11 phosphocellulose- and DNA-affinity chromatography is presented, and an enrichment table is drawn up. The purity of the final preparation is assessed and suGF1 is sized on silver-stained SDS-PAGE gels. The major band identified as suGF1 on these gels is proven to indeed be the DNA-binding activity originally defined as suGF1 in nuclear extracts (see section 1.2). This is accomplished by Southwestern blotting with and without specific and nonspecific competitors, and recovery and renaturation of specific suGF1 DNA-binding activity from a SDS-PAGE gel slice containing the major band. DNase I footprinting is carried out with the purified preparation to further verify the identity of the isolated protein as suGF1. Ultraviolet protein-DNA crosslinking with crude (nuclear extract) and purified suGF1 is compared.

4.2 Purification of suGF1

4.2.1 P11 phosphocellulose chromatography

Nuclear extract from approximately 18 litres of fourteen-hour embryonic culture (see section 2.5) was applied to a 180 ml P11 phosphocellulose column which was subsequently washed with 0,1 buffer C (see section 2.11.1). Bound proteins were eluted with a stepwise increase in ionic strength (buffer C containing 0,3, 0,5 and 0,8 M KCl respectively; see section 2.11.1).

The elution of protein was monitored by measuring the absorbance of the fractions at 280 nm against a reagent blank (6). This protein elution profile is shown in Figure 4.1 from which it can be seen that a large proportion of the protein did not bind to the column at 100 mM KCl (fractions 5 to 20). The nuclear extract was fractionated efficiently, since a substantial amount of protein eluted in just over one column volume (represented by 12 fractions) with each

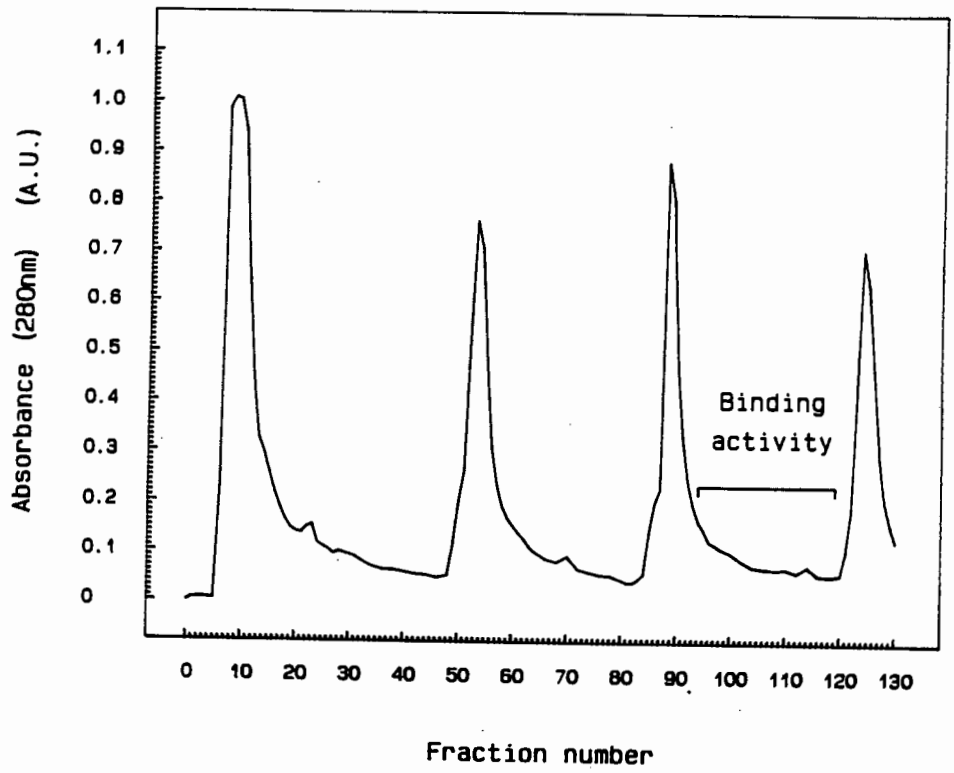


Figure 4.1 P11 phosphocellulose protein elution profile

P. angulosis fourteen-hour embryonic nuclear extract (see text for details) was loaded onto a P11 phosphocellulose column with bed volume 180 ml. The column was washed with 2,5 column volumes 0,1 buffer C (see section 2.11.1) and fractions with a volume of 15 ml were collected. Bound proteins were eluted stepwise with 3 column volumes 0,3 buffer C (started after fraction number 42), 3 column volumes 0,5 buffer C (started after fraction number 78) and 1 column volume 0,8 buffer C (started after fraction number 114). The absorbance of 250 µl aliquots of each fraction (or when readings were fairly constant every second fraction) was read against a reagent blank at 280 nm and plotted in arbitrary units (A.U.; ordinate) against fraction number (abscissa), to produce the protein elution profile. The bracket labeled "Binding activity" denotes the fractions (94 to 119) which contained the bulk of suGF1 activity. These fractions were pooled for further manipulation.

stepwise increase in KCl concentration. Every fifth fraction[3] was monitored for suGF1 activity in an EMSA, the result of which is shown in Figure 4.2.

When nuclear extract was incubated with the E/H fragment in an EMSA (Figure 4.2 lane 1), at least three factor-DNA complexes (referred to as complexes C1, C2 and C3) distinct from those known to be specific suGF1-DNA complexes (B1 and B2; see section 1.2) could be identified. These complexes, either alone or in combination with each other, have been observed in EMSAs with numerous nuclear extract preparations (J.P. Hapgood, D. Patterton, unpublished). The occurrence of these complexes may be explained by nonspecific binding to the E/H fragment. Alternatively, the complexes may contain factors distinct from suGF1, but binding specifically to the 11 bp G-string, or binding to other sequences in the E/H fragment. They may also represent multimers (slower migrating than B1 and B2) or degradation products (faster migrating than B1 or B2) of suGF1. The bands do not seem to be an artifact of electrophoresis as their presence coincides with the presence of nuclear extract proteins, and they can be fractionated by P11 phosphocellulose chromatography (discussed below). Although determination of the nature of these complexes was not an experimental aim of this investigation, additional results, which will be noted as encountered, either discount or strengthen some of the possible explanations.

The second important feature visible in Figure 4.2 is that no suGF1 activity (defined in section 2.8) could be detected in the flow-through (lane 4) or in lane 13, which corresponds to the 0,3 buffer C protein peak. The position of migration of complex C3 in lane 13, will be shown to be different from that of suGF1 multimeric species in a subsequent section. These multimeric species are only detected in EMSAs with highly enriched suGF1 preparations (after DNA-affinity chromatography), and when protein is in excess. Fraction 89 (lane 21) of the 0,5 buffer C protein peak of Figure 4.1, yielded a substantial amount of complex C2, as well as a high molecular weight smear of shifted probe. Fraction 95 contained some suGF1 activity (complexes B1 and B2), as well as complexes C1, C2 and C3. The amount of suGF1 activity increased in fractions 100 and 105 (while C1 decreased) and decreased again in fractions 110 and 115. It was unclear if fraction 125 (0,8 buffer C protein peak) contained any suGF1 activity.

To resolve the question of possible suGF1 activity in fraction 125 and other fractions (e.g. fraction 89), an EMSA was carried out using smaller volumes of the fractions. In addition, every second fraction from number 86 to 112 was assayed by EMSA for the presence of suGF1, in order to decide which fractions to pool. The results of these EMSAs are shown in

[3] Note that fractions 54 and 89 were tested instead of 55 and 90, to coincide with protein elution peaks in Figure 4.1.

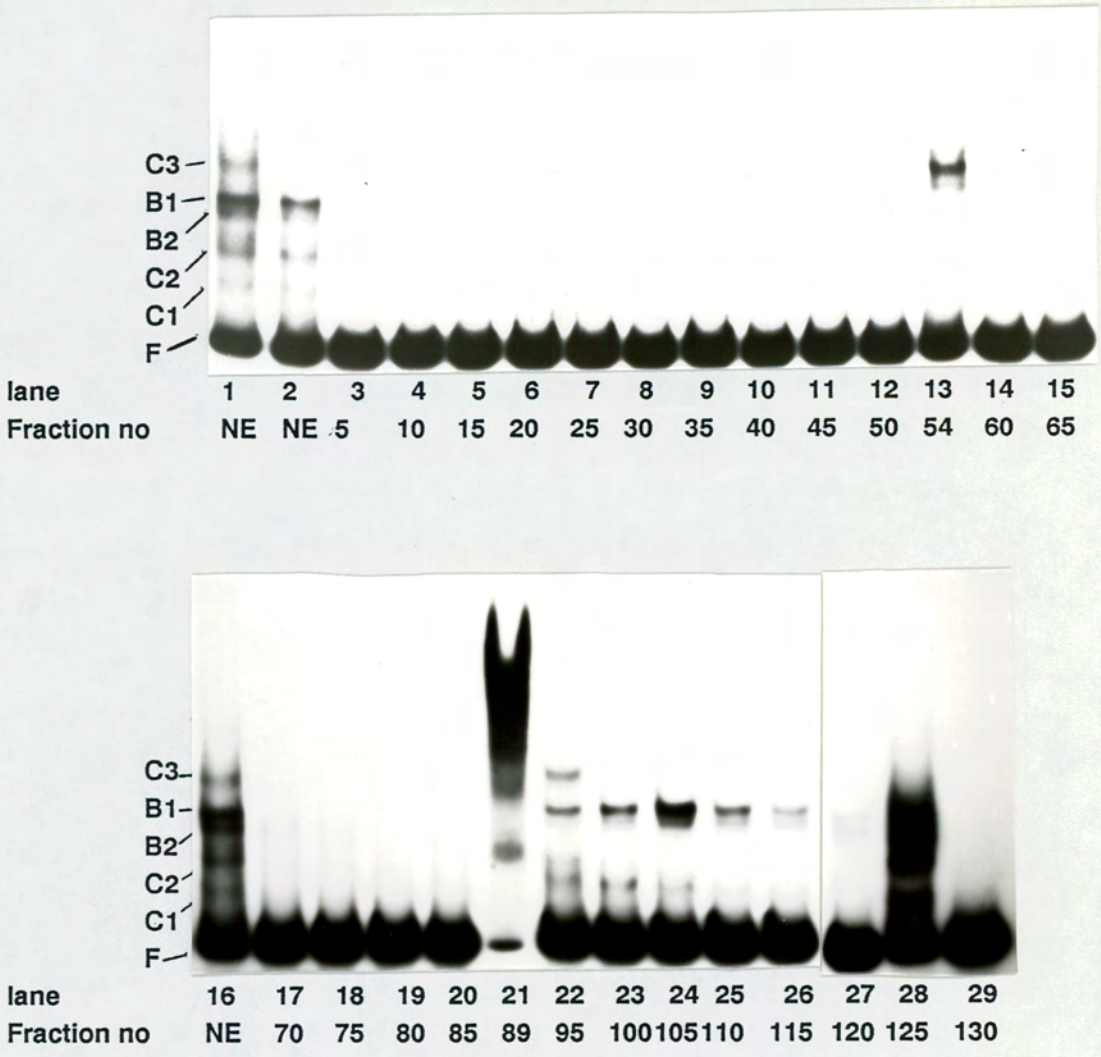


Figure 4.2 Elution profile of DNA-binding activities from the P11 phosphocellulose column

Aliquots of nuclear extract (NE; 0,1 µl in lane 2 or 0,2 µl in lanes 1 and 16), or 1 µl of every fifth fraction collected from the P11 column (indicated by fraction number), were incubated with 1 ng of end-labeled 335 bp E/H fragment in EMSA incubation buffer (see section 2.8; 250 mM KCl). Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). F is free end-labeled E/H fragment. B1 and B2 are suGF1-DNA complexes. C1, C2 and C3 are factor-DNA complexes distinct from B1 and B2. The fraction numbers refer to the same fractions as in Figure 4.1.

Figure 4.3.

suGF1-DNA complex B1 was first detected in fraction 90 (Figure 4.3(B), lane 5), was most abundant in fraction 102 (Figure 4.3(A) lane 10) and decreased towards fraction 112 (Figure 4.3(A) lane 15). Complex B1 could still be detected in fraction 117 (Figure 4.3(B) lane 15) and very faintly in fraction 119 (Figure 4.3(B) lane 16). suGF1-DNA complex B2 was detectable from fraction 94 (Figure 4.3(B) lane 11) to 119. Interestingly the DNA-binding protein involved in complex B1, eluted from the P11 column before that involved in complex B2. (The ratio of complex B1 to B2 as estimated by eye changes from almost 1:0 in fraction 94 to at least 1:1 in fraction 117.) This point will be raised later in discussion (see chapter 7).

Fraction 88 (Figure 4.3(A), lane 3) contained two complexes, C4a (denoted by bottom arrow) and C4b (denoted by top arrow), migrating between complexes C2 and B2. A third complex migrating between complexes C3 and B1, as well as higher molecular weight complexes which were not well resolved, could also be detected. From Figure 4.3(B) lanes 1, 2 and 3, however, it appeared likely that complexes in fraction 88 were not distinct DNA-binding species, but multimers or aggregates of C4a. (Complex C4b could not be detected in fraction 88 in Figure 4.3(B), and could possibly be a factor distinct from C4a and present in lower amounts.) Fraction 90 (lane 4 in Figure 4.3(A) and 4, 5 and 6 in Figure 4.3(B)), seemed to predominantly contain[4] the factor forming complex C2, with a small amount of factors forming complexes B1 and C3. The smear between complexes B1 and C3 in lane 4 of Figure 4.3(A) could most probably be ascribed to multimers or aggregates of complexes C2 or C3. Complexes C1, C2 and C3 were most prevalent in fractions number 96 to 98 (Figure 4.3(A) lanes 7 and 8), 90 (Figure 4.3(A) lane 4) and 92 to 94 (Figure 4.3(A) lanes 5 and 6), respectively.

The only remaining question was if suGF1 complexes B1 and B2 were present in the 0,8 buffer C protein peak. Figure 4.3(B) lanes 12 to 14 revealed the presence of a large amount of complexes C1 and C2, and only a very small amount of complex B1. In addition, a doublet migrating faster than the doublet of suGF1-DNA complexes (B1 and B2) was visible. This doublet may be ascribed to degraded suGF1, which only elutes at higher ionic strength from the cation exchange column.

It was therefore decided to pool fractions number 94 to 119. The large amounts of contaminating factors which could bind to the E/H fragment in the presence of pdIdC in fractions number 88 to 93, which coincided with the large 0,5 buffer C protein peak (see Figure 4.1), were thus excluded. Furthermore, DNA-binding proteins in the 0,8 buffer C peak

[4] In the context of this discussion only the DNA-binding species and not total protein is being referred to.

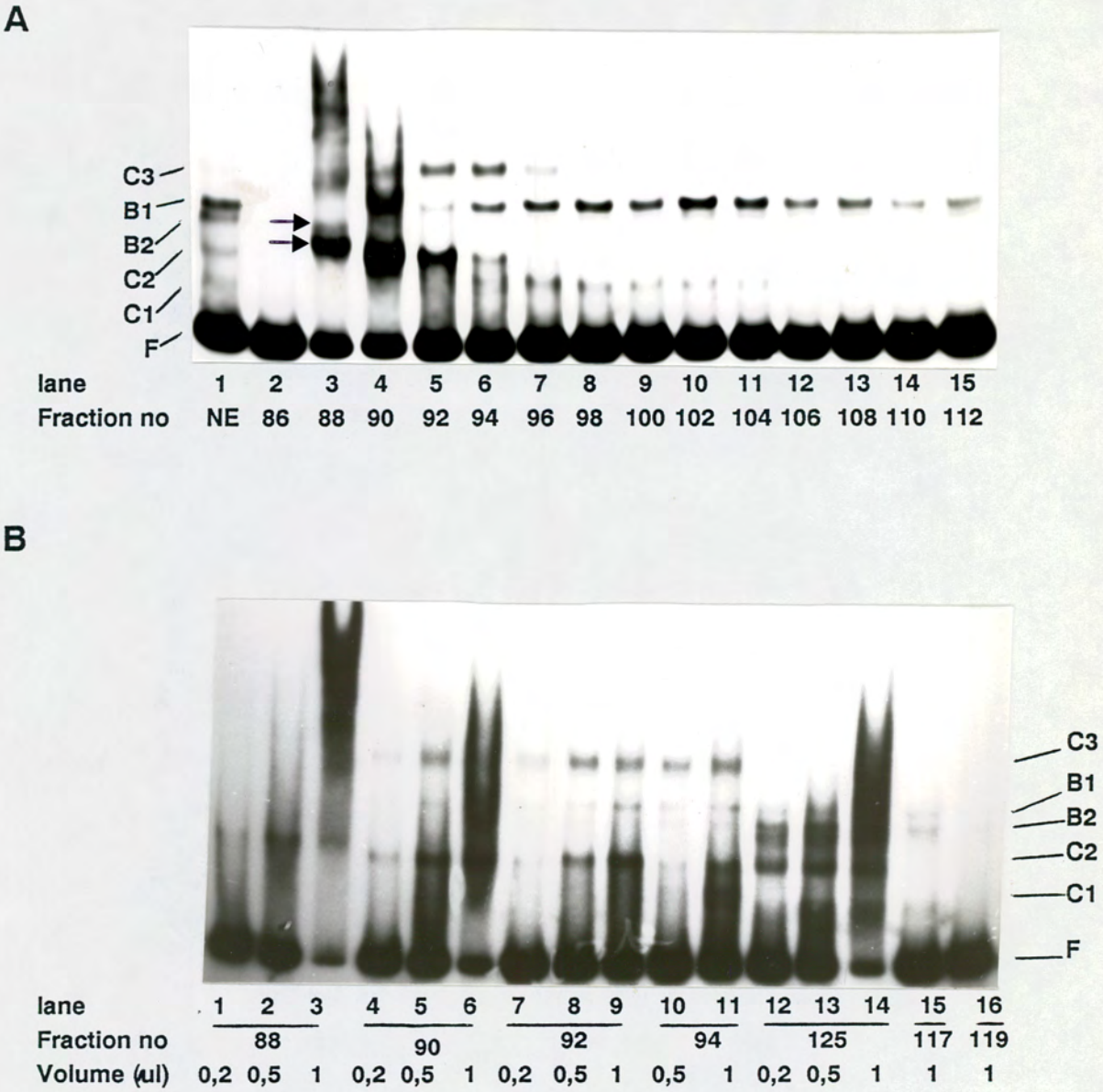


Figure 4.3 Elution profile of DNA-binding activities from the P11 phosphocellulose column : more detailed analysis

(A) An aliquot of nuclear extract (0,1 µl, lane 1) or 1 µl of the P11 fractions as indicated, were incubated with 1 ng of the end-labeled 335 bp E/H fragment in EMSA incubation buffer (see section 2.8; 250 mM KCl). Complexes were fractionated and detected as described in section 2.8. F refers to free end-labeled E/H fragment and B1 and B2 to suGF1-DNA complexes. C1, C2 and C3 are factor-DNA complexes distinct from B1 and B2. The top arrow denotes complex C4b in lane 3 and the bottom arrow complex C4a in the same lane. The fraction numbers refer to the same fractions as in Figures 4.1 and 4.2.

(B) Small volumes (0,2, 0,5 or 1 µl) of the indicated P11 fractions were incubated with the end-labeled E/H fragment as described in (A). F, B1, B2, C1, C2 and C3 are as in (A). The fraction numbers refer to the same fractions as in Figure 4.1.

(fraction 125) were excluded. The pooled fractions thus contained a selection of proteins including the bulk of the suGF1 activity and a small amount of contaminating factor(s) responsible for complexes C1, C2 and C3.

4.2.2 Poly(dG).Poly(dC)-affinity chromatography

Pooled active fractions from the P11 phosphocellulose purification step were diluted with buffer C to a final KCl concentration of 0,35 M and incubated with pdIdC (see section 2.11.2). The solution was applied to a 9 ml poly(dG).poly(dC)-affinity column pre-equilibrated in 0,35 buffer C. The flow-through was collected, followed by washing with 5 column volumes of 0,35 buffer C. Bound proteins were eluted with 8, 5 and 3 column volumes 0,55, 0,7 and 1,0 buffer C, respectively. Fractions of 1 column volume each were collected. Aliquots of 1 µl were assayed for suGF1 activity in an EMSA, the autoradiograph of which is shown in Figure 4.4.

suGF1 was eluted from the column in 3 column volumes of 0,7 buffer C (fractions 14, 15 and 16). No suGF1 activity could be detected in the flow-through (lanes 2 or 5). Multimers of the B1 and B2 complexes (doublet of bands B3 and B4, and doublet of bands B5 and B6) were resolved in this and other highly enriched preparations of suGF1. This is in contrast to the situation with nuclear extracts, where a high molecular weight smear is observed in the presence of excess protein (results not shown). This multimerization is discussed in later sections (see section 4.2.7 and chapter 7). Complex C3 (very faintly visible in lane 3) can now be seen not to be a multimer of suGF1, since it migrates in-between complexes B1B2[5] and B3B4 (compare lanes 2 and 3 with lane 1).

Contaminants responsible for complexes C1, C2 and C3 did not bind to the poly(dG).poly(dC)-affinity column. Although only small amounts of C1, C2 and C3 were visible in the flow-through and fraction number 1, a substantial total amount of these contaminants were present, due to the large volume of the fractions. (The volume of the flow-through was approximately 520 ml.). These complexes may contain factor(s) binding with relatively low affinity to poly(dG).poly(dC), thus not being able to bind to the column at 350 mM KCl and in the presence of pdIdC. Alternatively, the complexes may contain factors binding specifically to sequences other than the suGF1 binding site in the E/H fragment, or abundant nonspecific polyanion-binding proteins.

[5] In referring to the doublet represented by complexes B1 and B2, the notation "B1B2" is used. Similarly, complexes B3B4 refers to complexes B3 and B4, and so forth.

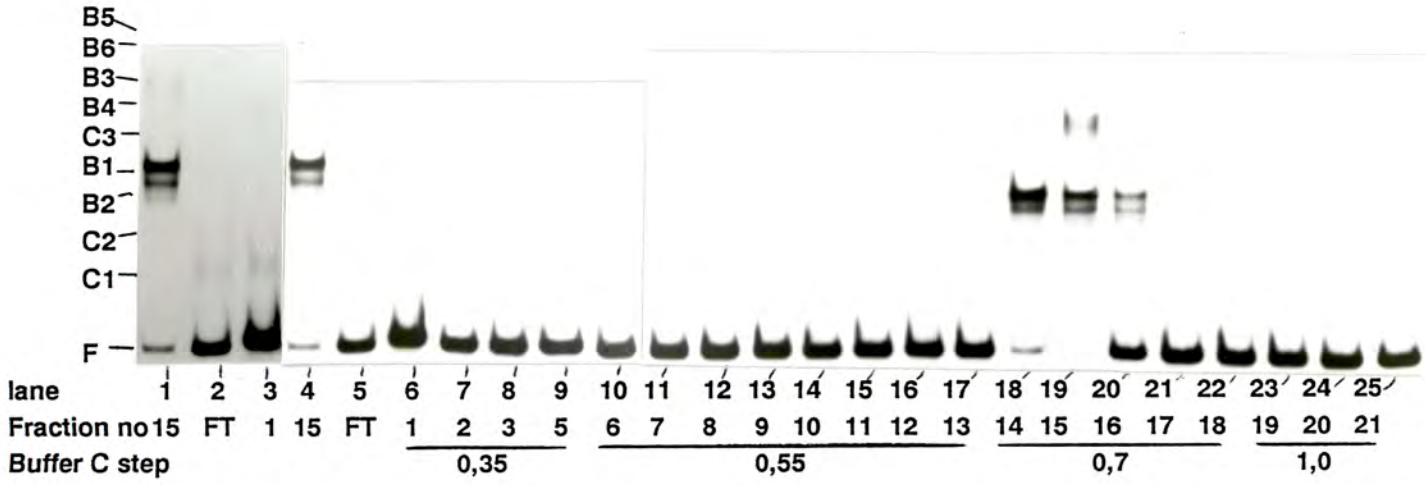


Figure 4.4 Elution profile of DNA-binding activities from the poly(dG).poly(dC)-affinity column

Aliquots (0,5 µl (lanes 1 and 4) or 1 µl (lanes 2, 3, and 5 to 25)) of the flow-through (FT) and each fraction collected from the poly(dG).poly(dC)-affinity column were assayed for the presence of suGF1 in an EMSA with 1 ng end-labeled 335 bp E/H fragment (see section 2.8; 250 mM KCl). Fraction numbers and KCl concentration in the elution buffers (buffer C step) are indicated. F is free end-labeled E/H fragment. B1 to B6 are suGF1-DNA complexes. C1 to C3 refer to the same complexes (distinct from B1 to B6) as in Figures 4.2 and 4.3. Lanes 1, 2 and 3 are derived from an overexposed photographic print of lanes 4, 5, and 6.

4.2.3 suGF1 Binding Site-affinity chromatography

Pooled active fractions from the poly(dG).poly(dC)-affinity chromatography step were diluted to 0,35 buffer C, incubated with pdIdC (see section 2.11.3) and one eighth loaded onto each 1 ml BS-affinity column. The flow-through was collected and the columns washed with 6 ml (6 column volumes) 0,35 buffer C. Bound proteins were eluted with 8, 6 and 4 ml of 0,5, 0,65 and 1,0 buffer C, respectively. Figure 4.5 depicts the results of an EMSA (see section 2.8; 250 mM KCl) wherein 1 µl of every second fraction was monitored for suGF1 activity. The only detectable DNA-binding activity was that of suGF1, which eluted at 0,65 buffer C (fractions 1 to 5) (lanes 9, 10 and 11).

The active fractions from all eight affinity columns were pooled, diluted with buffer C to a final KCl concentration of 0,35 M, and incubated with pdIdC. The solution was divided into four and passed over 1 ml BS-affinity columns a second time (see section 2.11.3). The elution profile of suGF1 determined by EMSA, is shown in Figure 4.6 (A). suGF1 activity eluted from the second BS-affinity column at the same ionic strength as the first pass, namely in the 0,65 buffer C fractions. No suGF1 activity could be detected in the flow-through or any fractions other than the 0,65 buffer C step.

4.2.4 Identification of a 59,5 kDa species on SDS-PAGE with the same elution profile as the suGF1 activity

The fractions from the final BS-affinity column which had been assayed for DNA-binding activity in an EMSA (Figure 4.6(A)) were analyzed in terms of protein components by SDS-PAGE and silver-staining (see section 2.12). The silver-stained SDS gel is shown in Figure 4.6(B). One major band (approximately 60 kDa) could be detected which co-eluted with the suGF1 activity (lanes 9 and 10). The overstained gel (lanes 14 and 15) revealed the presence of at least two smaller contaminating species (migrating at 49,5 kDa and 45,1 kDa respectively), visually estimated to constitute 15% or less of the total protein. The major band, which will subsequently be shown to be suGF1 (see section 4.3), was sized on this, and two independent silver-stained gels. Polynomials were fitted to the distances migrated by molecular weight standards (measured on enlarged photographic prints) *vs.* the log(MW), by the least squares method (see section 2.25 and Table 4.1). suGF1 was found to migrate on SDS-PAGE with an apparent molecular weight (MW_A) of 59,53 kDa (standard deviation 0,31).

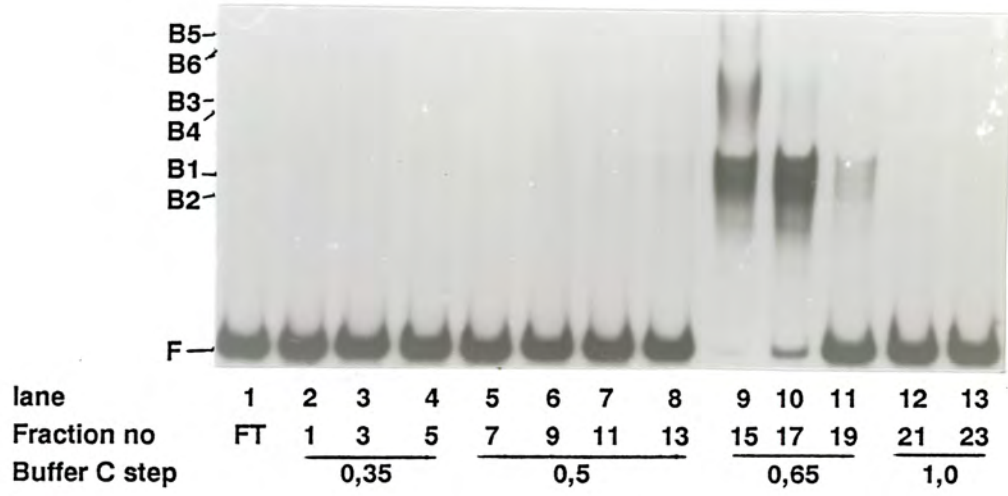


Figure 4.5 Elution profile of DNA-binding activities from the Binding Site-affinity chromatography column (First pass)

A 1 µl aliquot of the flow-through (FT) or every second fraction eluting from the BS-affinity column was assayed for the presence of suGF1 in an EMSA with 1 ng end-labeled 335 bp E/H fragment (see section 2.8; 250 mM KCl). Fraction numbers and KCl concentration in the elution buffers (buffer C step) are indicated. F is free end-labeled E/H fragment. B1 to B6 are suGF1-DNA complexes.

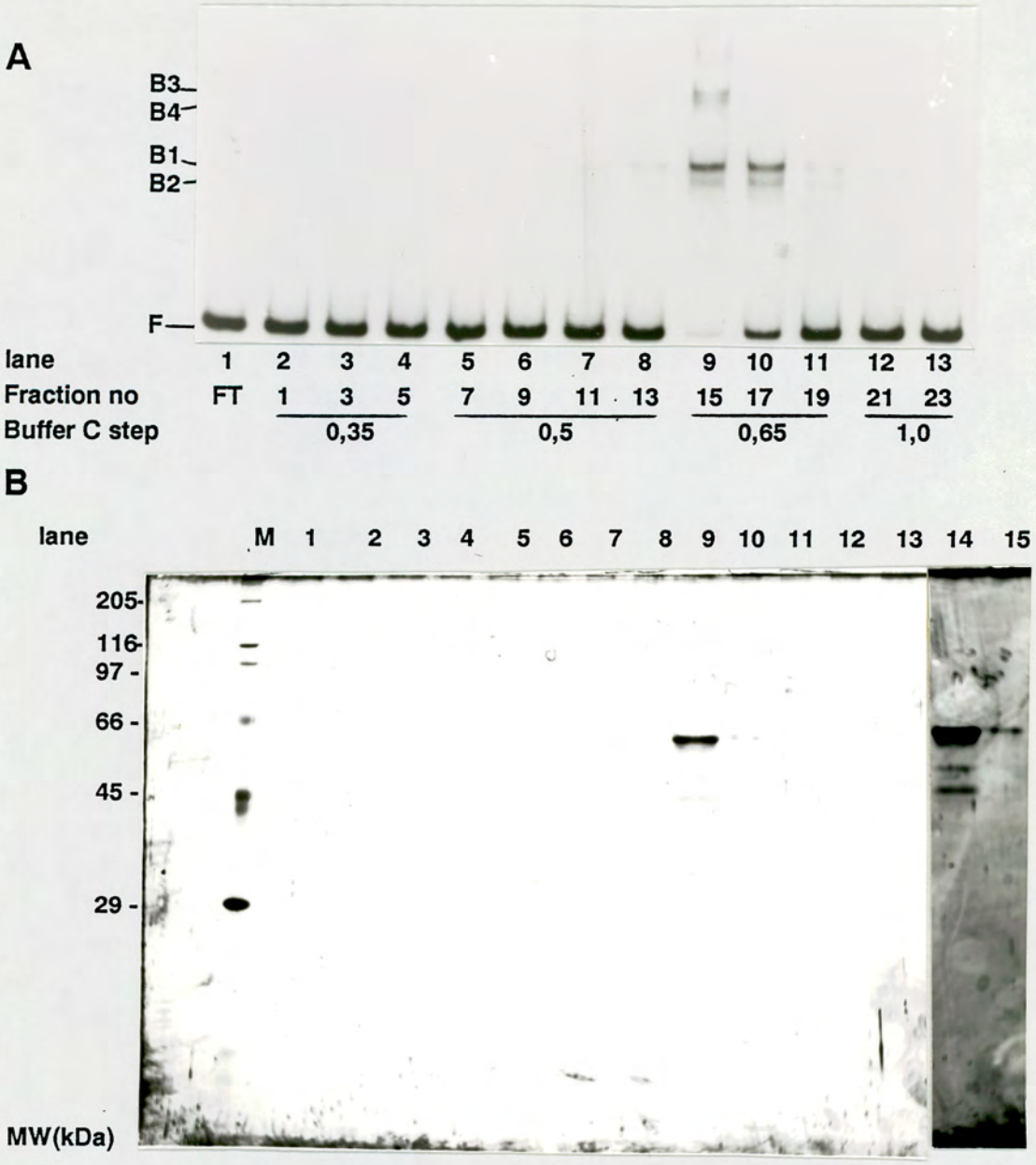


Figure 4.6 Elution profile of DNA-binding activities and total protein from the Binding Site-affinity chromatography column (Second pass)

(A) A 1 µl aliquot of the flow-through (FT) or every second fraction eluting from the second pass over the BS-affinity column was assayed for the presence of suGF1 in an EMSA with 1 ng end-labeled 335 bp E/H fragment (see section 2.8; 250 mM KCl). Fraction numbers and KCl concentration in the elution buffers (buffer C step) are indicated. F is free end-labeled E/H fragment. B1 to B4 are suGF1-DNA complexes.

(B) An aliquot (1 ml) of the flow-through (FT) or every second fraction eluting from the second pass over the BS-affinity column was precipitated with trichloroacetic acid, dissolved in SDS loading dye and neutralized with NaOH (see section 2.9.1). Electrophoresis on a 10% SDS polyacrylamide gel and silver-staining (method 2) were as described in Materials and Methods. Sizes (kDa) of molecular weight markers (lane M) are indicated in the left margin. Numbering of the lanes in (B) correspond to the same fractions as the numbers of the lanes in (A). Lanes 14 and 15 show lanes 9 and 10 where the gel has been overstained.

Table 4.1 : Sizing of suGF1 from SDS - PAGE

Gel	Order of polynomial	RMSE*	MW_A of suGF1 (kDa)
1	5th	$2{,}70 \times 10^{-6}$	59,8
2	4th	$3{,}65 \times 10^{-3}$	59,6
3	5th	$1{,}00 \times 10^{-5}$	59,2

* RMSE denotes the root mean square error of the least squares fit

4.2.5 Drawing up an enrichment table for the purification of suGF1

The purification is summarized in an enrichment table (see Table 4.2), from which several points can be noted. Firstly, the stepwise yields were consistently high, especially for the affinity chromatography steps, leading to a high overall yield of 29% of the initial suGF1 activity in the starting material. These yields will be compared in chapter 7 with those obtained by others for the purification of G-binding factors.

The approximately 10-fold stepwise enrichment for the P11 phosphocellulose step is followed by an enormous enrichment of 150- to 180-fold for the poly(dG).poly(dC)-affinity chromatography step, but only an additional 1,6- to 2-fold combined enrichment for the two BS-affinity chromatography steps. This suggests that the selectivity of the BS-affinity column is probably not significantly different from that of the poly(dG).poly(dC)-affinity column. The 11 contiguous G's may be the main feature recognized by suGF1 and the remaining contaminants. Although the BS-affinity steps increase the total enrichment from about 2200- to 3400-fold, the total yield is decreased by 20%. This represents a substantial loss, considering the small amount of total protein obtained at the last step (28 µg). More than one affinity chromatography step may not be the method of choice for future purification of large amounts of suGF1 to obtain peptide sequence information. It might rather be advisable to try a preparative EMSA (70) or SDS-PAGE (30) step after poly(dG).poly(dC)-affinity chromatography.

The most important point to be noted from Table 4.2 is that the overall enrichment of suGF1

Table 4.2 Enrichment Table for the Purification of suGF1[1]

Fraction	Volume (ml)	Activity[2] (U/µl)	Total Activity (U x 10^{-3})	Protein Conc[3] (µg/µl)	Total Protein (mg)	Specific Activity[4] (U/µg)	Step Yield[5] (%)	Total Yield[6] (%)	Step Enrichment[7]	Total Enrichment[8]
Nuclear extract	41,0	384,0	15744	8,03	329	47,8	100	100	1	1
P11 Phosphocellulose	370,0	33,9	12543	0,0601	22,2	564	79,7 (80)	79,7 (80)	11,8 (9,8)	11,8 (9,8)
Poly(dG).poly(dC)-affinity chromatography	26,2	322,0	8436	0,00310	0,0812	103871	67,3 (59)	53.6 (47)	184 (152)	2173 (1490)
1st pass BS-affinity chromatography	40,6	161,0	6537	0,00135	0,0548	119259	77,5 (66)	41,5 (31)	1,15 (1,5)	2495 (2234)
2nd pass BS-affinity chromatography	20,0	230,0	4600	0,00141	0,0282	163121	70,4	29,2	1,37	3413

1 Data presented in this table is from one complete purification. Calculations for a second, independent purification are shown in brackets.
2 suGF1 activity was measured in binding units (U). One unit was defined as the amount of suGF1 needed to shift 0,01 ng of E/H fragment in an EMSA. Quantitation was as described in section 2.8 and was carried out at approximately 10% total shift (see section 3.3).
3 Protein concentration was determined by the bicinchoninic acid method (see section 2.9.1).
4 Specific activity (U/µg) was obtained by dividing the binding activity (U/µl) by the protein concentration (µg/µl).
5 The total activity (U) for each step is given as a percentage of the total activity of the previous step, with the stepwise yield of the nuclear extract assigned a value of 100%.
6 The total activity (U) for each step is given as a percentage of the total activity of the nuclear extract.
7 The specific activity (U/µg) for each step is given as a percentage of the specific activity of the previous step. The nuclear extract is assigned a stepwise enrichment value of 1.
8 The specific activity (U/µg) for each step is given as a percentage of the specific activity of the nuclear extract.

from nuclear extracts is very high (3400-fold). Furthermore, if one considers the total enrichment of suGF1 from nuclei, this value can be multiplied by a factor of approximately 10. This is because the nuclear extract preparation is already considerably enriched in suGF1 by ammonium sulfate fractionations and precipitation of protein during dialysis (29; see section 3.2).

It should, however, be noted that EMSAs are not a very accurate way of quantitating DNA-binding activities (210). This is especially true in the case of crude preparations such as nuclear extracts (20; see section 3.3), thus making it very difficult to determine relative values at different stages of purification, which is exactly what is required for determination of the yield and enrichment. Accurate determination of the extremely low protein concentrations may lead to additional errors. The values for yield and enrichment should thus be regarded as guidelines, and viewed in conjunction with silver-stained SDS-PAGE gels to assess the purification. The possibility of different sensitivity of proteins to silver-staining serves to discount a categorical statement on purity estimated from a silver-stained SDS-PAGE gel. However, in many cases of purification of DNA-binding proteins, the purity of the final preparation is simply assessed from silver-stained SDS-PAGE gels. These silver-stained gels often reveal the presence of minor contaminants in pure preparations. Examples of this are found in the purification of BGP1 (37), NFKB (237) and P3A2 (29). The minor contaminants visible in overstained SDS-PAGE gels of the final suGF1 preparation are present in similar or lower abundance compared to contaminants in the examples mentioned above. suGF1 seemed to constitute at least 85% of the total protein on silver-stained SDS-PAGE gels of the purified preparation (see section 4.2.4).

A rough estimation of final purity based on calculated mass of suGF1 compared to mass of total protein[6], with the assumption of a molecular weight of 59,5 kDa, yields values of 47% (w/w) and 94% (w/w), when it is assumed that suGF1 binds as a monomer or dimer, respectively (see section 7.1). The significance of these values is debatable if the probability of error inherent in the EMSAs, and the possibility of loss of activity due to heat or mechanical stress such as repeated freezing and thawing during purification, is taken into account. The values should therefore not be regarded as evidence that suGF1 binds to DNA as a dimer.

[6] The value for protein concentration in the final preparation given in Table 4.2 was used for this calculation. Activity was estimated at 242 U/µl. This value was determined by EMSA at the optimum ionic strength for DNA-binding (175 mM KCl; see chapter 5) instead of at 250 mM KCl, which is the concentration used in EMSAs to obtain the data for Table 4.2. 1 ng E/H fragment is approximately $4{,}63 \times 10^{-15}$ moles fragment; 1 U of activity indicates binding of 0,01 ng fragment; the final volume of the purified preparation is 20 ml; the activity is 242 U/µl; therefore the final preparation contains 224 pmoles suGF1, i.e. 13,3 µg; the total protein is 28,2 µg

4.2.6 SDS-PAGE analysis of suGF1 purification

Aliquots of the nuclear extract and pooled active fractions for the different steps of the purification were electrophoresed on a SDS polyacrylamide gel, in order to compare the protein components as the purification proceeded. The silver-stained gel is shown in Figure 4.7. If the relative amount of units of suGF1 activity which was loaded in lane 1 is arbitrarily set at 9, then 5, 30, 540, 470 and 1000 relative units of suGF1 activity were loaded in lanes 2 to 6, respectively (as calculated from Table 4.2). Although there seemed to be some discrepancy between the calculated units and the intensity of the suGF1 band when comparing lanes 4 and 5 to lane 6, the general conclusions drawn from the enrichment table could be verified by this gel : the step of highest enrichment was the poly(dG).poly(dC)-affinity chromatography step, and the overall enrichment was substantial (several thousand-fold).

The final preparation is known to contain contaminants (see Figure 4.6(B), lane 14), but these seem to be present in very low amounts as judged by the detection of only one intense band on the silver-stained gel shown in Figure 4.7. The small enrichment brought about by the BS-affinity chromatography steps is presumably caused by a decrease in the percentage of contaminants present in quantities below the detection limit of silver-staining, or of the two major contaminants previously observed (see Figure 4.6(B) lane 14). These contaminants are detectable upon silver-staining of larger amounts of the active fractions after poly(dG).poly(dC)-affinity chromatography (results not shown).

4.2.7 Highly enriched suGF1 forms multimeric species in EMSA

It was necessary to concentrate the purified suGF1 preparation for further experiments such as DNase I- and hydroxyl radical footprinting. Pooled active fractions were concentrated by ultrafiltration in centricon devices (Amicon) as described in section 2.11.3. The filtrate and retentate were analyzed in an EMSA. Figure 4.8 shows a typical result obtained in such an EMSA. No suGF1 activity could be detected when 6 µl of the filtrate was incubated with radiolabeled E/H fragment (lane 3), whereas the starting material (lane 1) was concentrated approximately 10-fold (compare lanes 1 and 4). BSA and pdIdC are essential components of the EMSA incubation buffer with purified suGF1 (compare lane 6 with lanes 10, 11 and 12).

An interesting feature of highly enriched suGF1 preparations is the occurrence of multimerization of suGF1-DNA complexes (see for example Figure 4.8). As the ratio of purified suGF1 to DNA is increased, complexes B1 and B2 decrease in conjunction with the

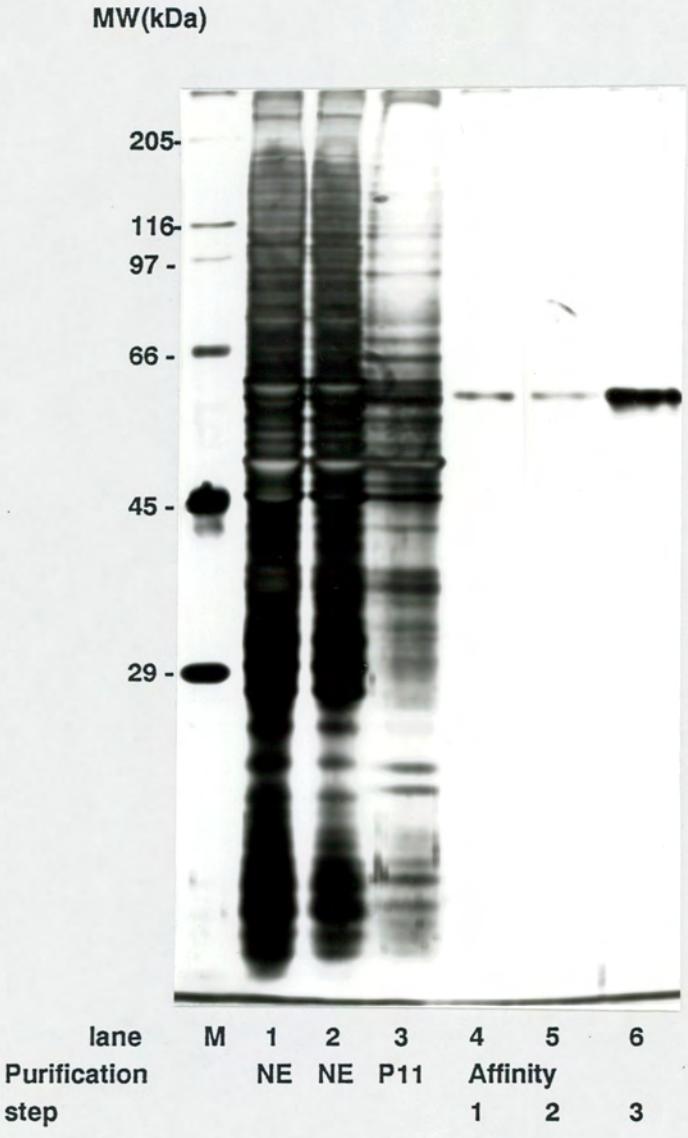


Figure 4.7 Affinity chromatography results in a highly enriched preparation of suGF1

Nuclear extract (NE; 7 µl in lane 1 and 4 µl in lane 2), P11 pooled active fractions (P11; 300 µl in lane 3), and 500 µl (lane 4), 900 µl (lane 5) and 1300 µl (lane 6) pooled active fractions from the poly(dG).poly(dC), first and second pass BS-affinity columns respectively, were TCA precipitated, resuspended in SDS gel loading buffer, neutralized with NaOH, boiled and electrophoresed on a 10% SDS gel which was subsequently stained with silver (see section 2.12, silver-staining method 2). Sizes of molecular weight markers (kDa; lane M) are indicated in the left margin.

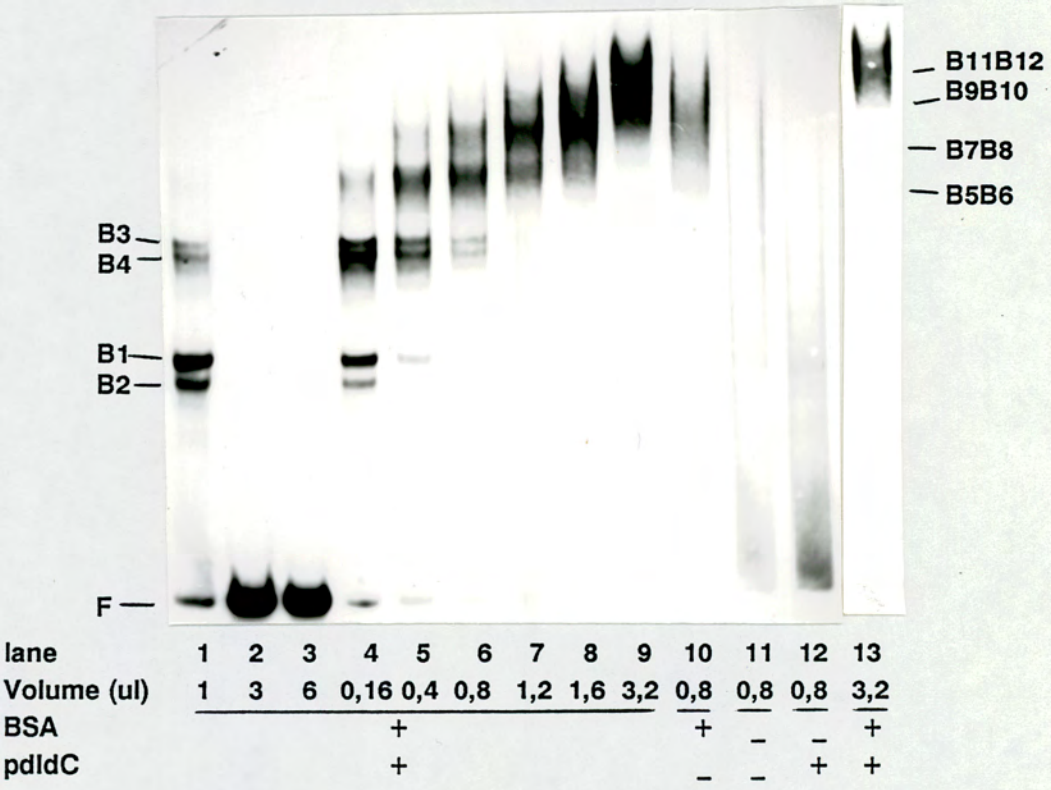


Figure 4.8 Concentrated purified suGF1 forms multimers and requires pdIdC and BSA for sequence-specific DNA-binding

Purified suGF1 before concentration (1 µl, lane 1), ultrafiltration filtrate (3 µl, lane 2 or 6 µl, lane 3), or different volumes of the retentate (as indicated in lanes 4 to 12) were incubated with the E/H fragment in a standard EMSA assay (see section 2.8; 175 mM KCl). The incubation was in the absence of BSA in lanes 11 and 12 and in the absence of pdIdC in lanes 10 and 11. Lane 13 is a lighter photographic print to show detail in lane 9. F is free end-labeled E/H fragment. B1 to B4, B5B6, B7B8, B9B10 and B11B12 are suGF1-DNA complexes.

migrating) complexes. Although the doublet character of the complexes is detectable only for the first two sets (in complexes B1B2 and in complexes B3B4), the relative distances migrated between the complexes seem to suggest that the slower-migrating bands represent two complexes each. The resolution of the electrophoretic system is, however, not sufficient to resolve complexes of similar size larger than B3B4. Lane 13, which is a lighter photographic print of the same negative of lane 9, clearly shows that at least six sets of complexes are resolved in this gel system. This multimerization will be discussed in chapter 7.

4.3 Identification of the 59,5 kDa species as suGF1

A prominent band representing a 59,5 kDa protein co-eluting with the suGF1 activity was detected on SDS-PAGE (see Figure 4.6). From Figure 4.6 and the estimated purity of at least 47% (see section 4.2.5), it appeared that this prominent protein corresponded to suGF1.

Four experiments were carried out to obtain unequivocal proof that the 59,5 kDa band on SDS-PAGE was indeed suGF1. These included Southwestern blotting, in which proteins are transferred from a SDS-PAGE gel to a nitrocellulose membrane. The membrane is then incubated with a radiolabeled DNA oligonucleotide containing a sequence known to bind the protein of interest. This is followed by stringent washing, and detection of bands by autoradiography. The second experiment involved isolation of the protein in the 59,5 kDa band from a SDS-PAGE gel, followed by renaturation and detection of specific binding to a fragment containing the recognition sequence of suGF1. Thirdly, the DNase I footprint obtained with the purified preparation on the E/H fragment was compared to the DNase I footprint obtained with nuclear extracts, which had been ascribed to the protein(s) named suGF1. Lastly, ultraviolet protein-DNA crosslinking was performed, in an attempt to verify the size(s) of the protein(s) involved in the sequence-specific suGF1-DNA complexes observed in EMSAs.

4.3.1 Southwestern blotting of suGF1 preparations

Aliquots (approximately 750 U of suGF1 activity) of two independent purified and concentrated suGF1 preparations were electrophoresed on a 10% SDS-PAGE gel, together with an aliquot of nuclear extract and molecular weight markers (see section 2.14). The gel was blotted onto nitrocellulose and probed with radiolabeled oligo-S (see Figure 2.1), which contains the binding site for suGF1 (see section 2.14 and the legend to Figure 4.9). Positions of the molecular weight standards were identified by Amido Black staining of the blot. Note that the 205 kDa standard did not show up (Figure 4.9 lane M), presumably due to poor transfer

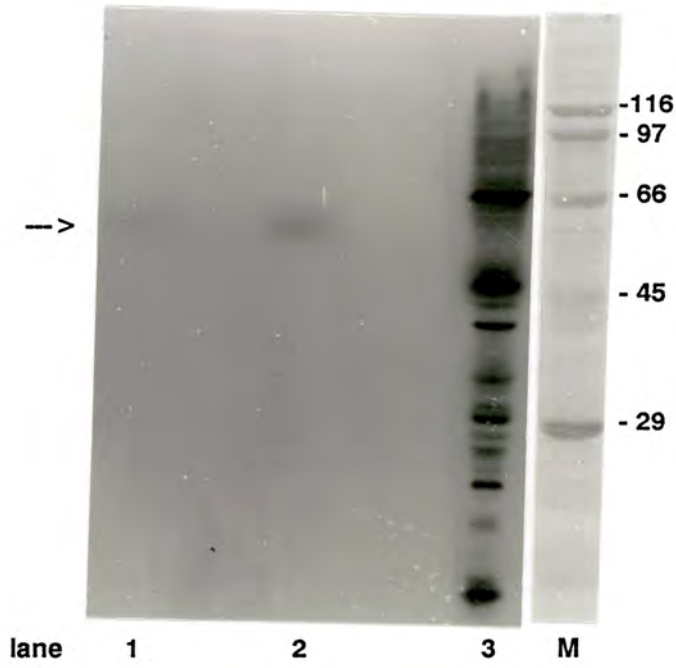


Figure 4.9 A 60 kDa band is detected upon Southwestern blotting of purified suGF1

Approximately 750 U of suGF1 activity contained in nuclear extract (lane 3), concentrated purified suGF1 of the same preparation (lane 2) shown in Figures 4.6, 4.7 and 4.8, and concentrated purified suGF1 of an independent preparation (lane 1) were adjusted to 80 µl in SDS loading dye, boiled, and electrophoresed with high molecular weight standards on SDS-PAGE (see section 2.12). The proteins were blotted onto a nitrocellulose membrane (see Materials and Methods), whereafter the blot of the standard lane was cut off and stained in Amido Black (lane M). The membrane containing the blotted samples was blocked, washed, probed with radiolabeled oligo-S (see Figure 2.1; approximately 400 ng, 20 000 dpm/ng), washed, blotted dry and subjected to autoradiography as described in Materials and Methods. Sizes of molecular weight standards (kDa) are indicated in the right margin. The arrow denotes the position of migration of the protein bound by oligo-S on the membrane.

onto the membrane due to the protein's large molecular mass.

Oligo-S bound to a protein migrating at the expected position for suGF1 (Figure 4.9, lanes 1 and 2, band denoted by an arrow) could be detected in both preparations of purified suGF1. The apparent molecular weight of the protein was determined after fitting a fourth order polynomial to the log(MW) *vs.* the distances migrated by the molecular weight standards, as measured directly on the stained blot aligned with the autoradiograph of the radioactively probed blot. The root mean square error of the fitted polynomial was $1{,}397 \times 10^{-3}$ and the apparent molecular weight of the protein was found to be 60,4 kDa. This value agrees very well with that of the major species seen upon silver-staining of the active fractions (only a 1,5% difference), and very poorly with the apparent molecular weights of the contaminating bands in the final preparation. The 1,5% discrepancy in size is within the limits of experimental error, considering that measurements were complicated by the blotted protein not being a sharp band.

A large number of factors present in the nuclear extract ranging in apparent molecular weight from about 116 kDa to smaller than 29 kDa , were also bound by oligo-S (Figure 4.9 lane 3). Although an excess of calf thymus DNA was present as nonspecific competitor in the probing solution (see section 2.14), specific binding of the 60 kDa protein and the factors in the nuclear extract to oligo-S could be determined unequivocally by probing nitrocellulose blots in the presence or absence of unlabeled specific (oligo-S) or nonspecific (oligo-NS) competitors. EMSAs using as radiolabeled probe oligo-S, and as unlabeled competitors oligo-S or -NS, had previously been performed (J.P. Hapgood, unpublished). It was shown that unlabeled oligo-S inhibited 50% of the radiolabeled suGF1-oligo-S complexes at a molar excess of 2,5-fold (unlabeled oligo-S : labeled oligo-S), whereas oligo-NS inhibited 10% or less of the radiolabeled suGF1-oligo-S complexes at a molar excess of 62,5-fold (unlabeled oligo-NS : labeled oligo-S) (J.P. Hapgood, unpublished).

Figure 4.10 depicts the result of probing the blotted proteins in the nuclear extract (lanes 1, 3, 5 and 7) or purified preparation (lanes 2, 4, 6 and 8) in the absence (lanes 1 to 4) or presence of specific (lanes 5 and 6) or nonspecific (lanes 7 and 8) competitors at a 200-fold molar excess over labeled probes[7]. These results show that the protein in the major band (suGF1; indicated by an arrow in Figure 4.9) formed a specific complex with oligo-S, since it did not bind to labeled oligo-NS in the absence of competitors (lane 4), or to labeled oligo-S in the presence of

[7] Note that the same preparations of purified suGF1 and nuclear extract as in lanes 1 and 3 of the Southwestern blot shown in Figure 4.9 were used in this experiment, the only differences being electrophoresis on a 7% SDS gel, and probing with oligonucleotides of 50% higher specific activity. These two changes seemed to greatly improve the amount of protein transferred to and detected on the blot. The mobilities of the nuclear extract and purified proteins were increased in the 7% gel.

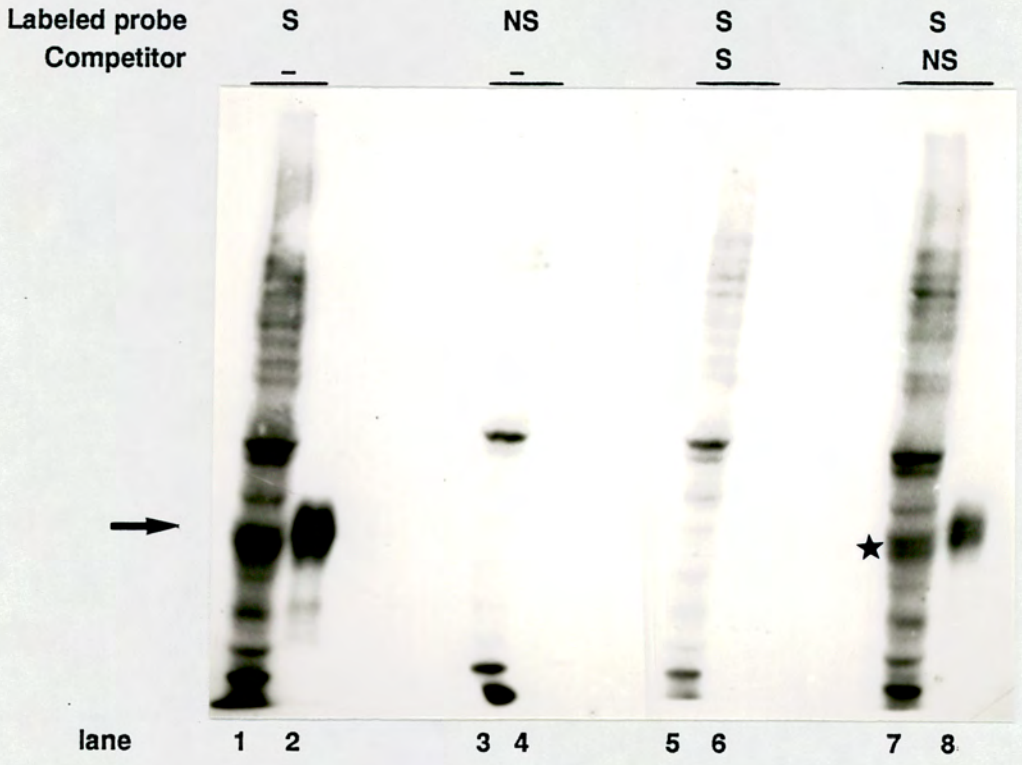


Figure 4.10 Southwestern blot of suGF1 with competitors

Approximately 750 U of suGF1 activity contained in the same nuclear extract (lanes 1, 3, 5 and 7) and purified suGF1 preparations (lanes 2, 4, 6 and 8) as used in lanes 3 and 2 in Figure 4.9 respectively were used. The samples were electrophoresed by SDS-PAGE and blotted onto nitrocellulose as in Figure 4.9, except that the percentage acrylamide was 7% compared to 10% in Figure 4.9. The membrane was then cut into four strips, each containing a lane with blotted nuclear extract and purified suGF1. The blots were probed with either radiolabeled oligo-S (denoted by S; lanes 1,2,5,6,7 and 8) or radiolabeled oligo-NS (see Figure 2.1; indicated by NS; lanes 3 and 4) at the same specific activity of approximately 30 000 dpm/ng, in the absence (lanes 1 to 4) or presence of a 200-fold molar excess of unlabeled specific (oligo-S; lanes 5 and 6) or nonspecific (oligo-NS; lanes 7 and 8) competitors. Probing and autoradiography were as described in Materials and Methods. The arrow denotes the position of the 60 kDa suGF1.

a 200-fold molar excess of oligo-S (lane 6), but bound to labeled oligo-S when oligo-NS was present at the same molar excess (lane 8). (Note that a scratch on the surface of the autoradiograph causes the band in lane 8 to appear as a doublet.) A minor (<5%) faster migrating contaminating species could be detected in addition to the major band obtained with the purified suGF1 preparation (lane 2).

Other factors present in the nuclear extract and binding with high specificity to oligo-S are expected to exhibit the same binding pattern as suGF1, namely being present in lanes 1 and 7 (labeled oligo-S in the presence and absence of nonspecific competitor respectively) and absent in lanes 3 and 5 (labeled oligo-NS, and labeled oligo-S respectively, with specific competitor). The band(s) in the nuclear extract migrating in a position just below suGF1 (marked with an asterisk) would fit these criteria. The two prominent bands in lane 3 could best be explained by two very abundant factors, exhibiting a low specificity for oligo-S. This would explain the bands being present in the absence of competitors with both oligo-S (lane 1) and oligo-NS (lane 3), but slightly more intense in the case of oligo-S. The high abundance would prevent these bands from being competed away by the 200-fold molar excess used in lanes 5 and 7, although oligo-S would be a slightly better competitor.

No factors binding specifically to oligo-NS (expected in lane 3) seem to be present in the nuclear extract within the limits of detection. Furthermore, factors binding nonspecifically to DNA do not seem to be present. Such factors would be expected in lanes 1 and 3 when present in small amounts. If such nonspecific factors were however fairly abundant, their presence would be expected in lanes 1, 3, and to a lesser extent in lanes 5 and 7. Very abundant nonspecific factors would be expected in equal amounts in lanes 1, 3, 5 and 7. All the other factors detected in lane 1 of the nuclear extract therefore seem to fall into the category of specific binding to oligo-S, but with lower affinity or specificity for oligo-S than suGF1 and the band(s) marked with an asterisk. This is deduced from the presence of these bands in lane 5, where suGF1 and the factor in the band marked with an asterisk are competed away completely by the specific competitor. These results strengthen the argument that the factors eluting just before the suGF1 activity from the P11 phosphocellulose column (detected in an EMSA, see Figures 4.2 and 4.3) might also be binding to the oligo(dG).oligo(dC) sequence contained in the H1-H4 intergenic fragment.

4.3.2 Recovery and renaturation of suGF1 activity from the 59,5 kDa band on a SDS gel

An aliquot (1 ml) of the final purified suGF1 preparation was TCA precipitated and electrophoresed by SDS-PAGE. The precise position of the 59,5 kDa band was determined by

silver-staining of parallel lanes containing purified suGF1 and molecular weight standards. The gel slice containing the 59,5 kDa protein(s) was excised from the unstained gel, and the protein was eluted and precipitated (29; see section 2.11). The protein(s) was denatured by resuspension in 100 µl of a buffer containing 6 M guanidinium hydrochloride. This solution was loaded onto a gel filtration column (P-6 Bio-Gel from Bio-Rad) with an exclusion limit of 6000 kDa. The column had been equilibrated with buffer F (see section 2.11), containing amongst other components 60 µg/ml BSA and 5 µM $ZnCl_2$. BSA was used to decrease nonspecific binding of protein to the column, whereas Zn(II) was thought to possibly be required in trace amounts for refolding of the protein to a conformation able to bind DNA (see chapter 5).

Fractions (2 drops, approximately 50 µl) were collected from the gel filtration column and 6 µl of every second fraction was assayed in an EMSA. The fractions were adjusted to 50% glycerol and stored at -70°C. A peak of suGF1 activity was detected in the void volume fractions and some suGF1 activity also eluted in a peak in the inner volume fractions. The suGF1 activity eluting in the inner volume yielded two bands (B1 and B2) with E/H labeled probe in an EMSA which could be competed specifically with oligo-S (results not shown) and might be ascribed to nonspecific adsorption of suGF1 to the resin. The void volume fractions displaying suGF1 activity were pooled and tested for specific binding to the suGF1 recognition site in the H1-H4 spacer of the early histone gene battery of *P. miliaris* in a competition EMSA, the result of which is shown in Figure 4.11.

Two complexes migrating at the same position as the suGF1-DNA complexes B1 and B2 were clearly visible in the renatured sample (lane 1), when compared to the control (lane 6) in which an aliquot of the purified suGF1 was incubated with the E/H fragment. The equally poor resolution of the renatured and control bands may be ascribed to a modified incubation buffer used in the EMSA (see section 2.13) to accommodate the solution in which the fractions eluted from the column. The two complexes migrating at the positions of B1 and B2 in the incubations with renatured sample were strongly competed by a 10-fold molar excess of specific competitor (oligo-S; lane 2) but could not be removed by a 50-fold molar excess of a nonspecific competitor (oligo-NS; lane 5), which is in agreement with the behaviour of control suGF1 in the B1 and B2 complexes in lanes 7 to 10. The major 59,5 kDa protein species present on SDS-PAGE of the purified suGF1 preparation therefore contains the protein(s) named suGF1 in the B1 and B2 protein-DNA complexes observed in an EMSA (see chapter 7 for discussion).

The yield of suGF1 activity after isolation from the SDS polyacrylamide gel was found to be less than 0,1%. This is 10-fold lower than that reported by Calzone *et al.* (29) for the protein

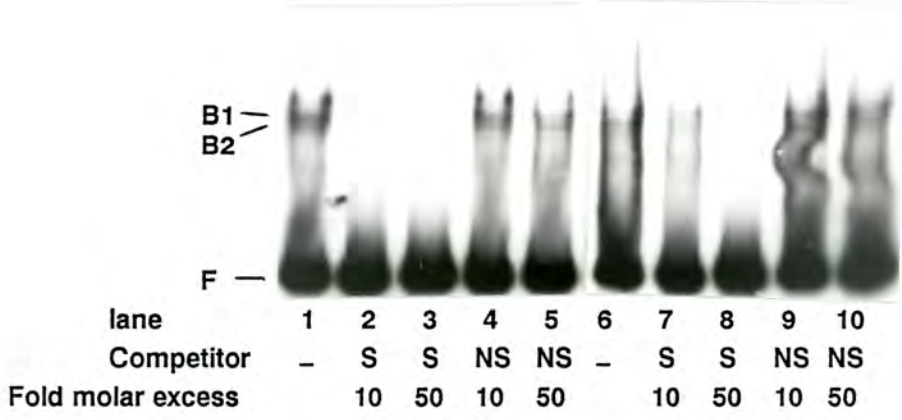


Figure 4.11 suGF1 sequence-specific DNA-binding activity can be renatured from the 59,5 kDa band on a SDS-PAGE gel

The 59,5 kDa protein(s) were eluted from a SDS-PAGE gel slice, denatured in guanidinium hydrochloride, passed over a Bio-Gel P-6 gel filtration column (Bio-Rad), and fractions assayed for binding to the E/H fragment (see section 2.13). Aliquots (6 µl) of the pooled active fractions eluting from the gel filtration column in the void volume (lanes 1 to 5), or 0,2 µl purified suGF1 of the same preparation that was initially loaded onto the SDS gel (lanes 6 to 10) were incubated with 1 ng end-labeled E/H fragment and specific or nonspecific competitor DNA in a total volume of 25 µl in modified EMSA incubation buffer (16 mM Hepes (pH 8) 0,024% (v/v) Nonidet-P40, 0,8 mM DTT, 175 mM KCl, 60 µg/ml BSA, 1,2 µM $ZnCl_2$, 0,4 mM PMSF, 1 mM EDTA, 1,6 mM $MgCl_2$) for 30 min at 4°C. Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). Specific and nonspecific competitors were oligo-S (lanes 2, 3, 7 and 8) and oligo-NS (lanes 4, 5, 9 and 10), respectively (see Figure 2.1). Competitors were present at a 10- (lanes 2, 4, 7 and 9) or 50-fold (lanes 3, 5, 8 and 10) molar excess over labeled E/H fragment. F is free end-labeled E/H fragment. B1 and B2 are specific suGF1-DNA complexes.

P3A2. In addition to possible large losses at the elution, precipitation and chromatography steps, renaturation of suGF1 might have been complicated by divalent cation requirement other than Zn^{2+} for correct folding (see chapter 5), or other factors. TCA precipitation could also have had an irreversible adverse effect on suGF1 DNA-binding activity.

4.3.3 DNase I footprint of purified and crude suGF1

A DNase I footprint assay was carried out with the E/H fragment to determine if the purified protein interacted with the same sequence in the H1-H4 intergenic region of the early histone gene battery of *P. miliaris*, as did the binding activity defined to be caused by suGF1 in the *P. angulosis* nuclear extracts. 1 ng E/H fragment was therefore subjected to limited DNase I digestion in the absence (Figure 4.12, lanes 4 and 5) or presence of 6,5 µg nuclear extract proteins (lane 3) or 6 ng purified (lane 2) proteins as described in section 2.18.1. The cleavage products were isolated and electrophoresed on a sequencing gel with a Maxam-Gilbert G-sequencing standard (lane 1; 151; see section 2.17).

The footprint was clearly visible in lanes 2 and 3 as a "window" in the digestion products where DNase I digestion was impeded by association of suGF1 with the DNA or by steric hinderance of bound suGF1, when compared to the digestion pattern of free DNA shown in lanes 4 and 5. The footprint extended over the same length and sequence of DNA (indicated in the margins of Figure 4.12) in the lanes with nuclear extract and purified suGF1, confirming that the isolated protein was suGF1.

4.3.4 Protein-DNA ultraviolet crosslinking

DNA containing bromodeoxyuridine is considerably more sensitive to UV-induced crosslinking compared to DNA containing thymidine (6). UV crosslinking had been used to estimate the size of suGF1 in nuclear extracts (J.P. Hapgood, personal communication). These earlier studies had revealed two crosslinked species of about 70 and 95 kDa on SDS-PAGE upon UV irradiation of *P. angulosis* nuclear extracts which had been incubated with BrdU substituted, radiolabeled oligo-S (J.P.Hapgood, personal communication). The two species had been shown to contain protein, were dependent on UV irradiation, and could be competed by specific but not nonspecific DNA competitors (J.P.Hapgood, unpublished). Since purified suGF1 had been sized as a 59,5 kDa species by SDS-PAGE, UV crosslinking was repeated with nuclear extracts and carried out for the first time with purified suGF1, in an attempt to clarify the earlier result and reconcile it with the results presented here. The latter included the fact that both complexes B1 and B2 could be obtained from a gel slice containing the 59,5 kDa species (see section 4.3.2), that this was the major (at least 85%) species detected on silver-

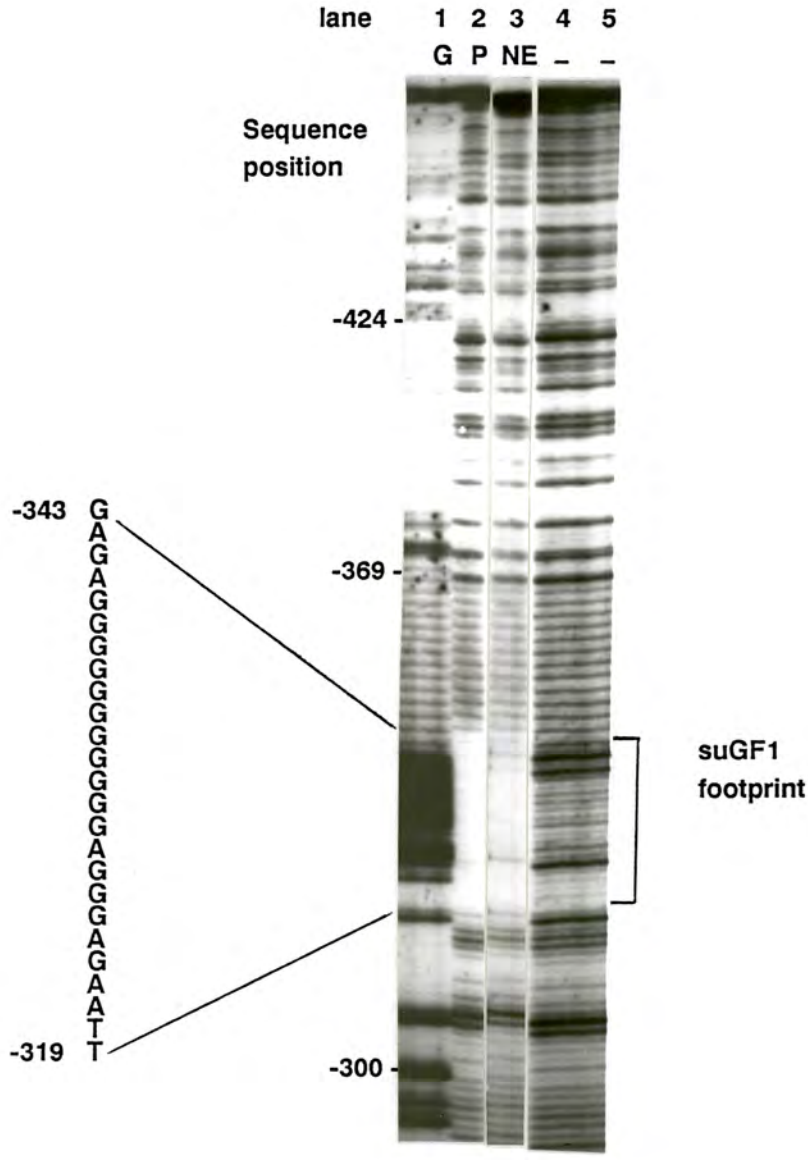


Figure 4.12 The same DNase I Footprint is obtained with crude (nuclear extract) and purified suGF1 on the H1-H4 intergenic region

E/H fragment (1 ng) 3' end-labeled only on the coding strand (see section 2.4) was incubated without (lanes 4 and 5) or with 6,5 µg nuclear extract (NE; lane 3) or 6 ng purified suGF1 (P; lane 2) in EMSA incubation buffer (see section 2.8; 175 mM KCl). Limited DNase I digestion was carried out and the cleavage products were isolated and electrophoresed together with a Maxam-Gilbert G-sequencing standard (G; lane 1; see section 2.17) on a 6% polyacrylamide, 7 M urea sequencing gel as described in sections 2.18.1 and 2.21 of Materials and Methods. The DNase I footprint of suGF1 is indicated by a bracket in the right margin, and the corresponding DNA sequence is given in the left margin. Sequence position is relative to the main mRNA cap site at +1 (92).

stained gels (see section 4.2.4), and that oligo-S bound specifically to this species and not to higher molecular weight proteins upon Southwestern blotting (see section 4.3.1).

The BrdU probe was prepared by a Klenow fill-in reaction, in the presence of bromodeoxyuridine and [α-^{32}P]dCTP, on a template (noncoding strand of oligo-S) annealed to a 9 bp complimentary primer as described in section 2.15. The full-length extension product was isolated from a gel and used in analytical EMSAs to determine optimum conditions for the UV crosslinking experiment (see section 2.15). Nuclear extract or suGF1 was then incubated with the radiolabeled probe under these optimum conditions, with 5 µg pdIdC, and in the absence or presence of specific (oligo-S) or nonspecific (oligo-NS) competitors. The incubations were irradiated with or shielded from (controls) UV light. Although BSA (MW 66 kDa) which is normally used as carrier protein in EMSAs with purified suGF1 does not compete for binding to oligo-S, it was substituted with Egg Albumin which has a molecular weight of 45 kDa and thus could not interfere with the detection of crosslinked species around 60 to 70 kDa (see section 2.15).

After crosslinking in solution, a tenth of each sample was analyzed on an EMSA gel, the autoradiograph of which is shown in Figure 4.13(A), while the remainder was electrophoresed on an identical gel for preparative purposes (see section 2.15). Two protein-DNA complexes can be detected in Figure 4.13(A). These two complexes were always obtained in EMSAs with oligo-S and were not well resolved in the EMSA gel system. The complexes always migrated slower than complexes B1 and B2 (with the E/H fragment), most likely due to only approximately one tenth of the negative charges of the E/H fragment being present in oligo-S (results not shown). The upper band was found to always be more prominent than the lower band, analogous to the situation with complexes B1 and B2 (J.P. Hapgood, unpublished). The complexes obtained with oligo-S were formally called BO1 and BO2 instead of B1 and B2 to indicate possible differences between these complexes obtained with the 30 bp oligonucleotide, and those with the 335 bp E/H fragment.

As shown in Figure 4.13(A), no shifts were observed in the absence of protein (lanes 1 and 2), whereas a large percentage of the BrdU probe was shifted into complexes BO1 and BO2 by nuclear extract (lanes 3 and 4) or suGF1 (lanes 5 and 6) regardless of UV irradiation. Complexes BO1 and BO2 were specific suGF1-DNA complexes since they were competed away by 10 µg specific competitor (oligo-S; lanes 7 and 9), but not competed to the same degree by the same amount of nonspecific competitor (oligo-NS; lanes 8 and 10)[8].

[8] Protein-DNA complexes in lane 10 appeared diminished on this gel probably due to an artifact of electrophoresis or gel drying, as these complexes were very clearly visible on the identical preparative gel.

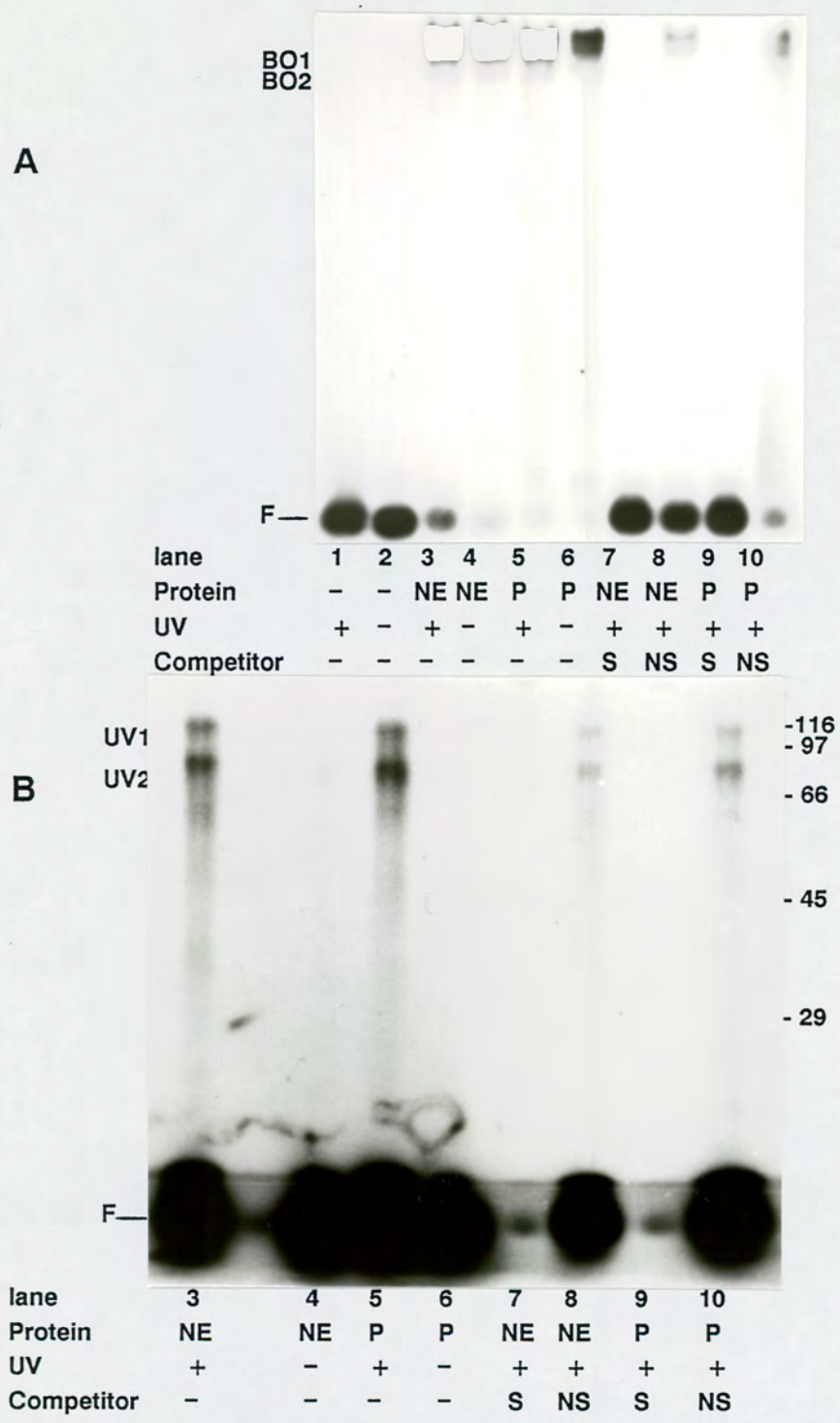


Figure 4.13 Protein-DNA ultraviolet crosslinking of crude (nuclear extract) and purified suGF1

(A) Nuclear extract (6,5 µl, approximately 52 µg protein) or purified suGF1 (approximately 9 ng protein) was incubated with 200 000 dpm of the internally labeled BrdU double-stranded oligo-S probe (see section 2.15) and 5 µg pdIdC in crosslink buffer (see section 2.15). Incubation was in the absence (lanes 1 to 6) or presence of specific (oligo-S; lanes 7 and 9) or nonspecific (oligo-NS; lanes 8 and 10) competitors at 4°C for 30 min in a final volume of 50 µl. Incubations of purified suGF1 contained 3 µg egg albumin (MW 45 000 kDa) as carrier protein. The samples were then irradiated with ultraviolet light as described in section 2.15 (lanes 1, 3, 5 and 7 to 10), or shielded from the UV source by foil (lanes 2, 4 and 6). Aliquots of 5 µl were analyzed on an identical nondenaturing polyacrylamide gel as was used to fractionate the remaining 45 µl (see sections 2.8 and 2.15). The autoradiograph of this analytical gel is shown in (A). NE is nuclear extract. P is purified suGF1. F is free BrdU probe. BO1 and BO2 are protein complexes with the BrdU probe.

(B) Gel slices containing BrdU probe-protein complexes (or containing no visible complexes) were excised from the preparative EMSA nondenaturing polyacrylamide gel, soaked in SDS loading dye for 30 min, placed in the wells of a 10% SDS gel and subjected to electrophoresis, staining, drying and autoradiography as described (see section 2.15). Lane numbers correspond to the lanes from which gel slices at the position of the BrdU probe-protein complexes were excised from the preparative gel as shown in (A). Abbreviations are as in (A). UV1 and UV2 are the major crosslinked species. Molecular weight and position of migration of standards as determined on the stained gel are shown in the right margin.

Gel slices containing either both the visualized BO1 and BO2 complexes, or the region where these complexes are normally expected to appear, were excised from the equivalent lanes on the preparative gel. The gel slices were soaked in SDS loading dye and placed in the wells of a 10% SDS-PAGE gel. The gel was subjected to electrophoresis, stained with Coomassie Brilliant Blue to visualize the size standards, and dried (see section 2.15). The autoradiograph is shown in Figure 4.13(B). Two UV-dependent bands (UV1 (major band) and UV2 (minor band)) were observed in the absence of competitors for the nuclear extract and suGF1 (lanes 3 and 5). These bands were specific suGF1-DNA species in the nuclear extract and purified preparation as both were competed away completely by specific competitor (lanes 7 and 9) but only slightly by nonspecific competitor (lanes 8 and 10). The efficiency of crosslinking was very low, as judged by the large amount of free BrdU probe migrating with the dye front.

The complexes were sized by fitting a 3rd order polynomial (RMSE $3{,}3\times10^{-3}$) through the log(MW) *vs.* the distance migrated by the molecular weight markers on the stained, dried gel. UV1 and UV2 migrated with apparent molecular weights of 69 and 88 kDa respectively. The 69 kDa band most likely contains the 59,5 kDa protein identified as suGF1 in earlier sections, because covalently bound DNA which is not removed by nuclease digestion prior to electrophoresis, is known to increase the apparent molecular weight (210). Attempts to digest away the DNA with DNase I resulted in a loss of the radioactive signal (J.P. Hapgood, unpublished). The identity of the 88 kDa band is, however, difficult to explain. Although formally possible, it seems improbable that this band is a contaminant present in roughly equal proportions to suGF1 before and after purification, binding specifically to oligo-S, and migrating at the same position as the suGF1-DNA complexes isolated from the preparative EMSA (BO1 and BO2). The possibility of the 88 kDa band (UV2) being a suGF1-species found in complex BO1 (or B1) and the 66 kDa band (UV2) a suGF1-species found in complex BO2 (or B2) is difficult to reconcile with the fact that both complexes B1 and B2 can be reconstituted from the 59,5 kDa band in an EMSA. This hypothesis has also been ruled out by the observation that both UV1 and UV2 were generated after excision of the indicated bands B01 and BO2 from a preparative gel in a UV-crosslinking experiment with crude nuclear extract (J.P. Hapgood, personal communication). The 88 kDa band may represent an artifact of the crosslinking method (see reference 180 for a review). UV crosslinking experiments of other factors have indeed yielded results which are difficult to explain (see for example reference 135).

CHAPTER 5

INVESTIGATION OF THE ION DEPENDENCE OF suGF1 DNA-BINDING

5.1 Introduction

Sequence-specific protein-DNA interactions are usually stable at relatively high ionic strengths in comparison to nonspecific protein-DNA interactions, the latter associations being based mainly on ionic interactions and thus disrupted more easily by masking of charge *via* increased ion concentrations (15). Knowledge of ionic strength required for optimal *in vitro* protein-DNA binding is crucial in the design of experiments involving the assay of DNA-binding activity and purification of DNA-binding proteins. The concentration and nature of divalent cations required for DNA-binding or function can even provide clues as to the structure of the protein or the mode of its interaction with DNA (see section 1.4).

Zinc(II) ions have been shown to be an important structural element of many DNA-binding proteins (12, 14, 87, 157, 205). Three classes of Zn-binding domains have been identified (reviewed in section 1.4). TFIIIA was the first identified member of the Zn finger DNA-binding family. This protein contains nine Cys_2His_2 Zn fingers (21, 158). The other two classes are the double loop-Zn-helix domains which contain $(Cys_2Cys_2)_2$ Zn-binding domains (65, 87, 205), and the Cys_6 Zn-binding domains, such as those found in yeast transcription factor GAL4 (87, 176, 177).

Merkle and Berg (157) recently reviewed some of the methods that can be used to determine if DNA-binding proteins bind metal ions and/or if binding of metal affects activity. Total metal content in purified protein preparations can be determined by direct methods such as atomic absorption spectroscopy (see for example reference 83), or reconstitutive methods, for example blotting proteins separated by SDS-PAGE onto nitrocellulose and probing with radioactive zinc. A protein function requiring metal ions can be monitored before and after removal of, and subsequent supplementation with metal ions (157). Examples of such functional methods involve site-directed mutagenesis of putative metal-binding amino acids (157), and inactivation by chelators followed by reconstitution of *in vitro* transcription or DNA-binding activity (157; see for example references 43, 83, 109, 139, 237).

Removal of Zn^{2+} from TFIIIA resulted in a complete loss of DNA-binding as assayed by DNase I footprinting. This could be accomplished either by dialyzing the factor against EDTA or 1,10-phenanthroline (*o*-phenanthroline, abbreviated OP), or by a 10 minute incubation in the presence of either of these chelators followed by removal of chelated metal by gel filtration, or by chelator treatment without subsequent removal of the chelator-Zn^{2+} complexes (83). Minimum concentrations of only 2 mM EDTA or 0,2 mM OP were required, whereas 4,7-phenanthroline (*p*-phenanthroline, abbreviated PP), which is incapable of chelating Zn^{2+}, did not inhibit DNA-binding (83). DNA-binding of EDTA-treated TFIIIA could be restored by addition of 15 µM $ZnCl_2$ to the chelator-treated samples, but not by the same concentration of $CoCl_2$, $MnCl_2$ or $FeCl_2$ (83).

The results of similar experiments with selected DNA-binding proteins, including experiments involving suGF1 that are presented later in this chapter, are summarized in Table 5.1. If the results of such experiments involving proteins known to contain Zn-binding domains are considered (TFIIIA, Sp1, GR and GAL4), it is clear that these results should be approached with some caution. DNA-binding of transcription factor Sp1 which is known to contain three zinc fingers is, for example, not inhibited by 1 mM OP (109). Although incubation of Sp1 with 50 mM EDTA followed by dialysis and dilution to 0,2 mM EDTA is sufficient to remove all Sp1 DNA-binding activity, binding can only be restored to 10 or 20% of the original activity by addition of $ZnSO_4$ (109). DNA-binding of transcription factor GAL4, which is known to contain a class 3 Zn-binding domain (8, 127, 148; see section 1.4) can similarly not be fully restored after removal of divalent cations by EDTA (33; see Table 5.1). Furthermore, the addition of Zn^{2+} to buffers used in the purification of proteins known to contain Zn-binding domains, is not always required (20).

A difference in the ability to demonstrate metal-dependence of proteins for DNA-binding by removing metal ions with chelators may reflect differences in affinity for Zn^{2+} between various DNA-binding proteins that require maintenance of the structures of Zn-binding domains. These results may however also simply reflect differences in the sensitivity of the experiments. This could be due to differences in the chemical environment such as different concentrations of buffer components which may influence the solubility of Zn^{2+} salts. It is often difficult to remove trace amounts of divalent cations from all protein-containing samples, solutions and glassware. Contaminating trace amounts of Zn^{2+} may be sufficient for the maintenance of DNA-binding when the protein sample is assayed for DNA-binding in the absence of exogenously added Zn^{2+}. Furthermore, an inability to reconstitute DNA-binding by addition of metal ions cannot be interpreted as a lack of a requirement for the metal ion for DNA-binding. Such a result might be ascribed to irreversible denaturation or enhanced sensitivity to air oxidation at cysteine residues of the protein in the absence of metal

Table 5.1 : Effect of metal-chelating agents on DNA-binding of selected proteins

Factor	Chelator	Chelator Concentration (mM)	Cation (Concentration)	Binding Activity[4]	Reference
TFIIIA (purified; Contains nine Zn fingers)	EDTA[1] or OP[2]	2 and 0,2 respectively	-	-	83
			Zn^{2+} (15 µM)	+	
			Co^{2+}, Ni^{2+}, Mn^{2+}, Fe^{2+} (15 µM each)	-	
	PP[3]	0,2	-	+	83
Sp1 (purified; Contains three Zn fingers)	EDTA	50	-	-	109
			Zn^{2+} (500 µM)	10% to 20% restored	
			Co^{2+}, Ni^{2+} (500 µM)	-	
	OP	1	-	+	109
BGP1 (DNA-cellulose enriched; Not shown to contain a Zn-binding domain)	OP	1,5	-	-	139
			Zn^{2+} (5, 10 µM)	+	
			Co^{2+} (100 µM)	Partially restored	
			Mg^{2+}, Mn^{2+}, Fe^{2+} (10 µM, 100 µM, 1 mM)	-	
	PP	1,5	-	+	139
H4TF-1 (Phospho-cellulose P11 enriched; Not shown to contain a Zn-binding domain)	EDTA	5	-	-	43
	OP	0,5	-	-	43
			Zn^{2+} (50, 100, 500 µM)	+	
			Fe^{2+}, Mg^{2+}, Ca^{2+}, Cu^{2+} Cd^{2+}, Mn^{2+}, Co^{2+} (500 µM)	-	

Factor	Chelator	Chelator Concentration (mM)	Cation (Concentration)	Binding Activity[4]	Reference
H4TF-2 (Phosphocellulose P11 enriched; Not shown to contain a Zn-binding domain)	OP	0,5	-	-	43
			Zn^{2+} or Fe^{2+} (500 μM)	+	
			Mg^{2+}, Ca^{2+}, Cu^{2+},Cd^{2+} Mn^{2+}, Co^{2+} (500 μM)	-	
Ectoderm G-string factor (nuclear extracts; Not shown to contain a Zn-binding domain)	EDTA	4	-	+	236
Endoderm G-string factor (nuclear extracts; Not shown to contain a Zn-binding domain)	EDTA	0,5	-	-	236
			Zn^{2+}, Mg^{2+}, Mn^{2+}, Ca^{2+} Co^{2+} (no details)	-	
Glucocorticoid receptor DNA-binding domain (expressed; Class 2 Zn-binding domain)	EDTA pH 3[5]	10	-	-	65
			$ZnCl_2$ (250 μM) or $CdSO_4$ (5 μM)	+	
GAL4 (Amino acids 1-147 expressed; Class 3 Zn-binding domain)	EDTA pH 3.5[6]	10	-	-	33
			$ZnCl_2$ (25 μM) or $CdCO_4$ (5 to 25 μM)	5 to 20% restored	
	EDTA pH 5[7]	10	-	-	176
			Zn^{2+}, Cd^{2+} [8]	+	
			Co^{2+} [9]	Partially restored	

ions (14, 157). The latter is a very common problem (14). It is therefore important to perform such experiments under anaerobic conditions and in a reducing environment to prevent oxidation of cysteine residues (see Table 5.1 for examples).

Table 5.1 also summarizes results of Zn^{2+}-depletion and reconstitution of DNA-binding of some proteins with very G-rich recognition sequences (Sp1, BGP1, H4TF-1, and the ecto- and endoderm G-string factors). It should be noted that of these, only Sp1 has been unequivocally shown to contain a Zn-binding domain (three Zn fingers (109)). The information may however be relevant to experiments with suGF1, since the DNA-binding domains of these other proteins could be similar to that of suGF1 by virtue of the similar DNA recognition sequences. The chicken erythrocyte factor BGP1 (see section 1.3) for example, does not bind to the G-string in the β^A-globin promoter in the presence of 1,5 mM OP (139). Binding is restored in the presence of low concentrations of Zn^{2+} but, as is the case for many of these proteins (157), binding is inhibited by Zn^{2+} at concentrations of 1 mM or higher (139). Whereas 0,5 mM OP added prior to DNA binding is sufficient to remove metal ions from human transcription factor H4TF-1, the metal ions are relatively inaccessible to chelation when H4TF-1 is complexed to DNA (43). DNA-binding of H4TF-1 is completely abolished in the presence of 5 mM EDTA, and unlike BGP1 (37), addition of zinc to all buffers used during the purification is required, which greatly enhances DNA-binding activity (43, 44). As mentioned above, this may reflect (amongst other factors) the different degrees of affinity with which Zn^{2+} is bound by different metal-binding proteins.

In conclusion, when investigating the Zn^{2+} requirement for DNA-binding of a novel protein, it is important to investigate the effects of more than one type of chelator under different conditions. In addition, reconstitution of binding activity should be tested with concentrations of Zn^{2+} from the micromolar to millimolar range (157; See Table 5.1) in a reducing and preferably anaerobic environment. A negative result obtained in these experiments cannot be interpreted as unequivocal proof of a lack of a requirement for metal ions for DNA-binding.

In this chapter optimal KCl and NaCl concentrations for suGF1 DNA-binding were firstly determined, whereafter the requirement for magnesium as opposed to monovalent cations was investigated. Because suGF1 binds specifically to poly(dG).poly(dC) and certain Zn fingers are known to contact GGG 3 bp subsites (see section 1.4), it was thought that the protein might contain Zn fingers in a DNA-binding domain. Possible removal of Zn^{2+} by OP and EDTA was therefore investigated.

5.2 Monovalent cation requirement for suGF1 DNA-binding

The optimal KCl and NaCl concentrations for DNA-binding of purified suGF1 were determined in an EMSA, the result of which is shown in Figure 5.1. The top panel representing the distribution of labeled E/H fragment into free DNA and suGF1-DNA complexes at increasing KCl concentrations, demonstrates a marked dependence of suGF1-DNA complex-formation on KCl concentration. A small amount of suGF1 is seen to bind at 50 mM KCl, with increased complex-formation up to approximately 175 to 200 mM KCl, and a decrease in the amount of suGF1-DNA complexes from 200 to 400 mM KCl, where virtually no binding can be detected. The same trend of DNA-binding is apparent in the presence of increasing concentrations of a different monovalent cation, NaCl (shown in the bottom panel of Figure 5.1). It seems that the optimum NaCl concentration for DNA-binding might be decreased by a maximum of 25 mM compared to KCl. The DNA-binding thus seems to be a function of the concentration of monovalent ions in general (or of ionic strength).

Dependence of suGF1 DNA-binding on KCl concentration was compared in nuclear extracts and purified suGF1 preparations by EMSAs. Densitometric tracings of lanes on the autoradiographs are shown in Figure 5.2. The KCl concentration for optimal suGF1 DNA-binding was determined by densitometry (see section 2.23) and found to be 175 mM for both preparations. Although the general dependence of the amount of suGF1-DNA complexes on KCl concentration was the same for nuclear extracts and purified suGF1, suGF1-DNA complexes were more prevalent in the purified preparation relative to those in the nuclear extract at 50 and 100 mM KCl. This can most simply be ascribed to the presence of nuclear extract factors with lower binding affinity or specificity than suGF1 to the suGF1-binding site in the E/H fragment, thus competing more effectively for binding to the E/H probe at low ionic strength than at high ionic strengths where such factor-DNA complexes should exhibit decreased salt-stability. The factor-DNA complex denoted by C in the 175 mM KCl trace of the nuclear extract is thought to contain such a factor.

In the nuclear extract incubations complex C is almost as abundant as the suGF1 complexes at low ionic strength (100 mM) and seems to be competing for the suGF1-binding site, since at least twice the amount of suGF1 complexes are formed with the purified suGF1 at the same ionic strength. At high ionic strength (300 mM), the competition by the presumed lower affinity or specificity factor C is greatly decreased and the amount of suGF1-DNA complexes are comparable in the pure and crude fractions. This observation demonstrates the importance of assaying relative binding activity during purification at ionic strengths where the relative signal of the desired protein is high so as to reduce the effects of multiple

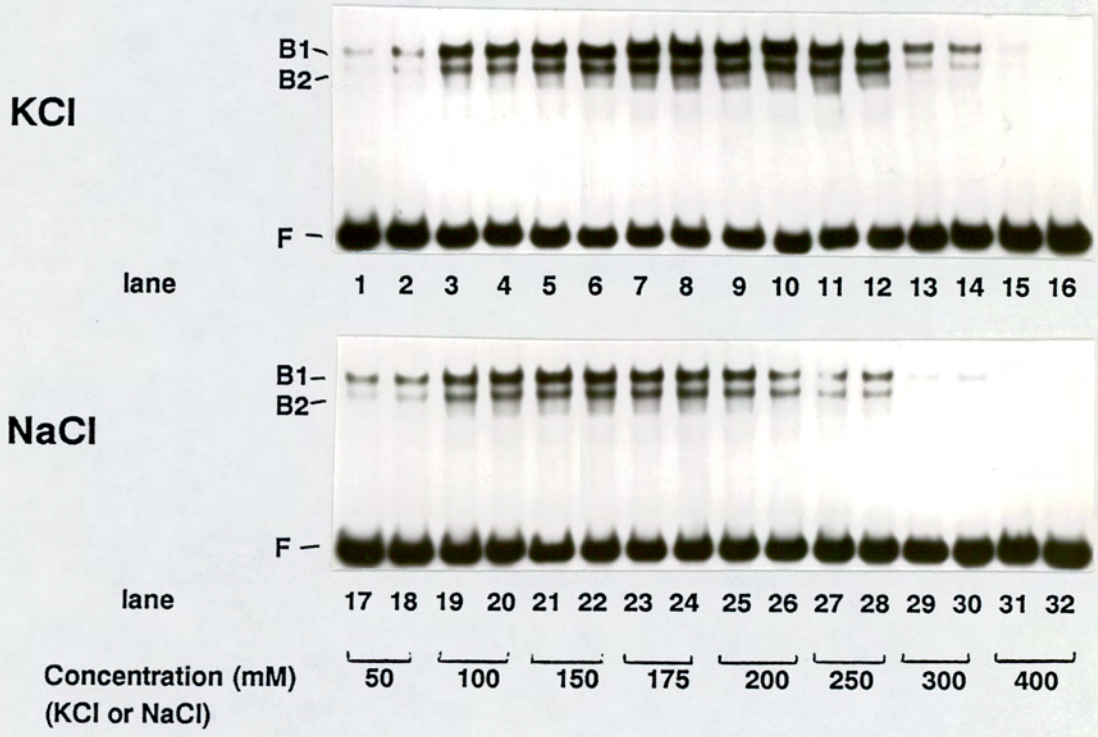


Figure 5.1 suGF1 DNA-binding is dependent on ionic strength

End-labeled 335 bp E/H fragment (1 ng) was incubated with an aliquot of purified suGF1 in standard EMSA incubation buffer (see section 2.8), at final KCl (top panel) or NaCl (bottom panel) concentrations (mM) as indicated. Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). F is free labeled DNA probe. B1 and B2 are suGF1-DNA complexes.

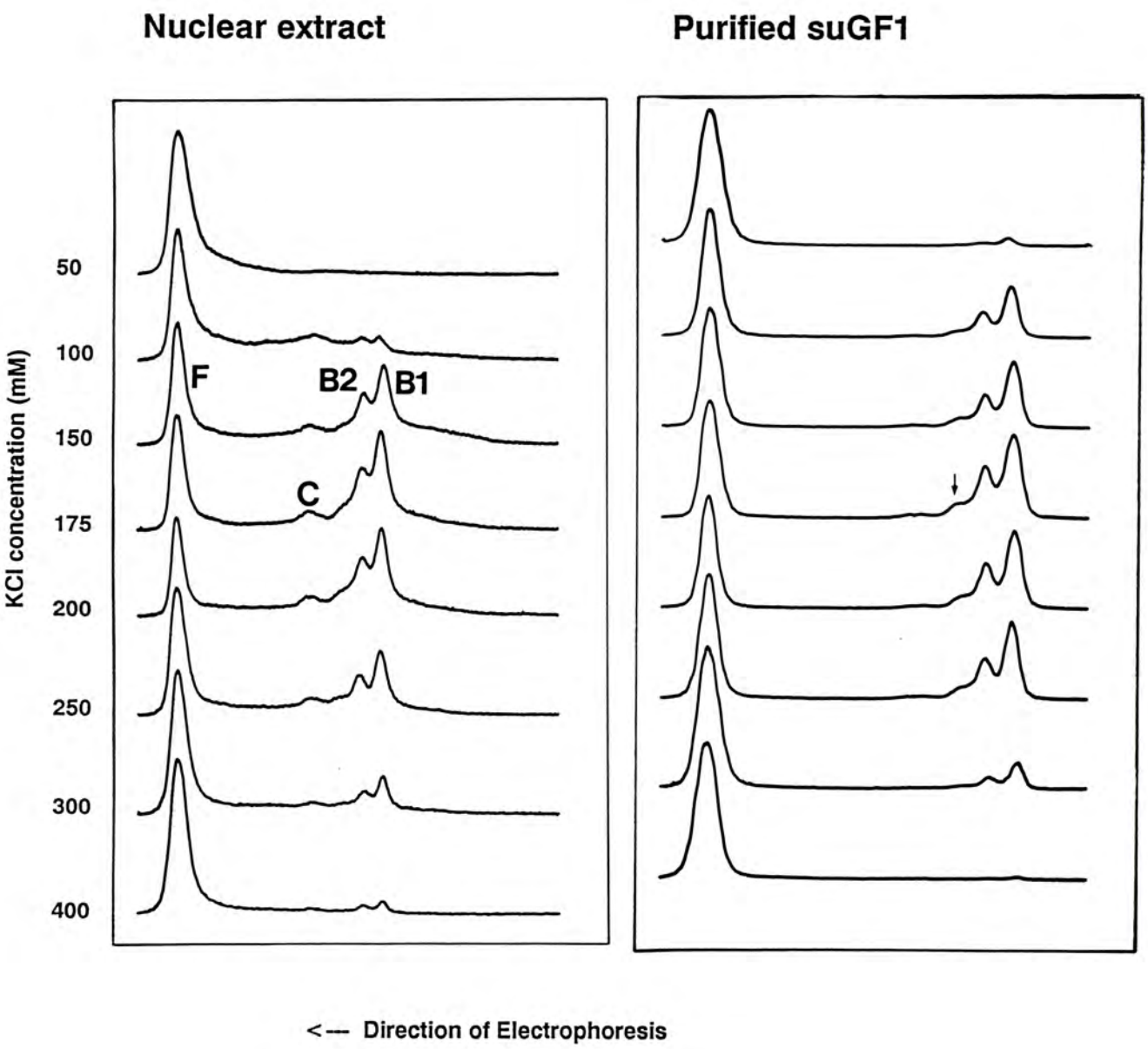


Figure 5.2 Dependence on ionic strength for suGF1 DNA-binding is the same in crude and purified preparations

End-labeled 335 bp E/H fragment (1 ng) was incubated with an appropriate amount of nuclear extract (left panel) or purified suGF1 (right panel) in EMSA incubation buffer (see section 2.8) at final KCl concentrations of 50, 100, 150, 175, 200, 250, 300 or 400 mM as indicated. Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). Lanes on the autoradiographs were scanned (see section 2.23) along the direction of electrophoresis as indicated. F denotes free DNA fragment. B1 and B2 are suGF1-DNA complexes. C is a factor-DNA complex distinct from B1 and B2. The activity denoted by the arrow is thought to be a suGF1 degradation product.

competing factors likely to be present in the nuclear extract.

5.3 Divalent cation requirement for suGF1 DNA-binding

Incubation of fractionated nuclear extracts containing transcription factors H4TF-1 or H4TF-2 in a binding buffer with 40 mM KCl and 1 to 6 mM $MgCl_2$, revealed a great enhancement in DNA-binding of H4TF-1 by raising the $MgCl_2$ concentration to 6 mM, but optimal binding of H4TF-2 at 0 to 1 mM $MgCl_2$ (42). Requirement for Mg^{2+} for suGF1 DNA-binding was investigated in an EMSA with enriched suGF1, which had been extensively dialyzed against 0,1 buffer C containing no $MgCl_2$ or EDTA (see legend to Figure 5.3). The autoradiograph of the EMSA gel is shown in Figure 5.3.

The ionic strength[9] of the binding incubation was increased from 160 (160 mM KCl; Figure 5.3, lanes 2 and 7) to 220 by addition of KCl (lanes 3 to 6) or 1, 5, 10 or 20 mM $MgCl_2$ (lanes 8 to 11) respectively. There was no significant difference in the percentage of DNA shifted when comparing lanes with the same ionic strength (lanes 3 and 8, 4 and 9, 5 and 10, 6 and 11), except for lane 11 (160 mM KCl, 20 mM $MgCl_2$, I=220) which showed a decreased shift as compared to lane 6 (220 mM KCl, 0 $MgCl_2$, I=220), possibly due to aggregation caused by the high Mg^{2+} concentration. This shows that the amount of suGF1 DNA-binding depends on the final ionic strength in the reaction mixture, and is insensitive to supplementation with at least 10 mM of a divalent cation ($MgCl_2$) as opposed to supplementation to the same final ionic strength with a monovalent cation (KCl). Variation of the ratio of KCl to $MgCl_2$ at the same ionic strength (175 mM; lanes 12 to 16), did not cause a variation in percentage binding (as determined by densitometry; see section 2.23). This is in support of the finding that DNA-binding is sensitive to ionic strength rather than the presence of the exogenously added divalent cation Mg^{2+}, except at very high concentrations of $MgCl_2$ (20 mM; lane 16). If $MgCl_2$ is required for suGF1-DNA binding it is thus only needed in trace amounts, which might be removed by dialysis against a divalent cation chelator.

The appearance of higher order complexes at ionic strengths of 160 and 163 (B3 and B4 in lanes 2, 3, 7 and 8 of Figure 5.3) but not at ionic strengths of 175 and higher, might indicate increased stability of suGF1 protein-protein interactions at lower ionic strengths.

The next aspect investigated was the possible Zn^{2+}-dependence of suGF1 DNA-binding. In contrast to H4TF-1 (43) and TFIIIA (83), supplementation of buffers with $ZnCl_2$ caused an inhibition of suGF1 DNA-binding (see section 3.7 and Figure 3.1). It was decided to

[9] Ionic strength (I) was calculated from the relation $0,5[\Sigma(c_i z_i^2)]$ where c_i and z_i denote the concentration and charge of species i, respectively.

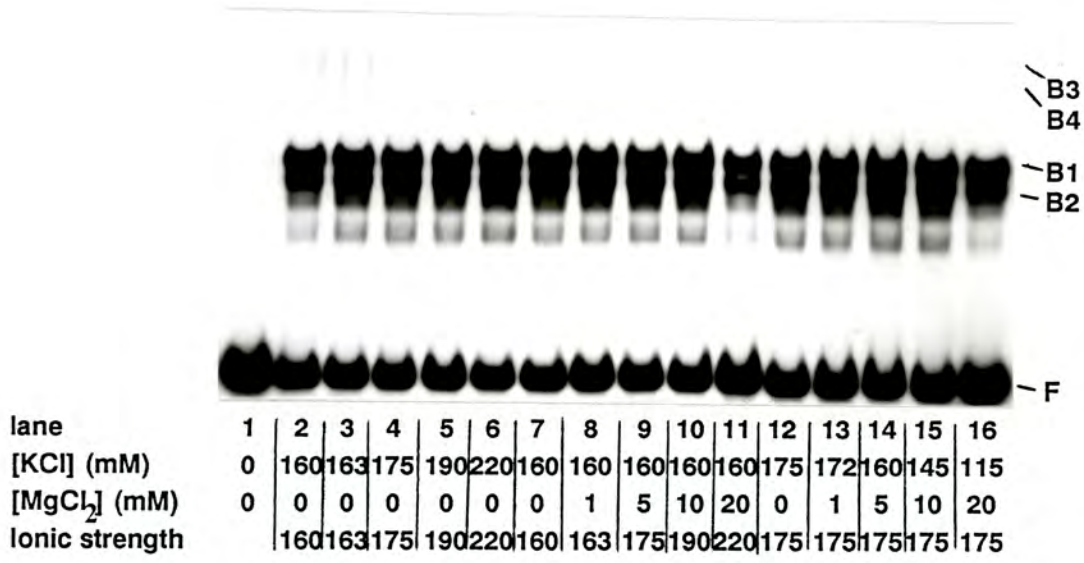


Figure 5.3 suGF1 DNA-binding is insensitive to the concentration of Mg^{2+}

A 2 ml suGF1 preparation which had been obtained by poly(dG).poly(dC)-affinity chromatography of fourteen-hour *P.angulosis* embryonic nuclear extract was extensively dialyzed against 5 changes of 1 litre 0,1 buffer C containing no $MgCl_2$ or EDTA. End-labeled 335 bp E/H fragment (1 ng) was incubated with an appropriate aliquot of dialyzed protein in a standard EMSA (see section 2.8) with final KCl and $MgCl_2$ concentrations and ionic strength (I) as indicated. Ionic strength was calculated as stated in the text. Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). F is free DNA. B1, B2, B3 and B4 are suGF1-DNA complexes.

investigate the effect of 1,10-phenanthroline (OP) on suGF1 DNA-binding. This chelator can be used efficiently in the presence of Mg^{2+} to chelate Zn^{2+} (63, 83) and has been shown to totally abolish DNA-binding for some proteins at low concentrations (0,5 to 2 mM) by metal-chelation as discussed in the introduction of this chapter.

The results of two EMSAs to test the effect of OP on suGF1 DNA-binding are shown in Figure 5.4. Crude (nuclear extract) or purified suGF1 was incubated with increasing concentrations of OP for 5 min at room temperature followed by 5 min on ice in EMSA buffer (see section 2.16.1). This was followed by the addition of end-labeled E/H DNA fragment and nonspecific competitor DNA (pdIdC), to start the DNA-binding reaction. A concentration of 2,3 mM OP (lane 2) did not decrease suGF1 DNA-binding significantly in nuclear extracts, whereas a marked decrease in DNA-binding could be detected at 4,7 mM OP. suGF1 DNA-binding using a purified protein preparation was substantially decreased (but could not be completely abolished) by incubation with OP up to a concentration of 20 mM. Duplicate lanes were however not consistently comparable (see for example lanes 6 and 7). It should be added that this was by no means a reproducible result, as no or only a slight decrease in DNA-binding could be detected in several attempts to repeat this experiment using the same or freshly-prepared solutions and suGF1 preparations. For this reason it was not possible to investigate the cause of the observed inhibition in DNA-binding by OP by for example attempting reconstitution of binding by addition of $ZnCl_2$, within the time constraints of this investigation. It is possible that the inhibition might have been an artifact of the experiment. A precipitate has for example been observed at OP concentrations higher than 1 or 2 mM in the case of the DNA-binding protein NF-κB (237). In the experiments with suGF1 it was shown however that OP could chelate Zn^{2+} under the reactions conditions, as the presence of 2 mM OP in an incubation reaction containing 1 mM $ZnCl_2$ alleviated the normal inhibition of DNA-binding by Zn^{2+} (results not shown). Incubation of binding reactions at room temperature instead of 4°C was not tried. Incubation of suGF1 with OP (10 mM) followed by the removal of possible Zn^{2+}-OP complexes by gel filtration did not result in a loss of DNA-binding activity (results not shown).

Analogous to the experimental procedure followed for Sp1 (109), it was decided to try a different metal-chelator, namely EDTA. Initial experiments showed that incubation of suGF1 with up to 50 mM EDTA, without a dialysis step, followed by addition of DNA and dilution to a final concentration of 34 mM EDTA had no effect on suGF1 DNA-binding in an EMSA (results not shown). In Figure 5.5 it is shown that when purified suGF1 was dialyzed against a buffer containing 50 mM EDTA (see section 2.16.2), most of the DNA-binding activity was abolished (lane 2) in comparison to a control sample which had been dialyzed against buffer lacking 50 mM EDTA (lane 1). These two protein preparations were dialyzed further into a

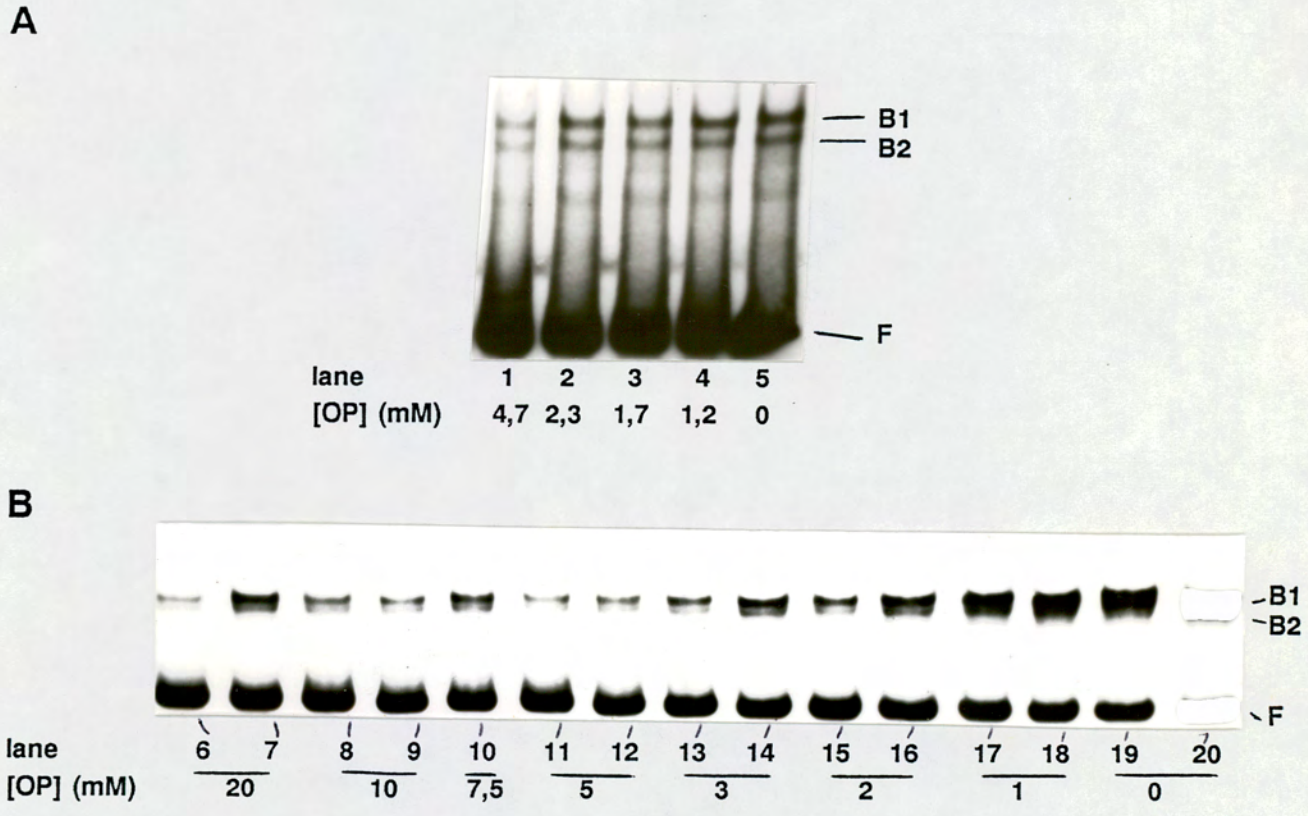


Figure 5.4 Effect of incubation of suGF1 with o-phenanthroline on suGF1 DNA-binding

Aliquots of nuclear extract (A) or purified suGF1 (B) were incubated in EMSA buffer (see section 2.8 and 2.16; 175 mM KCl) in the presence of OP from a 250 mM stock in ethanol (lanes 1 to 4 and 6 to 18) or ethanol as control (lanes 5, 19 and 20) for 5 min at room temperature and 5 min on ice. This was followed by the addition of 1 ng end-labeled E/H fragment and 0,5 µg pdIdC in a volume of 5 µl (final volume 25 µl). Reaction mixtures were incubated on ice for 30 min, and electrophoresed and detected in 4% polyacrylamide gels as described (see sections 2.8 and 2.22). Final OP concentrations in reaction mixtures were as indicated by [OP] in mM. F is free DNA. B1 and B2 are suGF1-DNA complexes.

Figure 5.5 suGF1 DNA-binding activity is removed by dialysis against EDTA and cannot be reconstituted by addition of Mg^{2+} or Zn^{2+}

An aliquot (440 µl) of dilute purified suGF1 was adjusted to 1 ml as described in section 2.16.2 and dialyzed against 0,1 buffer C containing 50 mM EDTA, no $MgCl_2$ and 0,1% (v/v) NP-40. This sample is indicated by S1. A control sample (C1) was dialyzed against 0,1 buffer C containing 0,1% (v/v) NP-40. Sample S1 was then dialyzed against 0,1 buffer C containing 0,2 mM EDTA, no $MgCl_2$ and 0,1% (v/v) NP-40 to yield a protein preparation referred to as sample S2. The control sample C1 was dialyzed against the same buffer containing 2 mM $MgCl_2$ to yield control sample C2. Aliquots (5 µl) of C1 (lane 1), S1 (lane 2), C2 (lane 3) or S2 (lanes 4 to 16) were incubated with 1 ng end-labeled E/H fragment in a modified EMSA binding buffer (16 mM Hepes, pH 8,0, 16% (v/v) glycerol, 0,8 mM DTT, 0,4 mM PMSF, 0,02% (v/v) NP-40 and 175 mM KCl) as described in section 2.8. The binding buffers were supplemented with $ZnCl_2$ and $MgCl_2$ to final concentrations as indicated by $[Zn^{2+}]$ and $[Mg^{2+}]$ in concentration units of µM and mM respectively. The final EDTA concentration in the EMSA was 10 mM in lane 2 and 0,04 mM in all other lanes. Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). F is free DNA. B1, B2, B3 and B4 are suGF1-DNA complexes.

buffer containing either 0,2 mM EDTA (sample S2) or 2 mM $MgCl_2$ without EDTA (control sample C2). This was done to allow for subsequent reconstitution experiments with low concentrations of Zn^{2+}. suGF1 DNA-binding in the control sample (C2; lane 3) without EDTA, was once again not affected by the dialysis step, whereas the sample dialyzed into 50 mM EDTA and then into 0,2 mM EDTA, still exhibited minimal binding activity (lane 4). This strongly suggested that the EDTA-treatment had resulted in the removal of divalent cations essential to suGF1 DNA-binding, as was found in the case of Sp1 (109). Reconstitution of DNA-binding activity could unfortunately not be accomplished by the addition of $ZnCl_2$ (lanes 5 to 12) or $MgCl_2$ (lanes 13 to 16) at the indicated concentrations. $ZnCl_2$ was found to inhibit rather than enhance DNA-binding as the concentration was increased (see for example lanes 11 and 12). The inability to reconstitute binding may be due to the fact that the dialysis experiments were not carried out under anaerobic conditions, even though a reducing environment was present (see section 5.1).

Although the results presented above do not prove that suGF1 requires a divalent metal for example Zn^{2+} for DNA-binding, they strongly suggest that binding might indeed be dependent on trace amounts of divalent cations which may be tightly bound to the protein. The results show certain similarities with those obtained for Sp1, which is known to contain three Zn fingers (109) : Both Sp1 and suGF1 DNA-binding activities are unaffected by 1 mM OP (109; Figure 5.4). Sp1 DNA-binding activity as assayed by DNase I footprinting is totally abolished by dialysis against 50 mM EDTA at low pH followed by dialysis and dilution to 0,2 mM EDTA (109). Similarly, almost all suGF1 DNA-binding activity can be removed by dialysis against 50 mM EDTA at pH 8, dialysis into 0,2 mM EDTA and dilution to 0,04 mM EDTA in the EMSA binding reaction (Figure 5.5). Only 10% to 20% of Sp1 DNA-binding activity can be restored by addition of 500 µM $ZnCl_2$ to the EDTA-treated sample containing 0,2 mM EDTA (109). Addition of $ZnCl_2$ up to 500 µM to the EDTA-treated suGF1 samples containing 0,04 mM EDTA did not restore DNA-binding activity (Figure 5.5). suGF1 may be more sensitive to inactivation of a metal binding site upon removal of divalent cations with EDTA than Sp1 (see section 5.1). It is also possible that suGF1 has a requirement for some other divalent cation other than Zn^{2+}. Time did not allow further experiments to be carried out.

These results may serve as an indication that suGF1 is distinct from certain other factors binding specifically to oligo(dG).oligo(dC), or closely related DNA sequences, such as BGP1 (see chapter 7).

CHAPTER 6

INVESTIGATION OF suGF1 DNA-BINDING BY CHEMICAL AND NUCLEASE PROBING

6.1 Introduction

Nuclease and chemical protection (footprinting) and interference experiments are widely used in the study of DNA-protein interactions (see section 1.4). The binding of purified suGF1 to a G-rich sequence in the H1-H4 intergenic region of the early histone gene battery of *P. miliaris* was investigated by DNase I and hydroxyl radical footprinting and methylation interference. These three probes were chosen as their sizes and DNA sequence- or structural specificities compliment each other and provide information on different structural features of the protein-DNA interaction.

In the presence of Ca^{2+} and Mg^{2+}, DNase I (MW 30,4 kDa) binds in the minor groove of double-stranded DNA, contacting the phosphate backbone on both strands, and cleaves the O3'-P bond of only one strand (53, 129, 171, 216). The structure of a DNase I-DNA complex solved by X-ray crystallography revealed that DNase I induces bending of the DNA towards the major groove and widening of the minor groove across which it binds (129, 171, 216). The sequence-dependent manner in which DNase I cleaves DNA (55, 132) is thought to arise mainly from the variation in minor groove width and bendability of the DNA (129). A decrease in the extent of DNase I cleavage at a specific position in the DNA in the presence of bound protein does not necessarily indicate direct protein-DNA contact, but may result from steric hinderance of the DNase I by the bound protein because of the large size of DNase I. Since the DNase I molecule extends across the adjacent major groove and approaches the backbone approximately 5 bp to the 3' side of the scissile bond, occupation of the major groove by a DNA-binding ligand is also expected to sterically interfere with DNase I cleavage (129). This is depicted schematically in Figure 6.1. In this figure, a structure located in the major groove above a potential scissile bond on the Watson strand, could interfere sterically with binding of DNase I. Similarly, occupation of the major groove below the potential scissile bond in the Crick strand would interfere with DNase I binding.

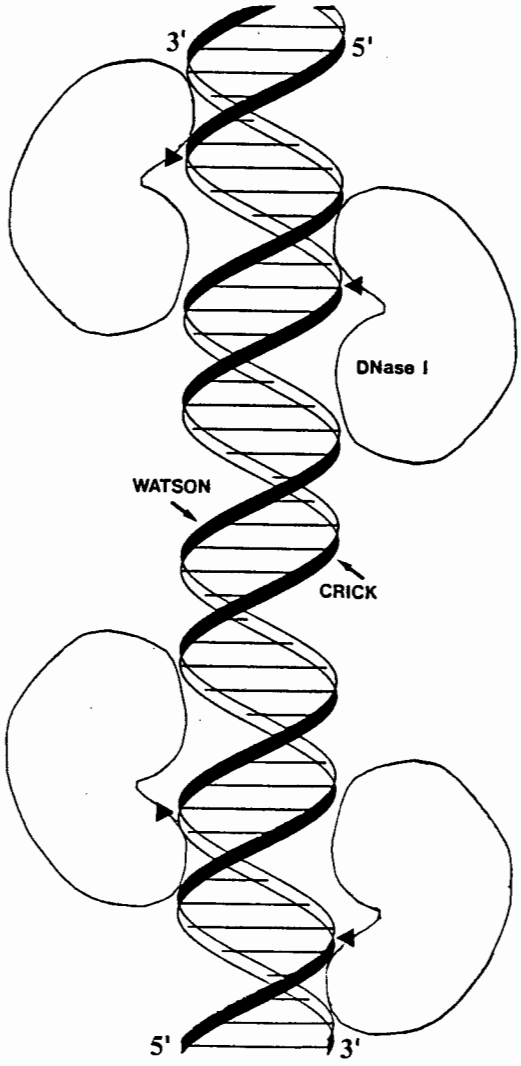


Figure 6.1 Schematic representation of the DNase I-DNA complex formed to cleave the Watson and Crick strands respectively

The sugar phosphate backbones of a double-stranded DNA helix are schematically represented by ribbons. Horizontal lines represent base-pairs. The "top" and "bottom" strands are (arbitrarily) denoted as the Watson and Crick strands respectively. The polarity of the strands is indicated at the top and bottom of the helix. Two bound DNase I molecules are shown schematically on either side to demonstrate steric hinderance encountered by DNase I when cleaving the Watson (bottom molecules on either side) and Crick (top molecules on either side) strands respectively (129). In each case, the scissile bond is indicated by a closed triangle.

Hydroxyl radicals, produced by the reduction of hydrogen peroxide by iron(II)EDTA, cleave DNA by abstracting hydrogen atoms from deoxyribose residues, causing the resulting sugar radicals to decompose and leave gaps in the DNA molecule (51, 223, 224). Although both DNase I and hydroxyl radicals approach and cleave the DNA from the minor groove side, the small size and lack of DNA-binding of the hydroxyl radicals provide information on protein-DNA contacts at every position in the sequence without complications of long range steric hinderance. A change in the extent of hydroxyl radical cleavage in the presence of bound protein either indicates very close proximity of protein to the phosphate backbone, or distortions in the local geometry of the DNA molecule. It is important to note that the absence of hydroxyl radical protection does not discount the possibility of protein-DNA contacts located in the major groove (223). Although hydroxyl radicals cleave DNA with low sequence-specificity, the extent of cleavage has been observed to decrease at positions where the minor groove is narrow (25). In the case of curved DNA, for instance, a periodic pattern of enhanced protection and cleavage with a period corresponding to the helical repeat is seen (25). Because of the small size of hydroxyl radicals, the differences in extents of cleavage caused by protein binding are often subtle. Careful analysis of the cutting data in the form of difference probability plots is therefore essential (see below).

Dimethylsulfate can be used to methylate G's at the N-7 position projecting into the major groove. Methylation interference of a specific G is therefore usually taken as an indication that the protein contacts the specific G from the major groove side. Methylation interference is limited to delineating major groove contacts to only G's (and in some cases to adenine residues). The combination of this major groove contact interference experiment with the minor groove footprinting at high (hydroxyl radical) and low (DNase I) resolution, can be used to construct a model for the interaction of suGF1 with a G-rich binding site (see chapter 7).

Footprinting experiments always require the comparison of cleavage in the absence and presence of the pertinent protein. Such comparisons can best be accomplished by the construction of difference probability plots (36, 53, 55, 62, 145, 190, 191; see section 2.24). Since the ln(probability cleavage) of the DNA in the absence of protein (free DNA) is subtracted from that in the presence of protein (bound DNA), the high background of sequence- or structural-specific cleavage of a probe can be corrected for in such plots. This enables direct detection of the effects on cleavage of the DNA by a probe, solely due to the bound protein. In addition, calculation of ln(difference probabilities) of cleavage corrects for the presence of secondary cleavage of the DNA molecule between the radiolabeled end and the primary site of cleavage (see section 2.24).

Two general strategies can be used in footprinting experiments to obtain the free and bound cleaved DNA populations required for the construction of the difference probability plots. Firstly, the DNA can be incubated in the absence or presence of saturating amounts of protein, subjected to digestion, recovered, and directly analyzed on a sequencing gel. It is, however, often advantageous to rather separate free and bound DNA in an EMSA after incubation and limited cleavage has been performed. This often results in a lower background of contaminating free cleavage products in the bound population (see for example reference 51). In addition, saturating concentrations of protein (which may lead to artifacts), are not required in the incubation in the event of separation of free and bound complexes by EMSA. A further important advantage of this strategy, is that bound DNA from individual complexes can be analyzed after separation and recovery from gel slices.

When isolating bound cleaved DNA from an EMSA gel in a footprinting experiment, the population of free cleaved DNA is usually also recovered from the gel, in order to correct for any possible effect of the EMSA electrophoresis step on the DNA. Such recovered free DNA may sometimes be expected to be contaminated with a small percentage of bound DNA, due to dissociation of the bound complexes, or exchange of protein between DNA originally present in the bound and free populations at the time of cleavage. Such possibilities were kept in mind in the experiments presented in this chapter (see sections 6.2.2 and 6.4.1).

As opposed to footprinting (protection) experiments, separation of bound and free DNA is essential in interference experiments (for example methylation interference). This is because the DNA is modified prior to incubation with protein. The protein then selects for a population of DNA (bound) which is modified at positions not interfering with binding, and selects against the interfering population (free). In this case the bound and free populations can however not be compared, since both these populations are enriched (free) or depleted (bound) in the selection of DNA molecules interfering with protein binding. Both bound and free populations are thus compared to the population of DNA which had not been subjected to incubation with protein. This "unselected" population is referred to as the "naked" DNA (see section 6.3).

In this investigation, methylation interference could be detected for suGF1-DNA complexes which had been incubated at 37°C, but not after incubation at 4°C (see section 6.3). suGF1-DNA incubations for DNase I and hydroxyl radical footprinting were therefore carried out at 4 and 37°C to determine if the structure of the suGF1-DNA complex differered significantly between these two temperatures.

6.2 DNase I footprinting of the suGF1-DNA complex

6.2.1 Characterization of the DNase I footprint of the suGF1-DNA complex

DNA fragments containing part of the H1-H4 intergenic region of the early histone gene battery of *P. miliaris* (see section 2.4) were 3' end-labeled on the Watson or Crick strands, incubated in the absence or presence of purified suGF1 for 30 min at 4°C, and subjected to limited DNase I digestion as described in section 2.18.1. The DNA was recovered from the solutions and electrophoresed on sequencing gels (see sections 2.17, 2.18.1 and 2.22). Autoradiographs of DNase I nicking in the Watson and Crick strands of the duplex DNA are shown in Figure 4.12 and 6.2 respectively. Comparison of lanes containing DNA which had been incubated in the absence and presence of suGF1 show a clear "window" (footprint) where DNase I digestion was hindered by bound suGF1. The footprint occurred over the same region of the H1-H4 intergenic region in the Watson and Crick strands (compare lanes 4 and 2 of Figure 4.12 and lanes 1 and 12 of Figure 6.2).

In order to compensate for the sequence specificity of DNase I cleavage, the cutting data was interpreted as difference probabilities of cleavage of free and protein-associated DNA. The difference probability plots of the DNase I footprints on the Watson and Crick strands are shown in Figure 6.3 (see section 2.24). In these plots a ln(difference probability) value of 0 (base-line) indicates that the probability of DNase I cleavage is the same in the absence and presence of protein at the specific sequence position. Negative and positive differences in the ln(difference probability) of cleavage indicate protection and enhanced reactivity (or hypersensitivity), respectively. Small fluctuations from the base-line can be disregarded as experimental error. The products of DNase I digestion carry a 5' phosphate, thus migrating in phase with the G-sequencing standard lane (55, 151). The regions protected from DNase I digestion by suGF1 are clearly visible as minima[10] in Figure 6.3. The borders of the footprint can be assigned to sequence position -345 and -317 on the Watson strand and positions -346 and -315 on the Crick strand. suGF1 thus protects 29 and 32 bp on the Watson and Crick strands respectively, although the actual region of proximity of the protein to the DNA is expected to be a few bp shorter on either side, due to steric hinderance caused by the large size of DNase I (see section 6.1).

An interesting feature of the DNase I footprint on the Watson strand is a modulation in the ln(difference probability) of cleavage (local maxima and minima) within the borders of the

[10] A minimum refers to a negative value for ln(difference probability).

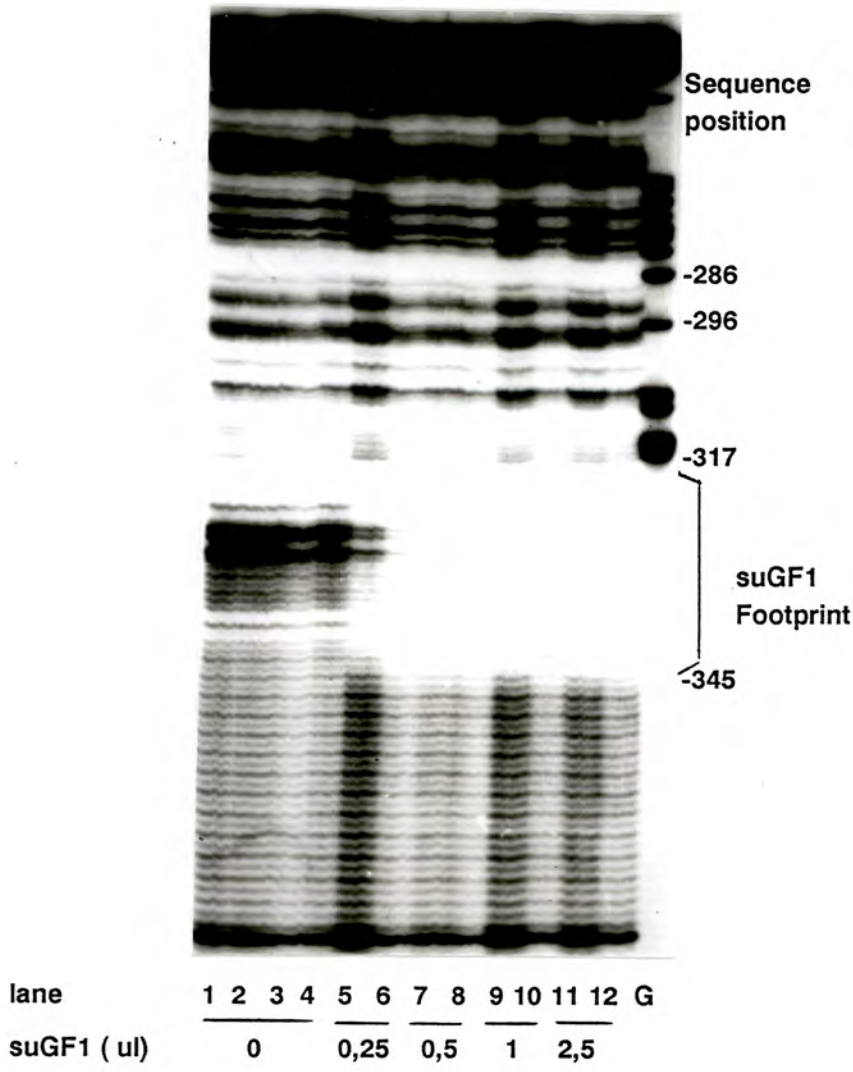


Figure 6.2 DNase I footprint of suGF1 on the Crick strand of the P. miliaris H1-H4 intergenic region

End-labeled A/X fragment (1 ng) was incubated in the absence (lanes 1 to 4) or presence (lanes 5 to 12) of aliquots of purified suGF1 (volume (µl) as indicated) and subjected to limited DNase I digestion as described in section 2.18.1. DNase I stock solution was diluted 1/40 (lanes 1, 2, 5, 7, 9 and 11) or 1/80 (lanes 3, 4, 6, 8, 10 and 12) in DNase I dilution buffer (see section 2.18.1). DNA was isolated, electrophoresed on a sequencing gel together with a Maxam-Gilbert G-sequencing standard (G) and autoradiographed (see sections 2.17, 2.18.1, 2.21 and 2.22). The suGF1 footprint is indicated by a bracket. Sequence positions are indicated (see Figure 2.1 for the sequence).

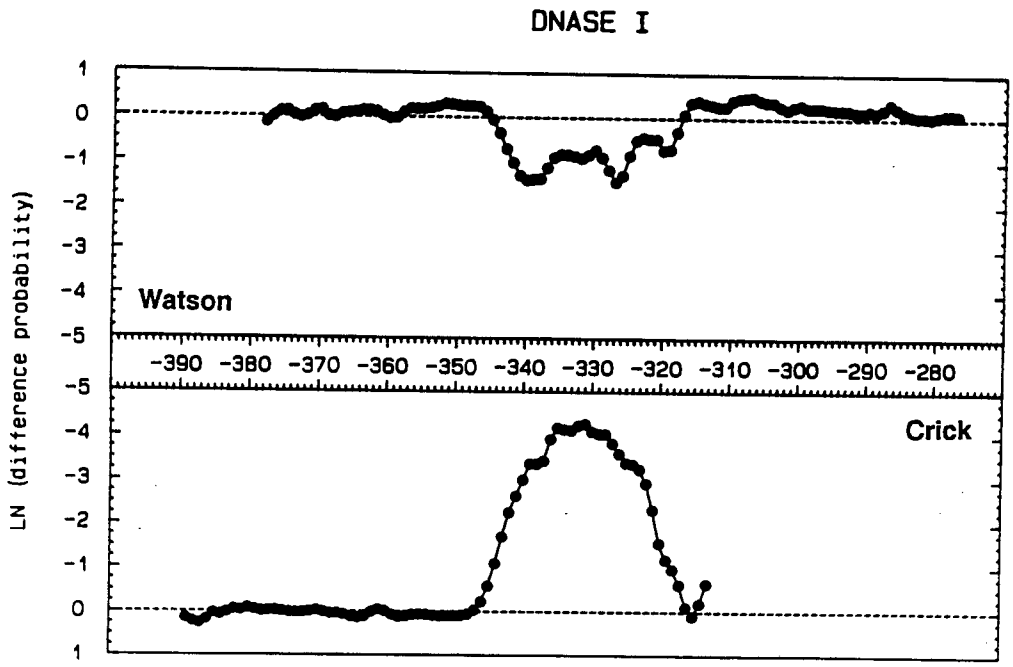


Figure 6.3 Difference probability plot of the suGF1-DNA complex probed with DNase I at 4°C

The natural logarithm of the probability of cleavage by DNase I at 4°C in the absence of suGF1 was subtracted from that in the presence of suGF1, calculated as a three point running average at each position in the sequence indicated by a datum point (filled circle) (see section 2.24). The resulting difference probability plots for the Watson and Crick strands are shown in the top and bottom panels respectively. The individual lanes which were scanned in order to calculate the difference probability plots were lanes 2 and 4 shown in Figure 4.12 (Watson strand) and lanes 2 and 12 shown in Figure 6.1 (Crick strand).

footprint (global minimum). Local maxima are present at positions -339, -327 and -319, while local minima are observed at -323 and between positions -335 to -330. When a DNA double helix is adsorbed onto a flat surface, the minor groove is most accessible to DNase I binding and cleavage when orientated on the opposite side of the helix from the side contacting the surface (192). This leads to a periodic pattern of maximum cleavage and protection sites following each other at approximately 5 bp intervals (half the helical repeat of the helix in question). The local maxima and minima in the DNase I difference probability plot of the Watson strand appear approximately in phase with the helical repeat of a DNA double helix (10 to 11 bp), and are thus consistent with the protein-DNA complex being more accessible to DNase I from one side of the helix and more protected by suGF1 on the other. A similar modulation can however not be detected on the Crick strand. The degree of protection on the Crick strand appears to be greater than that on the Watson strand. These results will be discussed further in chapter 7 where a model for the suGF1-DNA interaction is proposed.

6.2.2 DNase I footprinting of complexes B1 and B2

In chapter 4 it was shown that two bands (B1 and B2) representing specific protein-DNA complexes were obtained in EMSAs with purified suGF1. In addition, titration with increasing concentrations of purified suGF1 did not lead to a decrease in complex B2 and simultaneous increase in complex B1, but rather resulted in the disappearance of both complexes B1 and B2 in conjunction with the appearance of a slower migrating doublet (bands B3 and B4). This doublet was in turn replaced by even slower migrating doublets at higher concentrations of suGF1 (see Figure 4.8). These results seemed difficult to reconcile with B1 and B2 being a dimer and monomer of suGF1 (see chapter 7). It was also shown that the two specific complexes B1 and B2 could be obtained by renaturation of the 59,5 kDa band (suGF1) from a SDS-PAGE gel. To obtain further proof that complexes B1 and B2 both contain the same binding activity to the G-rich sequence in the H1-H4 intergenic region defined to be caused by suGF1, DNase I digested complexes B1 and B2 were separated in an EMSA gel, excised, and the DNA from each complex recovered and resolved in a sequencing gel (see section 2.18.2).

Prior to presenting the result of the experiment detailed above, it should be noted that the degree of dissociation or redistribution of suGF1-DNA complexes did not seem significant during the period of DNase I digestion and gel loading. This was shown in an EMSA where suGF1 was incubated with labeled 335 bp E/H fragment for 30 min at 37°C. After the incubation period, the samples were adjusted to the same volume and composition as for DNase I digestions and supplemented with an aliquot of labeled 216 bp A/X fragment (5:1 (mole : mole) E/H fragment to A/X fragment). From Figure 6.4 it can be seen that no

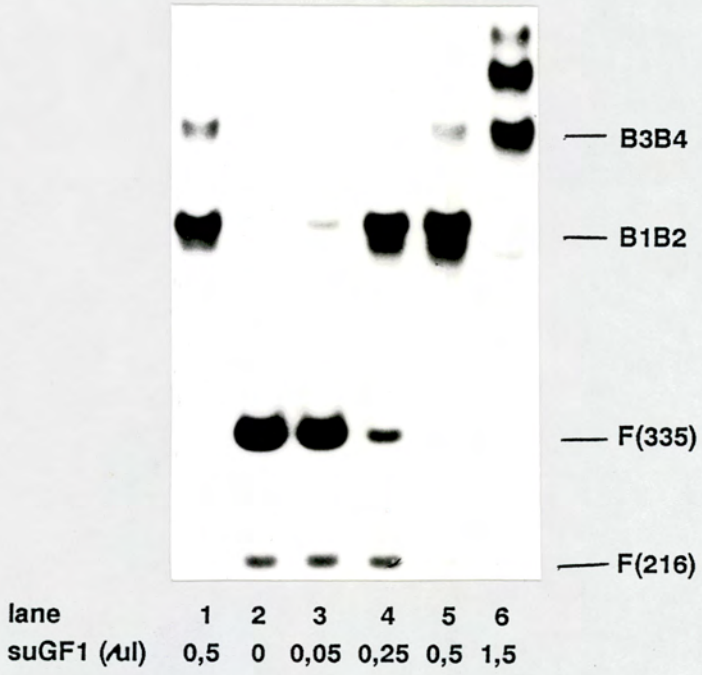


Figure 6.4 Significant suGF1-DNA dissociation does not take place during DNase I digestion and gel loading

An aliquot of E/H fragment (13,5 fmol) was incubated without (lane 2) or with (lanes 1, 3, 4, 5 and 6) purified suGF1 (volume (µl) as indicated) for 30 min at 37°C in standard EMSA buffer (see section 2.8; 175 mM KCl and 3 µg pdIdC per incubation) in a volume of 50 µl. The incubations were placed on ice for 2 min and adjusted to 15 mM $MgCl_2$ and 15 mM $CaCl_2$. DNase I dilution buffer (3 µl) and DNase I stop solution B (nondenaturing; 8,4 µl; see section 2.18.2) containing 2,7 fmol 216 bp A/X fragment was added to the incubations. Electrophoresis and autoradiography were as described (see sections 2.8 and 2.22). F is free DNA fragment. B1 to B4 are suGF1-DNA complexes. The length (bp) of the free radiolabeled DNA fragments are indicated in brackets.

exchange of suGF1 occurred between prebound suGF1-E/H fragment complexes and free A/X fragment added just before loading the gel. In lanes 2 and 4 the 335 bp fragment was pre-incubated with 0 and 0,25 µl suGF1 respectively, before addition of the 216 bp fragment. If redistribution of suGF1 took place between the labeled species, it would be expected that the 335 and 216 bp probes would decrease by the same percentage when comparing lanes 2 and 4. Instead, no visible decrease in the amount of free 216 bp fragment is observed, and at least 70% of the 335 bp fragment is bound by suGF1. In lanes 5 and 6 where the amount of 335 bp fragment was limiting in the pre-incubation with suGF1, the 216 bp fragment is bound by excess suGF1 during the period of time required for DNase I digestion and gel loading. These results implied a fast association and slow dissociation rate for suGF1 DNA-binding. It thus seemed that the DNA recovered from a gel slice containing a certain complex truly represented the DNA population which was found in that complex at the time of nuclease digestion.

Returning to the DNase I footprints of complexes B1 and B2 (shown in Figure 6.5), protection of the same G-rich sequence of the H1-H4 intergenic region in both complexes is clearly visible. The same protein (suGF1) thus appeared to be involved in complexes B1 and B2. The different electrophoretic mobilities of B1 and B2 might be caused by posttranslational modifications of suGF1, or a truncated form of the protein being present (see chapter 7). It is also strictly possible that the conformation of the DNA and/or suGF1 differs in complexes B1 and B2. A difference cannot, however, be detected in the Watson strand of the DNA by DNase I footprinting (Figure 6.5). Hydroxyl radical footprints of both DNA strands shown in subsequent sections also seem to rule out a significant difference in DNA conformation capable of changing the electrophoretic mobility between complexes B1 and B2 (see section 6.4 and chapter 7).

6.2.3 DNase I footprinting of complexes B1B2 and B3B4 at 4 and 37°C

DNase I footprinting of the slower migrating complexes B3 and B4 was carried out to prove that these complexes contained suGF1. The results of the preparative EMSAs following DNase I digestion are shown in Figure 6.6 (see section 2.18.2). Unbound DNA and protein-DNA complexes B1B2 and B3B4 were detected by wet autoradiography, and excised from the gel as described in section 2.22. DNA was recovered from the gel slices, resolved in a sequencing gel and autoradiographed (see sections 2.18.2, 2.21 and 2.22). No difference in the extent of the DNase I footprint produced in the lower (L; B1B2) or upper (U; B3B4) set of complexes could be detected (see Figure 6.7). The slower migrating doublet B3B4 thus also contains bound suGF1. The formation of complexes B3 and B4 from complexes B1 and B2 can most likely be ascribed to protein-protein (suGF1-suGF1) contacts, since no protection of the

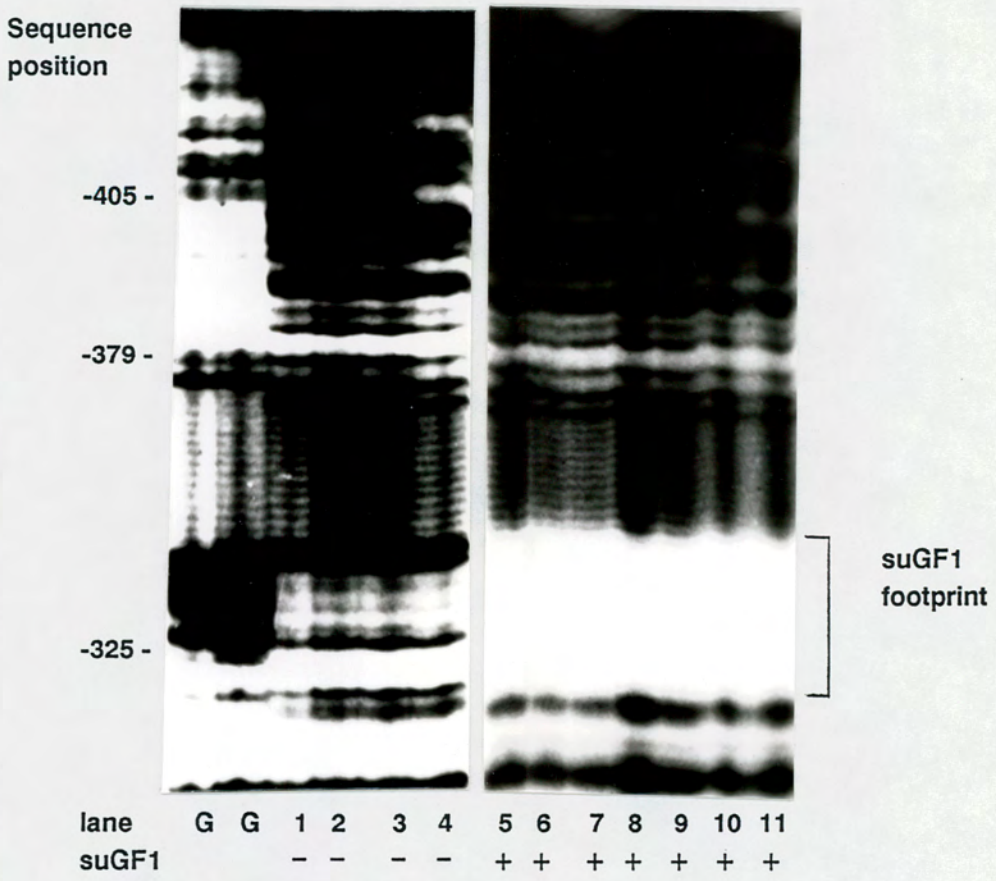


Figure 6.5 DNase I footprint on the Watson strand of suGF1-DNA complexes B1 and B2

Aliquots of end-labeled E/H fragment (6 ng) were incubated with 0,6 µl purified suGF1 and 6 µg pdIdC in EMSA incubation buffer (see section 2.18.2; 153 mM KCl and 72 µg/ml BSA) for 30 min at 4°C, subjected to limited DNase I digestion as described in section 2.18.2 and electrophoresed on 4% nondenaturing polyacrylamide gels (see section 2.8). suGF1-DNA complexes B1 (lanes 9 to 11) and B2 (lanes 5 to 8) were excised separately after wet autoradiography of the gel. DNA was recovered and electrophoresed on a sequencing gel together with a Maxam-Gilbert G-sequencing standard (lanes marked G) and DNA digested with DNase I in the absence of suGF1 (lanes 1 to 4; see section 2.18.1) (see sections 2.17 and 2.18.2). The DNase I stock solution had been diluted 1/160 (lane 5), 1/80 (lanes 6 and 9), 1/40 (lanes 7 and 10) and 1/20 (lanes 8 and 11) (see section 2.18.1). The suGF1 footprint is indicated by a bracket. Sequence positions are indicated (see Figure 2.1 for the sequence).

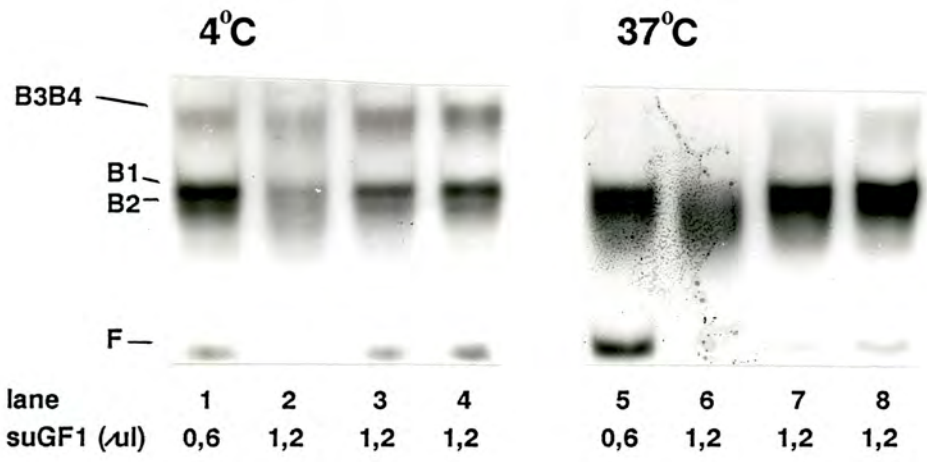


Figure 6.6 Preparative electrophoretic mobility shift assay of DNase I footprinting

Aliquots of E/H fragment (6 ng, 250 000 to 300 000 dpm) were incubated with purified suGF1 (volume (µl) as indicated) for 30 min at 4°C (lanes 1 to 4) or 37°C (lanes 5 to 8), subjected to limited DNase I digestion, electrophoresed in 4% nondenaturing polyacrylamide gels and radioactive species identified by wet autoradiography as described in sections 2.18.2, 2.8 and 2.22. DNase I stock solutions were diluted 1/20 (lanes 2 and 6), 1/40 (lanes 1, 3, 5 and 7) or 1/80 (lanes 4 and 8) (see section 2.18.1). F is free DNA fragment. B1 to B4 are suGF1-DNA complexes.

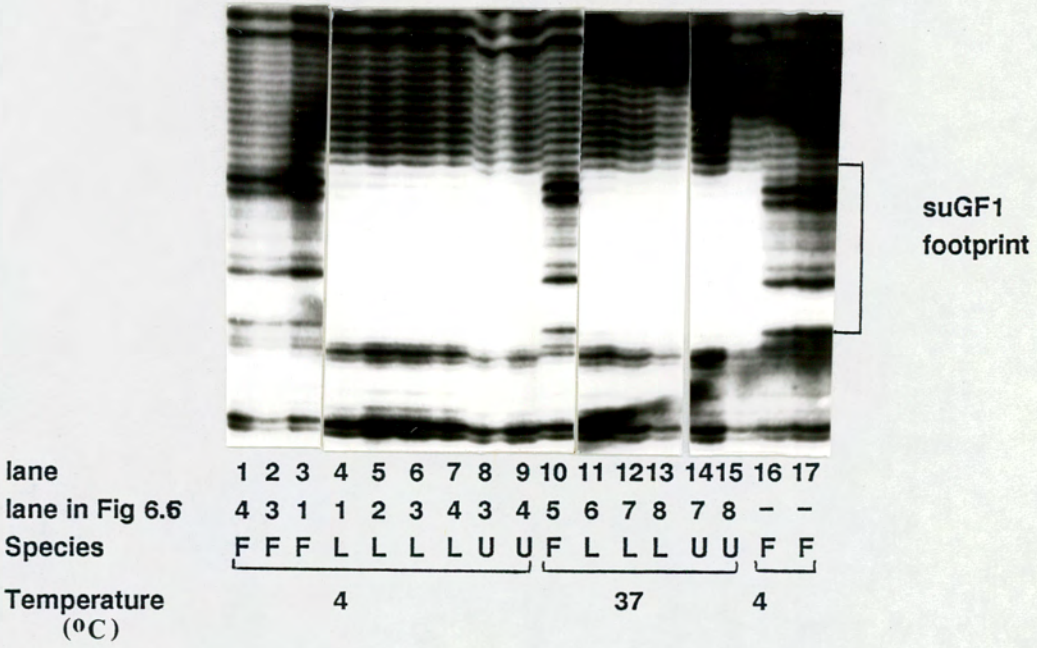


Figure 6.7 DNase I footprint on the Watson strand of suGF1 complexes B1B2 and B3B4 at 4 and 37°C

DNase I digested free DNA (F), and suGF1-DNA complexes B1B2 (denoted by L for lower) and B3B4 (denoted by U for upper) were excised from the indicated lanes in the preparative EMSAs shown in Figure 6.6. DNA was recovered and electrophoresed on a sequencing gel (see sections 2.18.2, 2.21 and 2.22). The temperatures at which the suGF1-DNA incubations were carried out are indicated. The bracket denotes the suGF1 footprint. Lanes 16 and 17 contain DNase I digestion products of DNA in the absence of suGF1 without subsequent isolation *via* a preparative EMSA (see section 2.18.1).

DNA was observed in complexes B3B4 apart from that obtained in the suGF1-DNA footprint of complexes B1B2. No significant differences in the structure of the DNA or in the borders of the footprints were observed in the DNase I digestion products of the Watson strand at 4 and 37°C, for complexes B1B2 or B3B4 (Figure 6.7). From Figure 6.6 it seemed that the ratio of complexes B3B4 to complexes B1B2 were increased at 4°C compared to 37°C. This might be due to increased stability of protein-protein interactions at 4°C. This result was however not reproducibly observed.

6.3 Methylation interference of the suGF1 DNA-binding site

Guanines of the E/H fragment were methylated at the N-7 position by treatment with DMS, as described in section 2.19. The resulting DNA population was methylated on average at one G per strand. Aliquots (12 ng) of the methylated DNA fragments, were incubated at 4 or 37°C with appropriate amounts of purified suGF1, as determined from an analytical EMSA (see section 2.19). The incubations were electrophoresed on a 4% nondenaturing polyacrylamide gel and radiolabeled species detected by wet autoradiography as described in Materials and Methods. The autoradiograph of the preparative EMSA is shown in Figure 6.8.

Gel slices containing free DNA fragments or protein-DNA complexes were excised from the preparative EMSA gel. The methylated DNA fragments were recovered from the gel slices, cleaved with piperidine at positions of methylated G's, lyophilized, and resolved on a sequencing gel (see sections 2.19 and 2.22). A sample of the unselected methylated DNA fragments (naked DNA) was cleaved with piperidine and co-electrophoresed on the gel (see section 6.1).

The autoradiograph of the resolved populations of fragments isolated from the preparative EMSA gel is shown in Figure 6.9(A). From lanes 1 and 2 which represent the unselected population of fragments (naked DNA), it is obvious that certain G's were methylated to a greater extent than others, probably due to sequence-dependent local favourable DNA structure enabling easy access of DMS. In the absence of protein, the extent of methylation of G's in the $(GA)_{16}$ repeat (-371 to -340) was, for example, much less than that of G's in the 11 bp G-string (-339 to -329). The population of fragments obtained from the EMSA gel slices containing free (unbound) DNA is expected to be enriched in fragments cleaved at methylated G's which interfere with suGF1 DNA-binding, whereas the population of fragments obtained from the protein-DNA complexes is expected to be depleted of such fragments.

Visual inspection of Figure 6.9(A) and (B) reveals differences in the populations of fragments only in lanes 3 to 6. Densitometric traces of these lanes and of a lane containing the population

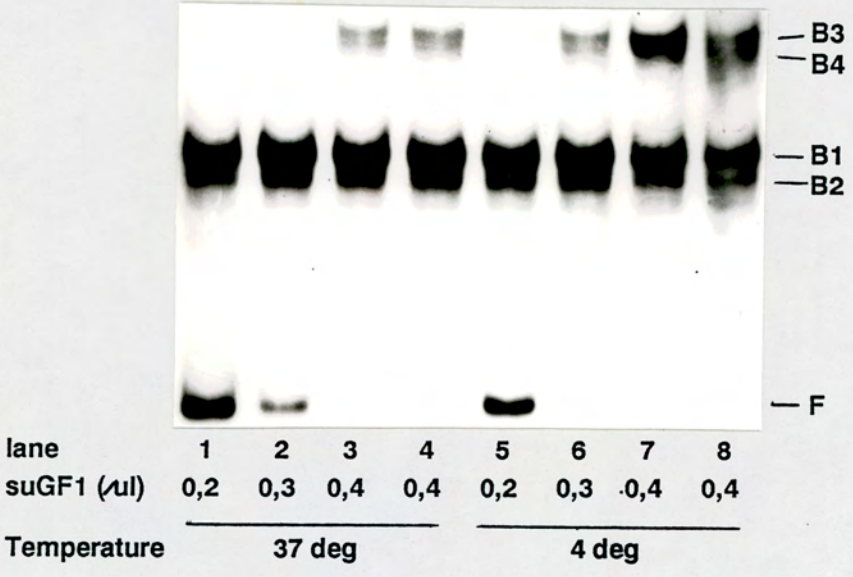
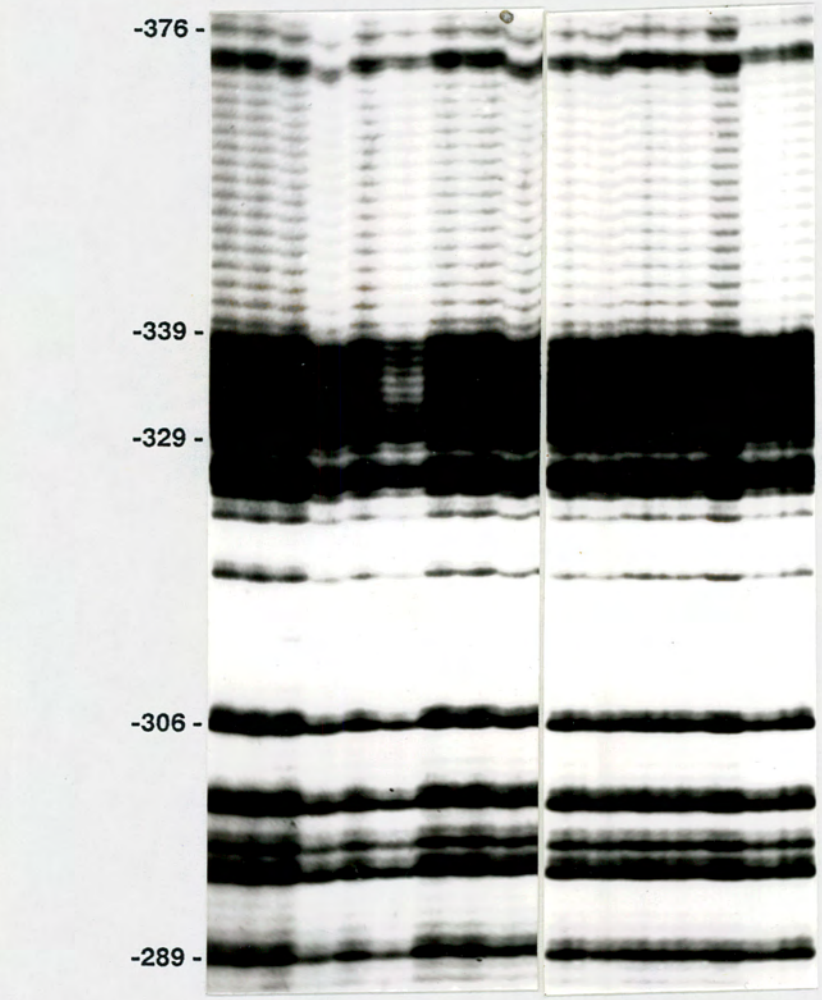


Figure 6.8 Preparative electrophoretic mobility shift assay for methylation interference

End-labeled E/H fragment methylated with DMS (12 ng, approximately 500 000 dpm) was incubated with purified suGF1 (volume (µl) as indicated) in EMSA incubation buffer (see section 2.19; 175 mM KCl, 60 µg/ml BSA and 9 µg pdIdC) for 30 min at 37°C (lanes 1 to 4) or 4°C (lanes 5 to 8). Electrophoresis and wet autoradiography were as described (see sections 2.8 and 2.22). F is free DNA fragment. B1 to B4 are suGF1-DNA complexes.

A


Sequence position
-376 -
-339 -
-329 -
-306 -
-289 -
lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
lane of Fig 6.8 – – 1 2 1 1 2 3 3;4 5 6 5 5 6 7 7;8
Species N N F F B1 B2 L L U F F B1 B2 L L U
Temperature 37 deg 4 deg

B

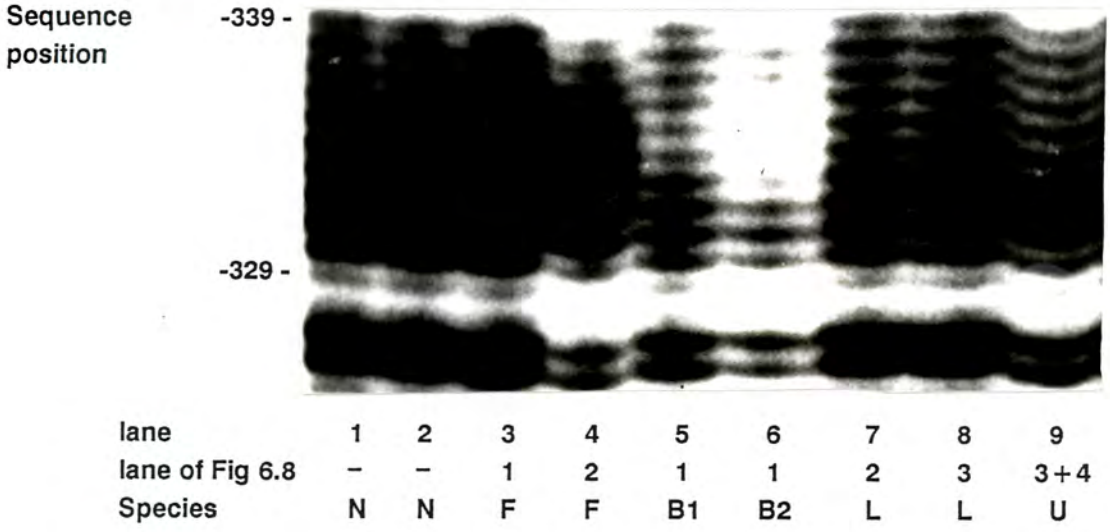


Figure 6.9 Effect of methylation on suGF1 DNA-binding

(A) Gel slices were excised from the indicated lanes of the preparative EMSA gel shown in Figure 6.8. Methylated DNA was recovered from the gel slices containing free (F) methylated DNA, and methylated DNA in protein-DNA complex B1 (B1), complex B2 (B2), complexes B1B2 (L; lower) and B3B4 (U; upper). The DNA was cleaved with piperidine at methylated positions, lyophilized, dissolved in sequencing gel loading dye and electrophoresed on sequencing gels (see sections 2.19 and 2.22). N refers to naked methylated and cleaved E/H fragment (in effect G-sequencing standard) not used in the preparative EMSA. Sequence positions are indicated (see Figure 2.1 for the sequence).

(B) Detail of the methylation interference over the 11 contiguous G's is shown in an enlargement of the first 10 lanes on the sequencing gel of (A).

of naked DNA fragments are shown in Figure 6.10. From these figures it is clear that fragments resulting from methylation of the central G's of the G-string (sequence positions -333, -334, -335 and -336) are more abundant in the population of free fragments (lanes 3 and 4 of Figure 6.9) than in the naked DNA, but depleted of the populations of bound fragments obtained from complexes B1 or B2 formed at 37°C (lanes 5 and 6 respectively of Figure 6.9). It thus seems that suGF1 contacts the G's located approximately in the center of the G-string in the major groove in both complexes B1 and B2 (see section 6.1). In addition G's located on the 5' side (sequence positions -337, -338 and -339) of the G-string appear to interfere with suGF1-DNA binding to a lesser extent than the central G's.

A similar depletion in fragments resulting from methylation at central G's cannot be visually detected in lanes 7 and 8 of Figure 6.9. The DNA populations shown in these two lanes were obtained from complexes B1 and B2 of lanes 2 and 3 of the preparative EMSA gel respectively (Figure 6.8). Inspection of these lanes indicates that all or a very large percentage of the methylated free probe had been shifted into suGF1-complexes. Methylation of G's in the binding site therefore does not cause complete inhibition of suGF1 DNA-binding. When present in limiting concentrations, suGF1 appears to bind firstly to the population of fragments which do not interfere with contacts to the central G's at 37°C (Figure 6.9, lanes 5 and 6), but interference is overcome at high concentrations of suGF1 (lanes 7 and 8). Slower migrating complexes B3 and B4 are probably formed by protein-protein contacts (see section 6.2.3), and only at protein concentrations where virtually all of the DNA fragment is bound. No selection against DNA methylation at contact positions is therefore expected to take place at this step, as is evidenced in lane 9.

An unexpected result of the methylation interference investigation is that no interference could be detected at 4°C in this and a second, independent experiment (see Figure 6.9, lanes 10 to 13). The reason for the absence of methylation interference at 4°C is not known. The small increase in binding observed in lane 5 of Figure 6.8 (EMSA at 4°C) compared to lane 1 (37°C), might be expected to weaken the observation of depleted fragment derived from methylation of central G's in the bound population. However, it could similarly be argued that observation of enrichment of these fragments should then be more clearly visible in the population of free DNA. The latter effect is observed at 37°C where the feature of central enhancement in the G-run is much more pronounced in the fragments produced from the smaller population of free DNA (lane 4 of Figure 6.9, obtained from the free DNA in lane 2 of Figure 6.8), than from the population isolated from lane 1 of the preparative EMSA (lane 3 of Figure 6.9).

It is likely that a combination of major groove contacts to several G's, as well as other contacts (involving for example the sugar phosphate backbone) are required for optimal suGF1 DNA-

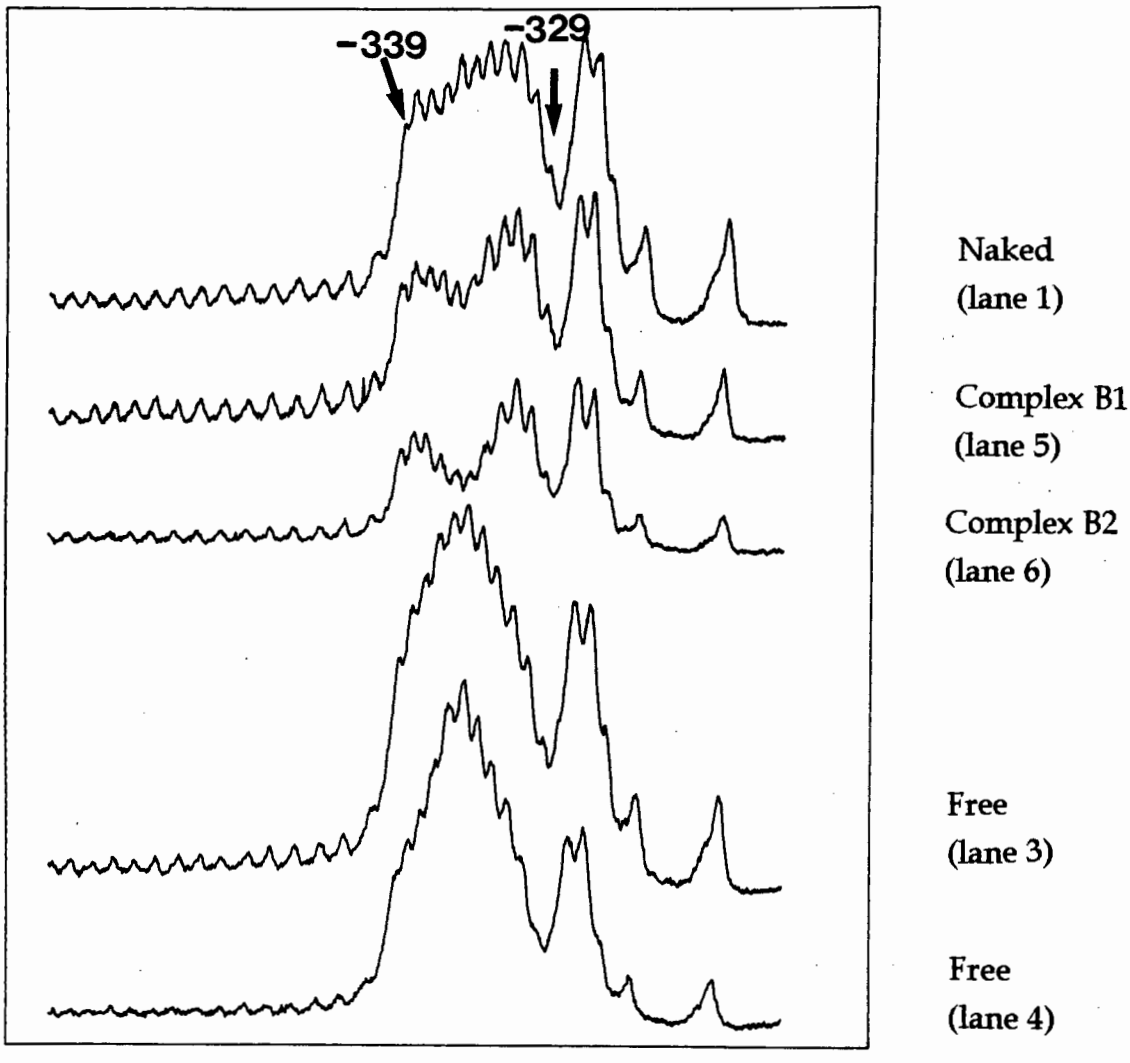


Figure 6.10 Guanines in the center of the G-string on the Watson strand interfere the most with suGF1 DNA-binding

Densitometric traces of the lanes of the sequencing gel shown in Figure 6.9 are shown. The individual lanes which were scanned are indicated in brackets. Numbers indicate sequence positions (see Figure 2.1).

binding. The fact that suGF1 binding is not completely inhibited by methylation of any one G in the G-rich binding site, seems to indicate that the loss of any one of all the possible suGF1 contacts to G's is not sufficient to disrupt sequence-specific suGF1 DNA-binding, both at 4 and 37°C. It is possible that subtle differences in the local structure of the DNA, such as a small increase in the twist, may occur at 4°C (48). Such small differences might affect the proximity of amino acid side chains to DNA bases or the sugar phosphate backbone, altering the relative importance of the various protein contacts. A possible explanation for the lack of any detectable methylation interference at 4°C might thus be increased relative importance of protein-DNA contacts at 4°C other than those to G's in the major groove.

The protein BGP1 has been shown to bind to oligo(dG).oligo(dC) sequences in different frames, with a minimum requirement of 7 contiguous G's (37). The possibility of suGF1 binding in several different frames within the G-string might explain the absence of a clear methylation interference pattern, or the increased probability of methylation interference of central G's (see chapter 7).

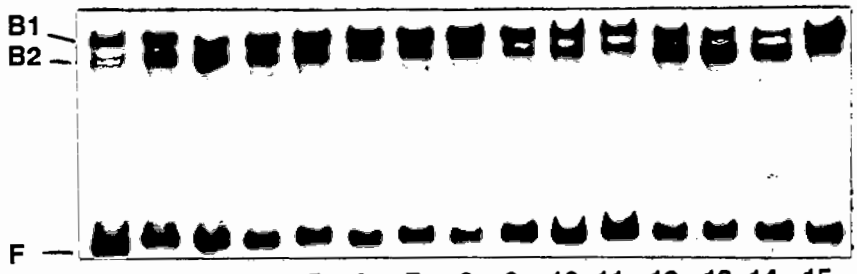
In conclusion, the methylation interference results suggest that suGF1 interacts with G-residues located approximately in the center and towards the 5' side of the G-string. A model for suGF1 DNA-binding incorporating these results will be presented in chapter 7.

6.4 Hydroxyl radical footprinting of the suGF1-DNA complex

6.4.1 suGF1 DNA-binding in the presence of hydroxyl radical footprinting reagents

The reagents used in hydroxyl radical footprinting (H_2O_2, Fe(II)EDTA and ascorbate) are often required at high concentrations to obtain a sufficient level of DNA cleavage (approximately 10% to 20%) (51). For some proteins, these reagents have however been found to decrease or abolish DNA-binding at certain concentrations (51, 224). It is thus essential to assay protein DNA-binding in conjunction with optimization of the amount of cleavage (51, 224). The effect of individual hydroxyl radical reagents on protein-DNA binding is usually assayed by DNase I footprinting. However, by using an EMSA, the combined effect of the reagents can also be monitored.

DNA-binding of suGF1 in the presence of hydroxyl radical reagents was assessed in an EMSA. Labeled A/X fragment was incubated with an appropriate volume of purified suGF1 in hydroxyl radical incubation buffer, followed by the addition of varying concentrations of H_2O_2, Fe(II)EDTA and ascorbate (see section 2.20 and the legend to Figure 6.11). The samples were adjusted to 5% (v/v) glycerol and electrophoresed on 4% nondenaturing polyacrylamide



lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
[Ascorbate] (mM)	1	10	10	1	10	1	10	1	10	1	10	1	1	1	–
% H_2O_2 (v/v)	2,3	0,2	0,03	–	–	–	–	–	–	2,3	2,3	0,6	0,2	0,03	–
[Iron(II)] (mM)	1	0,1	0,1	1	1	0,1	0,1	0,01	0,01	–	–	–	–	–	–
[EDTA] (mM)	2	0,2	0,2	2	2	0,2	0,2	0,02	0,02	–	–	–	–	–	–

Figure 6.11 suGF1 binds to DNA in the presence of hydroxyl radical footprinting reagents

An aliquot of A/X fragment (1 ng) was incubated with an appropriate volume of purified suGF1 in hydroxyl radical incubation buffer (see section 2.20) for 30 min at 4°C in a volume of 38,5 µl. H_2O_2, Fe(II)EDTA and ascorbate were then added to concentrations as indicated, in a final volume of 50 µl (see section 2.20). After two minutes, the incubations were adjusted to 5% (v/v) glycerol, and loaded onto a 4% nondenaturing polyacrylamide gel. Electrophoresis and autoradiography were as described in sections 2.8 and 2.22. F is free DNA fragment. B1 and B2 are suGF1-DNA complexes.

gels (see section 2.8). The autoradiograph of the EMSA is shown in Figure 6.11.

Fe(II)EDTA and ascorbate did not appear to have a significant effect on suGF1 DNA-binding (compare lanes 4, 5, 6, 7, 8 and 9 to lane 15). H_2O_2 at a concentration of 2,3% (v/v) and 1 mM ascorbate (lane 10), seemed to cause only a very small decrease in DNA-binding. An intriguing effect of lower concentrations of H_2O_2 in the presence of ascorbate was a marked change in the observed ratio of suGF1-DNA complexes B1 and B2 (lanes 12, 13 and 14). The increase in complex B2 and decrease in B1 upon treatment with low concentrations of H_2O_2 (and possibly only in the presence of ascorbate) might be explained by chemical modification of the DNA or protein in the complex (see chapter 7).

The final conditions at which hydroxyl radical cleavage was carried out were identical to those shown in lane 1 of Figure 6.11. For these conditions it can be seen that although the individual reagents alone did not significantly change the percentage DNA-binding, the same concentration of the combined reagents caused a decrease in binding (compare lane 1 to lane 15). It might be expected that the free DNA population could be contaminated with a small percentage of DNA found in suGF1-DNA complexes at the start of the cleavage reaction (see section 6.1). The significance of this effect would depend on the extent of protein-DNA dissociation during the time of the cleavage reaction.

Similar results were obtained in a DNase I footprinting experiment in the presence of hydroxyl radical footprinting reagents, although interference of lower concentrations of hydroxyl radical reagents was monitored than in the EMSA (results not shown). The DNase I footprint however established that glycerol (a potent hydroxyl radical scavenger) was not necessary for suGF1 DNA-binding during the incubation period. The possibility that suGF1 only bound to the DNA after hydroxyl radical cleavage and addition of glycerol (to enable EMSA gel loading), was thus discounted.

6.4.2 Hydroxyl radical cleavage of free and suGF1-DNA complexes : analysis of sequencing gels and densitometric traces

The preparative EMSA for hydroxyl radical footprinting of the Watson strand is shown in Figure 6.12. The EMSA for footprinting of the Crick strand was identical to that shown in Figure 6.12, the only difference being the identity of the radiolabeled strand. Incubation of suGF1 with end-labeled DNA fragments, and hydroxyl radical cleavage, was carried out at different temperatures (indicated in Figure 6.12, and described in section 2.20). It was noted that hydroxyl radical cleavage at temperatures other than 4°C adversely affected suGF1 binding. This might be due to extensive damage to the DNA by the relatively high

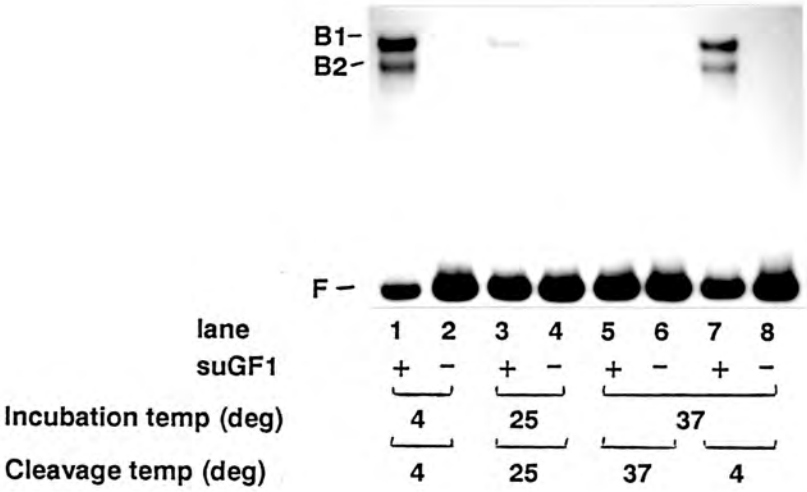


Figure 6.12 Preparative electrophoretic mobility shift assay for hydroxyl radical footprinting

Aliquots of A/X fragment (2 ng), 3' end-labeled on the Watson strand, were incubated with an appropriate volume of purified suGF1 in hydroxyl radical incubation buffer (see section 2.20) for 30 min at the indicated temperature in a volume of 76,9 µl. Limited hydroxyl radical cleavage was carried out by addition of H_2O_2, Fe(II), EDTA and ascorbate to concentrations of 2,3% (v/v), 1 mM, 2 mM and 1 mM, in a final volume of 100 µl (see section 2.20). The reactions were allowed to proceed for 2 min at the indicated temperatures and terminated by the addition of 10 µl stop solution (see section 2.20). Reaction mixtures were immediately electrophoresed on EMSA gels, and complexes identified by wet autoradiography (see section 2.20). F is free DNA. B1 and B2 are suGF1-DNA complexes.

concentrations of H_2O_2. Appropriate protein-DNA complexes and free DNA were identified by wet autoradiography and excised from the gel (see section 2.22). The DNA was recovered from the gel slices and electrophoresed on sequencing gels with appropriate Maxam-Gilbert G-sequencing standards (see sections 2.17 and 2.20). The autoradiographs are shown in Figure 6.13.

A large percentage (approximately 85 to 90%) of the radiolabeled fragments was not cleaved by hydroxyl radicals. It therefore follows that 99% of the cut fragments were nicked only once, assuming a Poisson distribution (19). It is thus not necessary to correct for secondary cleavage at a position between the radiolabel and the primary nick. The extents of cleavage can be directly compared from the intensities of the bands on the gel or the peak heights of the densitometric traces.

Several features of the hydroxyl radical footprints are immediately noticeable from Figure 6.13. In the Watson strand, two "patterns" of hydroxyl radical cleavage are evident. The lanes containing free DNA all show a similar cleavage pattern (lanes 3, 5, 6 and 9), irrespective of the incubation or cleavage temperature. A pattern distinct from that of the free DNA is observed in all the lanes containing DNA isolated from suGF1-bound species (lanes 1, 2, 4, 7 and 8). The free and bound DNA populations in the Crick strand similarly display different patterns, although not as easily observed by eye as in the Watson strand. No differences can be detected between the pattern of cleavage products at different incubation temperatures or between suGF1-DNA complexes B1 and B2. It thus seems that the structure of the phosphate backbone of the double helix does not undergo a significant irreversible change between 4 and 37°C. In addition, no significant differences can be observed between complexes B1 and B2 in terms of the conformation of the phosphate backbone, or possible contacts made to the backbone.

The densitometric traces of the lanes containing free DNA and that isolated from complex B1, are shown in Figure 6.14 (Watson strand) and 6.15 (Crick strand). Interestingly, a clear modulation in the extent of hydroxyl radical cleavage is present in the Watson strand, both in the free and suGF1-bound DNA. This modulation consists of consecutive local maxima (indicated by asterisks) and minima in the extent of hydroxyl radical cleavage, repeated in phase with the helical period of the DNA double helix (approximately 10,5 bp/turn (193, 232)). The maxima are located at sequence positions -296, -306, -316, -326 and -337 in the Watson strand of complex B1 (Figure 6.14, first and third traces). The positions of maximum cleavage are not as clear in the free DNA, but can nonetheless be assigned to approximately -296, -306 and -316. Although the modulation is less pronounced in the free DNA, it is not the result of contamination by cleavage products of the bound population, since this modulation

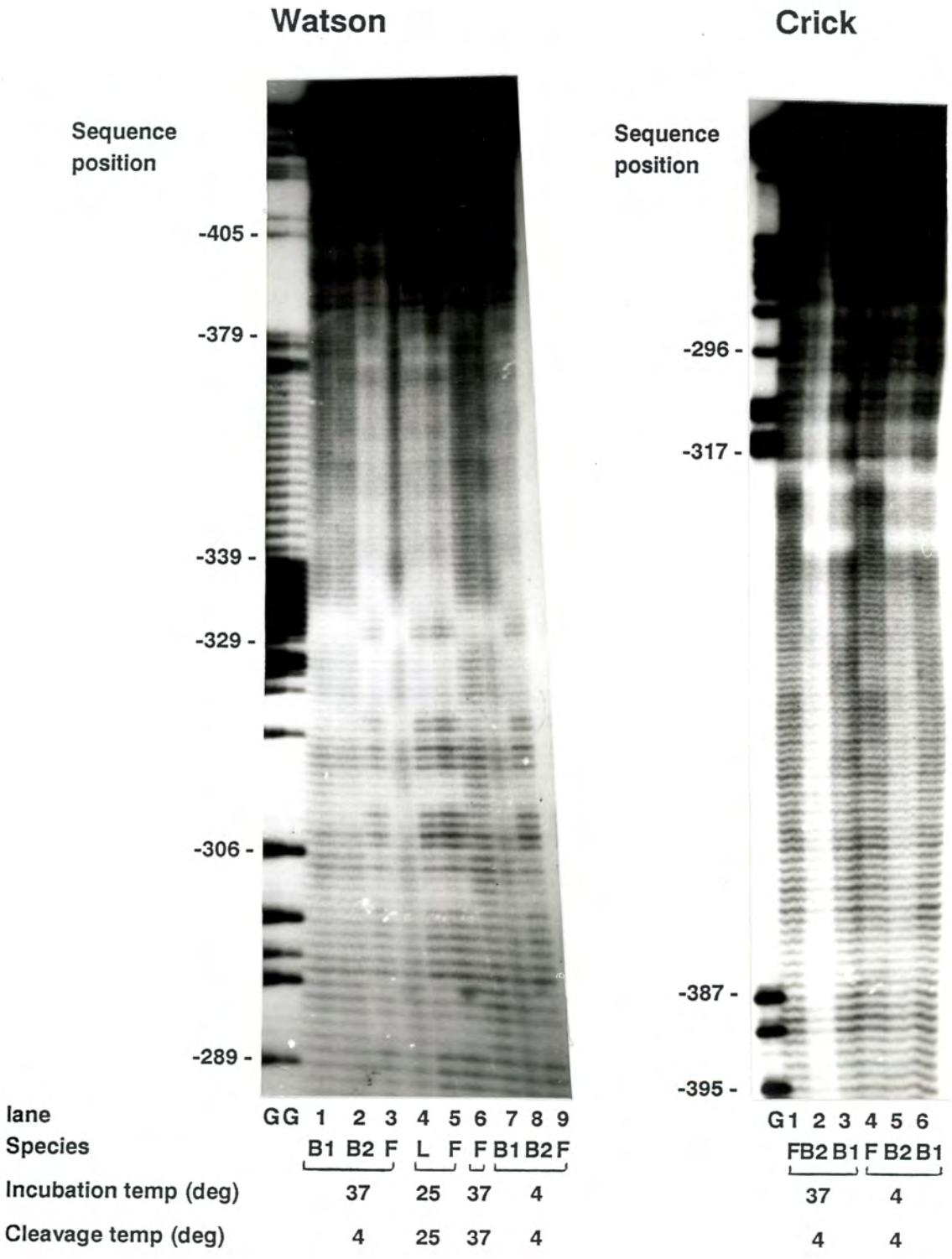


Figure 6.13 Hydroxyl radical footprints of suGF1 on the Watson and Crick strands of the P. miliaris early H1-H4 intergenic region

Hydroxyl radical cleaved free DNA (F), and suGF1-DNA complexes B1, B2, or B1B2 (denoted by L for lower), were excised from the appropriate lanes in the preparative EMSAs for the Watson and Crick end-labeled strands respectively. DNA was recovered and electrophoresed on sequencing gels (see sections 2.20 to 2.22). The temperatures at which the suGF1-DNA incubations and hydroxyl radical cleavage reactions were carried out are indicated. Lanes marked by G contain Maxam-Gilbert G-sequencing standards. The sequence positions are indicated (see Figure 2.1 for the sequence).

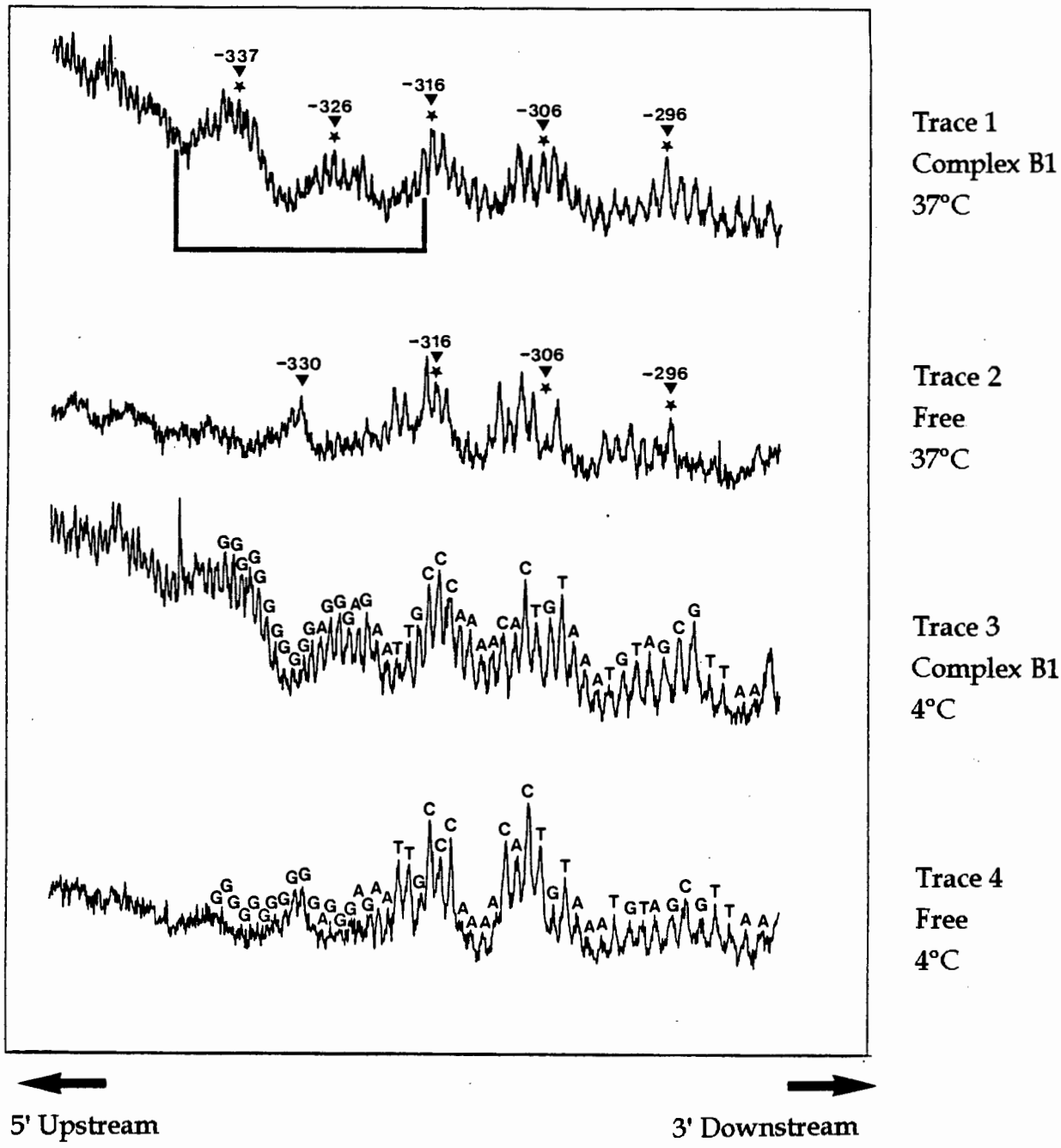


Figure 6.14 Densitometric traces of hydroxyl radical cleavage products of the Watson strand of duplex DNA, resolved on a sequencing gel

Lanes on the autoradiograph of the sequencing gel containing hydroxyl radical cleavage products of the Watson strand of the duplex DNA were scanned (see Materials and Methods). The densitometric traces (from top to bottom) were obtained from lanes 1 (complex B1, incubated at 37°C prior to hydroxyl radical cleavage), 3 (free DNA, incubated at 37°C), 7 (complex B1, incubated at 4°C) and 9 (free DNA incubated at 4°C) of the autoradiograph of the Watson strand shown in Figure 6.13. Local maxima in the extent of hydroxyl radical cleavage are indicated by asterisks. Triangles point to the indicated sequence positions (top traces), and the sequence is shown (bottom traces) (see Figure 2.1). The bracket denotes the extent of the DNase I footprint determined from the DNase I difference probability plot shown earlier. The 5' (upstream, more negative sequence positions) and 3' (downstream) directions are indicated.

has also been observed in the Watson strand cleaved by hydroxyl radicals in the absence of protein (results not shown).

The modulation in the free DNA is consistent with periodic narrowing and widening of the minor groove, suggesting that a section of the DNA molecule is curved in solution (25, 80, 213; see references 79 and 221 for reviews). This is in agreement with results indicating that the fragment is indeed curved in solution as shown by anomalous migration in polyacrylamide gels (H.-G. Patterton & C. von Holt, submitted). If wide and narrow minor grooves are repeated in phase with the helical period, a planar curve is obtained. Nonplanar curvature is the result of the narrowing and widening of the minor groove out of phase with the helical period. The local maxima in the extent of hydroxyl radical cleavage appear to have a repeat of approximately 10 bp ([-296-(-316)]/2) in the free (and over the same region in the bound) DNA, consistent with the occurrence of planar curvature.

A nucleosome core has previously been shown to be translationally and rotationally positioned *in vitro* over the H1-H4 intergenic region with translational borders at -387 and -247 (189, H.-G. Patterton & C. von Holt, submitted). Positions where the minor groove was narrow or wide, facing towards and away from the surface of the nucleosome respectively, were determined by DNase I probing (189, H.-G. Patterton & C. von Holt, submitted), and coincide with those determined in this study for the modulation in the extent of hydroxyl radical cleavage. The direction in which the DNA is bent around the nucleosome core is therefore the same as the direction of curvature[11] of the free fragment.

The structure of poly(dA).poly(dT) has been shown to be rigid with a narrow minor groove, caused by high propeller twisting enabling bifurcated hydrogen bonds in the major groove (see reference 54 for a review). In contrast, the minor groove in short runs of oligo(dG).oligo(dC) is wide or can easily become wide, as evidenced by X-ray crystallographic studies (54, see section 1.3). Although the physical basis of curvature of mixed sequence DNA such as the A/X fragment is not understood, it is well established that the repetition of short runs of oligo(dA) repeated in phase with the helical period is associated with planar curvature (see references 54, 79 and 221 for reviews). The direction of curvature in the region of the suGF1 DNA-binding site is probably related to the occurrence of A-rich regions repeated in phase with the helical repeat (H.-G. Patterton & C. von Holt, submitted). In Figure 6.14 the periodic narrowing of the minor groove in the 5' to 3' direction is clearly visible as a decrease

[11] Note that the use of the term curvature is usually reserved to indicate an inherent deflection of the helical axis of a DNA molecule to describe a curve in solution (planar or nonplanar) in the absence of an external force. Bending refers to a distortion of the helical axis into a curve due to the presence of an external force, such as the electrostatic interaction between the DNA phosphate backbone and the basic amino acid residues on the surface of the histone octamer, which folds the DNA molecule into a nucleosome core (221).

in consecutive extents of hydroxyl radical cleavage at sequence positions -322 to -319 of trace 3 (bound), and positions -314 to -311, -305 to -301, and -294 to -291 in traces 3 and 4 (bound and free). Such a narrowing of the minor groove in the 5' to 3' direction of each A-rich region of a curved DNA fragment was first detected by hydroxyl radical cleavage by Burkhoff and Tullius (25).

The modulation in the extents in hydroxyl radical cleavage of the Watson strand, is mirrored in the Crick strand of the duplex. Local maxima in the extent of cleavage are located at approximately -337, -326, and -318 in the free and bound DNA (Figure 6.15, traces 2,4 and 1,3 respectively; local maxima indicated by asterisks). In the Crick strand of the bound DNA, the modulation is clearly seen to continue with a repeat of approximately 10 bp over at least three helical turns downstream of the local maximum at -318. The corresponding modulation in the free DNA is less obvious. A strong periodic modulation similar to that found downstream of the $(GA)_{16}$-repeat (Watson strand) or corresponding $(TC)_{16}$-repeat (Crick strand), is not observed upstream of the 5' side of the $(TC)_{16}$-repeat in the Crick strand.

The modulation in the extent of hydroxyl radical cleavage extends across the G-string in the Watson strand of bound DNA (Figure 6.14, traces 1 and 3), but clearly does not extend across the G-string in the free DNA (Figure 6.14, traces 2 and 4). Instead, a local maximum is found at -330 in the free DNA. Interestingly, the modulation appears to extend upstream of position -318 and includes the C-string in the Crick strand of both the bound and free DNA (Figure 6.15). In the free DNA a local maximum is located at approximately -326 and a small local minimum at approximately -331. The maxima and minima in the Watson strand of the free DNA thus appear to be removed by half a helical turn from those in the Crick strand in this region. The offset of 5 bp seems too large to be explained by staggered cleavage across the minor groove in a curved DNA molecule (55). It therefore seems that the curvature is disrupted towards the G-string in the free DNA fragment.

The increase in the extent of hydroxyl radical cleavage towards the 3' end of the G-string in the Watson strand of free DNA (Figure 6.14, traces 2 and 4) is not necessarily an indication of widening of the minor groove. The local maximum at -330 might be caused by increased accessibility to hydroxyl radical cleavage at a "kink" in the DNA caused by a disruption in the stacking of the G-residues. The G-residues may exhibit high slide and low roll as described in the model of poly(dG).poly(dC) derived from an X-ray crystallographic study (154) (see section 1.3). Stacking of purine rings in the region of the G-string and in the adjacent AG-rich region could lead to a shortening of the Watson strand (54), presenting a less favourable environment for hydroxyl radical cleavage on average. A shortened Watson strand exhibiting little or no cross-strand stacking, may cause perturbations in the conformation of the Crick

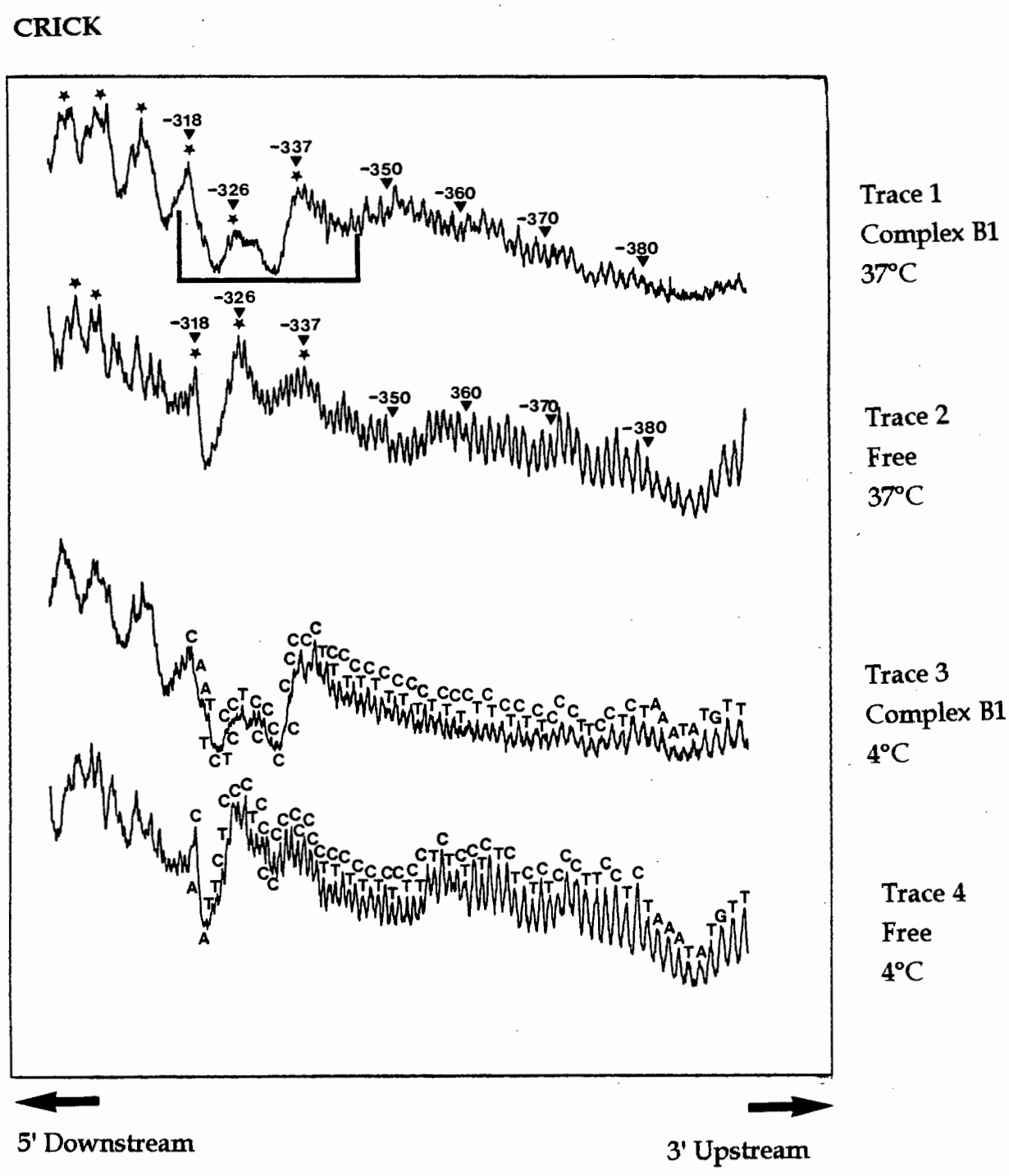


Figure 6.15 Densitometric traces of hydroxyl radical cleavage products of the Crick strand of duplex DNA, resolved on a sequencing gel

Lanes on the autoradiograph of the sequencing gel containing hydroxyl radical cleavage products of the Crick strand of the duplex DNA were scanned (see Materials and Methods). The densitometric traces (from top to bottom) were obtained from lanes 6 (complex B1, incubated at 37°C prior to hydroxyl radical cleavage), 4 (free DNA, incubated at 37°C), 3 (complex B1, incubated at 4°C) and 1 (free DNA incubated at 4°C) of the autoradiograph of the Crick strand shown in Figure 6.13. Local maxima in the extent of hydroxyl radical cleavage are indicated by asterisks. Triangles point to the indicated sequence positions (top traces), and the sequence is shown (bottom traces) (see Figure 2.1). The bracket denotes the extent of the DNase I footprint determined from the DNase I difference probability plot shown earlier. The 5'(downstream) and 3' (upstream) directions are indicated.

strand which is expected to contain very little if any same-strand stacking, at least in the region of the C-string (154). Such perturbations may be reflected by the large local maximum at approximately -326 (free DNA, Figure 6.15). However, in the absence of structural data in the form of X-ray crystallography or NMR, these points must remain speculative.

What information regarding suGF1 DNA-binding can be obtained from the hydroxyl radical footprinting? A periodic modulation in hydroxyl radical cleavage in protein-bound DNA such as that seen in the suGF1-DNA complex may be caused by protein contacts to one side of the helix, by periodic perturbation in the DNA conformation repeated in phase with the helical repeat, or by deformation of the helical axis to describe a curve in solution as discussed above. The cleavage pattern may also be due to a combination of these effects, and additional information is needed to distinguish between them. In the same way in which the modulation in hydroxyl radical cleavage can be at a period similar to the helical repeat (planar curvature) or at a different period (nonplanar curvature), a periodic modulation caused by proximity of protein to the backbone is not necessarily an indication of the helical period of the DNA. The modulation may, for instance, be caused by consecutive sites of protein-DNA contacts systematically rotating ("snaking") around the helix, without following only the major or minor groove, but rather crossing over grooves. The period of hydroxyl radical protection would thus be greater or smaller than the helical repeat of the DNA, depending on the period of rotation of the protein contact sites around the helix.

In the case of suGF1-bound DNA, the modulation in the hydroxyl radical cleavage pattern is due to curvature as opposed to direct suGF1 contacts to the backbone, for two reasons. Firstly, a modulation with the same period can also be detected in the free DNA, and secondly, the modulation extends downstream beyond the border of the DNase I footprint (-316) in the bound and free DNA. The fact that the modulation in hydroxyl radical cleavage is actually enhanced outside the borders of the DNase I footprint in the bound DNA, may be due to suGF1-induced bending of the DNA in the direction of the observed curvature of the free fragment. This, however, appears unlikely, since bending over a region beyond the DNase I footprint would require additional protein-DNA contacts to be made on the far side of the bent region. No such contacts are observed by DNase I footprinting.

A curved DNA molecule is not a rigid structure in solution. The observed curvature is rather thought to represent a time-averaged structure, reflecting the average conformation out of many conformations assumed during free motion. The most likely explanation for the enhancement in the modulation of hydroxyl radical cleavage in the presence of suGF1 seems to be the restriction of the degree of freedom of movement of the DNA due to the proximity of the suGF1 molecule. Binding of suGF1 could therefore limit the DNA molecule to a range of

structures closer to that of a curved molecule than would be assumed by a DNA molecule in the absence of suGF1. The observed enhancement in the modulation of the hydroxyl radical cleavage, as well as the extension of the modulation over the G-string in the bound but not in the free DNA, are consistent with suGF1-DNA contacts being made in the major groove approximately in the center of the G-string, causing the minor groove at the back of the helix to be narrowed in phase with the period of the preferred curvature of the free fragment.

An additional important feature evident from the densitometric scans should be noted. As a result of the DNA curvature, a local maximum is expected between the local minima at -333 and -323 in the Crick strand. Although a large maximum is observed at -326 in the free DNA (Figure 6.15), the corresponding local maximum is much smaller in the suGF1-bound DNA. The absence of a large local maximum in the suGF1-bound DNA at this position probably indicates that suGF1 contacts are made to the phosphate backbone in this region (see chapter 7).

6.4.3 Hydroxyl radical footprinting analysis of suGF1 DNA-binding : difference probability plots

Correction for secondary cleavage within one molecule is not strictly necessary in the case of hydroxyl radical footprinting, since the occurrence of such events should be very low (see section 6.4.2). Analysis of the data by the construction of difference probability plots is however useful, as the effect on the extent of digestion solely due to binding of suGF1 can be seen more clearly.

As expected from analysis of the densitometric traces, a periodic modulation in ln(difference probability cleavage) is clearly visible in the Watson strand of complex B1 at 4°C (top panel of the top plot of Figure 6.16). Local maxima occur at sequence positions -337, -326, -315 and -305, whereas local minima are found at positions -331, -320, -310 and -302. Because the modulation is seen in a difference probability plot where sequence or structural specificity of the probe for DNA in the absence of protein is corrected for, the modulation clearly supports the previous observation that the modulation in hydroxyl radical cleavage is enhanced upon suGF1 DNA-binding.

The difference probability plot of the Crick strand of complex B1 incubated at 4°C before hydroxyl radical cleavage (Figure 6.16, top plot, bottom panel), reveals two very prominent local minima. These local minima are located at sequence positions -333 and -323, within the borders of the DNase I footprint. The position where the minor groove is most protected from hydroxyl radical cleavage upon suGF1 binding, is at -333. This corresponds to the region

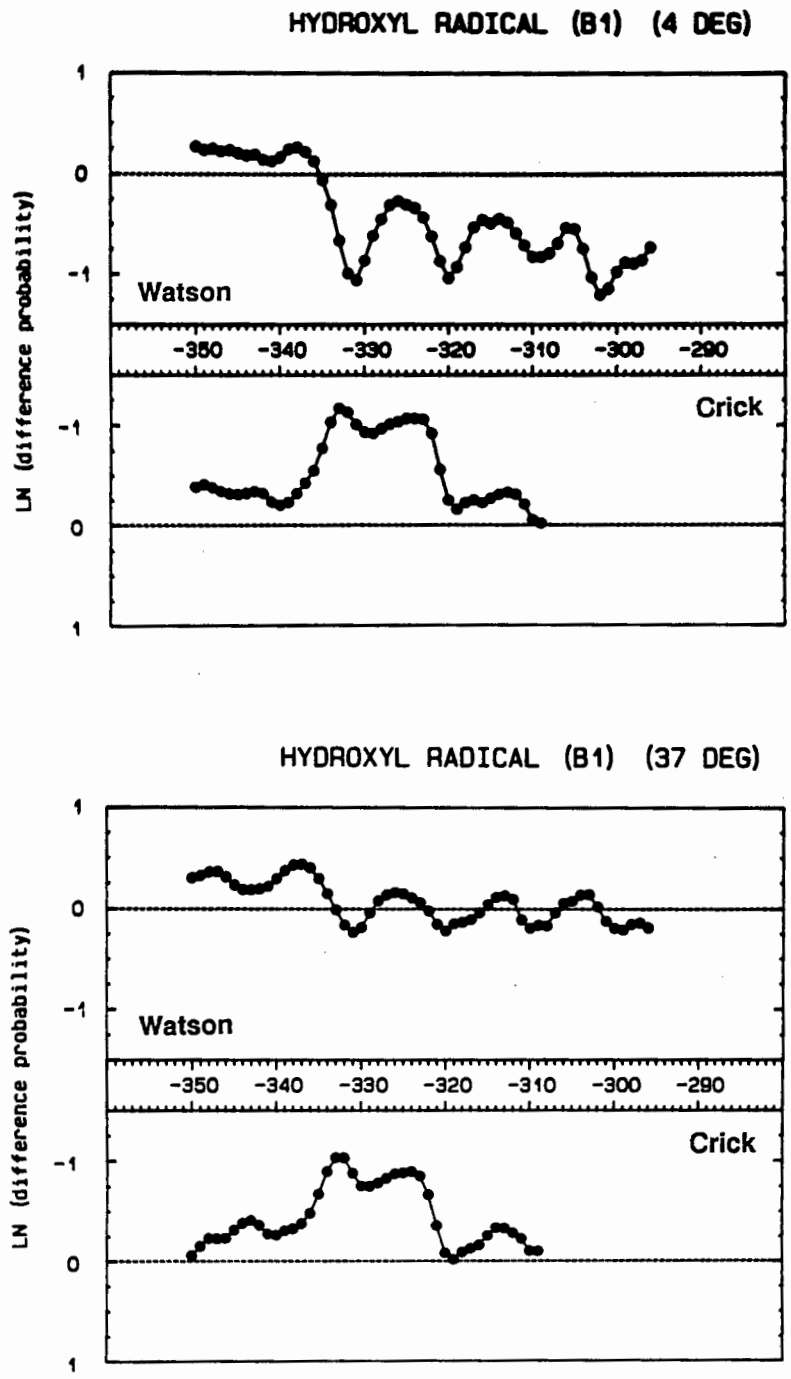


Figure 6.16 Difference probability plots of suGF1-DNA complex B1 formed at 4 and 37°C and probed with hydroxyl radicals

The natural logarithm of the probability of cleavage by hydroxyl radicals at 4°C, after incubation at 4°C (top plot) or 37°C (bottom plot) in the absence of suGF1, was subtracted from that in the presence of suGF1 in complex B1, calculated as a three point running average at each position in the sequence indicated by a datum point (filled circle) (see section 2.24). The resulting difference probability plots for the Watson and Crick strands are shown in the top and bottom panels of each plot respectively. The relevant lanes on the autoradiographs shown in Figure 6.13 were scanned in order to calculate the difference probability plots.

where major groove contacts to G's are proposed to occur (see section 6.3), possibly resulting in opening up of the major groove and narrowing of the minor groove at the back, consistent with a greatly decreased value in ln(difference probability) cleavage.

The protection of the Crick strand of the duplex in the region of -338 is in stark contrast to the expected modulation in the hydroxyl radical cleavage observed for free DNA. This confirms the observation made in section 6.4.2, that suGF1 seems to contact the phosphate backbone in this region (see chapter 7). Smaller local minima are found in the Crick strand of the duplex at -314 and approximately -344, consistent with the data for the Watson strand, and indicative of the suGF1-induced restriction of the degree of freedom of movement of the DNA molecule into conformations other than the time-averaged curvature observed for free DNA.

Comparison of the difference probability plots of complex B1 for incubation at 4 and 37°C (top and bottom panel of Figure 6.16) reveals no significant difference between the ln(difference probability) values for the Crick strands. The Watson strands of complex B1 incubated at 4 or 37°C also show very similar characteristics. Both the degree of protection (negative ln(difference probability)) and amplitude of the modulation are increased for incubation at 4°C (Watson strand). The meaning of this is currently not understood, and it may not be significant. Although the enhanced amplitude of the modulation in the Watson strand at 4°C may reflect the availability of less free energy of movement at lower temperatures, causing the DNA molecule to deviate less from the average (curved) structure than at 37°C, this effect is not mirrored when comparing the Watson strand of complex B2 at the two temperatures (Figure 6.17).

When comparing complexes B1 and B2 (Figures 6.16 and 6.17), the modulation in ln(difference probability cleavage) is once again very similar. These results clearly show that the same protein-DNA interaction is involved in the two complexes. Differences in the amplitude of the modulation and overall protection between complex B1 and B2 incubated at 4°C could be detected, but are once again difficult to explain. Further footprinting and interference experiments (for example ethylation interference) may detect differences between the suGF1-DNA interaction in the two complexes (see chapter 7). Given the data obtained in this investigation, however, no significant differences in the protein-DNA interaction which could explain the different electrophoretic mobilities of complexes B1 and B2 in EMSAs could be detected. The results presented here are discussed further in chapter 7, where a model for the suGF1-DNA interaction is presented.

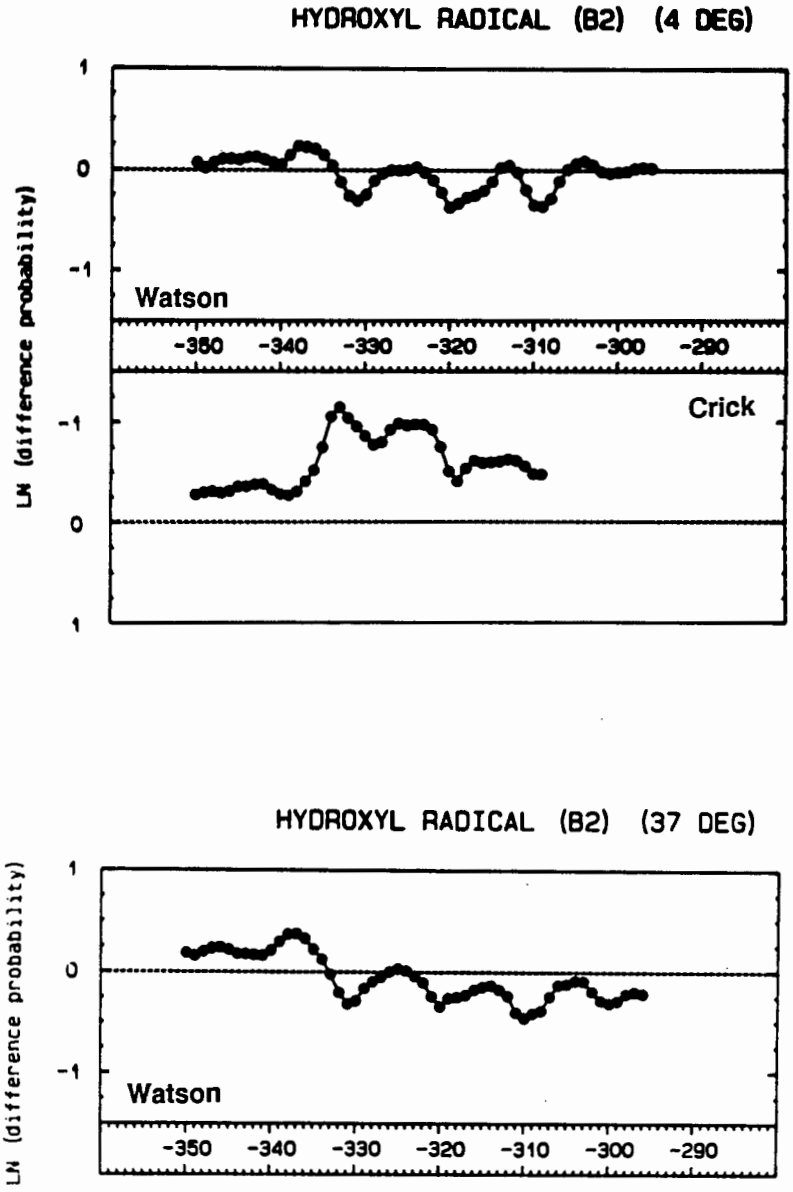


Figure 6.17 Difference probability plots of suGF1-DNA complex B2 formed at 4 and 37°C and probed with hydroxyl radicals

The natural logarithm of the probability of cleavage by hydroxyl radicals at 4°C, after incubation at 4°C (top plot) or 37°C (bottom plot) in the absence of suGF1, was subtracted from that in the presence of suGF1 in complex B2, calculated as a three point running average at each position in the sequence indicated by a datum point (filled circle) (see section 2.24). The resulting difference probability plots for the Watson and Crick strands are shown in the top and bottom panels respectively. The difference probability plot for the Crick strand at 37°C was not calculated as the bound lane on the gel was too light to scan. The relevant lanes on the autoradiographs shown in Figure 6.13 were scanned in order to calculate the difference probability plots.

CHAPTER 7

DISCUSSION

7.1 Purification to near homogeneity of a 59,5 kDa poly(dG).poly(dC)-binding protein from sea urchin

In this investigation, a poly(dG).poly(dC)-binding protein (suGF1) was purified from sea urchin embryos (*P. angulosis*) by sequence-specific DNA-affinity chromatography. suGF1 was enriched approximately 3 400-fold, with a final yield of 29% from *P.angulosis* embryonic nuclear extracts (see Table 4.2). The enrichment of suGF1 from total nuclear protein is estimated to be as high as 34 000-fold (29; see section 3.2).

The first step in the purification (cation exchange : P11 phosphocellulose) gave a 10-fold enrichment and yield of approximately 80% (see Table 4.2). Poly(dG).poly(dC)-affinity chromatography, resulted in a very high stepwise enrichment of 150- to 180-fold, compared to 100-fold typically expected (see Table 4.2). This result shows that suGF1 binds with high affinity to poly(dG).poly(dC), and that the column exhibits high specificity for binding of suGF1. The additional enrichment obtained by two BS-affinity steps was only about 1,6-fold, suggesting that the selectivity of the two columns for proteins in the nuclear extract was very similar, as discussed in chapter 4. The yield and enrichment provided by the poly(dG).poly(dC)-affinity chromatography step, followed by one pass over the BS-affinity column, was approximately 52% and 200-fold, respectively. This compares well with the yield and enrichment of 30% and 100- to 500-fold respectively, typically expected for two sequential affinity chromatography steps (108). The high yield of suGF1 may be explained by less nonspecific binding to the matrix.

The purification of suGF1 is comparable to that of other G-binding factors. Properties of such factors are summarized in Table 7.1. (This table will be referred to as needed, before discussing it in more detail in section 7.6) Table 7.1 contains a summary of the steps used in the purification of G-binding factors which may be similar to suGF1. suGF1 showed a very similar enrichment and yield from a P11 phosphocellulose column, compared to the G-binding transcription factor H4TF-1 (44). In addition, suGF1 and H4TF-1 both eluted in a 0,5 M KCl step from the respective columns (44). Further steps in the purification of H4TF-1

also involved affinity chromatography (44). The overall enrichment and yield for purification from HeLa nuclear extracts was found to be 3 550-fold (comparable to that of suGF1) and only 3%, respectively. The low overall yield could be traced to the sequence-specific DNA-affinity chromatography step, which gave an enrichment of 130-fold, but a yield of only 8%. Other characteristics of H4TF-1 and suGF1 will be compared in later sections. Transcription factor Sp1 was purified 6 000-fold with a yield of 10% from HeLa cells (20). The Binding Site-affinity chromatography step contributed a stepwise enrichment and yield of 83-fold and 56%, respectively, to these final values. The high overall enrichment may reflect low relative abundance of Sp1 compared to total nuclear extract protein.

Purification of the poly(dG).poly(dC)-binding protein BGP1 also involved a poly(dG).poly(dC)-affinity chromatography step (37). This affinity column was loaded at high ionic strength (see Table 7.1), and bound proteins were eluted by a stepwise increase in the concentration of KCl in the buffer, similar to the purification procedure for suGF1. BGP1 eluted at 0,5 M KCl from the poly(dG).poly(dC)-affinity column. The fact that suGF1 only eluted at 0,7 M KCl suggests that suGF1 may bind poly(dG).poly(dC) with higher affinity that BGP1. No values for yield or enrichment were provided for the purification of BGP1 (37).

The purity of suGF1 in the final preparation was assessed from the DNA-binding activity determined in EMSAs, and by estimation of proteins present on silver-stained SDS-PAGE gels. The possible errors inherent in these methods for determination of purity were discussed in chapter 4. From silver-stained gels, the purity of the final preparation was estimated to be at least 85% (see chapter 4). An estimation based on the activity of the final preparation determined by EMSA, however, yielded a value of 47% or 94% purity for suGF1. These two values are for suGF1 bound to DNA as a monomer or a dimer, respectively. It should be noted that there is however no evidence that suGF1 binds as a dimer.

Since the overall yield and total amount of nuclear extract protein are known, the abundance of suGF1 in the nuclear extract can be calculated from the above estimates of purity. Using values of 85% and 47% purity, the abundance of suGF1 is 83 µg/329 mg protein and 20 µg/392 mg protein, respectively. This translates to suGF1 constituting 0,025% (w/w) and 0,014% (w/w) of nuclear extract protein respectively. This estimate is in very good agreement with that determined by Calzone *et al.* (29) for a sea urchin embryonic nuclear protein P3A2 (0,02% (w/w)).

Isolation of a protein from a SDS-PAGE gel can usually be accomplished when the purity is only 5% (29). Isolation of pure suGF1 by preparative SDS-PAGE after passing the nuclear extract over a poly(dG).poly(dC)-affinity column, should therefore be possible. This

preparation can then be used to generate peptide sequence information in order to design probes to isolate the cDNA encoding suGF1 from a cDNA library. Such an investigation is currently underway (S.D. Scherer, J.P. Hapgood, personal communication). suGF1 cDNA may then be sequenced, and the protein expressed from a cloned cDNA. suGF1 purified by DNA-affinity chromatography may however be superior to the expressed protein for functional investigations. Stage-specific posttranslational modifications may for example take place in the sea urchin embryo but not in an expression system.

7.2 The same protein (suGF1) is involved in two sequence-specific complexes

Two sequence-specific protein-DNA complexes (B1 and B2) are always obtained in EMSAs with *P. angulosis* nuclear extracts, and the H1-H4 intergenic fragment (E/H fragment) or an oligonucleotide containing the binding site of suGF1 (S-oligo) (J.P. Hapgood, D. Patterton, unpublished) (see section 1.2). The occurrence of the two bands in the nuclear extract could be explained by several possibilities (see section 1.2). These possibilities include proteolysis or posttranslational modifications of the same protein (suGF1). Examples of posttranslational modifications include phosphorylation, glycosylation, and ADP-ribosylation. The size of suGF1 might differ between the two complexes due to a truncated from of the protein being present. Furthermore, two different proteins may be present in the two complexes, or suGF1 might be present in both complex B1 and B2 but an additional protein (involved in protein-protein interactions to suGF1) might be present in complex B1. Complexes B2 and B1 may represent a monomer and dimer respectively, or other multimers of suGF1. It is possible that only suGF1 was present in the two complexes, but the different electrophoretic mobilities of the complexes resulted from the presence of two different conformations. This could include a change in the conformation of the protein or in the conformation or chemical composition of the DNA. The size of the DNA could also be different in the two complexes. This could be brought about by the presence of an endogenous nuclease in the nuclear extracts, or suGF1 exhibiting nuclease activity. (An endonuclease specifically cleaving poly(dG).poly(dC) has been identified in higher eukaryotes (76, 77, 197).) Furthermore, the involvement of RNA in one of the complexes could result in a change in mobility in an EMSA.

Experimental evidence which discounts many of the possibilities stated above has, however, been obtained. This evidence includes the fact that the ratio of the two bands B1 and B2 are always the same for different preparations of nuclear extracts, or nuclear extracts prepared from embryos at different developmental stages (J.P. Hapgood, D. Patterton, unpublished). Furthermore, the binding activity(ies) in the two complexes were copurified in this investigation (see chapter 4). Complexes B1 and B2 could be obtained in an EMSA after renaturation of DNA-binding activity from a single protein band (59,5 kDa) on a SDS-PAGE

gel, and only one band of approximately 60 kDa could be detected in Southwestern blotting (see chapter 4). The same hydroxyl radical and DNase I footprints were obtained for the isolated complexes B1 and B2 with purified suGF1, and the same methylated G's interfered with binding in complexes B1 and B2 in nuclear extracts and purified preparations (see chapter 6; J.P. Hapgood, unpublished). No differences in the DNA-binding specificity or affinity of B1 compared to B2 could be detected with many different competitors in EMSAs (J.P. Hapgood, D. Patterton, unpublished; see section 1.2). No evidence could be found for differential phosphorylation of the proteins in the two bands, and RNaseA had no effect on the electrophoretic mobilities of the two complexes (J.P. Hapgood, unpublished; see section 1.2). The DNA in the two complexes was the same size, and no evidence for endonuclease activity causing DNA cleavage could be found (J.P. Hapgood, unpublished). When the DNA fragment was titrated with nuclear extract or purified suGF1, the occurrence of B2 at low protein concentrations, followed by the occurrence of B1 at higher protein concentrations (and corresponding decrease in B2), was never observed.

From this evidence, only a few of the above possibilities remain. It seems extremely unlikely that two different proteins are involved, since these two proteins would be expected to both have a molecular weight of 59,5 kDa, both exhibit the same DNA-binding specificity, and both be characterized by the same mode of interaction with DNA, as deduced from footprinting and interference experiments. Although very unlikely, it is formally possible that a change in the conformation of the DNA may be responsible for the different electrophoretic mobilities of the two bands, and that such a change is not detected by the probes used in experiments presented in chapter 6. It is also possible that a RNA structure resistant to digestion with RNaseA could be associated with suGF1. The possibility that suGF1 can only bind as a dimer in complexes B1 and B2 cannot be excluded from the experimental evidence (see below). The two most likely explanations for the two bands detected in EMSAs, are firstly the presence of an additional moiety (amino acids or posttranslational modifications other than phosphorylation) in one of the suGF1 complexes, or secondly a change in the conformation of suGF1 in one of the complexes compared to the other.

In chapter 4 it was noted that protein(s) forming complex B1 in an EMSA eluted at slightly lower ionic strength from the cation exchanger P11 phosphocellulose than did the protein(s) forming complex B2. The differential elution of proteins forming B1 and B2 on cation exchange chromatography is difficult to explain in terms of the different electrophoretic mobility of the complexes in EMSAs. Whereas elution from cation exchange is based mainly on the effect of the net surface charge of the protein, the mobility on EMSA gels is an effect of both charge (direction of electrophoresis is from negative to positive) and size (smaller complexes migrate the furthest) of the combined protein-DNA complex..

The differential elution on cation exchange chromatography suggests that the protein in complex B1 may contain an additional negatively charged moiety, or the protein in complex B2 may contain an additional positively charged moiety. If the latter suggestion is correct, B2 would be expected to have a lower electrophoretic mobility on an EMSA gel and migrate above B1. (This would be due to an increased molecular weight, and an accompanying decrease in net negative charge.) Complex B2 was however found to have a higher electrophoretic mobility than complex B1 on EMSA gels, discounting the possibility that the protein in complex B2 contains an additional positively charged moiety. The elution from the cation exchange column is thus probably due to the occurrence of an additional negatively charged moiety on the protein forming complex B1. If this negatively charged moiety contributed very little to the size of B1, B1 would however be expected to migrate below B2 on an EMSA gel, based on charge effects. This suggests that B1 may contain a relatively large posttranslational modification or additional amino acid residues contributing additional negative charges to the protein. In order to be consistent with the mobility in EMSAs, the additional negative charge on B1 would have to contribute very little to the overall negative charge of the protein-DNA complex (compared to the contribution to the size), such that B1 would migrate slower than B2 mainly due to the increased size.

The existence of two forms of suGF1 consistent with the elution from the cation exchanger and the mobility on EMSAs, may result from two different genes encoding suGF1, or by differential RNA processing of a single primary transcript. Alternatively, a posttranslational modification such as glycosylation may be involved (see above). Indeed, it is known that glycoproteins can be fractionated by ion exchange (95). The decrease in the ratio of complexes B1:B2 in the presence of dilute concentrations of H_2O_2 might be explained by chemical modification of the protein (including damage to a posttranslational modification such as glycosylation), or by modification of the DNA. It is known that H_2O_2 causes damage to DNA bases, for example deoxycytidine (231). The elution of the protein involved in complex B1 from the P11 phosphocellulose column before that involved in B2 may also be explained by a conformational change altering the net surface charge of the protein (net surface charge of protein in B1 less positive). If this were the case, a difference in the overall shape of complex B1 compared to B2 could result in their differential migration in EMSAs.

The fact that both suGF1-DNA complexes B1 and B2 can be obtained by renaturation of the 59,5 kDa species isolated from a SDS-PAGE gel, and that only one major band (approximately 60 kDa) is obtained in Southwestern blotting of purified suGF1, is still consistent with the proposal of the location of a relatively large negative moiety on suGF1 forming complex B1. The molecular weight contributed by such a moiety need not be great enough to result in

separation of the proteins in complexes B1 and B2 by SDS-PAGE in 10% gels. Different species of suGF1 may be resolved on 4% EMSA nondenaturing gels when complexed to DNA.

The fact that DNA-binding activity can be reconstituted from the 59,5 kDa protein(s) and in the Southwestern blot, does not necessarily indicate that proteins in complexes B1 and B2 bind to DNA as monomers of approximately 59,5 kDa. It is possible that dimers of approximately 119 kDa are disrupted during the SDS-PAGE steps, and subsequently re-dimerize under the conditions of renaturation. Moreover, the protein(s) may be present as monomers in solution, and only dimerize in the protein-DNA complex (examples of this were given in chapter 1). No evidence in support of dimers in complexes B1 and B2 was, however, found (see also section 7.5). Furthermore, it appears unlikely that complex B1 contains a dimer of the protein involved in comlex B2, in view of the repeating doublet arrangement of higher multimers of suGF1. If complex B3 and B4 contain dimers of the suGF1 species involved in complexes B1 and B2 respectively, complex B1 cannot contain a dimer of the protein involved in complex B2, since this would result in complexes B1 and B4 having the same electrophoretic mobility.

Although the results presented here have discounted certain obvious possibilities for the presence of complexes B1 and B2, and shown that suGF1 was involved in both complexes, further experiments are needed to establish the difference between the two complexes. The most likely explanations as discussed above are posttranslational modifications, a truncated form of suGF1 being present, or the existence of two conformations of suGF1. Cloning and sequencing of suGF1 as well as functional studies and DNA-binding experiments will be needed to ascertain the possible structural or functional differences between complexes B1 and B2.

It should be noted that suGF1 does not represent the only example of a DNA-binding protein giving two bands on an EMSA gel. In fact, it is very common for multiple complexes to be obtained. Transcription factor Sp1 occurs as two polypeptides of molecular weight 105 kDa and 95 kDa (20;. see Table 7.1 under characteristics). Sp1 has been shown to be glycosylated and phosphorylated (100, 101; see section 3.1). The latter modification was shown to be responsible for the different mobility of the two complexes on EMSA gels, and the presence of two proteins on SDS-PAGE (100). The two Sp1 species are able to bind sequence-specifically to DNA in an independent manner (20). Two protein-DNA complexes are also observed in EMSAs using purified H4TF-1 (44). These complexes were proposed to contain two polypeptides of sizes 105 kDa and 110 kDa as shown by UV crosslinking results (44). The 110 kDa polypeptide was shown to elute at higher ionic strength from a DNA-affinity column, implying a small difference in binding affinity or specificity of the two polypeptides (44). The poly(dG).poly(dC)-binding protein BGP1 may represent another example of two very similar

protein species being present. A doublet at 66-67 kDa has sometimes been resolved on silver-stained SDS-PAGE gels with purified BGP1, but it is not known if both species represent BGP1 (37). A sea urchin factor shown to bind to short oligo(dG).oligo(dC) regions (ectoderm G-string factor; see Table 7.1) also resulted in two major complexes in an EMSA. This factor may be related to or identical to suGF1 (see section 7.6).

7.3 Enriched suGF1 forms multimeric complexes with DNA probably by protein-protein interactions

An interesting feature of highly enriched suGF1 is the occurrence of multimeric species in EMSAs in the presence of excess protein (see section 4.2.7). The distribution of these higher complexes is dependent on the amount of purified suGF1 (see Figure 4.8). The DNase I footprint remains unaltered in complexes B1B2 and B3B4, indicative of the presence of suGF1 in complexes B3 and B4 (see section 6.2.3). The borders of the DNase I footprint are unchanged in complexes B3B4 compared to complexes B1B2. This indicates that no additional significant protein-DNA contacts are made in complexes B3B4 compared to complexes B1B2. The multimerization thus probably involves protein-protein contacts. The conservation of the doublet character (detectable at least in the lower multimers) is intriguing. It was proposed in section 7.2 that complexes B1 and B2 contained suGF1 differing somewhat in size and net charge. If these two suGF1 species could multimerize in any combination through protein-protein interactions, a distribution of complexes in the ratio of 1:2:1 (protein B1 dimer : protein B1,protein B2 heterodimer : protein B2 dimer) would be expected in the position where complexes B3 and B4 are found, based purely on statistical considerations. A third species corresponding to the heterodimer (proteins of B1 and B2) was however not observed. This suggests that the proteins involved in complexes B1 and B2 may show a *bias* against multimerizing with each other.

Multimerization of DNA-binding proteins is a common occurrence. Transcription factor Sp1, for example, has been shown to form large complexes consisting of tetramers stacked in register (150, 179, 215). Such complexes were shown to involve Sp1 molecules bound far apart on the DNA, causing the intervening DNA to be looped (150). Looping may play an integral role in regulation of transcription (see reference 94 for a review). By implication, multimerization may be involved in the function(s) of suGF1. It is interesting to note that suGF1 eluted in the void volume from a Sephacryl S300 gel filtration column, implying the existence of multimeric species in fairly dilute nuclear extracts. Similar behaviour was reported for Sp1 (see Table 7.1 under "characteristics").

7.3 suGF1 DNA-binding is optimal at relatively high ionic strength and is sensitive to chelation of divalent cations by EDTA

suGF1 in nuclear extracts or from purified fractions was shown to bind optimally to the H1-H4 intergenic fragment at 175 mM KCl (see chapter 5). Optimal DNA-binding at relatively high ionic strength has also been shown for other DNA-binding proteins (see for example references 236 and 237), and may reflect the high affinity and specificity of these protein-DNA interactions, needed to bring about the recognition of cognate sequences *in vivo* (15). In addition, hydrophobic interactions may be stabilized and ionic interactions destabilized in the protein-DNA complex at relatively high ionic strength. The optimal ionic strength of the suGF1-DNA interaction is similar to that of the ectoderm G-string factor form sea urchins, but not to that of the endoderm G-string factor (see Table 7.1). The DNA-binding of suGF1 seems to be dependent on ionic strength, rather than being sensitive to the ratio of the divalent cation Mg^{2+} to the monovalent cation K^{1+} (see chapter 5).

The possible requirement for a divalent cation for DNA-binding of purified suGF1 was investigated. It was shown in chapter 5 that dialysis of suGF1 against 50 mM EDTA resulted in a complete loss of DNA-binding activity. This suggests that suGF1 indeed requires a divalent cation for DNA-binding. Such a cation is expected to be complexed very tightly by suGF1, since DNA-binding activity was shown to be insensitive to concentrations of OP which are normally sufficient to result in a removal of all DNA-binding activity from certain other factors. These factors are either known to contain Zn^{2+} (for example TFIIIA (83)), or are proposed to contain Zn^{2+} (for example BGP1 (139) and H4TF-1 (43)). Tight complexing of a divalent cation by suGF1 is supported by the fact that buffers do not need to be supplemented with Zn^{2+} during or after the purification. This is once again in agreement with certain factors (for example BGP1 (37) and Sp1 (20)), but not with others such as H4TF-1 (43, 44). As discussed in chapter 5, the nature and concentration of chelators required to remove DNA-binding activity from different proteins may reflect differences in the affinity constant for bound Zn^{2+} in these proteins. This could be a result of differences in the structures of the DNA-binding domains of these proteins. It was noted in chapter 5 that treatment required for removal of DNA-binding activity from suGF1 was similar to that resulting in a loss of DNA-binding activity of transcription factor Sp1 due to removal of Zn^{2+}. Thus it is possible that suGF1 contains a Zn finger DNA-binding domain, similar to that of Sp1 (see also section 7.5.6).

If suGF1 indeed requires a divalent cation such as Zn^{2+}, the failure to reconstitute DNA-binding of suGF1 after removal of such a divalent cation by dialysis against EDTA can most

probably be ascribed to oxidation of cysteine residues required to complex the metal. Future investigations should be conducted under anaerobic conditions. Alternatively, reconstitution of activity after denaturation by guanidinium chloride in the presence of reducing agents, and subsequent renaturation by gel filtration in the presence of Zn^{2+}, may be investigated.

7.5 A model for suGF1 DNA-binding

In chapter 6 the protein-DNA complex was analyzed by difference probability plots showing the accessibility of the DNA molecule to DNase I and hydroxyl radicals in the suGF1-DNA complex. The individual guanine residues where methylation interfered with suGF1 binding were also identified. These results are summarized in Figure 7.1(A)[12].

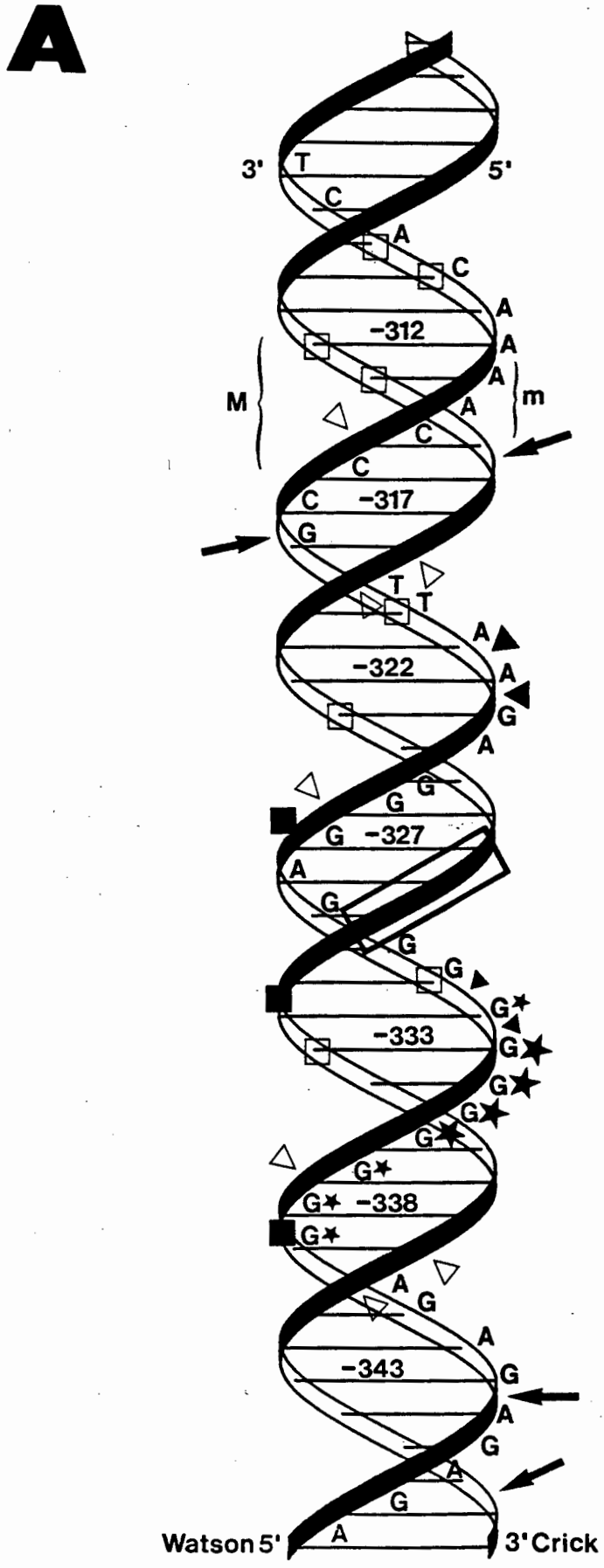
In Figure 7.1(A) the DNA helix is shown schematically as B-type DNA, with a helical repeat of 10,5 bp, and the base-pairs not tilted relative to the helical axis. For the sake of clarity, the values of the ln(difference probability cleavage) for the different probes are not indicated at every position, but local maxima and minima are rather shown. It should thus be kept in mind when referring to Figure 7.1(A) that the accessibility of the DNA to a probe changes gradually between a local maximum and minimum (see difference probability plots in chapter 6).

7.5.1 DNA conformation in the suGF1-DNA complex : evidence consistent with DNA curvature

It is immediately clear from Figure 7.1(A) that local minima of hydroxyl radical cleavage (open squares) occur on only one side of the helix (in phase with the helical repeat of the DNA). These local minima are located on the phosphate backbone of the Watson and Crick strands, opposite each other across the minor groove at the "back" of the helix (the side beneath the plane of the paper). Corresponding local maxima in hydroxyl radical cleavage (open triangles) are located across the minor groove at the "front" of the helix (the side above the plane of the paper), except at position -328 on the Crick strand (see section 7.5.3).

As discussed in chapter 6, the results suggest a periodic narrowing of the minor groove at the back of the helix (decreased accessibility to hydroxyl radical cleavage), and a similar periodic widening at the front of the helix (increased accessibility to hydroxyl radical cleavage) upon

[12] Note that the methylation interference results obtained at 37°C are used to derive a model in this chapter. It should thus be kept in mind that small differences may occur in the complex at 4°C. The general model for the structure is, however, expected to be very similar at 4 and 37°C as evidenced by no significant differences in the hydroxyl radical and DNase I footprints for incubation at the different temperatures (see chapter 6).


A
3'
5'
T
C
A
C
A
-312
A
M
m
A
C
C
C
-317
G
T
T
A
-322
A
G
A
G
G
G
-327
A
G
G
G
G
-333
G
G
G
G
G
G
-338
G
A
G
A
-343
G
A
G
A
G
Watson 5'
3' Crick

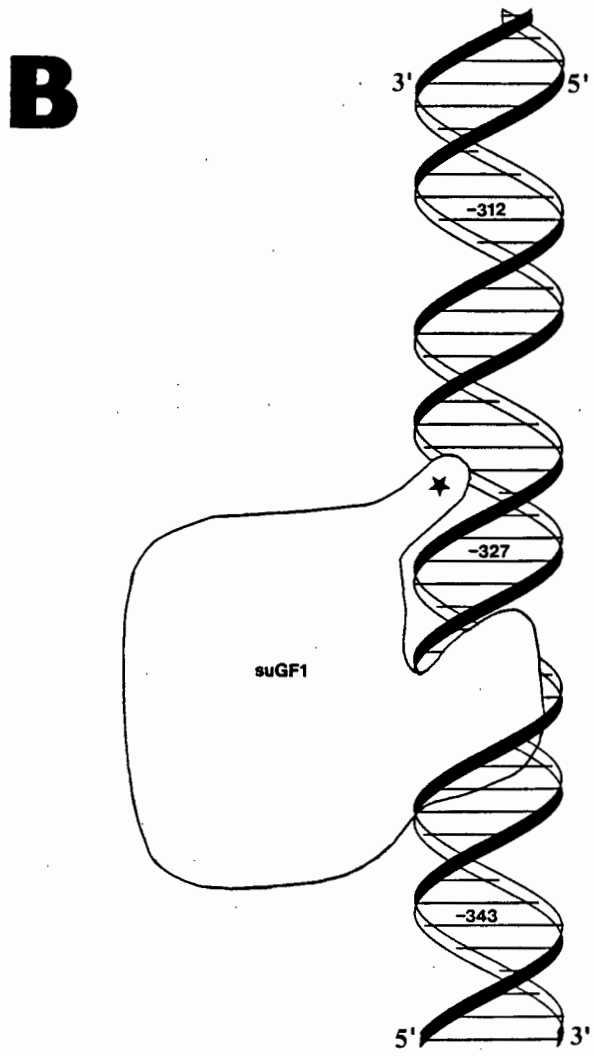


Figure 7.1 A model for suGF1 DNA-binding

(A) Summary of nuclease and chemical protection and methylation interference data presented in chapter 6

The sugar phosphate backbones of a double-stranded DNA helix are schematically represented by ribbons. Horizontal lines represent base-pairs. The polarity of the Watson and Crick strands is indicated at the top and bottom of the helix. The minor and major groove are indicated by "m" and "M" respectively. The sequence of the Watson strand is given. Numbers shown above lines (base-pairs) indicate the sequence position of base-pairs directly underneath, as given in Figure 2.1 (200). DNase I and hydroxyl radical cleavage data are from difference probability plots. Methylation interference data are from results shown in chapter 6. Local minima and maxima of hydroxyl radical cleavage are indicated by open squares and triangles respectively. The open rectangle indicates extended local inaccessibility to hydroxyl radical cleavage. Local minima and maxima of DNase I cleavage are indicated by closed squares and triangles respectively. Small closed triangles indicate small local maxima in DNase I cleavage. Arrows indicate the borders of the DNase I footprint (where the ln(difference probability) returns to zero on either side of the footprint). It is important to note that the entire region between these borders is extensively protected on both strands from DNase I cleavage. Asterisks denote G's which interfere with suGF1 DNA-binding when methylated on the N-7 position. The size of asterisks indicate the relative interference. (Large asterisks indicate more pronounced interference.)

(B) Schematic representation of suGF1-DNA complex

The sugar phosphate backbone and base-pairs of the double helix are presented schematically as in (A). Deviations of the structure of the DNA from B-type are not shown (see text). Sequence positions are indicated above base-pairs, and correspond to those given in (A). The proximity of suGF1 to the DNA was deduced from the nuclease and chemical probing experiments, which are summarized in (A). suGF1 is presented schematically by the structure located to the left of the helix. The presence of the protrusion marked by an asterisk has not been proven unequivocally from the data (see text).

suGF1 binding. Since these results were obtained from a difference function, they must reflect an effect of suGF1 on the DNA conformation, rather than an inherent property of the free DNA molecule. The modulation in the hydroxyl radical cleavage was shown to extend beyond the downstream border of the DNase I footprint (at approximately -317; indicated by arrows), suggesting that suGF1 is not in close contact with the DNA molecule over the entire length of the hydroxyl radical modulation (see chapter 6). The most likely explanation for this periodic modulation in the extent of hydroxyl radical cleavage is a restriction in the freedom of movement of the DNA molecule in the suGF1-DNA complex, where the DNA is limited to conformations closer to that of a curved molecule. The hydroxyl radical cleavage pattern, which gives an indication of the average conformation in the population of molecules, would therefore suggest that the inherent curvature of the free DNA is enhanced in the suGF1-DNA complex (see chapter 6). The radius of curvature of the DNA is in a plane perpendicular to the plane of the page, with the imaginary midpoint of the circle located below the plane of the page (i.e. the ends of the fragment can be thought of as located below the plane of the paper).

It must be emphasized that a sharp local bend at the major groove where suGF1 is thought to contact the DNA (sequence position -335; see section 7.5.3) is not excluded. Although such a bend will not lead to the long range curvature discussed above, the location of a large suGF1 structure in the major groove (-335) may cause local perturbations in DNA structure. This major groove may for example be opened up (widened) and the minor groove "squeezed" at the back, in phase with the deduced curvature. It is difficult to ascribe the exact position (within one or two base-pairs) of the local minima in hydroxyl radical accessibility in the minor groove on the Watson and Crick strands at approximately -333 to any one cause. These local minima are consistent with apparent narrowing of the minor groove due to the restriction of the direction of curvature caused by steric hinderance of bound suGF1. They are also consistent with suGF1-induced narrowing of the minor groove (a sharp bend) as mentioned above. The exact position of these local minima could, however, also be an indication of direct contact of suGF1 from the major groove side to the backbone at position -333 (Crick strand) and/or -331 (Watson strand)[13]. Alternatively, the sugar phosphate backbone could be distorted by suGF1, causing for example a change in sugar pucker at position -333 on the Crick strand, and an accompanying decrease in accessibility to hydroxyl radicals.

The maxima of hydroxyl radical protection located opposite each other across the minor

[13] In view of the probable location of a suGF1 structure in the major groove (at -335; see section 7.5.3), it seems unlikely that suGF1 approaches and contacts the helix from the back (minor groove) at this position. Contacts may however be made to the backbone from the side (see section 7.5.3) or from the major groove where G's are contacted.

groove on the Watson and Crick strands, are displaced in a 3' direction (3'stagger) from the central base-pair. A similar stagger appears in the minima of cleavage on the two strands. The modulation in ln(difference probability cleavage) by hydroxyl radicals cannot be interpreted as a direct indication of the local helical period of the DNA, especially within the borders of the DNase I footprint, for reasons mentioned above. Similarly, the size of the stagger in hydroxyl radical cleavage across the minor groove in the region of suGF1-DNA interaction cannot be used to infer local base-pair tilt. However, if one considers that it is very likely that suGF1 is in close proximity to the DNA molecule only on the left side of the helix (see section 7.5.2), the modulation in ln(difference probability cleavage) by hydroxyl radicals may indeed accurately reflect the local stagger at some positions. The stagger at these positions (for example -312, -317 and -322) is 3 bp in the 3' direction. This would indicate a negative tilt, implying general narrowing of the minor groove and widening of the major groove.

7.5.2 suGF1 seems to be located on one side of the helix in the suGF1-DNA complex

Although the entire region of the Watson strand between -317 and -343[14]is strongly protected from DNase I digestion (arrows denote the borders of the footprint), local maxima (increased accessibility) and minima (increased protection) were detected within the borders of the DNase I footprint in the difference probability plot (see Figure 6.3). It is important to note that these maxima (closed triangles, Figure 7.1(A)) and minima (closed squares) appear in phase with the helical repeat of the DNA, but do not appear in phase with the modulation in the hydroxyl radical cleavage (see Figure 7.1(A)). The reason for this lies in the differences in size and sequence- or structural specificity of the two probes (see section 6.1). The sugar phosphate backbone is protected from the small hydroxyl radicals only when protein is in very close proximity to the backbone, or causes local or global changes in the geometry of the DNA. The large nuclease is, however, much more sensitive to steric hinderance of bound suGF1 when approaching the helix to bind across the minor groove. This leads to decreased cleavage over the whole region of protein proximity (a DNase I footprint). The local maxima in ln(difference probability) occur where DNase I encounters the least steric hinderance in approaching and binding across the minor groove, whereas the local minima are found where the proximity of suGF1 effectively blocks access of the enzyme to the duplex. Because the modulation in DNase I protection is in phase with the helical repeat, it thus seems that the bulk of suGF1 approaches the helix from one side only. This side of approach is delineated by the sites of maximum protection from DNase I cleavage (closed squares), and is approximately from the left side (in the plane of the paper), as the helix is drawn in Figure

[14]Note that DNase I cleaves the O-3'P bond, leaving a 5' phosphate. The position of cleavage is thus indicated 3' to the sequence number.

7.1(A).

The modulation in ln(difference probability cleavage) of the Watson strand is not mirrored in the Crick strand. Instead, the accessibility to DNase I decreases rapidly from the borders of the footprint (denoted by arrows) towards the center of maximum protection (-332)[15] (see Figure 6.3). This position of maximum protection of the Crick strand (-332), together with the large local minimum in cleavage on the Watson strand at -339, is consistent with suGF1 docking into the major groove centered at -335 from the left side of the helix (see Figure 7.1(A)). The opposite side (right side) of the helix is more accessible, as evidenced by the small local maximum in DNase I accessibility at -333 and -332 on the Watson strand.

7.5.3 A large suGF1 structure seems to be located in the major groove at -335 and contacts the phosphate backbone above it

The inaccessibility of the backbone to hydroxyl radical cleavage on the Crick strand around position -328 (indicated by a large open rectangle), suggests that suGF1 contacts the sugar phosphate backbone in this region (see chapter 6). The presence of a local maximum in DNase I cleavage on the Watson strand at approximately -322 or -323, indicates that DNase I can still bind fairly efficiently across the minor groove centered at -324[16]. The proposed direct suGF1-DNA contact to the backbone (Figure 7.1(A), open rectangle) thus probably does not extend beyond -327.

No contacts are made to the phosphate backbone of the Watson strand across the minor groove (approximately -326). The protected phosphate backbone of the minor groove (Crick strand position -328), flanks the major groove where methylation of the N-7 positions of G's was found to interfere with suGF1 DNA-binding (see chapter 6; strongest interference (large asterisks) at positions -333, -334, -335 and -336; significant interference (smaller asterisks) at positions -332, -337, -338 and -339). It thus seems that a structure of suGF1 is located in the major groove around position -335, presumably contacting G's by hydrogen bonding, and also making contacts to the flanking sugar phosphate backbone on the Crick strand located immediately above the site of protein interaction with the G's (refer to Figure 7.1(A)). The location of suGF1 on the left side of the helix (see section 7.5.2) can easily be reconciled with the protrusion of a large protein structure into the major groove at -335 as postulated in this

[15] Note that only the center of protection (-322) and borders of the footprint of the Crick strand are indicated in Figure 7.1(A), due to the lack in a local modulation within the footprint. It should be kept in mind that the whole region of the footpint (Crick strand) is well protected from DNase I digestion in the presence of suGF1 (see Figure 6.3).

[16] Due to the geometry of the DNA molecule, nicks appear on the DNA backbone offset by approximately 1 bp in the 3' direction relative to the position most accessible to DNase I binding across the minor groove.

section.

The absence of a local maximum in accessibility at -326 or -327 on the Crick strand may be ascribed to a combination of steric hinderance of suGF1 in the major groove below the scissile bond, and of steric hinderance or obstruction of binding across the minor groove due to suGF1 contacts to the backbone in the region of -327. The fact that the minor groove (centered at -323) can more easily be bound by DNase I to cleave the Watson strand at -322 or -323, might be explained by less steric hinderance of the suGF1 minor groove contact site in the complex with the Watson strand. The DNase I molecule does not display a two-fold symmetry, and a less accessible environment could be sensed at the minor groove when DNase I is rotated on the DNA to cleave the Crick strand, than to cleave the Watson strand.

7.5.4 The region of suGF1 proximity to the helix seems to be smaller than that delineated by the upstream and downstream borders of the DNase I footprint

The downstream border of the DNase I footprint on the Watson strand is located at position -318 (arrows in Figure 7.1(A)). This bond is cleaved when DNase I binds across the minor groove centered at approximately -319. This local maximum firstly indicates that the minor groove is unprotected by suGF1 at this position. In addition, the major groove (centered at -314) above the scissile bond in the Watson strand (-318), does not contain a bulky protein structure which could interfere with the approach of the extended structure of the DNase I molecule to the phosphate backbone around position -311 (see section 6.1 and Figure 6.1). Similarly, the border of the footprint on the Crick strand (-315; denoted by an arrow) is an indication of minor groove accessibility to DNase I at this position, lack of occupation of the major groove (approximately -319) by a bulky protein group, and no steric hinderance at the Watson strand at approximately -322. The presence of a local maximum in DNase I cleavage on the Watson strand at approximately -322 or -323, supports the absence of interference in the major groove centered at -319.

Accessibility to DNase I at position -321 (Crick strand), is already restricted (see Figure 6.3). This is not due to proximity of suGF1 to the minor groove at position -321, since DNase I binding is unhindered over this groove when the Watson strand is cleaved (border of footprint on Watson strand at -318; ln(difference probability) approximately zero). The nuclease is thus prevented from binding in a position enabling cleavage of the Crick strand by steric hinderance due to the proximity of suGF1. This steric hinderance can result from a structure located in the major groove centered at -324 (see section 7.5.5), or from suGF1 approaching the Watson strand in the region of -327 (see Figure 6.1). Steric hinderance at both locations may also be possible. The downstream border of proximity of suGF1 to the DNA is thus either

where suGF1 protrudes into the major groove at -324 (see section 7.5.5) and/or where suGF1 approaches the phosphate backbone in the region of -327.

The upstream border of the DNase I footprint of suGF1 on the Watson strand is located at -344 (arrow in Figure 7.1(A)). When bound across the minor groove centered at -345, DNase I is free to approach the Crick strand across the major groove in the region of -339. This also suggests that no interfering suGF1 structure is present in the major groove centered at -340. Furthermore, the backbone remains accessible to DNase I at positions located 5' to the upstream border on the Watson strand, suggesting the absence of any further suGF1 major groove occupation upstream of the major groove centered at -340. The inaccessibility of the Crick strand to cleavage at position -342 is therefore most likely due to steric hinderance by suGF1 located in the proximity of this minor groove on the left side of the helix. The accessibility increases rapidly towards the upstream border of the footprint on the Crick strand on the right side of the helix (-347), where no steric hinderance is encountered. The upstream border of proximity of suGF1 to the DNA can thus be mapped to the region approaching the minor groove centered at -340.

7.5.5 An additional suGF1 protrusion into the major groove is possible

When the downstream region of the border of suGF1 proximity to the DNA was discussed above, it was noted that relative inaccessibility to DNase I at position -321 on the Crick strand, could result from (amongst other possibilities) steric hinderance by a structure located in the major groove centered at -324 (see section 7.5.4). Similarly, the local minimum in DNase I cleavage at -327 on the Watson strand cannot be ascribed with certainty to steric hinderance at any one position only, and therefore does not discount the possibility of major groove contacts centered at -324. No contacts to the phosphate backbone of the minor groove flanking this possible protrusion at the front of the helix could be detected by hydroxyl radical footprinting. In fact, a local maximum in accessibility to hydroxyl radical cleavage was located at -326. In addition, if such a protrusion is present in this major groove, it does not appear to be extensive, since no steric hinderance is encountered in the major groove at -319 by DNase I (see section 7.5.4, downstream border of DNase I footprint).

It should be noted that the fact that methylation of the N-7 positions of the G's located at -325 to -327 does not interfere significantly with suGF1 DNA-binding (see chapter 6), cannot be interpreted as a lack of any hydrogen bonding in the major groove at these positions. Hydrogen bonds have for example been shown to be made to the O-6 position of G in the cocrystal of the GAL4-DNA complex. It has also been proposed that Sp1 contacts the G in the central CG bp of the GC-box by hydrogen bonding to the O-6 position (138).

7.5.6 A low resolution model for the suGF1-DNA interaction

The nuclease and chemical probing experiments presented above support a model for the suGF1 DNA-complex schematically presented in Figure 7.1(B)[17]. The essential features of the model represented in this figure are the following :

1) suGF1 approaches the helix from the one side (the left side; see section 7.5.2).

2) The protein is in close proximity to the DNA over a region of approximately 1,5 to 2 helical turns (approximately -340 to -322; see section 7.5.4 and Figure 7.1(B)).

3) A bulky suGF1 structure protrudes from the left side of the helix into the major groove in the region of -335, contacting G's by hydrogen bonding, and contacting the phosphate backbone above the major groove (see section 7.5.3). These contacts to the phosphate backbone may serve to position the major groove structure for specific interactions, as well as stabilizing the interaction.

4) An additional, small protrusion into the major groove at -324 may be present (marked by an asterisk in Figure 7.1(B)). Such a small protrusion may contact G's by hydrogen bonding to the O-6 position (see section 7.5.5). No evidence was found for contacts to the phosphate backbone below this putative protrusion.

5) The DNA-molecule in the suGF1-DNA complex is on average curved with the midpoint of the curvature located below the plane of the paper (see section 7.5.1; note that the possible deviation of the structure of the DNA from classical B-type DNA is not depicted in Figure 7.1(B)). The results are consistent with the curvature of the DNA molecule in the complex extending from a position of at least one helical turn beyond the downstream border of the DNase I footprint, and over the region of suGF1 binding towards the beginning of the $(AG)_{16}$-repeat. There may, however, be a sharp bend or "kink" in the direction of the inherent DNA curvature at the major groove centered at -335 where the main suGF1-DNA contacts are proposed to take place (see section 7.5.1).

6) It seems very unlikely that suGF1 binds to the H1-H4 intergenic fragment as a

[17] The helix is drawn in the same orientation as in Figure 7.1(A).

rotationally symmetric dimer. No obvious rotational symmetry can be detected in the suGF1 nuclease and chemical probing data (see Figure 7.1), and obvious two-fold symmetry is also absent from the sequence of the binding site. Because the ln(difference probability cleavage) values provide information on an average structure for the complex, it is, however, formally possible that suGF1 could bind in different frames to the G-string

The model for the suGF1-DNA complex can be compared to models or features of DNA-binding of other factors. It has been proposed that the factor BGP1 binds in different frames to G-strings (37). BGP1 was found to require 7 contiguous G's for DNA-binding. Binding studies with G-strings of different lengths indicated a simple relationship between binding affinity and the number of possible frames of binding (7 G's each) provided by the G-string. The increase in binding affinity with increased frames of binding was supported by methylation interference studies (37). The G's in the center of a string of 16 or 12 contiguous G's, interfered the most with BGP1 DNA-binding (37). This could be explained statistically by the central G's being present in the majority of possible binding frames (37). All G's were found to interfere with BGP1 DNA-binding in a mutant binding site containing only 7 contiguous G's (37).

Methylation interference results of suGF1 may reflect a similar situation. Methylation of the central G's of the 11 bp G-string were also shown to interfere the most with suGF1 binding (see Figures 6.9 and 6.10). This could be explained by binding in different frames. Based purely on statistical considerations the length of the hypothetical binding frames would be 5 to 7 contiguous G's, to result in the significant enhancement in the methylation interference signal of the central 3 to 4 base-pairs of an 11 bp G-string. The length of a hypothetical binding frame in the suGF1-DNA interaction cannot, however, be predicted solely from these results, especially since there seems to be some preference towards binding in frames located downstream to the central G's (see section 6.3). If suGF1 indeed binds in different frames, it is also probable that the binding affinity may differ for the different frames. Binding in certain frames could for example result in unfavourable interactions to the phosphate backbone, or could be energetically unfavourable by resulting in bending of the DNA in a direction different to that of the preferred curvature in solution. Possible binding of suGF1 in different frames can be assessed by mutations in the length of the G-string in future studies.

Comparison of the model of the suGF1-DNA complex with the protein-DNA complexes of the main DBDs known to date (reviewed in section 1.4) suggests Zn fingers as the most likely candidate for the suGF1 DNA-binding domain, for the following reasons : suGF1 is proposed

to contact DNA in the major groove; this major groove contains 11 contiguous G's, providing three 3 bp GGG subsites for the possible binding of three Zn fingers (see section 1.4); methylation interference implicates at least seven of these G's (5' region) in possible hydrogen suGF1-DNA contacts in the major groove; the suGF1 DNA-binding site does not contain consensus sequences or rotationally symmetrical sequences recognized by double-LZnH domains, Zn-binding class 3 domains, leucine zippers or homeodomains; and contacts to the phosphate backbone are reconcilable with Zn finger domains (see section 1.4). It should, however, be stressed that although suGF1 appears to require a divalent cation for DNA-binding, and the low resolution suGF1-DNA model does not discount the involvement of Zn fingers, the data by no means proves that suGF1 is a Zn finger protein. The suGF1 complex may even belong to an as yet uncharacterized class of DBD. The nuclease and chemical footprinting and interference data show some resemblance to that recently obtained for another DNA-binding protein, Ets-1, proposed to represent a novel DBD (166). The Ets-1-DNA backbone contacts all mapped to one side of the helix within a central 8 bp region. Strongest methylation interference was also observed for G's located on one side of the helix. The Ets-1 DBD is proposed to contact DNA base-pairs in the major groove and the backbone on either side of this major groove contact site (166). suGF1, however, only seems to contact the backbone on one side of the proposed major groove interaction. It is important to note from the Ets-1 study that protection of one strand of the minor groove (flanking the major groove interaction) was supported by ethylation interference. The latter experiment is not expected to be as sensitive to backbone distortions as hydroxyl radicals. Ethylation interference of the suGF1-DNA complex would be useful in confirming direct contacts to the sugar phosphate DNA backbone located above the major groove contact site (indicated by an open rectangle in Figure 7.1(A)).

In conclusion, it should be emphasized that information on the structure of the suGF1-DNA complex at high resolution can only be obtained by determination of the structure of the complex by two-dimensional NMR or X-ray crystallography.

7.6 Comparison of suGF1 with other G-binding proteins

The possible role of oligo(dG).oligo(dC) regions in gene regulation was discussed in chapter 1. Many DNA-binding factors with oligo(dG).oligo(dC)-containing binding sites have been identified. Table 7.1 contains a summary of the properties related to this investigation of selected G-binding factors. Some of these factors may be identical or belong to families of related proteins with similar DNA-binding specificities and functions conserved through evolution. Some of the factors contained in the table have been shown to be distinct from others, based on binding specificities, functional studies, or other properties. This is indicated

Table 7.1 Examples of factors known to bind to oligo(dG).oligo(dC)-containing DNA recognition sites

Factor	Promoter (and DNase I footprint)[1]	Recognition sequence or Consensus binding site[2] (Methylation Interference=*)[3]	Function and Crossreactivity	Purification	Characteristics	Refs
1. Examples of sea urchin proteins with oligo(dG).oligo(dC)-containing recognition sequences						
UHF-1	Early H4 S. purpuratus (-130 to -108)	* ** * CAAGGGGGCGCACTC GTTCCCCCGCGAGAG -122 -108	Positive transcription factor Not Spl or H4TF-1	Not purified	MW 85 kDa (from UV crosslinking)	225 135 226
	Late H4 S. purpuratus (-90 to -69)	* ** ** GTAGGCGGCTCACTC CATCCGCCGAGTGAG * * -84 -70				
NFH3-2	Early H3 of S.purpuratus (-116 to -106)	CCCCCTTCCCGTCA GGGGGAAGGGCAGT -119 -106	Negative transcription factor	Not purified	Requires Mg^{2+} Forms two complexes in EMSA	50

Factor	Promoter (and DNase I footprint)	Recognition sequence or Consensus binding site (Methylation Interference=*)	Function and Crossreactivity	Purification	Characteristics	Refs
Ectoderm and endoderm G-string factors	LpS1β gene of Lytechinus pictus	GGAGCCCCCCT (Only binds CCTCGGGGGGA ectoderm factor) ** * * -730 -720 TGGGGGGCG (Binds ecto- and ACCCCCCGC and endoderm -76 -68 factors)	Positive transcription factor Weak competition for DNA-binding by poly(dG).poly(dC) Might be related to IF-1	Not purified	Ectoderm factor forms two major bands (doublet) and one minor band in EMSA Endoderm factor forms similar but faster migrating doublet in EMSA; Endoderm factor DNA-binding abolished by 0,5 mM EDTA. Ectoderm factor DNA-binding unaffected by 4 mM EDTA Binding activity compared at 100, 200 and 400 mM KCl : Ectoderm factor shows optimal binding at 200 mM, endoderm factor at 100 mM.	236

2. Examples of proteins shown to contain Zn fingers, with oligo(dG).oligo(dC)-containing recognitions sequences

Sp1	Many; for example SV40 early promoter; see refs 110, 20, 58 for other examples	Degenerate; binds with varying affinity to many variations on the GC-box (GGGGCGGGGC); Different consensus sequences proposed : (G/T)GGGCGG(G/A)(G/A)(C/T) (110), (G/T)(G/A)GGCG(G/T)(G/A)(G/A)(C/T) (20)	Positive transcription factor; Ubiquitous in several vertebrates, not found in invertebrates; shown to activate transcription in vivo synergistically;	From HeLa nuclear extracts; -Sephacryl-S300 step, enrichment 15-fold, yield can not be accurately determined; Sp1 appears at 500 kDa -DEAE Sepharose step, no enrichment 75% yield (remove contaminants) -Heparin agarose step, enrichment	3 Zn fingers (sequenced); requires Zn^{2+}; two polypeptides of 105 and 95 kDa, dilute ammonium hydroxide or phosphatases convert all to 95 species, thus phosphorylation responsible for two species; proved to be phosphorylated and glycosylated; both 95 and 105 kDa species show specific DNA-binding when renatured from SDS gel; multimers seen	110 109 102 101 57 74 58 5

Factor	Promoter (and DNase I footprint)	Recognition sequence or Consensus binding site (Methylation Interference=*)	Function and Crossreactivity	Purification	Characteristics	Refs
(Sp1 continued)		Methylation interference not consistent; central C/G bp of GC-box does not interfere, other G's interfere	interacts directly with other transcription factors to mediate synergistic activation	2-fold, yield 53%; activity at 0,3 M KCl; -HPLC (FPLC Mono S), enriched 2,4-fold, yield 45%, activity at 0,25 M KCl -DNA Sp1 Binding Site-affinity chromatography, load in 0,1 M KCl, elute stepwise (0,1 KCl step) from 0,2 to 1,0 M, activity from 0,3 to 0,6 M KCl, Enrichment 83-fold, yield 56%; overall enrichment 6000-fold, overall yield 10%	in EMSA; forms tetramer on binding site and subsequently assembles multiple tetramers stacked in register at DNA loop junction;	150 140 20 111 179 100
Zif268 (EGR-1, Krox-24, NGF1-A)	Binding sites identified upstream of several serum-inducible genes (protects about 20 bp from DNase I)	* * *** * GCG(G/T)GGGCG CGC(C/A)CCCGC *	Positive transcription factor (Not Sp1 and does not bind appreciably to Sp1 site) Belongs to GSG element-binding family	(Cloned and sequenced)	Contains 3 Zn fingers (similar to those of EGR-2 and NGFI-C) Requires Zn^{2+} for DNA-binding Structure of complex with DNA solved by X-ray crystallography MW 82 and 88 kDa Glycosylated and phosphorylated (in virus-infected cells)	137 31 181 35

Factor	Promoter (and DNase I footprint)	Recognition sequence or Consensus binding site (Methylation Interference=*)	Function and Crossreactivity	Purification	Characteristics	Refs
EGR-2 (Krox-20)	Binding sites identified upstream of several serum-inducible genes (protects about 20 bp from DNase I)	* * *** * GCGGGGGCGG CACCCCCGCC *	Identical to that given for Zif268 above	(Cloned and sequenced)	Contains 3 Zn fingers (similar to those of Zif268 and NGFI-C)	34 106
NGF1-C	Binding sites identified upstream of serum-inducible genes	GCGGGGGCG CGCCCCCGC	Identical to that given for Zif268 above	(Cloned and sequenced)	Contains 3 Zn fingers (similar to those of Zif268 and EGR-2) MW approximately 50 kDa	40
WT-ZFP (WT1)	Binding sites identified upstream of serum-inducible genes	GCGGGGGCG CGCCCCCGC	Negative transcription factor Member of GSG element-binding family Does not bind to Sp1 site	(Cloned and sequenced)	Contains 4 Zn fingers Requires Zn^{2+} for binding MW approximately 50 kDa	23 188 147 28

Factor	Promoter (and DNase I footprint)	Recognition sequence or Consensus binding site (Methylation Interference=*)	Function and Crossreactivity	Purification	Characteristics	Refs
3. Examples of other factors with oligo(dG).oligo(dC)-containing recognition sequences, which may be similar to suGF1						
BGP1	chicken β^A-globin (-196 to -176)	GGGGGGGGGGGGGGGCGGGTGGTGGT CCCCCCCCCCCCCCCCGCCCACCACCA -196 ------- -170 Methylation Interference : No interference when wild type sequence (16 contiguousG's) is used; Central G's interfere in a mutant containing 12 G's; all G's interfere in a mutant with only 7 G's (Sp1 recognition sequence underlined) (Note : Numbering as in ref 60; some strains have 18 contiguous G's instead of 16)	May alter stability of a positioned nucleosome to make promoter accessible to other transcription factors; not Sp1, but possible structural resemblance; suGF1 gives footprint over BGP1 binding site	From chicken erythrocytes -Calf thymus DNA-cellulose step; activity elutes at 0,185 M ammonium sulfate) -Poly(dG).poly(dC)-affinity chromatography; load in 0,2 M NaOH, 25 mM ammonium sulfate; elute stepwise with 0,25, 0,3 0,5 M NaOH; Activity at 0,5 M. No yield or enrichment given	Requires Zn^{2+}; MW : doublet at 66-67 kDa (not resolved on all SDS gels; not known if both bands represent BGP1); requires minimum of 7 contiguous G's to bind; binding affinity directly related to number of sites containing 7 G's in a G-string (binds in different frames); G-string is major S1 cleavage site in supercoiled plasmids at low pH, or in nuclei where gene expressed; nucleosome positioned in vitro with borders at -68 and -212 (in clone with 18 G's)	155 212 202 52 133 164 59 60 99 184 113 139 61 37

Factor	Promoter (and DNase I footprint)	Recognition sequence or Consensus binding site (Methylation Interference=*)	Function and Crossreactivity	Purification	Characteristics	Refs
H4TF-1	human H4 (-105 to -82)	TTCCCCTCCCCCACCGGGGCGG AAGGGGAGGGGGTGGCCCCGCG *** ***** ------- -103 -82 (Stippled line indicates Sp1 consensus sequence)	Positive transcription factor Not Sp1	From S-phase HeLa nuclear extracts -Phosphocellulose P11 : stepwise elution (0,3, 0,5, and 0,8 M KCl); activity eluted in 0,5 M step; 80% yield, 10-fold enrichment -Salmon sperm DNA-sepharose; stepwise yield and enrichment of 46% and 2,6-fold respectively; activity eluted in 0,4 M step -Binding Site-affinity chromatography; stepwise yield and enrichment of 8% and 130-fold respectively Overall yield of 3%, and overall enrichment 3550-fold	Requires Zn^{2+} for DNA-binding; two polypeptides MW 105 and 110 kDa (from UV crosslinking); forms two complexes in EMSA; 110 kDa species elutes at higher ionic strength from DNA-affinity column	43 42 44 85
PuF	human *c-myc* P2 promoter (No DNase I footprint detected)	CCCCACCCTCCCCACCC GGGGTGGGAGGGGTGGG *** *** *** *** -138 -122	Positive transcription factor Not AP2	From HeLa nuclear extract; no purification table; -elutes at 0,2 M KCl from heparin-agarose -elutes at 0,4 M KCl from DEAE-sepharose	DNA from -138 to -122 involved in triplex formation *in vitro*, and by introducing an oligonucleotide into nuclei; DNase I hypersensitive site at -125; major S1 nuclease cleavage site at -115; ribonucleoprotein might be involved;	186 185 39 18 89 115

Factor	Promoter (and DNase I footprint)	Recognition sequence or Consensus binding site (Methylation Interference=*)	Function and Crossreactivity	Purification	Characteristics	Refs
(PuF continued)				-elutes at 0,3 M KCl from DNA-affinity column Overall yield and enrichment not given	protein (NSEP-1) has been obtained from HeLa cDNA expression library, claimed to bind specifically to double-and single-stranded site	45 125
"Factor IV"	chicken *c-myc* (-326 to -225)	CTCGGGGGGGGG GAGCCCCCCCCC -236 -225	Function not known Footprint obtained over recognition sequence with suGF1	Not purified	MW in the region of 100 kDa (roughly determined by gel filtration)	142
ME1a1 factor	murine *c-myc* P2 promoter (-57 to -48)	AGGGGAGGGA -57 -48	Positive transcription factor Seems distinct from Sp1	Not purified		81

1 The sequence given in brackets below the promoter shows the region protected from DNase I digestion where determined.

2 The double-stranded recognition sequence or concensus binding site is shown.

3 Guanines which interfere with factor DNA-binding when methylated are indicated by asterisks, where determined.

under "Function and Crossreactivity" in Table 7.1. suGF1 was compared to these proteins throughout this discussion. In this section, the possibility that suGF1 is very similar or identical to any factor in Table 7.1, is discussed.

The sea urchin G-binding factors UHF-1 and NFH3-2 (see Table 7.1) have not been purified. The size of UHF-1 (85 kDa, determined by UV crosslinking) is different to that of suGF1. NFH3-2 has been shown to require exogenously added Mg^{2+} for DNA-binding, whereas this is not the case for suGF1. In addition, the binding sites of these factors do not contain very long regions of contiguous G's. This suggests that suGF1 may not be very similar to these factors. suGF1 may however be related to the sea urchin G-string factors identified in the ectoderm or endoderm (236; see Table 7.1). Two specific factor-DNA complexes are obtained in EMSAs for these factors (236), similar to suGF1-DNA complexes B1 and B2. In addition, DNA-binding of the ectoderm factor has been shown to be insensitive to the presence of 4 mM EDTA, and this factor bound optimally to DNA at relatively high ionic strength (see Table 7.1). There is functional evidence that the G-string factors are involved in regulation of transcription of sea urchin LpSα and LpSβ genes (236). No data is available on the chromatin structure of these genes, and it is not known if the G-string factors play a role in transcription *via* interaction with nucleosomes. The ectoderm G-string factor is not present in unfertilized eggs, but can be found in blastula stage sea urchins. The endoderm factor is present in unfertilized eggs. If suGF1 is indeed identical to the G-string factors, it may be a general transcription factor found in sea urchins. It was however, reported that poly(dG).poly(dC) did not compete strongly for binding of the ecto- or endoderm factors to G-strings containing 6 contiguous G's. It remains to be established if these factors are the same as or related to suGF1.

The molecular weight of suGF1 differs from that of the GSG-element (GCGGGGGCG)-binding proteins contained in Table 7.1. In addition it is not known if suGF1 binds with high affinity to contiguous G's interrupted by C's as is the case for these proteins. Sp1 and suGF1 are similar in many aspects, for example the occurrence of multimers, similar sensitivity to metal ion chelators (see chapter 5), and DNA-binding to G-rich recognition sequences. Sp1 is a general, ubiquitous positive transcription factor (see Table 7.1), whereas the distribution or function of suGF1 is not known. The size of suGF1 has been determined as 59,5 kDa whereas Sp1 is known to exist as two polypeptides of 105 and 95 kDa. Furthermore, although Sp1 can bind to a GC-box with the central C mutated to a G, this does not represent a high affinity binding site for Sp1 (138).

Some similarities can also be found between suGF1 and H4TF-1. There is however a large difference in size between the two proteins, and H4TF-1 requires the addition of Zn^{2+} to all

buffers during the purification. suGF1 may be related to PuF or Me1a1, but no conclusions can be drawn from the available data. A DNase I footprint on the chicken *c-myc* promoter *in vitro* has been ascribed to a DNA-binding factor (factor IV). This footprint includes 9 contiguous G's, and sea urchin nuclear extracts containing suGF1 have been shown to give a similar footprint on the *c-myc* promoter over the region of the G's (D. Patterton, J.P. Hapgood, unpublished). It is possible that this factor is related to suGF1.

Many similarities can be found between suGF1 and BGP1. The molecular weights of the two proteins only differ by approximately 7 kDa (see Table 7.1). Both proteins bind specifically and with high affinity to oligo(dG).oligo(dC) sequences, and can be purified by poly(dG).poly(dC)-affinity chromatography. Neither BGP1 nor suGF1 requires the addition of Zn^{2+} to buffers during the purification. BGP1 appears to require Zn^{2+} for DNA-binding, and suGF1 may have a similar requirement. In addition, a clear DNase I footprint is obtained in vitro with sea urchin nuclear extracts containing suGF1 over the G_{16}-string in the β^A-globin promoter where BGP1 binds (D. Patterton, J.P. Hapgood, unpublished). BGP1 is a tissue- and developmental stage specific chicken factor. suGF1 is thought to be present at different developmental stages during embryonic development of sea urchins, based on DNA-binding activity (J.P. Hapgood, unpublished). The precise stage-specific distribution of suGF1 is however not known. The function of BGP1 and suGF1 may be similar, and may involve interactions with positioned nucleosomes as discussed in chapter 1. suGF1 may thus be the sea urchin equivalent of BGP1.

CHAPTER 8

REFERENCES

1) Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. & Gehring, W. J. (1990). DNA binding properties of the purified Antennapedia homeodomain. *Proc. Natl. Acad. Sci. USA*, 87, 4093-4097.

2) Alberts, B. & Herrick, G. (1971). DNA-cellulose chromatography. *Meth. Enzymol.* 21, 198-217.

3) Anderson, G. M. & Freytag, S. O. (1991). Synergistic activation of a human promoter *in vivo* by transcription factor Sp1. *Mol. Cell. Biol.* 11, 1935-1943.

4) Anello, L., Albanese, I., Casano, C., Palla, F., Gianguzza, F., Di Bernardo, M. G., Di Marzo, R. & Spinelli, G. (1986). Different micrococcal nuclease cleavage patterns characterize transcriptionally active and inactive sea-urchin histone genes. *Eur. J. Biochem.* 156, 367-374.

5) Arndt-Jovin, D. J., Jovin, J. M., Bähr, W., Frischauf, A. -M. & Marquardt, M. (1975). Covalent attachment of DNA to agarose. Improved synthesis and use in affinity chromatography. *Eur. J. Biochem.* 54, 411-418.

6) Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987). *Current Protocols in Molecular Biology*, John Wiley and Sons, New York.

7) Baldwin, A. S. (1990). Analysis of sequence-specific DNA-binding proteins by the gel mobility shift assay. *DNA and Protein Engineering Techniques.* 2, 73-88.

8) Baleja, J. D., Marmorstein, R., Harrison, S. C. & Wagner, G. (1992). Solution structure of the DNA-binding domain of Cd_2-GAL4 from *S. cerevisiae. Nature (London)*, 356, 450-453.

9) Barberis, A., Superti-Furga, G. & Busslinger, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell*, 50, 347-359.

10) Barberis, A., Superti-Furga, G., Vitelli, L., Kemler, I. & Busslinger, M. (1989). Developmental and tissue-specific regulation of a novel transcription factor of the sea urchin. *Genes & Dev.* 3, 663-675.

11) Bensadouan, A. & Weinstein, D. (1976). Assay of proteins in the presence of interfering materials. *Anal. Biochem.* 70, 241-250.

12) Berg, J. M. (1986). Potential metal-binding domains in nucleic acid binding proteins. *Science*, 232, 485-487.

13) Berg, J. M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci. USA*, 85, 99-102.

14) Berg, J. M. (1990). Zinc fingers and other metal-binding domains. *J. Biol. Chem.* 265, 6513-6516.

15) Berg, O. G. & von Hippel, P. H. (1988). Selection of DNA binding sites by regulatory proteins. *Trends Biochem. Sci.* 13, 207-211.

16) Berk, A. J. (1989). Regulation of eukaryotic transcription factors by post-translational modification. *Biochim. Biophys. Acta*, 1009, 103-109.

17) Blanks, R. & McLaughlin, L. W. (1988). An oligonucleotide affinity column for the isolation of sequence specific DNA binding proteins. *Nucl. Acids Res.* 16, 10283-10299.

18) Boles, T. C. & Hogan, M. E. (1987). DNA structure equilibria in the human *c-myc* gene. *Biochemistry*, 26, 367-376.

19) Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1985). Quantitative DNase footprint titration: A method for studying protein-DNA interactions. *Meth. Enzymol.* 130, 132-181.

20) Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor Sp1. *Science*, 234, 47-52.

21) Brown, R. S., Sander, C. & Argos, P. (1985). The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.* 186, 271-274.

22) Bryan, P. N., Olah, J. & Birnstiel, M. L. (1983). Major changes in the 5' and 3' chromatin structure of sea urchin histone genes accompany their activation and inactivation in development. *Cell*, 33, 843-848.

23) Buckler, A. J., Pelletier, J., Haber, D. A., Glaser, T. & Housman, D. E. (1991). Isolation, characterization, and expression of the Murine Wilms' Tumor gene (WT1) during kidney development. *Mol. Cell. Biol.* 11, 1707-1712.

24) Burgess, R. R. (1969). A new method for the large scale purification of *Escherichia coli* deoxyribonucleic acid-dependent ribonucleic acid polymerase. *J. Biol. Chem.* 244, 6160-6167.

25) Burkhoff, A. M. & Tullius, T. D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell*, 48, 935-943.

26) Busch, S.J. & Sassone-Corsi, P. (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet.* 6, 36-40.

27) Busslinger, M. & Barberis, A. (1985). Synthesis of sperm and late histone cDNA of the sea urchin with a primer complementary to the conserved 3' terminal palindrome: Evidence for tissue-specific and more general histone gene variants. *Proc. Natl. Acad. Sci. USA*, 82, 5676-5680.

28) Call, K. M., Glaser, T., Ito, C. Y., Buckler, A. J., Pelletier, J., Haber, D. A., Rose, E. A., Kral, A., Yeger, H., Lewis, W. H., Jones, C. & Housman, D. E. (1990). Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' Tumor Locus. *Cell*, 60, 509-520.

29) Calzone, F. J., Hoög, C., Teplow, D. B., Cutting, A. E., Zeller, R. W., Britten, R. J. & Davidson, E. H. (1991). Gene regulatory factors of the sea urchin embryo. I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein. *Development*, 112, 335-350.

30) Calzone, F. J., Thézé, N., Thiebaud, P., Hill, R. L., Britten, R. J. & Davidson, E. H. (1988). Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes & Dev.* 2, 1074-1088.

31) Cao, X., Koski, R. A., Gashler, A., McKiernan, M., Morris, C. F., Gaffney, R., Hay, R. V. & Sukhatme, U. P. (1990). Identification and characterization of the Egr-1 gene product, a DNA-binding zinc finger protein induced by differentiation and growth signals. *Mol. Cell. Biol.* 10, 1931-1939.

32) Carey, J. (1991). Gel retardation. *Meth. Enzymol.* 208, 103-117.

33) Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F. & Ptashne, M. (1989). An amino-terminal fragment of GAL4 binds DNA as a dimer. *J. Mol. Biol.* 209, 423-432.

34) Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dollé, P., Dúboule, D. & Charnay, P. (1990). The segment-specific gene *Krox-20* encodes a transcription factor with binding sites in the promoter region of the *Hox-1.4* gene. *EMBO J.* 9, 1209-1218.

35) Christy, B. & Nathans, D. (1989). DNA binding site of the growth factor-inducible protein Zif268. *Proc. Natl. Acad. Sci. USA*, 86, 8737-8741.

36) Churchill, M. E. A., Tullius, T. D. & Klug, A. (1990). Mode of Interaction of the zinc finger protein TFIIIA with a 5S RNA gene of *Xenopus*. *Proc. Natl. Acad. Sci. USA* 87, 5528-5532.

37) Clark, S. P., Lewis, C. D. & Felsenfeld, G. (1990). Properties of BGP1, a poly(dG)-binding protein from chicken erythrocytes. *Nucl. Acids Res.* 18, 5119-5126.

38) Colin, A. M., Catlin, T. L., Kidson, S. H. & Maxon, R. (1988). Closely linked early and late histone H2B genes are differentially expressed after microinjection into sea urchin zygotes. *Proc. Natl. Acad. Sci. USA*, 85, 507-510.

39) Cooney, M., Czernuszewics, G., Postel, E. H., Flint, S. J. & Hogan, M. E. (1988). Site-specific oligonucleotide binding represses transcription of the human *c-myc* gene *in vitro*. *Science*, 241, 456-459.

40) Crosby, S. D., Puetz, J. J., Simburger, K. S., Fahrner, T. J. & Milbrandt, J. (1991). The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGGGGGCG (GSG) element-binding protein family. *Mol. Cell. Biol.* 11, 3835-3841.

41) Crothers, D. M., Gartenberg, M. R. & Shrader, T. E. (1991). DNA bending in protein-DNA complexes. *Meth. Enzymol.* 208, 118-147.

42) Dailey, L., Hanly, S. M., Roeder, R. G. & Heintz, N. (1986). Distinct transcription factors bind specifically to two regions of the human histone H4 promoter. *Proc. Natl. Acad. Sci. USA*, 83, 7241-7245.

43) Dailey, L., Roberts, S. B. & Heintz, N. (1987). RNA polymerase II transcription factors H4TF-1 and H4TF-2 require metal to bind specific DNA sequences. *Mol. Cell. Biol.* 7, 4582-4584.

44) Dailey, L., Roberts, S. B. & Heintz, N. (1988). Purification of the human histone H4 gene-specific transcription factors H4TF-1 and H4TF-2. *Genes & Dev.* 2, 1700-1712.

45) Davis, T. L., Firulli, A. B. & Kinniburgh, A. J. (1989). Ribonucleoprotein and protein factors bind to an H-DNA forming *c-myc* DNA element. Possible regulators of the *c-myc* gene. *Proc. Natl. Acad. Sci. USA*, 86, 9682-9686.

46) Davison, B. L., Leighton, T., Rabinowitz, J. C. (1979). Purification of *Bacillus subtilis* RNA polymerase with heparin-agarose. *J. Biol. Chem.* 254, 9220-9226.

47) de Groot, P. (1982). Histone H1 of the sea urchin embryo, partial structures, enzymatic modifications and developmental programme. Masters Thesis, University of Cape Town.

48) Depew, R. E. & Wang, J. C. (1975). Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. USA*, 72, 4275-4279.

49) Dignam, J. D., Lebovitz, R. M., Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucl. Acids Res.* 11, 1475-1489.

50) DiLiberto, M., Lai, Z., Fei, H. & Childs, G. (1989). Developmental control of promoter-specific factors responsible for the embryonic activation and inactivation of the sea urchin early histone H3 gene. *Genes & Dev.* 3, 973-983.

51) Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A. & Tullius, T. D. (1991). Hydroxyl radical footprinting. *Meth. Enzymol.* 208, 380-413.

52) Dolan, M., Dodgson, J. B. & Engel, J. D. (1983). Analysis of the adult chicken β-globin gene. Nucleotide sequence of the locus, microheterogeneity at the 5'-end of β-globin mRNA, and aberrant nuclear RNA species. *J. Biol. Chem.* 258, 3983-3990.

53) Drew, H. R. (1984). Structural specificities of five commonly used DNA nucleases. *J. Mol. Biol.* 176, 535-557.

54) Drew, H. R., McCall, M. J. & Calladine, C. R. (1990). New approaches to DNA in the crystal and in solution. In *DNA Topology and its Biological Effects*. Cozzarelli, N. R. & Wang, J. C., eds. (Cold Spring Harbor Laboratory Press), pp. 1-56.

55) Drew, H. R. & Travers, A. A. (1984). DNA structural variations in the *E. coli tyrT* promoter. *Cell*, 37, 491-502.

56) Dynan, W. S., Sazer, S., Tjian, R. & Schimke, R. T. (1986). Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolase reductase promoter. *Nature (London)*, 319, 246-248.

57) Dynan, W. S. & Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*, 35, 79-87.

58) Dynan, W. S. & Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature (London)*, 316, 774-778.

59) Emerson, B. M. & Felsenfeld, G. (1984). Specific factor conferring nuclease hypersensitivity at the 5' end of the chicken adult β-globin gene. *Proc. Natl. Acad. Sci. USA*, 81, 95-99.

60) Emerson, B. M., Lewis, C. D. & Felsenfeld, G. (1985). Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin Gene: Nature of the binding domain. *Cell*, 41, 21-30.

61) Emerson, B. M., Nickol, J. M. & Fong, T. C. (1989). Erythroid-specific activation and derepression of the chick β-globin promoter *in vitro*. *Cell*, 57, 1189-1200.

62) Fairall, L., Rhodes, D. & Klug, A. (1986). Mapping of the sites of protection on a 5S RNA gene by the *Xenopus* transcription factor IIIA. A Model for the interaction. *J. Mol. Biol.* 192, 577-591.

63) Falchuk, K. H., Mazus, B., Ulpino, L. & Vallee, B. L. (1976). *Euglenia gracilis* DNA dependent RNA polymerase II: A zinc metalloenzyme. *Biochemistry*, 15, 4468-4475.

64) Frankel, A. D. & Kim, P. S. (1992). Modular structure of transcription factors: Implications for gene regulation. *Cell*, 65, 717-719.

65) Freedman, L. P., Luisi, B. F., Korszun, Z. R., Basavappa, R., Sigler, P. B. & Yamamoto, K. R. (1988). The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature (London)*, 334, 543-546.

66) Freemont, P. S., Hanson, I. M. & Trowsdale, J. (1991). A novel cysteine-rich sequence motif. *Cell*, 64, 483-484.

67) Fried, M. & Crothers, D. M. (1981). Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucl. Acids Res.* 9, 6505-6525.

68) Fronk, J., Tank, G. A. & Langmore, J. P. (1990). Chromatin structure of the developmentally regulated early histone genes of the sea urchin *Strongylocentrotus purpuratus*. *Nucl. Acids Res.* 18, 5255-5263.

69) Gabrielsen, O. D., Hornes, G., Korsnes, L., Rvet, A. & Øyen, T. B. (1989). Magnetic DNA affinity purification of yeast transcription factor J - a new purification principle for the ultrarapid isolation of near homogeneous factor. *Nucl. Acids Res.* 17, 6253-6267.

70) Gander, I., Foeckler, R., Rogge, L., Meisterernst, M., Schneider, R., Mertz, R., Lottspeich, F. & Winnacker E. -L. (1988). Purification methods for the sequence-specific DNA-binding protein nuclear factor I (NFI) - generation of protein sequence information. *Biochim. Biophys. Acta*, 951, 411-418.

71) Gardner, K. H., Pan, T., Narula, S., Rivera, E. & Coleman, J. E. (1991). Structure of the binuclear metal-binding site in the GAL4 transcription factor. *Biochemistry*, 30, 11292-11302.

72) Garner, M. M., Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucl. Acids Res.* 9, 3047-3060.

73) Gartenberg, M. R., Ampe, C., Steitz, T. A. & Crothers, D. M. (1990). Molecular characterization of the GCN4-DNA complex. *Proc. Natl. Acad. Sci. USA*, 87, 6034-6038.

74) Gidoni, D., Dynan, W. S. & Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature (London)*, 312, 409-413.

75) Giniger, E., Varnum, S. M. & Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, 40, 767-774.

76) Gottlieb, J. & Muzyczka, N. (1990). Purification and characterization of HeLa endonuclease R. *J. Biol. Chem.* 265, 10836-10841.

77) Gottlieb, J. & Muzyczka, N. (1990). Substrate specificity of HeLa endonuclease R. *J. Biol. Chem.* 265, 10842-10850.

78) Griffith, V. J. (1988). The role of phosphorylation in chromatin condensation. Honnours project, University of Cape Town.

79) Hagerman P. J. (1990). Sequence-directed curvature of DNA. *Ann. Rev. Biochem.* 59, 755-781.

80) Hagerman, P. J. (1985). Sequence dependence of the curvature of DNA: A test of the phasing hypothesis. *Biochemistry*, 24, 7033-7037.

81) Hall, D. J. (1990). Regulation of *c-myc* transcription *in vitro*.: dependence of the guanine-rich promoter element Me1a1. *Oncogene*, 5, 47-54.

82) Halvorsen, Y. -D. C., Nandubalan, K. & Dickson, R. C. (1991). Identification of base and backbone contacts used for DNA sequence recognition and high-affinity binding by GAL9, a transcription activator containing a C6 zinc finger. *Mol. Cell. Biol.* 11, 1777-1784.

83) Hanas, J. S., Hazuda, D. J., Bugenhagen, D. F., Wu, F. Y. -H. & Wu, C. -W. (1983). *Xenopus* transcription factor A requires zinc for binding to the 5S RNA gene. *J. Biol. Chem.* 258, 14120-14125.

84) Hanes, S. D. & Brent, R. (1991). A genetic model for interaction of the homeodomain recognition helix with DNA. *Science*, 251, 426-429.

85) Hanly, S. M., Bleecker, G. C., Heintz, N. (1985). Identification of promoter elements necessary for transcriptional regulation of a human histone H4 gene *in vitro*. *Mol. Cell. Biol.* 5, 380-389.

86) Härd, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J, -Å. & Kaptein, R. (1990). Solution structure of the glucocorticoid DNA-binding domain. *Science*, 249, 157-159.

87) Harrison, S. C. (1991). A structural taxonomy of DNA-binding domains. *Nature (London)*, 353, 715-719.

88) Harrison, S. C. & Aggarwal, A. K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Ann. Rev. Biochem.* 59, 933-969.

89) Hay, N., Bishop, M. & Levens, D. (1987). Regulatory elements that modulate expression of human *c-myc*. *Genes & Dev.* 1, 659-671.

90) Hayashi, S. & Schott, M. P. (1990). What determines the specificity of action of *Drosophila* homeodomain proteins? *Cell*, 63, 883-894.

91) Hentschel, C. C. (1982). Homocopolymer sequences in the spacer of the sea urchin histone gene repeat are sensitive to S1 nuclease. *Nature (London)*, 295, 714-716.

92) Hentschel, C. C. & Birnstiel, M. L. (1981). The organization and expression of histone gene families. *Cell*, 25, 301-313.

93) Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. *Science*, 234, 451-457.

94) Hochschild, A. (1990). Protein-protein interactions and DNA loop formation. In *DNA Topology and its Biological Effects*. Cozzarelli, N. R. & Wang, J. C., eds. (Cold Spring Harbor Laboratory Press), pp. 107-138.

95) Horowitz, M. I. (1977). Isolation and purification of glycoproteins. In *The Glycoconjugates*, Vol. 1. Horowitz, M. I. & Pigman, W., eds. (Academic Press), pp. 15-26.

96) Huber, P. W., Morii, T., Mei, H.-Y. & Barton, J. K. (1991). Structural polymorphism in the major groove of a 5S RNA gene complements the zinc finger domains of transcription factor IIIA. *Proc. Natl. Acad. Sci. USA*, 88, 10801-10805.

97) Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L. & Rosenfeld, M. G. (1990). The POU-specific domain of Pit-1 is essential for sequence-specific, high-affinity DNA binding and DNA-dependent Pit-1-Pit-1 interaction. *Cell*, 61, 1021-1033.

98) Ishii, S., Kadanoga, J. T., Tjian, R., Brady, J. N., Merline, G. T. & Pastan, I. (1986). Binding of the Sp1 transcription factor by the human Harvey *ras1* proto-oncogene promoter. *Science*, 232, 1410-1413.

99) Jackson, P. D. & Felsenfeld, G. (1985). A method for mapping intranuclear protein-DNA interactions and its application to a nuclease hypersensitive site. *Proc. Natl. Acad. Sci., USA* 82, 2296-2300.

100) Jackson, S. P., MacDonald, J. J., Lees-Miller, S. & Tjian, R. (1990). GC Box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. *Cell*, 63, 155-165.

101) Jackson, S. P. & Tjian, R. (1988). O-Glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. *Cell*, 55, 125-133.

102) Jackson, S. P. & Tjian, R. (1989). Purification and analysis of RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. *Proc. Natl. Acad. Sci. USA*, 86, 1781-1785.

103) Janasoff, A., Kochoyan, M., Fraenkel, E., Lee, J. P. & Weiss, M. A. (1992). Aromatic-Aromatic interactions in the zinc finger motif. Analysis of the two-dimensional nuclear magnetic resonance structure of a mutant domain. *J. Mol. Biol.* 225, 1035-1047.

104) Johnson, P. F. & McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. *Ann. Rev. Biochem.* 58, 799-839.

105) Jones, K. A., Kadanoga, J. T., Luciw, P. A. & Tjian, R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor Sp1. *Science*, 232, 755-759.

106) Joseph, L. J., Beau, M. M., Jamieson, G. A., Acharya, S., Shows, T. B., Rowley, J. D. & Sukhatme, V. P. (1988). Molecular cloning, sequencing and mapping of *EGR2*, a human early growth response gene encoding a protein with "zinc-binding finger" structure. *Proc. Natl. Acad. Sci. USA*, 85, 7164-7168.

107) Kadanoga, J. T. (1990). Sequence-specific DNA affinity chromatography. DNA and Protein Engineering Techniques, 2, 82-87.

108) Kadanoga, J. T. (1991). Purification of sequence-specific binding proteins by DNA affinity chromatography. *Meth. Enzymol.* 208, 10-23.

109) Kadanoga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell*, 51, 1079-1090.

110) Kadanoga, J. T., Jones, K. A. & Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem. Sci.* 11, 20-23.

111) Kadanoga, J. T. & Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* 83, 5889-5893.

112) Kaumeyer, J. F. & Weinberg, E. S. (1986). Sequence organization and expression of late embryonic H3 and H4 histone genes from the sea urchin, *Strongylocentrotus purpuratus*. *Nucl. Acids Res.* 14, 4557-4576.

113) Kefalas, P., Gray, F. C. & Allan, J. (1988). Precise nucleosome positioning in the promoter of the chicken β^A globin gene. *Nucl. Acids Res.* 16, 501-517.

114) Kerppola, T. K. & Curran, T. (1991). Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: Implications for transcription factor cooperativity. *Cell*, 66, 317-326.

115) Kinniburgh, A. J. (1989). A *cis*-acting transcription element of the *c-myc* gene can assume an H-DNA conformation. *Nucl. Acids Res.* 17, 7771-7778.

116) Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, J. B. & Pabo, C. O. (1990). Crystal structure of an *engrailed* homeodomain-DNA complex at 2.8Å resolution: A framework for understanding homeodomain-DNA interactions. *Cell*, 63, 579-590.

117) Klein-Hitpass, L., Tsai, S. Y., Greene, G. L., Clark, J. H., Tsai, M. -J. & O'Malley, B. W. (1989). Specific binding of estrogen receptor to the estrogen response element. *Mol. Cell. Biol.* 9, 43-49.

118) Klevit, R. E. (1991). Recognition of DNA by Cys_2 His_2 zinc fingers. *Science*, 253, 1367 & 1393.

119) Knowles, J. A. & Child, G. J. (1984). Temporal expression of late histone messenger RNA in the sea urchin *Lytechinus pictus*. *Proc. Natl. Acad. Sci. USA* 81, 2411-2415.

120) Kohwi Y. & Kohwi-Shigematsu, T. (1988). Magnesium ion-dependent homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc. Natl. Acad. Sci. USA*, 85, 3781-3785.

121) Kohwi, Y. & Kohwi-Shigematsu, T. (1991). Altered gene expression correlates with DNA structure. *Genes & Dev.* 5, 2547-2554.

122) Kohwi-Shigematsu, T., Gelinas, R. & Weintraub, H. (1983). Detection of an altered DNA conformation at specific sites in chromatin and supercoiled DNA. *Proc. Natl. Acad. Sci. USA*, 80, 4389-4393.

123) Kohwi-Shigematsu, T. & Kohwi, Y. (1985). Poly(dG).poly(dC) sequences, under torsional stress, induce an altered DNA conformation upon neighbouring DNA sequences. *Cell*, 43, 199-206.

124) Kohwi-Shigematsu, T. & Kohwi, Y. (1991). Detection of triple-helix related structures adopted by poly(dG).poly(dC) sequences in supercoiled plasmid DNA. *Nucl. Acids. Res.* 19, 4267-4271.

125) Kolluri, R. & Kinniburgh, A. J. (1991). Full length cDNA sequence encoding a nuclease-sensitive element DNA binding protein. *Nucl. Acids Res.* 19, 4771.

126) Kornberg, R. D. & Lorch, Y. (1991). Irresistible force meets immovable object: Transcription and the nucleosome. *Cell*, 67, 833-836.

127) Kraulis, P. J., Raine, A. R. C., Gudhavi, P. L. & Laue, E. D. (1992). Structure of the DNA-binding domain of zinc GAL4. *Nature (London)*, 356, 448-450.

128) Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)*, 227, 680-685.

129) Lahm, A. & Suck, D. (1991). DNase I-induced DNA conformation. 2Å structure of a DNase I-octamer complex. *J. Mol. Biol.* 221, 645-667.

130) Lai, Z. -C., Maxson, R. & Childs, G. (1988). Both basal and ontogenic promoter elements affect the timing and level of expression of a sea urchin H1 gene during early embryogenesis. *Genes & Dev.* 2, 173-183.

131) Lamb, P. & McKnight, S. L. (1991). Diversity and specificity in transcriptional regulation : the benefits of heterotypic dimerization. *Trends Biochem. Sci.* 16, 417-422.

132) Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988). The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science*, 240, 1759-1764.

133) Larsen, A. & Weintraub, H. (1982). An altered DNA conformation detected by S1 Nuclease occurs at specific regions in active chick globin chromatin. *Cell*, 29, 609-622.

134) Leblond-Francillard, M., Dreyfus, M. & Rougeon, F. (1987). Isolation of DNA-protein complexes based on streptavidin and biotin interaction. *Eur. J. Biochem.* 166, 351-355.

135) Lee, I. J., Tung, L., Bumcrot, D. A., Weinberg, E. S. (1991). UHF-1 a factor required for maximal transcription of early and late sea urchin histone H4 genes: Analysis of promoter-binding sites. *Mol. Cell. Biol.* 11, 1048-1061.

136) Lee, M. S., Gippert, G. P., Sodam, K. V., Case, D. A. & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, 245, 635-637.

137) Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R. & Charnay, P. (1990). The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptional activator. *Mol. Cell. Biol.* 10, 3456-3467.

138) Letovsky, J. & Dynan, W. S. (1989). Measurement of the binding of transcription factor Sp1 to a single GC recognition sequence. *Nucl. Acids Res.* 17, 2639-2653.

139) Lewis, C. D., Clark, S. P., Felsenfeld, G. & Gould, H. (1988). An erythroid-specific protein that binds to the poly(dG) region of the chicken β-globin gene promoter. *Genes & Dev.* 2, 863-873.

140) Li, R., Knight, J. D., Jackson, S. P., Tjian, R. & Botchan, M. R. (1991). Direct interactions between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. *Cell*, 65, 493-505.

141) Lichtsteiner, S. & Schibler, U. (1989). A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. *Cell*, 57, 1179-1187.

142) Lobanenkov, V. V., Nicolas, R. H., Plumb, M. A., Wright, C. A. & Goodwin, G. H. (1986). Sequence-specific DNA-binding proteins which interact with (G+C)-rich sequences flanking the chicken *c-myc* gene. *Eur. J. Biochem.* 159, 181-188.

143) Luisi, B. (1992). Zinc standard for economy. *Nature (London)*, 356, 379-380.

144) Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. & Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature (London)*, 352, 497-505.

145) Lutter, L. C. (1978). Kinetic analysis of Deoxyribonuclease I cleavages in the nucleosome core: evidence for a DNA superhelix. *J. Mol. Biol.* 124, 391-420.

146) Lyamichev, V. I., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1985). A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J. Biomolec. Struct. Dynam.* 3, 327-338.

147) Madden, S. L., Cook, D. M., Morris, J. F., Gashler, A., Sutkhatme, V. P. & Rauscher, F. J. (1991). Transcriptional repression mediated by the WT1 Wilms tumor gene product. *Science*, 253, 1550-1553.

148) Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. (1992). DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature (London)*, 356, 408-414.

149) Martin, K. J. (1991). The interactions of transcription factors and their adaptors, coactivators and accessory proteins. *Bio Essays.* 13, 499-503.

150) Mastrangelo, I. A., Courney, A. J., Wall, J. S., Jackson, S. P. & Hough, P. V. C. (1991). DNA looping and Sp1 multimer links: A mechanism for transcriptional synergism and enhancement. *Proc. Natl. Acad. Sci. USA*, 88, 5670-5674.

151) Maxam, A. M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* 65, 499-580.

152) Maxson, R., Cohn, R. & Kedes, L. (1983). Expression and organization of histone genes. *Ann. Rev. Genet.* 17, 239-277.

153) Maxson, R., Mohun, T., Gormezano, G., Childs, G. & Kedes, L. (1983). Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. *Nature (London).* 301, 120-125.

154) McCall, M., Brown, T. & Kennard, O. (1985). The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG).poly(dC). *J. Mol. Biol.* 183, 385-396.

155) McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D. & Felsenfeld, G. (1981). A 200 base pair region at the 5' end of the chicken adult β-globin gene is accessible to nuclease digestion. *Cell*, 27, 45-55.

156) McKinney, J. D. & Heintz, N. (1991). Transcriptional regulation in the eukaryotic cell cycle. *Trends Biochem. Sci.* 16, 430-435.

157) Merkle, D. L. & Berg, J. M. (1991). Metal requirements for nucleic acid binding proteins. *Meth. Enzymol.* 208, 46-55.

158) Miller, J., McLachlan, A. D. & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4, 1609-1614.

159) Miner, J.N. & Yamamoto, K.R. (1991). Regulatory crosstalk at composite response elements. *Trends Biochem. Sci.* 16, 423-4426.

160) Mitchell, P. J. & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, 245, 371-378.

161) Morris, G. F., Marzluff, W. F. (1983). A factor in sea urchin eggs inhibits transcription in isolated nuclei by sea urchin RNA polymerase III. *Biochemistry*, 22, 645-653.

162) Morris, G. F., Price, D. H. & Marzluff, W. F. (1986). Synthesis of U1 RNA in a DNA-dependent system from sea urchin embryos. *Proc. Natl. Acad. Sci. USA*, 83, 3674-3678.

163) Nardelli, J., Gibson, T. J., Vesque, C. & Charnay, P. (1991). Base sequence discrimination by zinc-finger DNA-binding domains. *Nature (London)*, 349, 175-178.

164) Nickol & J. M., Felsenfeld, G. (1983). DNA Conformation at the 5' end of the chicken adult β-globin gene. *Cell*, 35, 467-477.

165) Nye, J. A. & Graves, B. J. (1990). Alkylation interference identifies essential DNA contacts for sequence-specific binding of the eukaryotic transcription factor C/EBP. *Proc. Natl. Acad. Sci. USA*, 87, 3992-3996.

166) Nye, J. A., Peterson, J. M., Gunther, C. V., Jonsen, M. D. & Graves, B. J. (1992). Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes & Dev.* 6, 975-990.

167) O'Neil, K. T., Hoess, R. H. & de Grado, W. F. (1990). Design of DNA-binding peptides based on the leucine zipper motif. *Science*, 249, 774-778.

168) O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989). Evidence that the leucine zipper is a coiled coil. *Science*, 243, 538-542.

169) Oakley, M. G. & Dervan, P. B. (1990). Structural motif of the GCN4 DNA binding domain characterized by affinity cleaving. *Science*, 248, 847-850.

170) Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W. & Kim, P. S. (1990). Secondary structure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. *Biochemistry*, 29, 2891-2894.

171) Oefner, C. & Suck, D. (1986). Crystallographic refinement and structure of DNase I at 2Å resolution. *J. Mol. Biol.* 192, 605-632.

172) Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990). High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry*, 29, 9324-9334.

173) Otting, G., Qian, X. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wütrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* 9, 3085-3092.

174) Pabo, C. O. & Sauer, R. T. (1984). Protein-DNA recognition. *Ann. Rev. Biochem.* 53, 293-321.

175) Palla, F., Casano, C., Albanese, I., Anello, L., Gianguzza, F., Di Bernardo, M. G., Bonura, C. & Spinelli, G. (1989). *Cis*-acting elements of the sea urchin histone H2A modulator binds transcription factors. *Proc. Natl. Acad. Sci. USA* 86, 6033-6037.

176) Pan, T. & Coleman, J. E. (1989). Structure and function of the Zn(II) binding site within the DNA binding domain of the GAL4 transcription factor. *Proc Natl. Acad Sci. USA* 86, 3145-3149.

177) Pan, T. & Coleman, J. E. (1990). GAL4 transcription factor is not a "zinc finger" but forms a $Zn(II)_2$ Cys_6 binuclear cluster. *Proc Natl. Acad. Sci. USA* 87, 2077-2081.

178) Parker, C. S. & Topol, J. (1984). A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell*, 37, 273-283.

179) Pascal, E. & Tjian, R. (1991). Different activating domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes & Dev.* 5, 1646-1656.

180) Pashev, I. G., Dimitrov, S. I. & Angelov, D. (1991). Crosslinking proteins to nucleic acids by ultraviolet irradiation. *Trends Biochem. Sci.* 16, 323-326.

181) Pavletich, N. P. & Pabo, C. O. (1991). Zinc-finger DNA recognition: crystal structure of a Zif268-DNA complex at 2.1Å. *Science*, 252, 809-817.

182) Peck, L. J. & Wang, J. C. (1981). Sequence dependence of the helical repeat of DNA in solution. *Nature (London)*, 292, 375-378.

183) Percival-Smith, A., Müller, M., Affolter, M. & Gehring, W. J. (1990). The interaction with DNA of wild-type and mutant *fushi tarazu* homeodomains. *EMBO J.* 9, 3967-3974.

184) Plumb, M. A., Nicolas, R. H., Wright, C. A. & Goodwin, G. H. (1985). Multiple sequence-specific DNA binding activities are eluted from chicken nuclei at low ionic strengths. *Nucl. Acids Res.* 13, 4047-4065.

185) Postel, E. H., Flint, S. J., Kessler, D. J., Hogan, M. E. (1991). Evidence that a triplex-forming oligodeoxyribonucleotide binds to the *c-myc* promoter in HeLa cells reducing *c-myc* mRNA levels. *Proc. Natl. Acad. Sci. USA* 88, 8227-8231.

186) Postel, E. H., Mango, S. E. & Flint, S. J. (1989). A nuclease-hypersensitive element of the human *c-myc* promoter interacts with a transcription initiation factor. *Mol. Cell. Biol.* 9, 5123-5133.

187) Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wütrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: Comparison with prokaryotic repressors. *Cell*, 59, 573-580.

188) Rauscher, F. J., Morris, J. F., Tournay, O. E., Cook, D. M. & Curran, T. (1990). Binding of the Wilms tumor locus zinc finger protein to the EGR-1 consensus sequence. *Science*, 250, 1259-1262.

189) Retief, J. D., Sewell, B. T. & von Holt, C. (1987). Nucleosome cores assembled *in vitro* occupy two preferred frames flanking the histone H1 gene from *Psammechinus miliaris*. *Biochemistry*, 26, 4449-4453.

190) Rhodes, D. (1985). Structural analysis of a triple-complex between the histone octamer, a *Xenopus* gene for 5S RNA and transcription factor IIIA. *EMBO J.* 4, 3473-3482.

191) Rhodes, D. (1989). Analysis of sequence-specific DNA-binding proteins. In *Protein function : a practical approach.* Creighton, T. E. , ed. (IRL Press at Oxford Unviversity), pp. 177-198.

192) Rhodes, D. & Klug, A. (1980). Helical periodicity of DNA determined by enzyme digestion. *Nature (London)*, 286, 573-578.

193) Rhodes, D. & Klug, A. (1981). Sequence-dependent helical periodicity of DNA. *Nature (London)*, 292, 378-380.

194) Rhodes, D. & Klug, A. (1986). An underlying repeat in some transcriptional control sequences corresponding to half a double helical turn of DNA. *Cell*, 46, 123-132.

195) Roeder, R. G. (1991). The complexities of eukaryotic transcription initiation: Regulation of preinitiation complex assembly. *Trends Biochem. Sci.* 16, 402-408.

196) Rosenfeld, P. J. & Kelly, T. J. (1986). Purification of Nuclear Factor I by DNA recognition site affinity chromatography. *J. Biol. Chem.* 261, 1398-1408.

197) Ruiz-Carillo, A. & Renaud, J. (1987). Endonuclease G: a $(dG)_n.(dC)_n$-specific DNase from higher eukaryotes. *EMBO J.* 6, 401-407.

198) Sakonju, S. & Brown, D. D. (1982). Contact points between a positive transcription factor and the *Xenopus* 5S RNA gene. *Cell*, 31, 395-405.

199) Schaffner, W., Gross, K., Telford, J. & Birnstiel, M. (1976). Molecular analysis of the histone gene cluster of *Psammechinus miliaris*. 2. The arrangement of the five histone-coding and spacer sequences. *Cell*, 8, 471-478.

200) Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O. & Birnstiel, M. L. (1978). Genes and spacers of cloned sea urchin histone DNA analyzed by sequencing. *Cell*, 14, 655-671.

201) Scheidereit, C. & Beato, M. (1984). Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. *Proc. Natl. Acad. Sci. USA*, 81, 3029-3033.

202) Schon, E., Evans, T., Welsh, J. & Efstratiadis, A. (1983). Conformation of promoter DNA: Fine mapping of S1-hypersensitive sites. *Cell*, 35, 837-848.

203) Schuman, J. D., Vinson, C. R. & McKnight, S. L. (1990). Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science*, 249, 771-774.

204) Schwabe, J. W. R., Neuhaus, D. & Rhodes, D. (1990). Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature (London)*, 348, 458-461.

205) Schwabe, J. W. R. & Rhodes, D. (1991). Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *Trends Biochem. Sci.* 16, 291-296.

206) Scott, M. P., Tamkun, J. W. & Hartzell, G. W. (1989). The structure and function of the homeodomains. *Biochem. Biophys. Acta*, 989, 25-48.

207) Simpson, R. T. (1991). Nucleosome positioning: Occurrence, mechanisms and functional consequences. *Progress Nucl. Acid Res.* 40, 143-184.

208) Singh, H., Le Bowitz, J. H., Baldwin, A. S. & Sharp, P. A. (1988). Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. *Cell*, 52, 415-423.

209) Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Anal. Biochem.* 150, 76-85.

210) Sorger, P. K., Ammerer, G. & Shore, D. (1989). Identification and purification of sequence-specific DNA-binding proteins. In *Protein function : a practical approach.* Creighton, T. E., ed. (IRL Press at Oxford University), pp. 199-223.

211) Spinelli, G., Albanese, I., Anello, L., Ciaccio, M. & Di Liegro, I. (1982). Chromatin structure of the histone genes in sea urchin sperms and embryos. *Nucl. Acids Res.* 10, 7977-7991.

212) Stadler, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. & Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNase I, *Cell*, 20, 451-460.

213) Stellwagen, N. C. (1983). Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels. *Biochemistry*, 22, 6186-6193.

214) Stokrová, J., Vojtisková, M. & Palecek, E. (1989). Electronmicroscopy of supercoiled pEJ4 DNA containing homopurine.homopyrimidine sequences. *J. Biomolec. Struct. Dynam.* 6, 893-897.

215) Su, W., Jackson, S., Tjian, R. & Echols, H. (1991). DNA looping between sites for transcriptional activation: Self-association of DNA-bound Sp1. *Genes & Dev.* 5, 820-826.

216) Suck, D. & Oefner, C. (1986). Structure of DNase I at 2.0Å resolution suggests a mechanism for binding to and cutting DNA. *Nature (London)*, 321, 620-625.

217) Tasheva, B. & Dessev, G. (1983). Artifacts in sodium dodecyl sulfate-polyacrylamide gel electrophoresis due to 2-mercaptoethanol. *Anal. Biochem.* 129, 98-102.

218) Topol, J., Ruden, D. M. & Parker, C. S. (1985). Sequences required for *in vitro* transcriptional activation of a *Drosophila* hsp70 gene. *Cell*, 42, 527-537.

219) Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA*, 76, 4350-4354.

220) Travers, A. A. (1989). DNA conformation and protein binding. *Ann. Rev. Biochem.* 58, 427-452.

221) Travers, A. A. & Klug, A. (1990). Bending of DNA in nucleoprotein complexes. In *DNA Topology and its Biological Effects*. Cozzarelli, N. R. & Wang, J. C., eds. (Cold Spring Harbor Laboratory Press), pp. 57-106.

222) Tsai, S. Y., Carlstedt-Duke, J., Weigel, N. L., Dahlman, K., Gustaffsson J. -Å., Tsai, M. -J. & O'Malley, B. W. (1988). Molecular interactions of steroid hormone receptor with its enhancer element: Evidence for receptor dimer formation. *Cell*, 55, 361-369.

223) Tullius, T. D. & Dombroski, B. A. (1986). Hydroxyl radical "footprinting": High-resolution information about DNA-protein contacts and application to ½ repressor and cro protein. *Proc. Natl. Acad. Sci. USA*, 83, 5469-5473.

224) Tullius, T. D., Dombroski, B. A., Churchill, M. E. A. & Kam, L. (1987). Hydroxyl radical footprinting: A high-resolution method for mapping protein-DNA contacts. *Meth. Enzymol.* 155, 537-558.

225) Tung, L., Lee, I. J., Rice, H. L. & Weinberg, E. S. (1990). Positive and negative transcription regulatory elements in the early H4 histone genes of the sea urchin *Strongylocentrotus purpuratus*. *Nucl. Acids Res.* 18, 7339-7348.

226) Tung, L., Morris, G. F., Yager, L. N. & Weinberg, E. S. (1988). Sea urchin early and late histone genes bind a specific transcription factor in a stable preinitiation complex. *Mol. Cell. Biol.* 9, 1476-1487.

227) Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science*, 246, 911-916.

228) Vitelli, L., Kemler, I., Lauber, B., Birnstiel, M. L. & Busslinger, M. (1988). Developmental regulation of micro-injected histone genes in sea urchin embryos. *Developmental Biol.* 127, 54-63.

229) von Holt, C., Brandt, W. F., Greyling, H. J., Lindsey, G. G., Retief, J. D., Rodrigues, J. deA, Schwager, S. & Sewell, B. T. (1989). Isolation and characterization of histones. *Meth. Enzymol.* 170, 431-523.

230) Vrana, K. E., Churchill, M. E. A., Tullius, T. D. & Brown, D. D. (1988). Mapping functional regions of transcription factor TFIIIA. *Mol. Cell. Biol.* 8, 1684-1696.

231) Wagner, J. R., Hu, C. -C. & Ames, B. N. (1992). Endogenous oxidative damage of deoxycytidine in DNA. *Proc. Natl. Acad. Sci. USA*, 89, 3380-3384.

232) Wang, J. C. (1979). Helical repeat of DNA in solution. *Proc. Natl. Acad. sci. USA*, 76, 200-203.

233) Wilchek, M., Miron, T. & Kohn, J. (1984). Affinity chromatography. *Meth. Enzymol.* 104, 3-55.

234) Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. & Ueda, H. (1987). Purification and properties of *Drosophila* heat shock activator protein. *Science*, 238, 1247-1253.

235) Wu, T.-C. & Simpson, R. T. (1985). Transient alterations of the chromatin structure of the sea urchin early histone genes during embryogenesis. *Nucl. Acids Res.* 13, 6185-6203.

236) Xiang, M., Lu, S. -Y., Musso, M., Karsenty, G., Klein, W. H. (1991). A G-string positive *cis*-regulatory element in the LpS1 promoter binds two distinct nuclear factors distributed non-uniformly in *Lytechinus pictus* embryos. *Development*, 113, 1345-1355.

237) Zabel, U., Schreck, R. & Baeuerle, P. A. (1991). DNA binding of purified transcription factor NF-KB. Affinity, specificity, Zn^{2+} dependence and differential half-site recognition. *J. Biol. Chem.* 266, 252-260.

238) Zhang, J. -X. & Hulling, P. J. (1990). pH and buffering in the bicinchoninic acid (4,4'-Dicarboxy-2,2'-biquinoline) protein assay. *Anal. Biochem.* 188, 9-10.