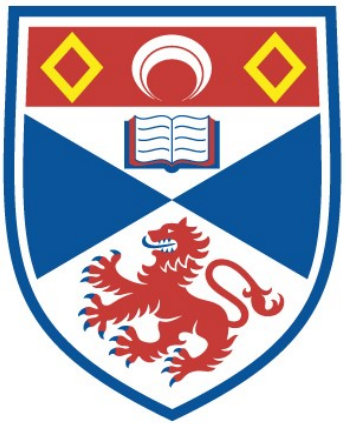
INVESTIGATION OF THE DNA BINDING DOMAIN OF NUCLEAR FACTOR I

Alasdair D. J. Freeman

A Thesis Submitted for the Degree of PhD at the University of St Andrews



1999

Investigation of the DNA Binding Domain of Nuclear Factor I

by

Alasdair D. J. Freeman

School of Biomedical Sciences

University of St. Andrews

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy

December 1998

Declaration

I, Alasdair Freeman, hereby certify that this thesis, which is approximately 50 000 words in length, has been composed by myself, that it is a record of my own work and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification

Signed Date 9/12/98

I was admitted as a research student in October 1992 and as a candidate for the degree of Doctor of Philosophy in October 1992; the higher study for which this is a record was carried out in the faculty of Science of the University of St. Andrews between 1992 and 1998.

Signed Date 9/12/98

I hereby certify that the candidate has fulfilled the condition of the resolution and regulation appropriate to the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified in submit this thesis in application for that degree.

Signed Date 9/12/98

In submitting this thesis to the University of St. Andrews, I understand that I am giving permission for it to be made available for use in accordance with the regulation of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published and that a copy of the work may be made and supplied to any bona fide library or research worker.

Acknowledgements

I would like to thank my supervisor, Prof. Ron Hay for his help during the course of this work and for the helpful criticisms during the writing of this thesis.

I thank all the people of "lab 28" and of the ex-Irvine building, especially Ian and Catherine for their company at the bench and sharing the unique experience of freezing in winter and burning in summer. I would like to thanks the technical support from Paul Talbot and Graham Kemps (protein sequencing), Alex Houston (DNA sequencing), Ellis Jaffray (football) and John Nicholson (cell culture).

I wish to thank in particular my family, Diane and John Murray, for their moral and financial support and thanks to Dr. Richard Freeman and Michelle Person. Many thanks to Rachel and Neil, David, Susie, Silas and Jeremy, Sarah and Marcus.
Merci aux amis francais qu'ils soient du Sud ou d'Alsace. Et la nana de Glasgow qui m'a tenu si longtemps compagnie. Allez la France! et Vive l'Ecosse!
Je remercie de tout coeur Laurent et Dominique de leur aide et hospitalite and Marion for being a member of her family. Merci a Sacha et a Louis. Merci Eric pour toutes ces Guiness, Fabienne, Huanting, Jill, Joanna, Margareth, Manuel, Jim and James. Thanks to the people in the laboratory for their invaluable supervision and in particular Angela and Ailsa.

Abbreviations

A	Adenine
Ad2	Adenovirus type 2
Amp	Ampicillin
APS	Ammonium persulphate
ATP	Adenosine triphosphate
dATP	deoxyadenosine triphosphate
Bis-Tris	2-bis[2-Hydroxyethyl]amino-2-[hydroxymethyl]1,3-propane
Bis-Tris Propane	1,3-bis[tris(Hydroxymethyl)methylamino]propane
bp	base pair
BSA	Bovine Serum Albumin
C	cytosine
C-terminus	Carboxyl Terminus
CAPS	(4-[Cyclohexylamino]-1-butanesulfonic acid)
cdk	cyclin dependent kinase
cDNA	complementary deoxyribonucleic acid
CHAPS	(3-[(3-Cholamidopropyl)dimethylammonio-]-1-propane sulfonate
CNBr	Cyanogen Bromide
CMC	critical micelle concentration
cpm	counts per minute
CTD	Carboxyl Terminal Domain
CTF	CAAT Transcription Factor
dATP	2'-deoxyadenosine 5'-triphosphate
dCMP	2'-deoxycytidine 5'-monophosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DBP	DNA Binding Protein
ddNTP	2',3'-dideoxy nucleoside 5'triphosphate
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
EBNA1	Epstein Barr virus nuclear antigen 1.
ECL	Enhanced Chemiluminiscence
E. coli	*Eschericia coli*
EDTA	Ethylene diamine tetracetic acid
FCS	Foetal calf Serum
G	guanine
GST	Glutathione-S-Transferase
Hepes	(N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid])
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
HRP	Horse Radish Peroxidase
HSV	Herpes Simplex Virus

Ig	Immunoglobulin
IPTG	Isopropyl-b-D-thioglalactopyranoside
iTP	intermediate Terminal Protein
ITR	Inverted Terminal Repeat
LTR	Long Terminal Repeats
MES	(2[N-Morpholino]propanesulfonic acid)
MMTV	Mouse Mamary Tumor Virus
MOPS	(3-[N-Morpholino]ethanesulfonic acid)
mRNA	messenger ribonucleic acid
MW	Molecular weight
NDSB	Non detergent sulphobetaine
NER	Nucleotide Excission Repair
NFI	Nuclear Factor I
NFIDBD	Nuclear Factor I DNA Binding Domain
NFII	Nuclear Factor II
NFIII	Nuclear Factor III
NLS	Nuclear Localisation Signal
NMR	Nuclear Magnetic Resonance
OD	Optical density
OPG	octylglucoside
PAGE	Polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
pfu	plaque forming units
p.i.	post infection
pI	Isoelectric point
PLP	Pyridoxal 5' Phosphate
PMSF	Phenylmethyl sulphonylfluoride
pol (Ad)	DNA polymerase (Adenovirus)
POUhd	POU homeodomain
POUs	POU specific domain
pTP	preTerminal Protein (Adenovirus)
PVDF	polyvinylidene difluoride
RPB	RNA polymerase B
S. cerevisiae (pombe)	Saccharomyces cerevisiae (pombe)
SDS	Sodium dodecyl Sulfate
Sf	Spodoptera frugiperda
(S)-HPMApp	(S)-9-(3-hydroxy-2-phosphonylmethoxypropyl)adenine
SRB	Suppressors of RNA polymerase B
ssDNA	single stranded deoxyribonucleic acid
SV40	Simian Virus 40
T	thymidine
TAF	TBP Associated Factor
TBE	Tris-borate-EDTA buffer
TBP	TATA Binding Protein
TCA	Trichloroacetic acid

TEMED	N,N,N',N'-tetramethylethylenediamine
TEN	Tris-EDTA-NaCl buffer
TFII	Transcription Factor of RNA polymerase II
tk	thymidine kinase
TP	Terminal Protein
Tris	(Tris[Hydroxymethyl]aminomethane)
ts	temperature sensitive
UV	ultra-violet
(v/v)	volume per volume ratio
(w/v)	weight per volume ratio
(w/w)	weight per weight ratio
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactoside

Physical units

℃	Temperature in degrees Celsius
g	gram for weight
g	gravitational acceleration ($9.81 ms^{-2}$)
m	meter
s	seconds
min	minutes
h	hour(s)
mol	mole
Da	Dalton
l	litre
Ci	Curie
M	Molar concentration (mol/l)

Order prefixes

T	tera	10^{12}
G	giga	10^{9}
M	mega	10^{6}
k	kilo	10^{3}
d	deci	10^{-1}
c	centi	10^{-2}
m	milli	10^{-3}
μ	micro	10^{-6}
n	nano	10^{-9}
p	pico	10^{-12}
f	femto	10^{-15}

Abbreviations for amino acids

A	Alanine
C	Cysteine
D	Aspartic Acid
E	Glutamic Acid
F	Phenylalanine
G	Glycine
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
N	Asparagine
P	Proline
Q	Glutamine
R	Arginine
S	Serine
T	Threonine
V	Valine
W	Tryptophane
Y	Tyrosine

Abstract

Nuclear Factor I is a cellular transcription factor involved in the initiation of adenovirus type 2 replication. It exists as a family of proteins and comprises two distinct domains. The N-terminal domain is highly conserved between species and is involved in DNA binding, dimerisation, and stimulates adenovirus replication. The C terminal domain is responsible for the transcriptional activation.

In this work, the N terminal, DNA binding domain of Nuclear Factor I, was cloned in two different expression systems. Using recombinant baculovirus, the unmodified protein and a GST fusion product were expressed in insect cells. Although both proteins were expressed at high levels, it was impossible to purify the fusion protein, while after purification, the unmodified protein remained heterogeneous. Expression in *E. coli* using the expression vector pET22b and pGEX2T produced inclusion bodies. The insoluble material was extracted, solubilised using guanidine HCl, then folded *in vitro* using new additives known as non detergent sulphobetaines. *In vitro* folding was optimised and yields of up to 8% could be obtained. The affinity of the refolded material for a specific DNA oligonucleotide was determined in a gel electrophoresis DNA binding assay and was identical to the native protein purified from baculovirus infected insect cells.

The structure of the Nuclear Factor I DNA Binding Domain was investigated using limited proteolysis followed by separation on SDS PAGE, N terminal sequencing and mass spectrometry. Residues 1 to 165 formed a compact domain. Residues 166 - 181 were extremely sensitive to degradation in the absence of DNA but fully protected in the presence of specific DNA. The C terminal region from residue 182 could be degraded in both conditions. Along with two other regions previously determined which bind DNA, the region between residues 166 and 181 is required for DNA binding. Differential labelling using iodoacetate showed that cysteine residues 95 and 111 were modified in the free protein which resulted in an inactive protein but were protected in the presence of DNA. This demonstrates their direct involvement in DNA binding.

TABLE OF CONTENTS

Introduction

I. Nuclear Factor I : A family of proteins

I.1 Introduction

Identification of proteins involved in the replication of adenovirus 2 and 5 DNA was facilitated by the development of an *in vitro* replication assay. The first cell free assay was developed by Challberg *et al.* in 1979 and could sustain the replication of protein linked adenovirus genomes. This system utilised HeLa cells infected with adenovirus in the presence of an inhibitor of DNA replication (hydroxyurea). When exogenous adenovirus genome purified from virion was added to nuclear extracts from infected cells, semi conservative viral DNA replication was detected. This assay was modified in order to fractionate the various elements required for replication (Ikeda *et al*, 1982, Lichy *et al*, 1982). Therefore a cytoplasmic extract from cells infected by adenovirus was complemented with nuclear extract from uninfected cells. By fractionating the nuclear extract and testing for the stimulation of replication, several factors were purified and shown to be involved in the replication of adenovirus 2 and 5 DNA. These included Nuclear Factor I (Nagata *et al*, 1982), Nuclear Factor II (Nagata *et al*, 1983) and Nuclear Factor III (Pruijn *et al*, 1986). Similarly the preterminal protein-DNA polymerase (pTP-pol) complex required for adenovirus DNA replication was purified by fractionation of the cytoplasmic extract.

I.2 Identification and isolation of Nuclear Factor I and related proteins.

The first protein purified from the nuclear extract and responsible for the stimulation of adenovirus replication was a novel protein with a molecular weight estimated at 47 kDa (Nagata *et al*, 1982). The protein was named Nuclear Factor I (NFI). Nuclear Factor I was purified by several ion exchange chromatography steps, followed by affinity chromatography on denatured DNA. The final step in the purification was the separation of proteins by centrifugation on a glycerol gradient. Using nitrocellulose filter binding assays and a DNA footprinting method, it was shown that NFI bound to a DNA sequence conserved among a number of adenovirus serotypes (Nagata *et al*, 1983). The DNA binding site of NFI was determined by the combination of deletion and mutation analysis. An NFI DNA binding site lies at the end of the adenovirus type 2 genome between nucleotides 23 and 40 (Leegwater *et al*, 1985).

Meanwhile, in the flanking region of the chicken lysozyme gene, four different DNA regions, which bound a protein possessing a similar DNA binding specificity were described (Nowock and Sippel, 1982). This protein recognised a specific DNA sequence and was named TGGCA binding protein (Borgmeyer *et al*, 1984). It could also bind to the adenovirus origin of DNA replication and enhance initiation of adenovirus DNA replication to the same degree as NFI (Nowock *et al*, 1985, Leegwater *et al*, 1985). Isolation of cDNAs encoding the TGGCA binding protein revealed that it was the chicken form of the human NFI protein (Rupp *et al*, 1990).

The function of NFI was still undetermined at that time. One function of NFI was found by Jones (1985) while studying the promoter of herpes simplex virus thymidine kinase (HSV tk). Two distinct proteins bound to the promoter of this gene : Sp1 and a protein binding to a CCAAT DNA sequence. This DNA sequence was required for efficient transcription *in vivo* and *in vitro*. The protein, although not yet purified, was named CTF (CAAT Transcription Factor). Purification and identification of the CTF protein revealed that CTF and NFI were similar (Jones *et al*, 1987). Both proteins

shared the same DNA binding properties and the ability to stimulate replication and transcription. Furthermore NFI and CTF 1 contained identical polypeptides.

Using an immobilised DNA with a specific NFI binding site, Rosenfeld and Kelly (1986) purified NFI. SDS PAGE analysis revealed a population of proteins of apparent molecular weight ranging from 52 to 66 kDa. This suggested that NFI was in fact a family of proteins. This was different from the protein purified by Nagata in terms of molecular weight (47 kDa). However the purification described by Nagata involved many steps where NFI could have been degraded. It is possible that this 47 kDa. protein is not NFI. NFI whose activity is observed through a gel retardation assay might be at a level below detection in Silver stained SDS PAGE.

I.3 Cloning of cDNAs encoding Nuclear Factor I and the emergence of a family of related proteins

cDNA encoding NFI was obtained from a variety of species including: human (Santoro *et al*, 1988, Apt *et al*, 1994), rat (Paonessa *et al*, 1988), hamster (Gil *et al*, 1988), chicken (Rupp *et al*, 1990), pig (Meisterernst *et al*, 1989), mouse (Inoue *et al*, 1990) and Xenopus (Roulet *et al*, 1995). Protein expressed from the cloned cDNA isolated from human, rat and pig bound to DNA and specifically recognised a CCAAT binding site.

I.3.1 Alternative splicing

From HeLa cells, three different clones were isolated : CTF 1, 2, 3 (the porcine NFI pCTF 1 to 3 were cloned and found almost identical to the human forms). CTF 4 and 6 were also cloned from HeLa cells but their functions are yet unknown. The porcine NFI gene was mapped and revealed eleven exons. pCTF 1 was the product of the translation of all eleven exons, while pCTF 2 and 3 lack exon 9 (figure 1).

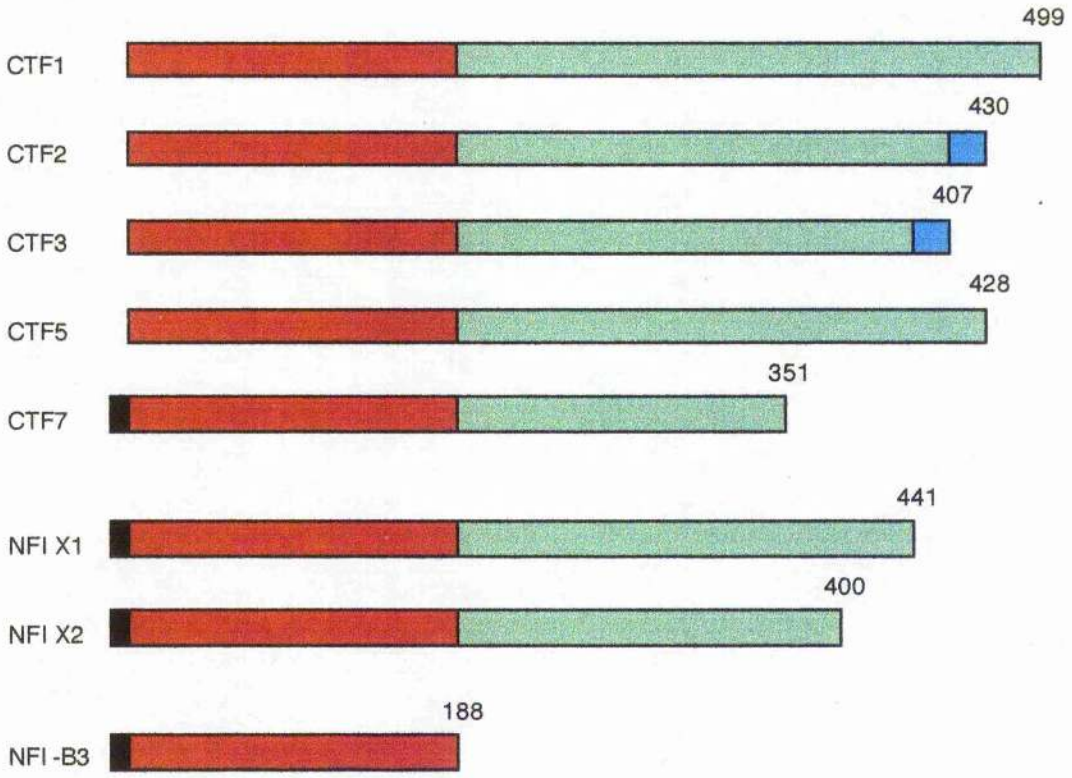


Figure 1 The Nuclear Factor I family

Isolated cDNA encoding NFI proteins were characterised by a conserved DNA binding region (red boxes). The variable C terminal region is produced by alternative splicing of the pre mRNA (green boxes) which can produce a frame shift in the open reading frame (lighter blue boxes). The NFI proteins represented in this figure are those isolated in humans except for CTF 7 (porcine).

The absence of exon 9 subsequently changed the reading frame of the mRNA. The shift in reading frame produced shorter proteins as a stop codon was present after encoding for 16 residues in exon 10 instead of the 33 residues in the reading frame used by pCTF 1. pCTF 3 also lacks exon 3. In other species such as chicken and Xenopus, several similar proteins were found differing only by the presence or absence of a number of residues at the C terminus. These variations seem to come from an alternative splicing of a pre mRNA. Therefore other unidentified NFI proteins might well exist. Recently two other forms of CTF protein were cloned : CTF 7 from a porcine library (Altmann *et al*, 1994) and CTF 5 from a cDNA library of HeLa cells (Wenzelides *et al*, 1996). CTF 7 lacks exons 9 and 10 while CTF 5 lacks exon 7 to 10. However these splicing events did not shift the reading frame of exon 11.

I.3.2 Gene encoding for NFI

Originally, in chicken, 4 different genes were found to code for NFI type proteins. In human and in hamster only 2 genes were known (NFI C and X). However chromosomal localisation of NFI genes in human using the FISH technique revealed four genes (NFI A, B, C and X) on different chromosomes (Qian *et al*, 1995).

I.3.3 NFI is a family of proteins

From these observations, a clearer picture of the various cloned proteins can be drawn. Figure 1 shows some of the NFI protein which have been isolated in human. To produce the various NFI proteins, cells have a choice of four genes. By differentially splicing the pre mRNA, cells can produce a plethora of different CTF proteins. In HeLa cells, seven different CTF proteins were found. Other cells produce other CTFs. However the mechanisms choosing which gene is expressed and how it is spliced are unknown. In one particular cell line (HeLa) it was observed that cells grown under different conditions produced different apparent molecular weight CTFs (Goyal *et al*,

1990) suggesting that these mechanisms are dynamic and are regulated by external conditions.

I.4 The domain organisation of the protein and its function.

Sequence analysis of the various NFI proteins clearly revealed the existence of two distinct regions : at the N terminus a domain rich in positively charged residues (20% in this domain compared to 11% on average), while the C terminus contains a high proportion of serine (16% compared to 8%), threonine (9% compared to 5.4%) and proline residues (13.4% compared to 6.1%). The prolines are clustered in the last hundred residues in the human CTF 1 protein. The C terminal region is not as well conserved as the N terminus and varies in length. This is due to the alternative splicing which affects this domain.

The N terminus region spans up to residue 179 (human CTF 1). Comparison between the protein sequences from different species (Figure 2) shows a highly conserved domain with at least 85% identity between the cDNA clones obtained. Also it shows that about two thirds of the basic residues are clustered between the arginine in position 28 and lysine at position 91 in the human NFI. Recently a NFI like sequence has been found in *Caernorhabditis elegans*. This protein was called Nuclear Factor I but has not yet been shown to specifically bind to DNA. Comparison between NFI from *C. elegans* and human shows at least 60% identity in the residues with the other NFI species. The comparison also revealed some highly conserved domains especially from an aspartic acid at position 102 to the valine in position 175 (numbering based on human NFI). Also the hydrophobic and hydrophilic regions are conserved (in red and blue in figure 2).

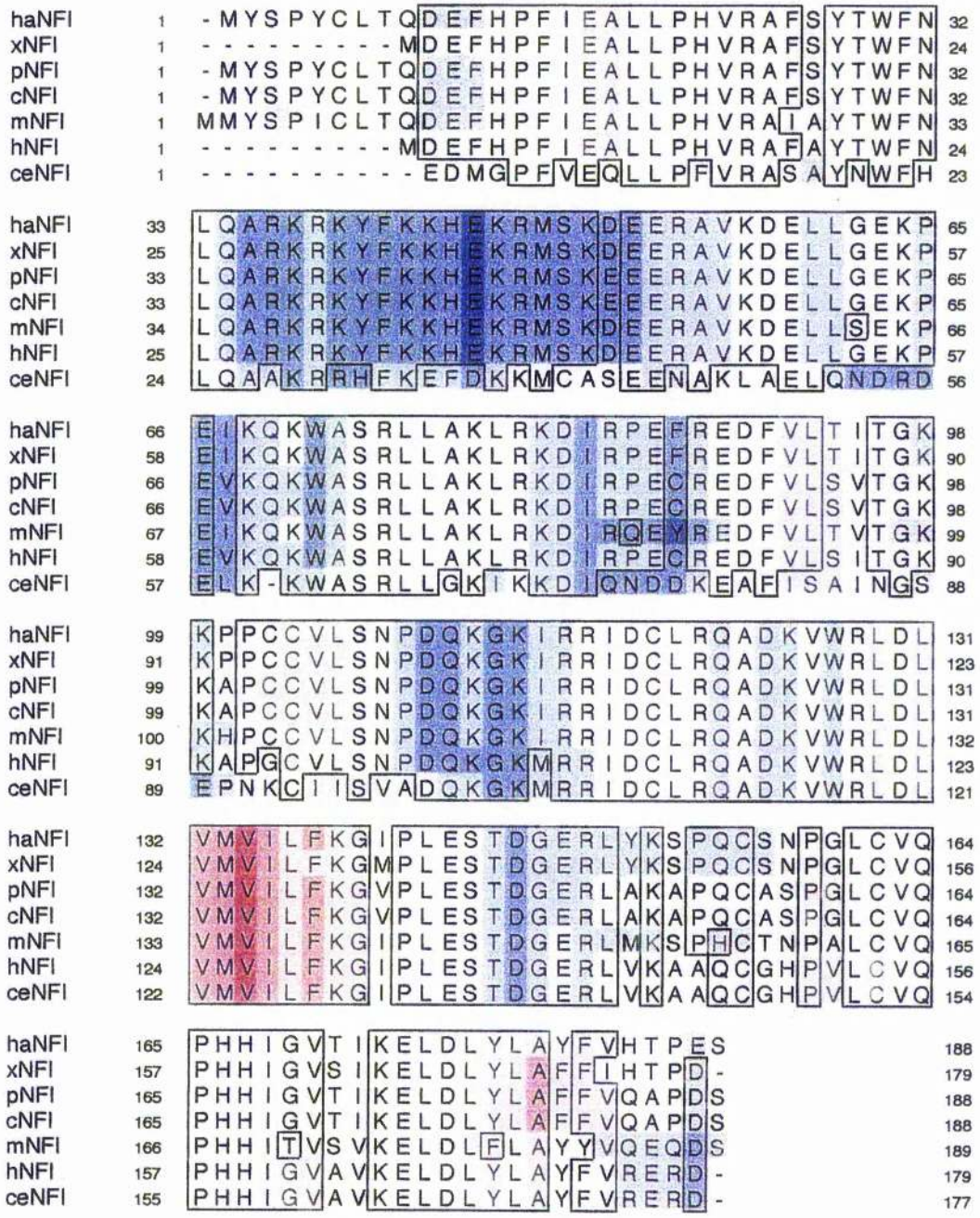


Figure 2 Comparison of the amino acid sequence of the N terminal region of human NFI with different animal species.

The deduced amino acid sequence of the 179 N terminal residues of the human Nuclear Factor I was compared to other species from pig (p), hamster (ha), mouse (m), chicken (c), Xenopus (x) and *C. elegans* (ce). Conserved residues are showed in boxes. Hydrophobic regions are in blue while hydrophobic regions are in red.

I.4.1 The DNA binding domain

The different regions in the protein possess different functions. From a rat clone, Gounari *et al* (1990) produced an NFI where the C terminal domain was deleted. The first 240 residues were expressed in a recombinant vaccinia virus. The purified protein was still dimeric and retained DNA binding activity.

The conserved 177 residue N terminal region is encoded in a single exon. It was thought that this region only was sufficient to confer a DNA binding activity to NFI. However the NFI B3 protein, which is translated from an unspliced mRNA and only contains the 178 N terminal residues, cannot on its own bind to DNA as a homodimer (Liu *et al*, 1997). Therefore some residues which do not belong to the N terminal region are also required for a proper DNA binding activity.

The specific DNA binding site for NFI was first examined by synthesising various oligonucleotides and by determining the DNA binding affinity for NFI for each one of them (Gronostajski 1988). The actual specific DNA site was discovered by sequencing oligonucleotides which were purified by gel retardation when bound to NFI. After seven rounds of purification, the oligonucleotides were cloned and sequenced. By comparing the various sequences, the best DNA site bound by NFI is a symmetrical dyad of the following sequence: 5'- TTGGCA N4 GCCAA - 3' (Osada *et al*, 1996). NFI bound tightly to a DNA oligonucleotide containing this site and could not be released easily by competition using the natural site found in the adenovirus genome. The DNA binding equilibrium constants of NFI were determined experimentally for a number of NFI binding sites. Equilibrium binding constants range from $2x10^{11}$ M^{-1} for a palindromic site found in the α globin gene promoter to 10^{7} M^{-1} for the half binding site in the HSV tk promoter (Meisterernst *et al*, 1988). The equilibrium binding constant for the binding site in the adenovirus genome is $8.4x10^{10}$ M^{-1} (Cleat *et al*, 1989). Binding to non specific DNA has an equilibrium constant of $5x10^{6}$ M^{-1}.

Methylation of DNA affects the guanine and adenine nucleotides. An oligonucleotide containing a NFI binding site was treated by random methylation. The methylated oligonucleotide was tested for its interference on the DNA binding activity of NFI. Therefore the various methylated oligonucleotides were incubated with NFI. Bound oligonucleotide and free probe were separated by filtration over a nitrocellulose membrane. Bound and free oligonucleotides were subsequently treated to cleave specifically the oligonucleotide at the point of the methylation. By running each fraction in a sequencing gel, the position of guanines in the oligonucleotide bound by NFI were compared to the unbound oligonucleotide. Unmethylated guanines in the TTGGA sequence were retained by NFI on the filter, hence showing that this sequence was bound to NFI. This also demonstrated that NFI bound to the major groove. A similar result was obtained by ethylation interference experiments in which ethylation affects the phosphate backbone of the DNA. Using methylation protection, NFI was bound to the oligonucleotide while methylation of guanine occurred. By comparing the guanine protected from methylation in the presence of NFI, the same result as the methylation interference was obtained.

Computational analysis suggested that the contacts between DNA and NFI occurred only on one side of the DNA molecule (De Vries *et al*, 1987). By ethylation interference assay, the region of contact between the protein and the DNA was mapped. Furthermore some protein-DNA interaction also occurred outside the DNA recognition sites as revealed by hydroxyl radical footprinting (Zorbas *et al*, 1989).

The primary structure of NF I does not contain any of the common DNA binding motifs such as zinc finger (Klug and Rhodes 1987), leucine zipper (Vinson *et al*, 1989), β sheets (Breg *et al*, 1990) or helix-turn-helix (Harrison and Aggarwal 1990). Deletion analysis of the NFI$_{DBD}$ revealed two different regions involved in DNA binding. A region between residues 1 and 78 is involved in non specific DNA binding while the region between residues 71 to 236 recognised specifically the NFI DNA binding site as a dimer (Dekker *et al*, 1996).

A new family of DNA binding proteins has been found from solving the structure of two unrelated proteins with no sequence homology : the Epstein-Barr virus nuclear antigen 1 (EBNA1) (Bochkarev *et al*, 1996) and the E2 protein of bovine papillomavirus (Hegde *et al*, 1992). NFI shares similar biochemical properties with the EBNA1 since EBNA1 also binds DNA as a dimer on a specific palindromic binding site (Ambinder *et al*, 1991, Frappier and O'Donnell 1991). Furthermore, EBNA1 DNA binding domain is also split into two regions : at the N terminus a region binding non specifically to DNA and another region which can bind specifically to DNA as a dimer.

Interestingly Suzuki (1993) compared the residues involved in DNA interaction of structurally known DNA binding proteins and showed the frequent existence of an helix in the protein residues recognising the DNA. In this helix, basic residues are conserved in positions 2, 7, 9, 11 and 12. In NFI, a putative α helix with a similar motif was identified in CTF 1 between residues 24 and 36. This region may be involved in the non specific DNA binding activity of NFI.

Site directed mutational analysis of NFI proteins in this region generally produced an inactive protein (Armentero *et al*, 1994). The DNA binding activity of NFI is also dependent upon its dimerisation. The role of the cysteine residues and of a possible red-ox regulation of the DNA binding activity has been demonstrated. NFI possesses four conserved cysteines at position 95, 111, 148 and 154. Treatment of NFI with cysteine modifying chemicals (such as diamide, N-ethylmaleinede) abolished DNA binding of NFI. Site directed mutagenesis of these residues produced misfolded protein except for the mutation of cysteine 111. The mutated NFI at position 111 from a cysteine to a serine or alanine was resistant to inactivation by oxidation (Novak *et al*, 1992, Bandyopadhyay *et al*, 1994). The regulation of NFI through the red-ox states of cysteines was also investigated : *in vitro*, a cellular enzyme thioltransferase (glutaredoxin) was able to restore activity from NFI inactivated by oxidation. *In vivo*, investigation of the cellular pathway of NFI reactivation suggests that the enzyme

thiotransferase, catalyses the reactivation in conjunction with GSH and GSH reductase (Bandyopadhyay *et al*, 1998).

I.4.2 The Transcriptional activation domain

Deletion analysis *in vivo* (Mermod *et al*, 1989), demonstrated that the C terminal domain is required for transcription stimulation. Using a reporter plasmid with DNA binding sites for NFI expressing the Chloramphenicol Acetyl Transferase (CAT), various CTF and deletions were co-transfected into Drosophila Schneider line-2 cells. These cells do not contain any known NFI protein. Deletion of any sequence in a region at the C terminus rich in proline residues, reduced dramatically the expression of the CAT. The region between residue 399 and 499 contained the transcription activation domain of CTF 1.

II. Function of Nuclear Factor I in transcription.

II.1 Promoters regulated by Nuclear Factor I.

The number of NFI binding sites in the HeLa genome was experimentally estimated at 30 000 (Gronostajski *et al*, 1985). Some of these potential DNA binding sites have been identified, but only in a limited number of instances have promoters been shown to be regulated by NFI.

The promoter of the mouse mammary tumour virus (MMTV) contains at least two NFI binding sites : one in the proximal region (Nowock *et al*, 1985; Buetti and Kuhnel, 1986; Cordingley *et al*, 1987 and 1988) and one in the distal region of the promoter (Mink *et al*, 1992). The role of the proximal region of the promoter has been extensively studied. It remains at this time the best example of a promoter for which both the role of NFI and its regulation are known.

The promoter of the MMTV is located in the Long Terminal Repeats (LTRs) at each end of the viral genome. The regulation of transcription is complex and involves many activator and inhibitor proteins binding along the whole length of the U3 region of the LTR (about 1200 bp). Although the effect of all proteins on the promoter is not fully understood and all transcription factors involved have not yet been characterised, a general picture starts to appear (Reviewed by Gunzburg and Salmons 1992).

The site of transcription initiation is located at the border between the U3 and R region in the LTR. In the proximal promoter which extends from the start site of transcription and 221 bp upstream, four different motifs were identified : a TATA box (-20 to -27), an NFIII or octamer binding site (-56 to -37), an NFI binding domain (-60 to -80) and a hormone responsive element binding domain which lies between -80 and -190.

Gene expression of the MMTV promoter can be induced by steroid hormones such as androgens, glucocorticoides and progesterones (Scheidereit *et al*, 1983, Chalepakis *et al*, 1988). Steroid hormones bind to their receptor which migrates to the nucleus and binds to specific DNA binding sites in the hormone responsive element of the MMTV LTR. Following glucocorticoid induction, five binding site were located at position -78 to -83, -93 to -98, -114 to -119, -144 to -156 and -170 to -175 (Bruggemeier *et al*, 1991, Chavez and Beato 1997). Under glucocorticoid induction, mutations which abolished NFI binding to the promoter of the MMTV resulted in a reduction of gene expression by six to eight fold. Similarly mutation of the NFIII binding site at position -49 to -56 reduced activation by glucocorticoid three to four times. Mutation of the other NFIII site (-37 to -44), had no effect on gene expression even when stimulated by steroid hormone (Cato *et al*, 1988). Similar results were obtained using androgen as inducer instead of glucocorticoids. Therefore NFI was required to stimulate gene expression under glucocorticoid hormone activation.

However using progestin as an inducer, the level of gene expression remained similar to the wild type even after mutation of the NFI DNA binding site. NFIII was required and was shown to bind DNA in a cooperative manner with the loaded progestin receptor. These results obtained *in vivo* were in agreement with transfection assays using mutated promoters and testing the induction at various concentrations of steroid hormone (Kalff *et al*, 1990, Cato *et al*, 1988). NFI and NFIII did not act in synergy but rather competed for their binding site on the MMTV promoter (Mows *et al*, 1994).

Although the MMTV promoter has been extensively investigated, other promoters which also contain specific NFI binding sites have been studied. Their specificity can be shown either using a competition assay with an oligonucleotide containing an NFI site or by DNA footprinting. The role of NFI in the promoter was usually studied using transfection of various promoter constructs and by mutation of the potential NFI binding elements followed by testing for the gene expression of a marker protein. The

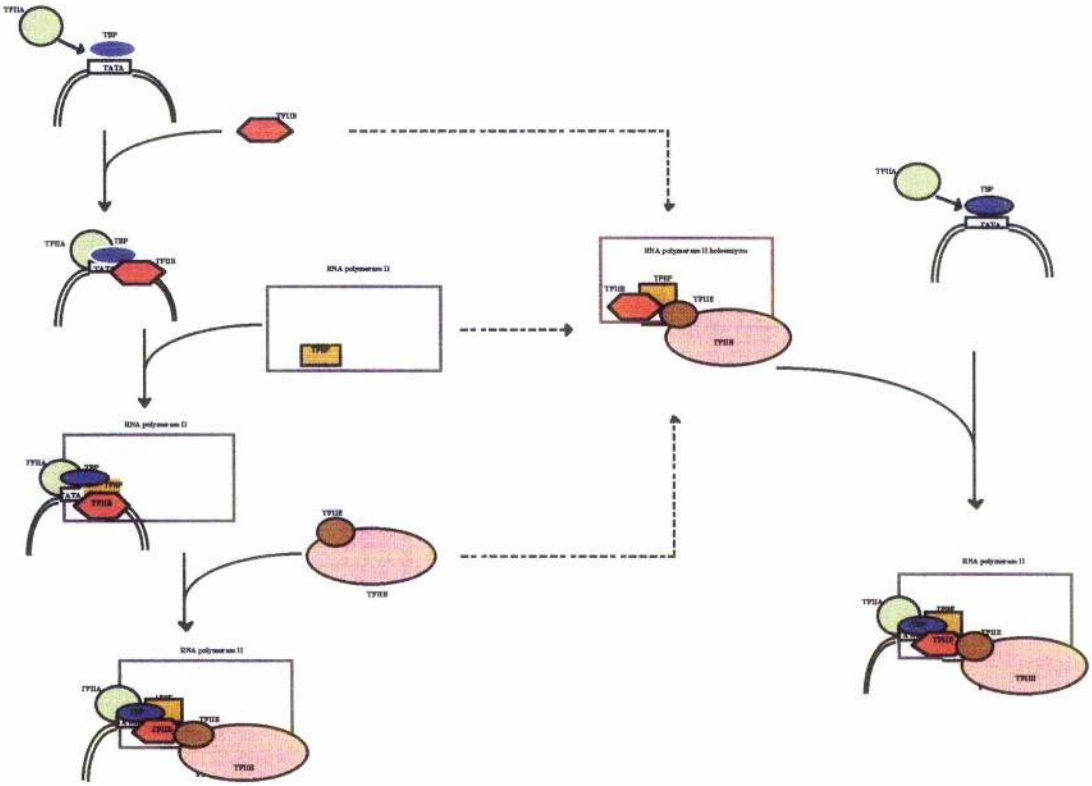
various promoters can activate extremely diverse functions and some are even cell specific. In some promoters, NFI activates gene expression such as in the promoter of the fibronectin gene (Alonso *et al*, 1996), the aspartate amino transferase gene promoter (Garlatti *et al*, 1996) which contains an NFI binding site between nucleotide -519 to-543, the phosphoenolpyruvate carboxykinase gene promoter (binding site at -100 to -115) (Crawford *et al*, 1998), in the production of milk protein such as α lactalbumin, β lactoglobulin and whey acidic protein (Mink *et al*, 1992; Li and Rosen 1995) and the rat pyruvate kinase L gene promoter (Yamada *et al*, 1997). However NFI can also act as a silencer as found in the regulation of expression of a DNA binding protein, the Pit-1/GHF-1 (Rajas *et al*, 1998). Other NFI binding sites have been located in other genes such as the promoter of α globin, IgM and α (I) collagen (Zorbas *et al*, 1992; Hennighausen *et al*, 1985; Oikarinen *et al*, 1987). However their role in the transcription process is still unclear.

A number of viruses contain recognisable NFI binding sites such as polyomavirus BK virus (Chakraborty *et al*, 1991) and hepatitis (Ori *et al*, 1994). The human polyoma JC virus contains at least 3 NFI sites in the late and early promoters for gene expression (Kumar *et al*, 1993). These NFI binding sites are involved in regulating transcription and replication of the virus (Sock *et al*, 1993). In brain cells, NFI was characterised for its transcriptional role specificity for gene expression of the late and early promoter of the human JC virus (Kumar *et al*, 1996).

II.2 Basal transcription and the formation of pre initiation complex.

A key stage in the control of gene expression is at the level of transcription and especially the initiation of the transcription. With recent advances in this area of research, the mechanism of basal transcription was extensively reviewed (Goodrich and Tjian 1994, Orphanides *et al*, 1996, Hampsey 1998).

In a promoter, the RNA polymerase II does not bind DNA alone but through a pre assembled protein complex (figure 3). Experiments isolated the various components required to form a preinitiation complex before the RNA polymerase is able to transcribe the gene. These various components are found in all species from yeast to human and are therefore known as general transcription factors. The different general transcription factors have been isolated and are composed of 6 protein complexes (TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH). The formation of the preinitiation complex has been studied *in vitro* (usually using binding kinetics and gel retention assay), in an orderly manner to understand the role of each protein and how they affect the assembly of the preinitiation complex. Once this preinitiation complex is formed, other events are required such as synthesis of the first phosphodiester bond, strand melting, and promoter clearance before the elongation occurs and a pre mRNA is synthesised. The preinitiation complex is also described as the basal transcription machinery since it does not require, for its formation, any activators binding to an upstream DNA element of the TATA box.



Stepwise assembly model Holoenzyme model

Figure 3 Model for the initiation of transcription by RNA polymerase II: holoenzyme model versus the stepwise assembly model

In the stepwise assembly model, TFIID binds first the TATA element of the promoter. This is achieved by the TATA Binding Protein (TBP). TFIIA and TFIIB are next to bind. TFIIF escorts the RNA polymerase II to the promoter. The preinitiation complex is completed by the binding of TFIIE and TFIIH. In the holoenzyme model, TFIIB, F, E and H preassembles with the RNA polymerase II before binding in a single step to the TFIID-TFIIA complex (the TAFII of the TFIID was omitted for clarity).

II.2.1 The structure of the TFIID complex

The first step in the formation of the preinitiation complex involves the recognition of the TATA box by the TFIID complex (Parker *et al*, 1984, Nakajima *et al*, 1988).

TFIID is formed of several proteins, the TATA binding protein (TBP, 27 kDa in yeast) and TBP associated factors (TAF) (Hernandez 1993). The binding activity of TFIID to the TATA box is confined to one protein in particular: the TBP. TBP alone is sufficient to catalyse the formation of the preinitiation complex *in vitro* but the whole TFIID complex is required for response to transcriptional activators.

TBP is made of two domains: the amino terminus is divergent between species and its function in the formation of the preinitiation complex is still unclear. However this region is known to regulate the RNA polymerase III transcription at the U6 promoter (Mittal *et al*, 1997). The carboxyl terminus region is conserved between species and consists of two imperfect repeats. This C terminal region is sufficient to direct transcription *in vitro* (Hoey *et al*, 1990; Zhou *et al*, 1993). The crystal structure of the carboxyl terminus revealed that TBP has a saddle shape (Nikolov *et al*, 1992). The protein has a two fold symmetry with each half of the symmetry made up of each repeat. Each repeat is composed of 2 α helices and 5 β sheets in the order of S1- H1- S2 - S5- H2. Crystal structure of TBP bound to the TATA box demonstrated that the protein recognised 8 base pair in the minor groove of the TATA box and that the DNA structure was dramatically affected by TBP binding (Kim J. *et al*, 1993; Kim Y. *et al*, 1993). The DNA binding surface in TBP is a curved, antiparallel β sheet and therefore to facilitate DNA recognition, the minor groove moulded itself to follow the curve of the β sheet of TBP which causes a bend in the structure of the DNA. As a result, downstream and upstream DNA elements were brought into closer contact. The binding of TBP to the TATA box also resulted in a partial unwinding of the DNA.

II.2.2 Regulation of the TBP binding to the TATA box.

As the recruitment of TBP to the promoter is the first step to the formation of the preinitiation complex, it is not surprising that TBP plays a critical role in the mechanism of transcriptional activation. The regulation of TBP binding to the DNA depends on three different factors :

1) TBP dimer formation: The recruitment of TBP to the TATA box is slow and potentially a rate limiting step in the formation of the preinitiation complex (Klein *et al*, 1994). This could be due to the fact that TBP was reported to form a homodimer thereby inactivating the DNA binding activity of the protein (Coleman *et al*, 1995). Dissociation of the TBP dimer is slow suggesting that dimer dissociation dictates the kinetics of TBP-DNA binding (Coleman *et al*, 1997). The TFIID also dimerises through the TBP protein (Taggart *et al*, 1996). Furthermore the yeast transcription factor GCN4 stimulates transcription by enhancing the kinetics of TBP recruitment (Klein *et al*, 1994).

2) Complex formation: TBP exists in a complex with at least eight other proteins known as TBP Associated Factors (TAFs) to form the TFIID complex (Tanese *et al*, 1991; Takada *et al*, 1992). TAFs are composed of proteins with molecular weights varying from 15 to 250 kDa in human. *In vitro* reconstitution experiments demonstrated that the TAFs were required for activation dependent, but not basal transcription (Tjian and Maniatis, 1994). Several roles have been associated with TAF proteins: mediation of transcription activation, DNA binding, and enzymatic activity. Several reports have demonstrated that TAF proteins such as the dTAFII 40 can interact with activators e.g. VP16 (Goodrich *et al*, 1993). However depletion or inactivation of several TAF did not affect transcriptional activation *in vivo* (Moqtaderi *et al*, 1996). The role of TAF in the transcription process remains a contentious issue.

A novel DNA sequence motif located 30 base pair downstream of the start site in TATA less promoter is recognised by TAF. A TAF was affinity purified with this

DNA motif (Burke and Kanodaga, 1996). In Drosophila, TAFII 150 was shown to bind DNA (Verrijzer *et al*, 1994).

Experimental evidence also demonstrated that TAFII 250 catalyses several enzymatic reactions: in human, TAFII 250 can specifically phosphorylate RAP74, a subunit of the TFIIF which stimulated its activity (Dikstein *et al*, 1996). TAFII 250 also contains a histone acetyl transferase activity which may affect chromatin structure (Mizzen *et al*, 1996).

A combination of biochemical and crystallographical experiments has provided strong evidence that a histone-like structure exists within the TFIID (Nakatani *et al*, 1996; Hoffmann *et al*, 1997). The amino termini of Drosophila TAFII 40 and 60 have strong homology to the histone proteins H3 and H4. The crystal structure of the amino termini of TAFII 40 and 60 showed that they form a heterotetramer which is similar to the corresponding region in H3 and H4 and exhibits a histone fold (Xie *et al*, 1996). A homologue of H2B was also identified as TAFII 20 forming a homodimer which can interact with TAFII 40 and 60 (Hoffmann *et al*, 1996). The role of the histone like structure may be to modify the interaction with DNA although the arginine residues in H3 and H4 are not conserved in TAFII 40 and 60 (Luger *et al*, 1997). Alternatively, the histone fold may function solely to mediate protein - protein interaction within TFIID.

3) role of TFIIA : TFIIA also recognises the TBP - promoter complex although the protein can also bind soluble TBP (Buratowski *et al*, 1989). TFIIA consists of three subunits in Drosophila (α, β, γ) (Yokomori *et al*, 1993) and two in yeast (TOA1 and TOA2, Ranish *et al*, 1992). The two largest subunits of TFIIA in Drosophila are encoded by the same gene but have different post translational modifications. The role of TFIIA in transcription was controversial. The requirement of TFIIA was different depending on the source of TBP. Using recombinant TBP, TFIIA was not required for transcription *in vitro* (Cortes *et al*, 1992). However using purified TBP from cells, TFIIA was needed to stimulate the same assay (Ozer *et al*, 1994; Sun *et al*, 1994). It has since been established that TFIIA acts as an activator by blocking transcriptional

repression. By binding to TBP, TFIIA displaces repressors associated with TBP such as MOTI (Auble *et al*, 1994) or topoisomerase I (Ma *et al*, 1996) and prevents TBP from forming other protein - protein interactions than those required for formation of the preinitiation complex. TFIIA also appeared to increase the affinity of TBP for the TATA box. In addition it was shown that TFIIA is required for TBP to bind the TATA box when in a nucleosome structure (Imbalzano *et al*, 1994a and b).

The spatial structure of yeast TFIIA has been determined by X-ray crystallography and shows that TFIIA is L shaped (Geiger *et al*, 1996; Tan *et al*, 1996). The carboxyl termini of both subunits forms a compact β sheet structure termed a β sandwich : three antiparrallel strands of each subunit form the hydrophobic core of a β barrel structure which can contact DNA. Another structural element identified is a four helix bundle made up of two helices from each subunit.

II.2.3 The TFIIB complex

The second step in the formation of the preinitiation complex is the recruitment of TFIIB which recognises the TFIID-promoter complex. TFIIB has a pivotal role in recruiting RNA polymerase to the TFIID-promoter complex. TFIIB can interact directly with TBP (Buratowski *et al*, 1989) and RNA polymerase II (Fang and Burton 1996), as well as with TFIIF (Ha *et al*, 1993), and TAF 40 of the TFIID complex (Goodrich *et al*, 1993). The binding of the TFIIB results in a more stable complex between TFIID or the TBP and DNA.

TFIIB is a monomeric protein of 38 kDa containing two distinct regions (Ha *et al*, 1991). The N-terminal region interacts with RNA polymerase and while the structure of this region has not been determined it is predicted to contain a zinc ribbon motif which is important for the function of the protein (Zhu *et al*, 1996). This region plays an important yet undefined role in selecting the start site for transcription *in vivo* (Bangur *et al*, 1997). The C-terminal region of TFIIB binds TBP and has been shown to contain two similar domains, each consisting of five α helices (Bagby *et al*, 1995).

Such a structure has also been found in cyclin A and in the retinoblastoma proteins (Gibson *et al*, 1994). The co crystal structure of the ternary complex containing the C-terminal region of TFIIB bound to TBP and DNA has been solved (Nikolov *et al*, 1995). This reveals that TFIIB binds to one side of TFIID downstream of the TATA box and underneath the TBP-DNA complex. In this complex, acidic residues of TBP interact with basic residues of TFIIB. The bend in the DNA caused by TBP allows TFIIB to contact promoter DNA both upstream and downstream of the TATA box sequence. TFIIA and TFIIB thus bind on opposite sides of TBP.

II.2.4 Recruitment of the RNA polymerase II - TFIIF complex

The TBP-TFIIB-DNA complex is finally recognised by the RNA polymerase II. Although in some cases RNA polymerase alone can be recruited to some promoters, in most cases the binding of the polymerase is escorted by TFIIF (Flores *et al*, 1989; Price *et al*, 1989).

TFIIF is a heterotetramer composed of two different subunits RAP30 and 74. Rap30 resembles the σ factor in bacteria (Conaway and Conaway 1990; Killeen and Greenblatt 1992). Rap30 alone can accomplish some of the functions of TFIIF such as interacting with TFIIB (Ha *et al*, 1993), RNA polymerase (McCracken and Greenblatt, 1991) and DNA (Tan *et al*, 1994). However for transcription activation RAP74 is needed. The role of RAP74 is to remove the RNA polymerase from non specific sites. RAP74 was found to position the RNA polymerase at the start site and had the ability to interact with TFIIB (Sun and Hampsey, 1995). RAP74 is a protein of 58 kDa which is highly phosphorylated. The TAFII 250 mediated phosphorylation of RAP74 seems to stimulate the activity of TFIIF.

In yeast, the RNA polymerase II is composed of 12 different subunits encoded by RPB genes 1 to 12 in yeast (Woychik and Young, 1994). 9 of the proteins encoded by these genes are either similar or common to all RNA polymerases (type I, II and III). Proteins Rpb 4, 7 and 9 are unique to the RNA polymerase II, and genes RPB 4 and 9

are dispensable for yeast growth. The role of each subunit is not yet fully understood. Some similarities exist between the subunit in yeast and the bacterial RNA polymerase. Therefore Rpb 1 and 2 are homologous to the β and β' subunits, while Rpb 3 is related to the α subunit (Young, 1991). Using yeast genetics and mutational analysis, it was demonstrated that, like the bacterial β' subunit, Rpb 1, the largest subunit (200 kDa) is also involved in DNA binding. Rpb 2 is known to bind nucleotide substrates. Rpb 4 and 7 are weakly bound to the polymerase and were found to easily dissociate and associate to other RNA polymerase II. These proteins could shuttle between RNA polymerase II molecules (Edwards *et al*, 1991). Yeast with a null RPB 4 gene could not grow in nutritionally deprived medium or at elevated temperatures. Furthermore these yeast are inactive in promoter directed transcriptional initiation. These proteins are implicated in recognition of promoters which regulate genes responsible for the stress response (Choder and Young, 1993).

II.2.5 The RNA polymerase, the Carboxyl Terminal Domain (CTD) and associated factors: The role of CTD in RNA polymerase II

The CTD is only found in RNA polymerase II and is composed of heptapeptide repeats of YSPTSPS. The number of repeats varies depending on the species. This varies from 26 to 27 in yeast to 34 in Drosophila and up to 52 repeats in the mouse and human RNA polymerase. The role of the CTD in RNA polymerase II mediated transcription is not well understood and was thought to be involved in phosphorylation dependent switching from the recruited RNA polymerase in the initiation complex to the actively elongating enzyme (Reviewed by Dahmus 1995). However other functions have been attributed to the CTD such as interaction with some basal transcription factors (Kang and Dahmus 1995) and interaction with DNA (Suzuki 1990). However a large number of proteins have been associated with the CTD and are known as Suppressors of RNA polymerase B (SRBs).

II.2.6 Proteins associated with the CTD of RNA polymerase II.

Using yeast genetics, a group of proteins have been identified which associate with the RNA polymerase II (review by Calson 1997). A yeast strain containing only 10 to 12 heptapeptide repeats in the CTD (instead of 26 to 27 in the wild type) was found to be cold temperature sensitive. Yeasts were selected which could suppress this defect by growing at a low temperature and were termed Suppressors of RNA polymerase B (SRB) (Nonet and Young, 1989). Nine different genes (SRB 2 and 4 to 11) were isolated, of which Srb 2, 4 and 7 are associated with purified RNA polymerase. A number of other proteins have been isolated by yeast genetics which are also thought to be components of an RNA polymerase II holoenzyme (Gal11, SIN4, Rgr1 and Rox3) and play similar roles to the Srbs.

These gene products have been linked to the mediation of gene expression. Srbs have been shown to affect repression and activation of gene expression. The repression of particular genes (HO, SUC2) are targeted by the Tup1 and Ssn6 (Cyc8) proteins which interact with site specific DNA binding protein (Mig1, α2) (Roth, 1995). Mutations in some of the SRBs (SIN4, Gal11, Srb 8 to 11) partly prevented the repression. Activation of genes under the Gal4 promoter control also required the activity of SRBs. *In vitro* experiments evidence demonstrate that Srbs are involved in the recruitment of the RNA polymerase to the promoter. This was shown by fusing Srb 2 to the DNA binding domain of LexA and demonstrating that the fusion protein activated transcription from a reporter containing a LexA DNA binding site in the promoter (Farell *et al*, 1996). Srbs 10 and 11 were shown to contain a kinase activity which could phosphorylate the CTD (Liao *et al*, 1995).

II.2.7 TFIIE and TFIIH

The complex formed at this stage is stable but in order to initiate transcription, two further general transcription factors need to be recruited : TFIIE and TFIIH.

TFIIE is a heterotetramer composed of two different subunits of molecular weight 34 and 56 kDa (Ohkuma *et al*, 1990; Inotroza *et al*, 1991). Both subunits are highly charged (with a pI of 4.5 and 9.5 respectively in the human form of TFIIE). The role of TFIIE is still unclear but TFIIE might act in the melting of DNA. Initiation of transcription could be achieved in the absence of TFIIE using a premelted DNA template. Furthermore, cross linking experiments showed that TFIIE lies on the DNA between the TATA box and the start site of transcription (Robert *et al*, 1996). A zinc ribbon motif, identified in an elongating factor TFIIS, is able to bind single stranded and double stranded DNA (Qian *et al*, 1993). Such a motif is also proposed in TFIIE (Maxon and Tijan 1994). As TFIIE is required for the binding of the TFIIH complex to the preinitiation complex, TFIIE might act as a checkpoint for the formation of the preinitiation complex via its control of TFIIH recruitment. This was elegantly demonstrated in a complementation assay : TFIIE from *S. cerevisia* could not replace the TFIIE in a preinitiation complex of *S. pombe*, except if exchanged as a complex TFIIE-TFIIH (Li *et al*, 1994).

TFIIH is a complex factor in which nine different subunits have been identified (Drapkin and Reinberg, 1994). The various subunits possess diverse enzymatic activities such as kinase, DNA helicase, DNA repair and might act during cell cycle regulation. Through its CTD kinase, TFIIH regulates the transition from transcription initiation to elongation. This was suggested by the observation that RNA polymerase II enters the complex in a unphosphorylated IIa form and is converted to the phosphorylated IIo form by CTD kinase (Dahmus, 1995). The kinase enzyme is located in the 40 kDa subunit which is also known as cdk7. This kinase associates with cyclin H (the 37 kDa subunit) and MAT 1 (the 32 kDa subunit of TFIIH) to form the cdk activating kinase factor. The cdk activating kinase is thought to have a key role in the regulation of cell cycle (Morgan, 1995). Although cdk7 was shown to be solely responsible for the phosphorylation of the CTD, its activity was not essential in a

promoter containing a TATA box (Maleka *et al*, 1995). As the cdk activating kinase complex is weakly bound to the TFIIH, it has been proposed that this may be another general transcription factor called TFIIK. In cells only a fifth of the existing cdk activating kinase is associated with TFIIH (Drapkin *et al*, 1996). In yeast, the subunit phosphorylating the CTD is known as Kin 28. This protein is not involved in regulating cell cycle. It is however unknown whether the cdk activating kinase does indeed regulate the cell cycle in mammalian cells.

Another biochemical activity of the TFIIH complex is the nucleotide excision repair (NER) mechanism. Five subunits of TFIIH are involved in DNA repair: ERCC2 and 3, the p62 subunit, the hSSLI (44 kDa) subunit and the p34 (Drapkin *et al*, 1994, Wang *et al*, 1995). This multienzyme complex repairs lesions in the DNA especially in transcribed genes (Mellon and Hanawalt, 1989). Furthermore, ERCC2 and 3 also have a DNA helicase activity.

II.2.8 The polymerase holoenzyme

In vitro, the general transcription factors and the RNA polymerase can be assembled onto a promoter in an ordered and stepwise manner. However the assembly, *in vivo*, of the preinitiation complex probably does not occur in such a way. Purification of the RNA polymerase from various cellular sources showed that the RNA polymerase is already preassembled with other general transcription factors (reviewed Greenblatt 1997). Purification of the RNA polymerase revealed that general transcription factors such as TFIIE, TFIIF, TFIIH co purified. However the general transcription factors which copurified with the RNA polymerase varied depending the source of material and the purification procedure. Some of these interactions seem to be very weak and therefore could easily be disturbed during the purification. Consistent with this idea, purification of the holoenzyme by affinity chromatography led to the isolation of holoenzyme with all five TFII (B, D, F, E and H) in near stochiometric amounts.

II.3 Effect of CTF 1 (NFI) on the formation of the preinitiation complex of transcription.

II.3.1 A CTD motif is present in CTF 1

The mechanism by which NFI stimulates transcription is poorly understood. CTF 1 was found to be the strongest activator in the NFI family of proteins and was therefore used to investigate the role of CTF 1 in transcription stimulation. The residues involved in the stimulation of transcription were mapped to a proline rich region which was sufficient to stimulate transcription when fused to the DNA binding domain of Gal 4 (Kim *et al*, 1993). NFI deleted of the proline rich region was unable to stimulate transcription suggesting that this region alone was responsible for transcription activation. Further mapping of the residues involved in transcription stimulation was achieved by deletion and point mutational analysis of the proline rich domain (Kim *et al*, 1994, Wendler *et al*, 1994). In the proline rich domain, a region spanning from position 460 to 480 was the main factor influencing transcription. This sequence has strong homology to the Carboxyl Terminal Domain (CTD) motif found in the RNA polymerase II and hence was called CTD motif.

II.3.2 Role of the CTD motif in CTF 1.

In yeast, the expression of only the CTD motif of CTF 1 (that is residue 460 to 480 found in the proline rich domain) as a fusion protein with the DNA binding domain of Gal 4 enhances the transcription of a Gal 4 dependent β galactosidase reporter vector. If this CTD motif was removed or replaced by a unrelated sequence, there was a sharp decrease in transcription stimulation. However transcriptional activity could be recovered by adding a single CTD repeat from the RNA polymerase. If several CTD repeats were added, the transcription stimulation was increased (Xiao *et al*, 1994). These workers demonstrated an interaction between the yeast TATA binding protein and the proline rich region. It is supposed that the Proline rich region stabilised the

interaction between TBP and the DNA. However by doing reconstitution experiments, Kim (1994) could not find any interaction between TBP and the proline rich sequence *in vitro*. Instead CTF 1 interacted with the TFIIB complex. Thus the role of the proline region in formation of the preinitiation complex is to facilitate the recruitment of TFIIB on to the TBP-DNA complex *in vitro* in humans and in yeast.

II.3.3 Transcription activation from other CTF proteins.

Recent work on the transcription activation of other CTF proteins have slightly altered the model described above. Vectors expressing either CTF 1, CTF 2, CTF 5 or CTF 7 were co transfected with an NFI dependent β galactosidase reporter. Although they lack elements of the proline region and the CTD motif, CTF 5 and CTF 7 were the strongest activators in the NFI family (Wenzelides *et al*, 1996). Similar results were obtained in Drosophila cells transfected by NFI X1 and X2. NFI X2 which lacks part of the proline region, was the most potent transcriptional activator (Apt *et al*, 1994). These experiments suggest that other regions in the proline region of NFI apart from the CTD motif are involved in stimulation of transcription.

II.3.4 Role of upstream binding factors in transcription

Other transcription factors act in a similar way to NFI. For example, Sp1 which has a glutamine region involved in the transcription also interacts with the TFIIB. TFIIB can associate with the TBP-TATA complex in the absence of activators although stable binding does require activators (Choy and Green 1993). The role of the transcription factor could be to stabilise the TFIIB-TBP interaction(s), and subsequently to alter the recruitment of the other factors in the following steps. The proline rich region also affected another factor in the TFIID complex: TAFII 55, a factor discovered in the human TFIID complex but not in Drosophila. TAFII 55 interacts with various transcription factors such as Sp1, adenoviral E1A and CTF (Chiang *et al*, 1995). Other transcription factors interact with other TAF subunits. For example, VP16 is known to

interact with TAFII 40. It is worth noting that many transcription factors function as dimers and hence it would not be surprising if these dimers had several targets.

Other transcription factors which bind to upstream elements of promoters have been cloned and placed, in three classes, following the existence in their sequence of specific regions: the proline rich region, the acidic rich domain (as found in VP16 activation domain) and the glutamine rich domain (found in Sp1).

These transcription factors could also be classified by their effect in activating transcription. The role of these different types in affecting the transcription has been studied by the "nuclear run on" method and through RNase protection analysis. Both sets of experiments showed that there are three type of transcription modulation domains depending upon their role in stimulating transcription : type I (such as Sp1 and NFI) stimulates initiation, type 2 stimulates only the elongation reaction (Tat) while type 3 stimulates both elongation and initiation (VP16). Mutation of VP16 could convert the transcription factor from a type 3 to type 2 (Blau *et al*, 1996).

II.4 Regulation of NFI activity

II.4.1 The regulation of NFI DNA binding through the structure of chromatin.

In eukaryotic cells, the DNA is structurally organised in chromatin. The DNA is compacted by a complex of proteins known as the histone. The histone is composed of 9 proteins: a core of 4 dimers of proteins H2A, H2B, H3 and H4 locked in place by H1. The DNA wraps around the core proteins and forms almost a double loop of 160 bp.

Early research on the promoter region of the c-myc gene showed that the structure of DNA is important for gene expression. The DNA structure in the promoter region was very different in transformed or quiescent Burkitts cells. The study of DNase I

hypersensitive sites showed a difference between the translocated allele and the non translocated allele. DNase sensitive sites were reported to be located next to NFI DNA binding sites. When these hypersensitive sites were lost, transcription decreased (Sienbenlist *et al*, 1982).

Many researchers used the promoter region of MMTV to study the role of chromatin in transcription. The long terminal repeat of the MMTV is tightly organised into six nucleosomes which prevent any transcription (Reviewed by Wolffe 1994). In the presence of glucocorticoids, the nucleosome was reorganised to stimulate transcription. The glucocorticoid receptors could bind the DNA even when the DNA was organised into a chromatin structure and stimulate recruitment of the SWI/SNF complex. The SWI/SNF complex then induced a restructuring of the chromatin (Farrants *et al*, 1997). This was demonstrated *in vitro* using DNA footprinting on an LTR promoter. The glucocorticoid receptor after induction of glucocorticoid, was able to bind to the LTR and recruited the SWI/SNF complex. The chromatin structure was altered and became more sensitive to nuclease. On the other hand the nucleosome structure inhibited NFI binding to the DNA. In the LTR of MMTV, the NFI site faces toward the inside of the nucleosome core and could not be bound by NFI. In fact Blomquist (1996) showed that as soon as an NFI binding site is found in a nucleosome, NFI was unable to bind the specific site irrespective of its localisation in the nucleosome. However NFI can bind DNA depending on which nucleosomal components were present. By reconstituting nucleosomes *in vitro* with the MMTV LTR, a tetramer nucleosome composed only of the proteins H3 and H4 could bind to NFI and Oct1 but these transcription factors could not bind to the octamer particle made up of the proteins H2A, H2B, H3 and H4 (Spangenberg *et al*, 1998).

Despite the fact that the compaction of the DNA into the nucleosome is detrimental to the transcription activation properties of promoters, surprisingly, the existence of a specific nucleosome organisation does enhance transcription. In the promoter of

Xenopus vitellogenin B1, a nucleosome organises itself between the two regulatory elements (Schild *et al*, 1993). The nucleosome forms itself between the oestrogen responsive unit located at a site of -300 to-330 with a NFI binding site in position -119 to -95. The reconstituted nucleosome stimulates the transcription by bringing into close proximity the various elements regulating transcription. Furthermore, the estrogen receptor and the proline rich region of CTF 1 act in synergy to promote transcription (Martinez *et al*, 1991).

In vivo, Chavez *et al* (1997), prepared a yeast strain producing the H4 protein under the control of an inducible galactose promoter. It was thus possible to modulate the degree of compaction in chromatin structures by growing the yeasts in media containing either glucose or galactose. Growing yeast in glucose media depleted histones from DNA, while using galactose media allowed nucleosome formation. Using an MMT virus LTR reporter plasmid, they measured the activation of the promoter by NFI. In the absence of histones, the promoter was activated while in the presence of histones it was repressed. This is in line with the results obtained previously. Surprisingly, the activation effect of glucocorticoids in the presence of NFI was diminished in the absence of histones suggesting that histones help NFI and glucocorticoids to act in synergy. This was in line with results obtained in-vitro on the promoter of vitellogenin B1. Therefore the role of chromatin is dual : it acts as a repressor but also as an activator depending on the concentration of nucleosome and its position in DNA.

II.4.2 Regulation of NFI activity by post translational modification

Glycosylation: Some transcription factors are glycosylated by linkage of an N-acetyl glucosamine to the protein via a OH group (serine, threonine or tyrosine group). NFI proteins possess sugars residues (Jackson *et al*, 1988). These glycosylated NFI forms were purified using a wheat germ agglutinin column. Using this method two CTF proteins were purified with apparent molecular weight of 62 and 64 kDa (non

glycosylated NFI range from 52 to 66 kDa) (Jackson *et al*, 1989). The function of NFI glycosylation has yet to be determined but the DNA binding activity is not affected by glycosylation. Experiments on the effect of glycosylation on transcriptional activation were mainly performed on Sp1. Wheat germ agglutinin inhibited the transcriptional activation of Sp1, however a further purification of Sp1 by wheat germ did not enhance transcription activation. Comparison of the transcription activation of Sp1 purified by affinity chromatography with Sp1 purified on a wheat germ agglutinin resin, did not show any significant differences. Thus the role of glycosylation is still unclear.

Phosphorylation: Little is known about the phophorylation of NFI. The DNA activated protein kinase is the only kinase to date known to phosphorylate NFI (Reviewed Anderson, 1993). This kinase recognises also a number of transcription factors such as Sp1 and cJun.

Dimerisation: NFI proteins only bind DNA as a dimer. Residues involved in dimerisation of the protein are located in the N terminal domain which is very well conserved. NFI forms homodimers and even heterodimers between a full length NFI and a C terminal truncated fragment of the protein (Kruse and Sippel 1994). These homo and heterodimers are stable and could only be obtained by co-translation. The transcription activation potential of these heterodimers has not been properly investigated yet. Chaudhry (1998) studied the effect on transcription of homo- and heterodimers. Transcription activation of the heterodimer is intermediate to those of the homodimers. Similarly, *in vivo*, NFI B3 regulates transcription by forming a heterodimer with transcriptionally active NFI proteins. The transcription activity of the heterodimer NFI B3 with other CTF (CTF 1, NFI-B1 and NFI-B2) proteins is decreased (2 to 6 times) compared to the homodimers of these same CTF proteins (Liu *et al*, 1997).

II.4.3 NFI regulation by Interaction with other nuclear proteins

A few proteins interact with NFI. Pirin was isolated by a yeast two hybrid system screen as an interactor with CTF 1 (Wendler *et al*, 1997). Pirin may be a cofactor of NFI. Also, the transcription of some promoters is affected by binding of NFI to other proteins. In the human papillomavirus type 16, a regulatory element in the promoter contains an Octamer binding element and a NFI binding site. A synergistic interaction between the two protein binding to these sites, promotes transcription (O'Connor and Bernard, 1995).

II.4.4 Regulation of the activity of NFI through its synthesis.

The synthesis of NFI in the cells has mainly been studied in embryonic cells. In primitive chicken erythrocytes, the NFI DNA binding site acts as a silencer in the synthesis of the α^{π} globin. In these cells, little NFI is present while in the adult synthesis of NFI led to the inhibition of the synthesis of the α^{π} globin (Knezetic and Felsenfeld, 1993). The same observation has been seen in rats which had a partial ablation of their liver. NFI was showed to activate transcription of the promoter coding for α1b adrenergic receptors. The regenerating liver markedly decreased the levels of the adrenergic receptor due to a decrease in the quantity of NFI protein (Gao *et al*, 1996). The role of NFI in transcription seems to affect growing tissue and embryonic cells. Therefore the amount of NFI protein synthesised during embryonic and post embryonic development has been studied (Puzianowska-Kuznicka and Shi, 1996) in Xenopus. The mRNA level and DNA binding activity was studied of NFI-C and NFI-B. NFI-B was not synthesised before stage 54 of the development while synthesis of NFI-C started at stage 33/34. This showed that different NFI proteins are synthesised at various times (sometimes for a short period of time) and in a tissue specific fashion.

II.5 Oncogeneity of NFI.

Overexpression of any NFI protein in chicken embryo fibroblast did not induce oncogenic transformation. Superinfection by an oncogenic retrovirus expressing jun or fos or junD of fibroblasts overexpressing NFI did not lead to transformation. However superinfection by an oncogenic retrovirus expressing ras (Schuur *et al*, 1995) resulted in transformation. However the reduced susceptibility of transformation with jun was not observed when using the C terminal truncated NFI. The effect of Ha-ras transformation on the action of NFI was investigated in murine cells. Transformation of these cells correlated with the down regulation of the expression of NFI protein. The mRNA encoding for NFI was less stable resulting in a low level of protein being encoded and affecting the gene expression of promoters relaying on NFI protein (Nebl *et al*, 1994). Since four genes are available and since alternative splicing occurs, a whole range of NFI protein can be synthesised. In recent years, the purification of NFI specific to some tissues was reported. NFI in the cerebellum is part of the NFI A family in human and NFI B in mouse. The response in gene expression in the presence of NFI A, is more pronounced than using co transfected NFI C protein (Krebs *et al*, 1996, Inoue *et al*, 1990). Studies of various tissues and cells have shown both that the mRNA synthesised and the protein obtained were tissue specific (Goyal *et al*, 1990, McQuillan *et al*, 1991, Paonessa *et al*, 1988). Furthermore other NFI proteins have been purified from other tissues such as in the mammary gland (Furlong *et al*, 1996) and in bone cells (Candeliere *et al*, 1996).

III. Role of Nuclear Factor I in the replication of Adenovirus DNA.

The mechanism of adenovirus replication and the role of NFI has been extensively studied and is now fairly well understood. Replication of small viral genomes was thought to give an insight into the replication mechanism of larger genomes. In this section, I will first briefly discuss adenovirus 2 and its genome and later, the mechanism of DNA replication and the role of NFI in this process.

III.1 Adenovirus: introduction.

Adenoviruses were first isolated while attempting to culture cells from tonsils and adenoidal tissue of children (Rowe 1953). Hilleman and Werner (1954) also found a variety of respiratory illnesses in military personnel due to this virus. Adenoviruses are a family of viruses with two different genera : the mastadenovirus which infects mammals and the aviadenovirus which infects birds (Hortwitz 1990). Over a hundred different serotypes of adenovirus have now been identified, among them 47 serotypes infect humans, causing illnesses such as respiratory, gastrointestinal, ocular and urinary diseases. In 1962, Trentin *et al* found that injection of adenovirus type 12 into new born hamsters induced malignant tumours. However, so far, no epidemiological evidence has been reported linking adenoviruses with malignant disease in human.

As adenoviruses are easily manipulated and obtained, they have been extensively used to study replication, gene expression, gene regulation and cell cycle control. There is now a renewed interest in these viruses since they may be used as a vector for gene therapy.

III.2 Structure of the adenovirus

Adenovirus are non enveloped, regular icosahedrons viruses (20 triangular surfaces and 12 vertices) of 70 to 100 nm in diameter. The virus is composed of an outer shell (the capsid) which surrounds an inner core containing the viral genome (Figure 4). The capsid is composed of 240 hexons and 12 pentons. Each penton makes up the base of a projecting fibre whose length varies among different serotypes (Shenk 1996). The capsid is made of a variety of different proteins (11 proteins identified to date). The hexon is formed of a trimer of protein II and associated with three other proteins VI, VIII and IX. Protein IIIa is also associated with the hexon units. Each penton is made up of five copies of the polypeptide III and is located at the vertex of the icosahedral particle. Three units of polypeptide IV forms the fibre which projects from the penton base. The fibre in most adenovirus species is encoded by one gene whereas 2 genes encode for the fibre in the adenovirus 40 and 41.

The core is composed of the viral DNA and four known proteins : V, VII, Mu and TP. V, VII and mu are highly basic proteins, rich in arginine residues (20 % of residues). Each core protein is in contact with the viral DNA. Polypeptide VII is the major core protein and probably functions as a histone-like protein on which the viral genome is wrapped. Polypeptide V can bind to the penton base and might bridge the core to the capsid. The fourth protein is a terminal protein which is covalently linked to the genome and will be discussed later.

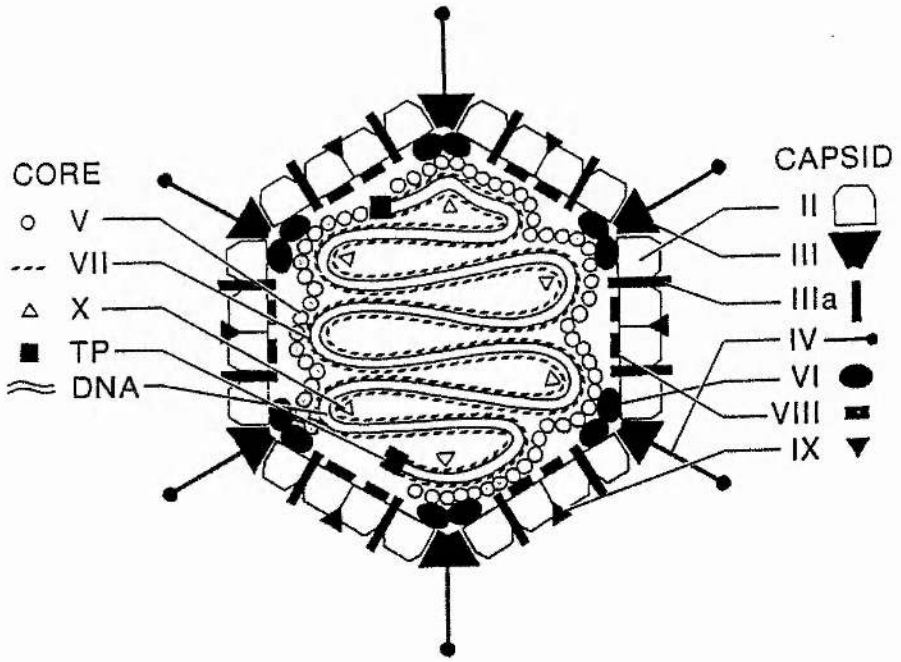


Figure 4 Model of the adenovirus particle with its polypeptide components and DNA.

Virion constituents are designated by their protein numbers with the exception of the terminal protein (TP).

III.3 The viral genome.

The genome is a linear, double stranded DNA of 35 to 36 kbp. Human adenovirus 12 DNA has 34 125 bp (Sprengel *et al*, 1984) while adenovirus 2 has 35 937 bp (Robert *et al*, 1986). The genome has the unusual feature of possessing a protein covalently linked to each 5' end of the DNA.

The terminal protein in adenovirus 2 makes a phosphodiester link between the serine residue in position 580 to a dCMP. Another particular feature is that an identical nucleotide sequence is present at each end of the viral genome. These sequences are known as Inverted Terminal Repeats (ITRs) (Wolfson *et al*, 1972, Garon *et al*, 1972). The length of these ITRs are specific of adenovirus serotypes and ranges from 63 bp in the chicken embryo lethal orphan virus (CELO) to 165 bp in human adenovirus 18 (Garon *et al*, 1982). Adenovirus types 2 and 5 have identical inverted terminal repeats in nucleotide sequence and in length (103 bp).

III.4 Adenovirus replication.

Adenovirus replication is estimated to begin 6 to 8 hours after infection and to reach a maximum 19 hrs after infection coinciding with a 90 % reduction in host DNA synthesis. After 24 hrs, 10^5 to 10^6 copies are synthesised from the original genome. However only 20% are packaged into mature virions.

The first attempts to obtain a cell free system for the replication of DNA were performed by isolating the adenovivus 2 replication complex with its DNA from infected cells (Frenkel 1978). However the isolated DNA was fragmented. The first replication assay *in vitro* was obtained after fractionation of nuclei using ammonium sulphate. The isolated complexes could synthesise DNA but these could only proceed from complexes that had already initiated the replication (Kaplan *et al*, 1977).

Chalberg and Kelly (1979) were the first to set up a fully functional *in vitro* system to study adenovirus replication. A nuclear protein extract from infected cells was isolated and used to obtain the replication of a DNA template purified from the virus. This *in vitro* system lead to the identification of the relevant nucleotide sequences in the viral genome and to the characterisation of three viral and three cellular proteins required for adenovirus 2 DNA replication. Cloning and overexpression of these proteins meant that replication could be accomplished *in vitro* using only purified proteins and DNA.

III.5 Origin of replication in adenovirus *in vivo* and *in vitro*.

Replication of adenovirus DNA was studied, *in vivo*, in infected cells using electron microscopy and pulse labelling techniques.

The first experiments used pulse labelling methods to identify the replicative intermediates in the presence of tritiated thymidine. A significant amount of DNA had a greater sedimentation rate than the virus DNA (Pearson *et al*, 1971). This band was composed of a single stranded DNA molecule as it disappeared after treatment with S1 nuclease (Pearson 1975). Studies from Lechner and Kelly (1977) isolated two different type of replicating DNA molecules which were seen under electron microscope named type I and II. Type I intermediates were described as adenovirus DNA molecules with one or more single stranded branches whereas type II intermediates were adenovirus DNA molecules consisting of a linear double stranded DNA molecule with a single stranded region extending from the other end of the genome.

Further observations led to the following model for the replication of adenovirus DNA *in vivo*. Replication initiates stochastically at either end of the viral DNA and proceeds in a semi conservative way in the 5' to 3' direction with concomitant strand displacement. Type I molecule was therefore a double stranded DNA molecule being replicated with a single strand being displaced. Other rounds of replication were started

on the same molecule before the first round was over therefore creating the existence of several branches of single stranded DNA molecules. DNA replication could also start from a single stranded molecule creating a hybrid molecule partly single stranded and partly double stranded as seen in type II molecules. The inverted terminal repeats region could anneal to one another (Daniell 1976), creating a local region of double stranded DNA. This inverted terminal repeats region could be used to initiate replication from a single stranded DNA molecule. Further evidence supporting this idea of a panhandle structure was given by experiment from Stow (1982) and Hay *et al* (1984). Stow showed that transfection of genomes with deletions of the left inverted terminal repeats generate virions containing the complete inverted terminal repeats at each end of the genome.

Further mapping, *in vitro*, of the origin of DNA replication was realised by *in vitro* DNA replication assay. As shown by the experiment of Lechner and Kelly (1977), DNA replication proceeded from one end of the genome to the other. Therefore the origin of replication was most likely to be in the inverted terminal repeats. The terminus of the viral genome was cloned. By assaying for replication on various deleted inverted terminal repeats and point mutations, the precise nucleotides required for initiation of DNA replication were delimited. To initiate DNA replication, only the first 18 nucleotides were needed. A region between nucleotide 9 and 18 was sensitive to a single nucleotide mutation. In the region upstream, only the first cytosine nucleotide was sensitive to mutation. Point mutation at any other nucleotide between 2 and 8 did not affect the initiation of replication (Tamanoi and Stillman 1983, Challberg and Rawlins 1984, Rawlins *et al*, 1984). Two further regions were delimited in the origin of replication. Both regions correspond to the DNA binding site of a cellular protein. None of these sites was necessary for *in vitro* replication but both stimulated the replication considerably : these were a NFI DNA binding site in the region between 23 and 40 nucleotides which stimulated the initiation of replication by 50 fold (Nagata *et*

al, 1983, Leegwater *et al*, 1985) and a NFIII DNA binding site between nucleotides 39 and 51 which stimulated the initiation of replication by 7 fold (Pruijn *et al*, 1986).

In vivo the requirement of each domain was tested by cotransfecting a wild type adenovirus DNA and various plasmid constructs containing 3' end deletion in the origin. The helper virus brought all the machinery required for replication in these transfected cells. These plasmids replicated efficiently as long as they contained the first 45 bp of the inverted terminal repeats. The plasmid containing only 36 bp replicated very poorly (Hay *et al*, 1984, 1985a 1985b). *In vivo*, adenovirus seems to require only the core origin and the NFI DNA binding site for efficient replication. Further studies using viral genome showed that inverted terminal repeats containing only the 45 bp of the 5' end inverted terminal repeats were as infectious as the virus containing the full inverted terminal repeats. However adenovirus 2 having just the first 18 bp of the inverted terminal repeats were not infectious (Hay and McDougall 1986).

III.6 Positioning and orientation of the DNA binding sites of the auxiliary region

The effect of the positioning and the orientation of these transcription factor binding sites was determined by comparing the relative enhancement obtained on the wild type origin of replication from adenovirus type 2 *in vitro*. Adhya *et al* (1986) reported that a NFI DNA binding site isolated from the human genome was as active in replication as the wild type adenovirus. This site was active in both orientations as long it remained centred between 30 to 36 nucleotides from the 5' end. Positioning of the NFI site from the core origin showed that there was little flexibility. The deletion of 2 nucleotides or the addition of 3 nucleotides between the core origin and the NFI DNA binding site reduces significantly the enhancement effect of replication obtained in the presence of NFI. Virtually no activation by NFI was found if six or more nucleotides were added between the core origin and the NFI binding site. Furthermore the nucleotide

composition in this spacer region was found to affect the stimulation of replication by NFI (Coenjaerts *et al*, 1991, Adhya *et al*, 1986). Similar results were obtained when a spacer region was added between the NFI and NFIII binding site. However the addition of at least 2 nucleotides or the deletion of 7 nucleotides (into the NFI DNA binding site) abolished the stimulation of replication by NFIII. All these results indicate that NFI and NFIII have direct interaction with the viral protein involved in the initiation of replication. *In vivo*, only the positioning requirements of NFI were studied: various plasmid construct with the insertion of 5, 10, 16 and 22 nucleotides between the core origin and the NFI binding site. These constructs and the wild type adenovirus genome were cotransfected and the synthesis of the plasmid DNA was measured. The level of replication of these various constructs was similar to the plasmid containing only the core origin. Therefore NFI lost the ability to stimulate replication in these constructs (Bosher *et al*, 1990).

Sequence comparison between the different serotypes of adenovirus in the inverted terminal repeats demonstrated a highly conserved region between nucleotide 9 and 18 and of the following sequence : 5' - ATAATATACC - 3' (Tolun *et al*, 1979). Also the first nucleotide in the adenovirus genome (a cytosine) is highly conserved among adenovirus serotypes. Furthermore the region between nucleotide 2 and 8 is also very well conserved but variation exists in adenovirus types 4, 7 and 9. However it is important to note that the first 3 or 4 nucleotides are always repeated in position 4 to 6 (for a trimer repeat) or 5 to 8 (for a the 4 nucleotide repeat). Overall this region is extremely rich in nucleotide A and T, probably being a reflection of the fact that this region must open to allow replication initiation. The NFI and NFIII binding sites are conserved between most adenovirus serotypes.

III.7 Viral proteins required for adenovirus replication.

Three viral proteins are needed for adenovirus replication *in vitro* : the DNA binding protein (DBP), the preterminal protein (pTP) and the viral DNA polymerase (Ad pol). All three are transcribed from the same promoter in the viral genome. Two types of mRNA are transcribed depending on their polyadenylation site : E2a and E2b. E2a encodes for the DBP while E2b was for Ad pol and pTP. The mRNA E2a and E2b are detectable 2 hours after infection and reach a maximum after 3 to 4 hrs.

III.7.1 The DNA binding Protein (DBP).

DBP is expressed in high abundance during the infectious cycle and therefore was easily purified and identified (Van der Vliet and Levine 1973). Initially, two proteins were purified (apparent Mr of 48 and 72 kDa) but the smaller protein was in fact a proteolytic product of the larger one (Van der Vliet and Levine 1977). The adenovirus 5 DBP is a protein of 529 residues with a calculated molecular weight of 59 049 Da but migrated oddly in a SDS PAGE at Mr = 72 kDa. Other serotypes of adenovirus encode a similar size DBP (Kruijer *et al*, 1981).

Temperature sensitive mutants (H5ts125, H5ts07) of adenovirus have shown the role of this protein during infection : These temperature sensitive viruses are defective in DNA replication at the restrictive temperature. DNA replication could be restored by adding wild type DBP *in vitro* (Friefeld *et al*, 1983). A subsequent extensive investigation of the phenotype of H5ts125 revealed that DBP is in fact a multifunctional protein involved in a variety of functions such as transcriptional control, transformation and virus assembly (Chase and William 1986, Van der Vliet 1988).

Partial chymotrypsin digest of purified DBP produced two distinguishable protein fragments (apparent Mr = 40 kDa and 27 kDa, Klein *et al*, 1979; Cleghon *et al*, 1993). The 27 kDa fragment is the N terminal region of the protein (residues from 1 to 173). This fragment is highly phosphorylated, not very conserved between the

different adenovirus serotypes and contains two nuclear localisation signals (Morin *et al*, 1989). This protein fragment is involved in the hostrange determination which interfere with the splicing of the late mRNAs (Anderson *et al*, 1983). The C terminal region (from residue 174 to 529) is conserved among the various adenovirus serotypes especially in three regions named CR1 to 3 (Kichingman 1985). The DNA binding activity of DBP is located in this region and this region is sufficient to stimulate replication *in vitro* (Tsernoglou *et al*, 1985).

Two crystal structures of the C terminal fragment of DBP were solved (Tucker *et al*, 1994, Kanellopoulos *et al*, 1996). DBP contains seven α helices (A to G) and three β sheets. Two atoms of zinc are coordinated by the loops that connect the strands of the β sheets. The structural role of the zinc finger was confirmed on temperature sensitive mutants H5ts111 and Ad2NDts23. These viruses have a mutation in the region of the cysteines from the zinc finger. DBP produced by these mutant virus at the restrictive temperature has a different proteolytic digestion pattern. Furthermore all cysteine residues involved in the coordination of the zinc are conserved between the different serotypes of adenovirus. A C-terminal extension of the final 17 residues hooks into residues in a pocket formed of the helix E and G via hydrophobic interactions to an adjacent DBP molecule. This enables DBP to form long chain of protein of 20 to 30 molecules. The role of this long chain of protein is to help cooperative binding of DBP to the DNA. Deletion of this C terminal extension decreases the ability of DBP to form these multimeric proteins in solution but did not abolish it (Dekker *et al*, 1997). Further protein-protein interactions between DBP have been reported involving residues 262, 406, 408, 506 and 511 (Kanellopoulos *et al*, 1996). Unfortunately although co-crystals of DNA-DBP were obtained, their structure has yet to be solved (Kanellopoulos *et al*, 1995). Therefore other methods were used to map the region involved in the DNA binding of DBP (Cleghon *et al*, 1992). By trypsin digestion, the DNA binding region was identified between residues 174 and 470. By crosslinking the protein to a polythymidine template, two peptides were found to be cross-linked to the

DNA at a region between 294 to 308 and 415 to 434. The methionine in position 299 and the phenylalanine in position 418 were directly crosslinked to the DNA. This gave an indication of the regions which may be involved in DNA binding. However the precise nature of all the residues involved in DNA binding is still to be determined.

To fully understand the role of DBP, one must investigate its DNA binding properties. DBP has different DNA binding depending on the structure of DNA. On ssDNA and RNA, DBP binds with higher affinity and in cooperative manner. A model indicates that the ssDNA winds around the DBP creating an extended structure with bases unstacked as predicted by circular dichroism measurements (Van Amerongen *et al*, 1987). These findings are also in agreement with electron microscope observations of DBP-ssDNA (Van der Vliet *et al*, 1978). *In vivo*, only DBP bound to single stranded DNA was isolated from infected cells (Kedinger *et al*, 1978). DBP was also found to bind dsDNA but with a lower affinity compared to ssDNA (Fowlkes *et al*, 1979, Schechter *et al*, 1980). The binding of DBP to dsDNA is not cooperative and dissociates more rapidly.

The role of DBP in replication affects both the initiation as well as in the elongation step. DBP stimulates at the level of initiation although initiation could occur without DBP (Kenny *et al*, 1988, Cleat *et al*, 1989). Kinetics studies showed that DBP lowered the Km of the initiation reaction for dCTP (Mul *et al*, 1993) while at the same time increasing the DNA binding affinity of NFI at the origin of adenovirus 2 replication. It is unclear how DBP affects the initiation of replication, this may be achieved either by affecting DNA structural changes in the origin of replication or via a direct interaction between the DBP and the complex pol-pTP. This latter possibility is more likely although no specific protein - protein contact(s) have yet been established. However pTP-pol binding to the origin is stimulated by DBP (Dekker *et al*, 1997) and the processivity of the polymerase was enhanced by DBP during the elongation stage especially on ssDNA (Lindenbaum *et al*, 1986). Furthermore the sensitivity of the Ad

pol changed to inhibitors such as S-HPMPA in the presence of DBP (Mul *et al*, 1989) suggesting that DBP interacts with the Ad pol.

Another role of DBP is to act in DNA unwinding to dissociate double stranded DNA into single stranded (Monagan *et al*, 1994, Zijderveld *et al*, 1994). Unlike ssDNA replication which does not require DBP, elongation of double stranded DNA is absolutely dependent on DBP. Furthermore adenovirus does not require a helicase. DBP act as the helicase but is ATP independent. The mechanism has been studied by Dekker *et al* and Monaghan *et al* (1994, 1997, 1998). Two deletion forms of DBP were produced : a C-terminal deleted form which lacks the last 17 residues and therefore could not produce multimerised protein chains, and a deleted DBP which lacks a flexible loop (between residue 297 and 331). The C-terminal deleted DBP could still stimulate initiation but could not support elongation probably due to a reduced helix destabilising activity. The DBP deleted of the flexible loop could not unwind DNA and stimulate the elongation process of adenovirus replication. However a mixture of each deleted form of DBP restored the activity of the protein in replication and could unwind the dsDNA. Therefore the results indicate that the C-terminal arm and the flexible loop are required for DNA unwinding and elongation but have distinct functions.

III.7.2 The preTerminal Protein (pTP) and the complex pTP-pol.

Early studies on the replication resulted in the purification of a protein complex of 220 kDa which was required for replication. The pTP and the polymerase form a stable complex in infected cells (Stillman *et al*, 1982). Purification of the pTP showed that its was complexed to the adenovirus polymerase (Enomoto *et al*, 1981) and that this complex was tightly bound and could only be separated in denaturing condition in the presence of 1.7 M urea (Lichy *et al*, 1981).

The terminal protein was first identified by its ability to mediate circularization of the viral DNA through a non convalent interaction and a protease sensitive agent. The

circularizing agent was shown to be a protein of Mr=55 kDa attached to the 5' end viral DNA (Robinson *et al*, 1973). However early replication assays showed that the terminal protein complexed to the Ad DNA polymerase was a protein with an apparent molecular weight of 80 kDa (Challberg 1980). In the initial step of replication, this 80 kDa protein was covanlently cross linked to the dCMP. By studying the gene encoding for the pTP, Stillman found that the adenovirus 2 mutant temperature sensitive (H2ts1) grown at the restrictive temperature contained only the 80 kDa protein while the same mutant grown at the permissive temperature contained three forms of terminal proteins (Mr: 80, 62 and 55 kDa). The mutant virus was able to replicate and form new virions (which were uninfectious) however the pTP remained at a molecular weight of 80 kDa while the protein from the mature virion from wild type adenovirus had a molecular weight of 55 kDa. By performing peptide sequencing from each different form of the terminal protein, these proteins were shown to derived from the pTP by proteolytic cleavage of the N terminus (Smart *et al*, 1982). The attachment point of the DNA to the pTP was also mapped at a serine in position 580. Further evidence showed that the mutation of serine 580 of pTP to a threonine or an alanine abolished initiation of DNA replication (Pettit *et al*, 1989).

Cloning and overexpression of the protein made it possible to identify some of the protein's roles in the infection (Pettit *et al*, 1988, Stunnenberg *et al*, 1988). pTP is synthesised as a 627 residues protein. The main role of pTP in replication involves protein priming. This was shown using an anti-serum raised against TP. This serum inhibited the labelling of pTP by dCMP using the viral genome (Rijinders *et al*, 1983) as well as linearised plasmid with the origin of replication (Van Bergen *et al*, 1983). This serum did not affect the DNA polymerase activity. Using purified pTP, this protein was found to bind DNA with some specificity to the origin of replication (Temperley and Hay 1992).

However other roles have been assigned to pTP. Transfection studies revealed that pTP facilitated the localisation of the Ad pol to the nucleus through a direct interaction

(Zhao *et al*, 1988). Mutants of the Ad pol which lacked the Nuclear Localisation Signal were transported to the nucleus if cotransfected with the pTP hence showing a transport activity from pTP. The NLS in the pTP was mapped between residues 360 and 412. It has also recently been reported that the complex pTP protected the viral genome from the nuclease activity of the DNA polymerase when complexed to the pTP (King *et al*, 1997).

Several approaches were used to define the functional domains in the protein : *in vivo*, mutant pTP was introduced into the viral genome followed by transfection of the virus. Mutant viruses that carried either N terminal or C terminal deletion were not viable, whereas the protein seems more tolerant to deletion and insertions in between these two domains (Roovers *et al*, 1991). Mutational analysis *in vitro* revealed that the gradual deletion of the N terminal residues resulted in a decrease of initiation of DNA replication. Deletion of 18 residues, at the N terminus, inactivates the pTP while pTP retained full activity if 17 residues were deleted at the C terminus (Pettit *et al*, 1989). Larger deletion at the C terminus inactivated pTP (more than 32 residues, Roovers *et al*, 1993). Most point mutation in the pTP resulted in an inactive protein (Friedman *et al*, 1991; Roovers *et al*, 1993).

Further results were obtained by using a combination of partial proteolysis and raising monoclonal antibodies. The DNA binding activity, its ability to form a complex with Ad pol to sustain DNA replication initiation were tested with the various proteolytic products. The first 175 residues are important for DNA binding, however other contact sites were found as monoclonal antibodies recognising the region between residues at position 184 to 200 and 342 to 354 inhibited DNA binding activity of pTP. Ad pol protected pTP from site specific proteases (such as the adenovirus protease, endoproteinase lys-C and endoproteinase Glu-C) the region between residues 175 and 184, but also at the C terminus around the serine involved in the initiation of DNA replication. Finally the C-terminal region (between residue 608 and 671) was found to regulate the DNA binding activity of pTP and dissociation of the pol-pTP complex.

Antibody binding to this region of the pTP increased the stability of the complex pTP-DNA and stabilised the complex pTP-pol (Webster *et al*, 1997a and 1997b)

pTP is degraded during the infection into two other forms known as iTP and TP. Both forms have been observed *in vivo*. The protease responsible for the digestion is encoded by the viral genome. This protease is a 23 kDa protein, which can already be detected 24 hrs after infection but is present at higher concentration later in the infection (after 36 hrs). Protease is present in both the nucleus and the cytoplasm. The protease is synthesised in an inactive form and is activated by a disulphide linked peptide derived from the carboxyl terminus of the pVI virion protein (Webster *et al*, 1993). The protease is a cysteine protease and is specific for the residues (M, I, L)XGX-X. The cleavage sites in adenovirus 2 pTP are after the phenylalanine 175 and the arginine residue 184 to produce iTP. TP is produced by cleavage after the glycine 349. iTP and TP were unable to bind DNA but iTP could still interact with the Ad pol. The iTP-pol complex was still able to sustain initiation and elongation of replication (Webster *et al*, 1997b). However the role of the processing of the terminal protein by the adenovirus protease remains unclear.

Binding and initiation of replication may be regulated by phosphorylation of pTP. *In vivo* pTP was found to be phosphorylated on several serine sites. pTP dephosphorylated by alkaline phosphatase was unable to initiate DNA replication and did not bind DNA (Kusukawa *et al*, 1994).

The terminal protein (TP) crosslinked to DNA should be regarded as a different functional entity with multiple roles . TP protects the viral genome from exonuclease degradation (Dunsworth-Browne *et al*, 1980). Sharp *et al* (1976) found that TP bound to the genome was more infectious than deproteinised DNA. TP itself may play an essential role in the initiation activity since initiation of replication is stimulated twenty fold when using TP-DNA compared to a protein free template. After chymotrypsin digestion, the TP bound to DNA reduced to an approximated molecular weight of 44 kDa is sufficient to promote replication (Pronk *et al*, 1992). The pTP-pol complex

prefers binding to the TP-DNA than naked DNA. This could be a result of the structural changes in DNA such as those observed by DNase footprint techniques in this template upon binding of NFI or NFIII when compared to a naked DNA template (Pronk and Van der Vliet 1993). The TP-DNA could also support correct initiation and direct the pol-pTP to the origin.

Finally adenovirus TP plays an important role in the localisation of the viral genome in the cell. Adenovirus is tightly bound to the nuclear matrix throughout the course of infection due to the interaction of TP with the nuclear matrix (Schaack *et al*, 1990). It has been suggested that the processing of DNA bound pTP into mature TP frees the viral genome from the nuclear matrix prior to packaging (Fredman and Engler 1993). However this theory is in total disagreement with experimental observation that, at the restrictive temperature, although the virus encodes for non active protease, the genome is packed into non-infectious virions.

III.7.3 The DNA dependent Adenovirus DNA polymerase

The existence of a viral DNA polymerase was first suggested by complementation experiment using temperature sensitive mutants of adenovirus type 5. These mutants were defective in both the initiation and the elongation of DNA replication *in vivo* (Stillman *et al*, 1982; Van Bergen *et al*, 1983). *In vitro*, nuclear extracts from cells infected with these mutant viruses could not initiate the replication. However addition of the wild type DNA polymerase restored the replication in these extracts.

cDNA encoding for Ad pol has been cloned by Shu *et al* (1987). This has enabled cloning both into vaccinia virus (Stunnenberg *et al*, 1988) and baculovirus (Watson and Hay 1990). Large quantity of proteins are available and have been purified. Ad pol possesses 1198 residues and catalyse three different reactions: - the initiation of replication by transferring a dCMP onto the serine in pTP, - replicating the entire adenovirus genome (in fact it is able to replicate any DNA strand from a primer), - a 3' to 5' exonuclease. The nuclease function helps increasing the fidelity of replication by

removing mismatched nucleotides (Reha-krantz and Bessman 1977). This exonuclease activity is common to other viral DNA polymerases and was ten times more active on single stranded than on duplex DNA (Field *et al*, 1984).

Sequence homologies were found in DNA polymerases α from human to yeast and shows the conservation of six regions. Five of these conserved regions are also found in the Ad pol (their order in the protein is also conserved). The conserved region VI is absent in Ad pol as well as in the DNA polymerases which uses a protein priming mechanism for replication (Wong *et al*, 1988, Wang *et al*, 1989). Instead a region around this position of the conserved region VI is conserved in protein priming DNA polymerase such as bacteriophage ψ29 and PRD1.

In *E. coli* DNA polymerase I, the 3' to 5' exonuclease active site has been characterised (Joyce and Steitz 1987). Based upon sequence comparison, similar regions were found in bacteriophage ψ29 and showed by mutational analysis to be important for 3' to 5' nuclease activity (Bernad *et al*, 1989) and similar sequences are found in the Ad pol. Since, the crystal structure of a DNA polymerase alpha from bacteriophage RB69 has been solved (Wang *et al*, 1997). This has enabled to further understand which residues are important in DNA polymerase.

Mutational analysis of the Ad pol especially in these regions has failed to produce distinct region involved in specific function (Chen and Horwitz 1989, Roovers *et al*, 1991 and 1993). An exception may be two regions containing several cysteines and histidines between residues 228 to 256 and 1056 to 1090. The region around residue 250 retains a basal DNA polymerase activity but is unable to initiate replication nor to recognise the initiation site on the DNA although it still can interact with pTP and NFI (Joung and Engler 1992). Some domain mapping was achieved using limited proteolysis. Partial proteolysis using endoproteinase lys-C of the Ad pol produced four different domains (from residue 1 to 236, 237 to 539, 521 to 1019 and 879 to 1198) (Parker *et al*, 1998). Using the west-western blotting, the three C-terminal regions

were shown to interact with the pTP. Several non-contiguous regions of the DNA polymerase are involved in interaction with the pTP.

However, enzymatic properties of the Ad pol has been investigated only to a limited extent. Ad pol also resembles the alpha DNA polymerase with respect to its sensitivity in the presence of cytosine beta D arabinose 5' triphosphate, NaCl and N ethylmaleimide and furthermore, both polymerases were active on activated DNA (Field *et al*, 1984). However Ad pol was less sensitive to aphidicolin and to ddNTPs. Using PLP, the DNA polymerase and the exonuclease activity were shown to exist as different domains. PLP could inhibit the DNA polymerase activity and the initiation of the replication without affecting the exonuclease activity of Ad pol. This experiment suggest that the DNA polymerase activity and initiation are carried out by the same domain (Monaghan and Hay 1996).

Recent biochemical experiments have shown the effect of various inhibitors and the affinity of the nucleotides on both the Ad pol alone and the complex pTP-Ad pol. The pTP-pol complex is extremely resistant to inhibitors such as aphidicolin, ddNTP and (S)-HMPA. The pTP-pol complex has also a different affinity to the trinucleotides than the Ad pol (King *et al*, 1997). These experiments suggest that when Ad pol is complexed to the pTP and catalyses the initiation reaction, the protein complex should be consider as a different enzyme to Ad pol catalysing the elongation reaction in the replication.

III.8 Cellular factors.

For an efficient initiation and elongation of the replication of adenovirus 2, three further protein were necessary in HeLa cells. Amongst them, two are DNA binding protein involved in regulating gene expression and one is an enzyme.

III.8.1 The role of Nuclear Factor I in adenovirus 2 replication

Nuclear Factor I was shown to stimulate the replication of adenovirus 2 and 5 *in vitro* (Nagata *et al*, 1982). *In vivo* the presence of its DNA binding site was required for virus replication (Hay 1985). A specific DNA binding site for NFI lies between nucleotide 25 and 40. The mechanism of stimulation of replication by NFI was thought to be complex (Mul *et al*, 1990). The degree of stimulation varied : it appeared to be strongly dependent on the concentration of the complex of pTP-pol. At low concentration of pTP-pol, NFI stimulates up to fifty fold. However at high concentration of pTP-pol, this stimulation is only two fold. This shows that the need for NFI, *in vitro*, can be overcome by high concentrations of pTP-pol. This result probably explains the discrepancies in stimulation level reported previously and strongly indicates the existence of an interaction between NFI and the pTP-pol complex.

NFI was shown to interact directly with Ad pol. Using an immobilised NFI for affinity chromatography, Ad pol and the complex pTP-pol were specifically bound to the resin. However pTP alone could not interact with NFI (Bosher *et al*, 1990). It is unclear yet which residues are involved in the interaction. Interaction between the two proteins were analysed mainly by deletion or site directed mutants. Deletion mutants in a region between residues 68 and 150 could not be coimmunoprecipitated while other mutants did precipitate with an anti DNA polymerase antibody (Chen *et al*, 1990). This region was further defined : double mutants in the region of cysteine 79 - arginine 80 and isoleucine 87 - threonine 88 were able to bind DNA and formed dimers and therefore

could be expected to fold properly. However these mutant were unable to interact with Ad pol and could not stimulate replication (Armentero *et al*, 1994).

Stimulation of replication by NFI was also enhanced in the presence of DBP. DBP increased the association rate of NFI for its site and decreased the rate of dissociation (Cleat and Hay 1989). DBP increased the initiation replication in the presence of NFI by increasing the affinity of NFI for its DNA binding site at non saturating amounts of NFI. However neither DBP nor DNA stabilised significantly the interaction between Ad pol and NFI.

In cells infected by adenovirus 2, immunofluorescence labelling of DBP and NFI showed that they co-localised into subnuclear site, the same sites as DBP which were in the replication areas. This was not observed in cells infected by in adenovirus 4 (Bosher *et al*, 1992). So one can conclude that, in infected cells, NFI is recruited to the replication site.

III.8.2 Nuclear Factor III.

A second replication stimulation factor, Nuclear Factor III (NFIII), was purified from nuclear extract of HeLa cells, which stimulated the initiation of replication in the presence of NFI (Pruijn *et al*, 1986). Using DNase footprint and methylation protection, a specific site on the DNA recognised by NFIII was mapped. This site lies between nucleotides at position 38 to 49 in the adenovirus 2 genome. NFIII stimulates the initiation of replication *in vitro* by three to seven fold (O'Neill and Kelly 1988). Although viruses whose NFIII binding site were deleted remained fully infectious (Hay and McDougall 1986), a later study does suggest that an intact NFIII site is important for efficient growth and DNA replication *in vivo* (Hatfield and Hearing 1993).

Purified NFIII is a protein of 92 kDa (O'Neill *et al*, 1988) and is in fact identical to the ubiquitous transcription factor Oct-1 (O'Niell *et al*, 1988, Pruijn *et al*, 1989). This protein binds preferentially to the octamer sequence ATGCAAAT present in many

promoter and enhancer such as the histone 2b, the immunoglobulin light and heavy chain, the SV40 enhancer and also to the adenovirus type 4 inverted terminal repeat (Pruijn *et al*, 1987). The NFIII binding site in adenovirus 4 has a 3 fold higher affinity for Oct1 than the Ad2/5 binding site (Verrijzer *et al*, 1990).

NFIII belongs, in fact, to a family of transcription factors characterised by a common DNA binding motif, (a variant of a helix turn helix motif forming an unique recognition domain known as POU domain, Herr *et al*, 1988, Cleary and Herr 1995). To date some forty genes with POU domains have been characterised in mammals, Xenopus, fish, Drosophila. On the basis of similarities in the POU domains, the family members have been classified into six subclasses (Stein *et al*, 1996).

Expression of the POU domain on its own was sufficient for DNA binding and recognition of its specific site. The POU domain can be subdivided in two subdomains which are folded independently with an linker domain connecting both subdomain. The linker region is variable in length and in sequence. The carboxy terminal, known as the POU homeodomain (POUhd) binds preferentially to the A/T AAT region of the specific DNA binding site. The N terminal domain recognises the nucleotides ATGC of the DNA binding site and is known as the POU specific domain (POUs) (Verrijzer *et al*, 1990 and 1992).

The structure of the homeodomain was determined by crystallography (Klemm *et al*, 1994) and by NMR (Dekker *et al*, 1993). The structure revealed that both subdomains bind DNA in an independent fashion. Both subdomains made contact to the DNA through a helix turn helix motif with the third helix making the contact to the specific base in the major groove.

The role of NFIII in the initiation is similar to NFI : NFIII affected the Vmax of the initiation of replication but did not affect the Km of the initiation of replication reaction. This result suggested that, like NFI, NFIII interacted with other proteins (Mul *et al*, 1993). Since the binding site of NFIII overlaps with NFI, it was suggested that both could act in a cooperative manner to stimulate replication, but Mul *et al* (1990) found

that NFIII and NFI bound the origin of DNA replication independently, without cooperative effect and did not enhanced replication in synergy.

NFIII interacts directly with the pTP-pol complex as shown using a immobilised GST-NFIII on a glutathione column after addition the pTP-pol complex. Interestingly some point mutations in the NFIII resulted in a stronger affinity to the pol-pTP, (i.e.: the replacement of a glutamine to an alanine at position 24, Coenjaerts *et al*, 1992). The site of interaction was further mapped to the homeodomain of the POU domain. The POU domain was showed to be sufficient to stimulate adenovirus replication (Verrijzer *et al*, 1990) but the homeodomain on its own was not sufficient to stimulate replication. It can also be noted the binding of NFIII to the DNA affected the structure of the DNA and induced a bend (Verrijzer *et al*, 1991).

III.8.3 Nuclear Factor II

Although the protein described previously were sufficient to initiate replication, these proteins could not sustain the synthesis of the total genome. *In vitro* the replication does not proceed beyond 25 to 30 %. A third factor was purified which helped to synthesise the total viral genome (Nagata *et al*, 1983). A complex protein of 30 kDa was purified with contained two polypeptides of 17 and 15 kDa. This protein had topoisomerase activity. Replication accumulates topological stress during the replication fork although the genome has free ends. This could be due to the binding of the TP to the nuclear matrix *in vivo*. *In vitro* the TP could interact making the genome to circularise as shown in the electron microscope (Robinson *et al*, 1973) which led to the discovery of the terminal protein. However more evidence suggest otherwise: the wild type genome was digested with Xba I which produced two DNA fragments one of 1339 and the other of 34 596 bp. *In vitro* replication showed that the replication stops after synthesis of 25 % of the genome of the second fragment (Van der Vliet 1995) which did not occur in the presence of topoisomerase 1. Since these templates

do contain a free end, the blockage is not due to TP-TP interaction. It is still not clear why the genome needs a topoisomerase for replication but this is the function of NFII

III.9 A model for the replication of Adenovirus DNA *in vitro*.

The replication could be subdivided into three main steps : first, the formation of a preinitiation complex and of its binding onto the template, secondly, the synthesis of a protein-priming complex with the pTP and a nucleotide, and in a third step elongation and synthesis of the complete gemone.

In the first step, The DBP is likely to bind the double stranded adenovirus template *in vivo* (figure 5). This has the effect of increasing NFI binding affinity to the origin of replication. By protein-protein interaction, NFI recruits the Ad pol which *in vivo* is also bound to pTP. The pTP-pol complex recognises the origin of replication between nucleotide 9 to 18. This interaction is further stabilised by interaction between the pTP and NFIII. The correct positioning of the polymerase could also be due to the various proteins and on the structure of the DNA (although this remained still unclear). Further stabilisation could come through a direct interaction between the bound TP and other proteins (such as pTP).

In the second stage, the DNA must unwind which takes place within the core of the viral origin of DNA replication and the complex pTP-pol repositions itself so that the protein start synthesising the first three nucleotides (CAT) at position 4 to 6.
This synthesis which is probably carried out by Ad pol is very different in nature from the DNA polymerase activity. The affinity of each nucleotide is different in the initiation step (or synthesis of the CAT) and is much more resistant to inhibitors such as ddNTP, aphidicolin and (S)-HPMPApp. The affinity for the first nucleotide (dCTP)

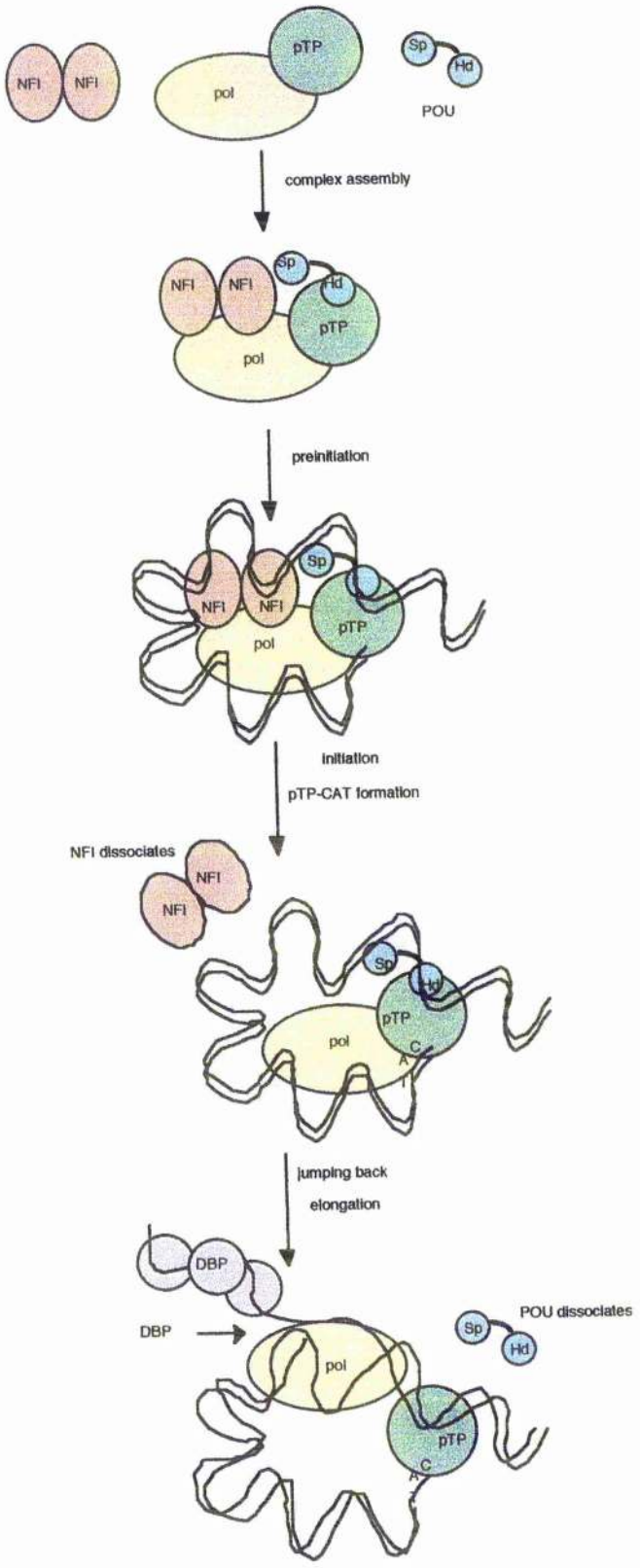


Figure 5 Model of the events occuring during and shortly after the initiation of adenovirus DNA replication.

is much higher than the affinity of Ad pol during elongation. Furthermore this step is carried out by Ad pol complexed to pTP. After synthesis of CAT, the complex dissociates (King *et al*, 1997). The start of synthesis also affect NFI DNA binding which dissociates at this stage (Coenjaerts *et al*, 1994). This was shown using immobilised GST-NFI on a glutathione column and adding the various components of the initiation complex.

Before synthesis, Ad pol must jump back to the start of the genome so that the newly synthesised CAT complement with the first three nucleotides. The event of the jumping back and of the dissociation of the complex pTP-pol is not know how they relate in time. It was clearly showed that both are not linked by using a mutant of the first three nucleotides. The dissociated polymerase is more efficient once dissociated.

In the third step, Ad pol synthesises a new strand of DNA. The synthesis is helped by unwinding of the DNA by DBP and by a topoisomerase. NFIII is dissociated during the elongating phase of the replication.

After synthesis, the replication leaves a single stranded molecule and a double stranded molecule. The non template strand is displaced during this reaction and DBP bound to this molecule. Although the inverted terminal repeats could annealed to form a panhandle structure, DBP is known to inhibit intramolecular interaction but favour intermolecular interaction. The ssDNA molecule has been demonstrated *in vitro* replication can be initiated from single stranded template.

To summarise, the main role of NFI during the replication of adenovirus is to recruit the pTP-pol complex to the origin of replication. The pTP-pol complex is brought to the origin by direct interaction between NFI and the polymerase. The role of NFI is limited to the first steps of the replication process. NFI quickly dissociates from the DNA as soon as the initiation of replication proceeds (Coenjaert *et al*, 1994).

IV. Aim of the project

The aim of this project was to determine the DNA binding domain of NFI. Such a project required large quantity of pure and active protein. Therefore the production and purification of large amounts of active protein was first investigated. Although NFI could be purified using recombinant baculovirus technology, bacterial systems were also studied and used. NFI produced in bacteria formed inclusion bodies, but these could be easily purified and could be renaturated into an active protein . Folding *in vitro* of NFI was further investigated and found to be a convenient method to obtain large quantity of protein.

In most transcription factors, the DNA binding region is localised to one specific region of the protein. To map the minimal region in NFI required to bind DNA, NFI was subjected to limited proteolysis in the presence and absence of DNA. The 180 N terminal residues was found to contain the DNA binding activity. A further region around residues 160 to 180 was also discovered to be critical for the DNA binding activity.

As the DNA binding domain remains quite large and it is expected that only a few residues will obviously be involved in making specific contact with the DNA, we used the differential labelling technique to identify these residues. Using iodoacetate, cysteine residues in NFI involved in close contact with DNA, were radioactively labelled and then identified.

Materials and Methods

I. Composition of commonly used buffers

1xTBE : 90 mM Tris-Borate, 2 mM EDTA

TEN300 : 10mM Tris pH8, 1mM EDTA, 300mM NaCl.

PBS (phosphate-buffered saline) : 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4

Buffer A : 25 mM Hepes NaOH pH7.5, 0.25 mM NaCl

II. Cloning of Nuclear Factor I

II.1 Oligonucleotides

Oligonucleotides used for the construction of the various plasmids for protein expression were obtained either in house (synthesis on an Applied Biosystems oligonucleotide synthesiser, followed by ethanol precipitation), or ordered from Oswell's DNA service (University of Southampton)

Oligonucleotides used in the cloning of pET 22b were

FOR : 5' - AGA G**CC ATG** GAT GAG TTC CAC CCG TTC A - 3'

BACK : 5' - AGA G**GA ATT** CAG AGC TCA GTG ACG CTG AA - 3'

Oligonucleotides used in the cloning of NFI in pGEX 2T cPk were :

FOR : 5' - GT**G GAT CC**A TGT TCC AGG CGG CCG AGG ATG AGT TCC ACC CGT TC - 3'

BACK : 5' - TA**G AAT TC**C GAG CTC AGT GAC GCT GA - 3'.

Oligonucleotides used for insertion of NFI cDNAs into transfer plasmids used for obtaining recombinant baculovirus were :

FOR : 5' - GT**G GAT CC**A TGT TCC AGG CGG CCG AGG ATG AGT TCC ACC CGT TC - 3'

BACK : 5' - TT**G GAT CC**T CAG GTG GAG TCC AAT CC - 3'.

The digestion sites of the restriction enzyme used in cloning are in bold.

II.2 Plasmids

The plasmid used for the bacterial expression system were : pET 22b(+) purchased from Novagen and plasmid pGEX 2T cPk which was a gift from Dr R. E. Randall (Hanke *et al*, 1994).

The plasmid for the construction of recombinant baculovirus were pVL1393 (gift from Dr A. Dawson) and pAcG2T (gift from Dr I. M. Jones (Davies *et al*, 1993)).

II.3 Agarose gel electrophoresis

DNA samples were analysed using a horizontal agarose gel apparatus (10 cm x 10 cm). The concentration of the gel ranged between 0.8 and 2%. DNA samples were mixed 1/6 with loading buffer (6x : 0.25% bromophenol blue, 30% glycerol) and loaded into the wells of the gel. Electrophoresis was carried out in 1xTBE at 80 V till the dye migrated two thirds of the way down the gel. The DNA was stained by ethidium bromide (1 mg/l) and visualised on a UV transilluminator.

To recover DNA fragments from the agarose gel, the desired band was sliced and enclosed in a dialysis bag with a molecular cut off of 3 500 Da (Pierce) with 200 to 400 µl of 1xTBE. The dialysis bag was placed in a electric field of 80 V in 1xTBE till the DNA migrated out of the agarose (usually 20 minutes). The liquid in the dialysis bag was recovered and the DNA fragment was concentration by butanol 2, precipitated by ethanol and the DNA pellet was dried. The DNA was resuspended in water and a fraction of it was used to estimate the level of DNA recovery by gel electrophoresis.

II.4 Amplification of cDNA encoding NFI by Polymerase Chain Reaction (PCR).

The cDNA encoding for the N terminal 222 residues of NFI was obtained by PCR. 100 µl reaction mixtures consisted of PCR buffer (10mM Tris HCl pH 8.8, 50mM KCl, 1.5 mM $MgCl_2$, 0.1% (v/v) Triton X 100), 0.2mM of each nucleotides (dATP, dCTP, dGTP, dTTP), 0.25mM of each primer and 1 unit of Vent DNA polymerase (New England Biolab). The amount of template varied from 1 to 10 ng. When possible, the template used was the original CTF1 clone obtained from Dr R. Tijan (Santoro *et al*, 1989). The conditions for PCR were : A first step incubation for 5 minutes at 94°C then 30 cycles of : 94°C for 2 minutes, 55°C for 1.5 minutes and 72°C for 1.5 minutes. The PCR ended with a final step of 72°C for 9 minutes. The PCR product was checked by agarose gel electrophoresis and purified using a Qiagen PCR purification kit following the manufacturer instructions.

II.5 Enzymatic reactions with DNA

Restriction digest of DNA : PCR products were digested using the appropriate restriction enzyme (usually Eco RI, Bam HI or Nco I) for 3 hours at 37°C following the manufacturer's instructions.

Dephosphorylation of the 5' ends : after endonuclease digestion, plasmid 5' end termini were dephosphorylated by further incubating the digestion product in 10 units of alkaline phosphatase (Boehringer Mannheim) in 50 µl for 1 hour at 37°C.

DNA Ligation : a 3 to 1 molar ratio of insert to vector was combined in a total of 10 µl of ligation buffer (1x : 50mM Tris HCl pH 7.8, 10mM $MgCl_2$, 10mM DTT, 1mM ATP and 25 µg/ml BSA), 200 units of T4 DNA ligase (New England Biolab) were then added and the ligation proceeded at 16°C for a minimum of 12 hours. The ligation mixture was used directly for the transformation of *E. coli* competent cells.

Removal of protruding ends in DNA : the DNA was digested by the mung bean nuclease with 20 units of nuclease for 5 µg of DNA in 20 µl digestion buffer (30mM

sodium acetate pH 4.6, 50mM NaCl, 1mM $ZnCl_2$ and 5% glycerol) at 37°C for 5 minutes.

II.6 Preparation and transformation of competent *E. coli* cells

E. coli XL-1 blue strain (in house supply) was used for transformation. Cells were made chemically competent following the method of Kushner. A 50 ml bacterial culture was grown to an OD of 0.15 to 0.2 at 650 nm. 30 ml of the culture was centrifuged at 1000 g for 10 minutes at 4°C. The bacterial pellet was resuspended in 10 ml of ice cold MOPS buffer I (10mM MOPS NaOH pH 7, 10mM RbCl) and again collected by centrifugation. The cells were gently resuspended in 10 ml ice cold MOPS buffer II (100mM MOPS NaOH pH 6.5, 100mM $CaCl_2$, 10mM RbCl) and kept on ice for a further 1 hour before use.

For each transformation, cells from 0.5 ml of suspension were collected by centrifugation as above, resuspended in 0.2 ml ice cold MOPS buffer II and incubated with plasmid DNA or ligation mix for 30 minutes on ice. Bacteria were heat shocked at 42°C for 90 seconds, then left to recover in LB broth supplemented with 20mM glucose at 37°C for 1 hour. Each transformed bacterial aliquot was spread on LB plates with the appropriate antibiotic (usually 100 µg/l of ampicillin). Two controls were transformed in the same way by using either 1 ng of pUC19 (positive control) or no DNA at all (negative control). Efficiency of transformation was at least 10^6 colonies per µg of pUC19.

II.7 Preparation of plasmid DNA

The alkaline-lysis method of Birnboim and Doly (1979) was chosen for small scale preparation of DNA (minipreps). Large scale preparation were obtained using a Quiagen DNA maxi-prep ion exchange column according to the manufacturer instructions. Quality and quantity of DNA were analysed by spectrophotometry at 260 and 280 nm and through visualisation by agarose gel electrophoresis.

II.8 DNA sequencing

Recombinant plasmids were sequenced using a ABI Prism 377 DNA sequencer (Perkin Elmer).

III. Construction of a recombinant baculovirus expressing NFI and isolation of the virus

III.1 Preparation of the baculovirus DNA

The DNA of a recombinant virus (AcRAK6) expressing β-galactosidase under the control of the polyhedrin promoter was co transfected with the transfer vector for the construction of a NFI expressing recombinant baculovirus. This viral DNA contains two Bsu36 I endonuclease sites one situated in the β-galactosidase gene and one in the ORF1629, a gene necessary for baculovirus infection. The DNA was obtained from Prof. R. T. Hay.

10 µg of the viral DNA was subjected to a digestion with 10 units of the endonuclease Bsu36 I (New England Biolab) in 50mM Tris HCl pH 7.9, 100mM NaCl, 10mM MgCl, 1mM DTT and 0.1 mg/ml BSA in a final volume of 50 µl at 37°C overnight. The mixture was heat treated at 70°C for 20 minutes after the addition of EDTA to a final concentration of 10mM.

III.2 Transfection of Sf9 cell

Recombinant baculovirus expressing NFI cPk and GST-NFI cPk were generated by co transfecting the viral genome of AcRAK6 digested with Bsu36I and the transfer vector pVL1393 NFI cPk or pAcG2T NFI cPk into *Spodoptera frugiperda* cells (Sf9 or Sf21) using lipofectin. 10^7 cells were left to attach in a 25 cm^2 flask in TC100 medium for 2 hours. Prior to transfection, the monolayer was washed twice with 5 ml of OPTIMEM medium (BRL-GIBCO) and incubated with 3 ml of the same medium.

1 µg of Bsu36 I digested AcRAK6 was mixed to 4 µg of transfer vector in 50 µl of water and combined with an equal volume of lipofectin (BRL). This mixture was incubated at 20°C for 15 minutes. The 100 µl lipofectin/DNA solution was added dropwise directly on the cell layer and incubated for 4 hours at 28°C. The medium was then replaced with TC100 supplemented with 10% FCS (Foetal Calf Serum) and incubated 4 days at 28°C. The supernatant of these cells was harvested and used to isolate NFI or GST-NFI expressing baculovirus.

III.3 Purification of a recombinant baculovirus expressing NFI or GST-NFI

Recombinant virus was purified by plaque assay as described by O'Reilly *et al* (1992). The media from the transfected cells was serially diluted a hundred to a million in TC100. 150 µl of each dilution was used to infect Sf9 cell monolayer in a 60mm^2 dish with the cell monolayer being 60-70% confluent. After 1 hour, the monolayer was overlaid with 4 ml of TC100 supplemented with 5% FCS, 1.5% low melting point agar at 40°C and 0.120 mg/ml of 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (Sigma). The plates were incubated in a humid compartment at 28°C. When blue plaques appeared (usually 6 days post infection), the cells were stained with 4 ml of TC100 containing 0.5% neutral red for 2 hours. Ten colonies were picked and the viruses of each plaque assayed for expression of NFI and of β-galactosidase. Virus stocks were gradually built up by infection of fresh Sf9 cells.

IV. Expression and purification of Nuclear Factor I

IV.1 Expression of NFI and of GST-NFI in *E. coli*

The *E. coli* BL21 (DE3) strain was used for the expression of NFI using the pET 22b plasmids. Expression of the fusion protein GST-NFI was achieved in *E. coli* strain XL1-blue. A single bacterial colony from a LB agar plate containing 100 mg/l of

ampicillin was picked and used to inoculate 50 ml of LB broth supplemented with 100 mg/l of ampicillin. The culture was left overnight at 37°C under agitation. The overnight culture was scaled up to 1 litre in LB-broth with 100 mg/l of ampicillin and grown till the OD of the culture reached 0.6 at 600nm. The bacterial colony was induced by 1mM IPTG, for 3 to 4 hours at 20°C for the production of GST-NFI or at 37°C for NFI. Cells were harvested by centrifugation at 6 000 g for 10 minutes at 4°. Pellets were washed once with PBS and stored at -70°C until needed.

IV.2 Expression of NFI and GST-NFI in insect cells

IV.2.1 Cell culture

Spodoptera frugiperda cells were maintained in exponential growth in TC100 medium (Gibco) supplemented with 7% foetal calf serum (Sera Lab Ltd).

IV.2.2 Virus stock

Recombinant baculovirus (*Autographa californica*) expressing the DNA binding domain of NFI was obtained from R. T. Hay (Bosher *et al*, 1991). Titres of virus stocks were determined by plaque assay as described above.

IV.2.3 Optimisation of expression of protein from a stock solution of baculovirus

To determine the optimal expression of NFI and GST-NFI from a particular virus stock, the following method was used. $5x10^5$ Sf9 cells were added to each well in a 24 well plate. In each individual well, cells were infected with various volumes of the virus stock and made up to 150 µl with TC100 and left to infect for 1 hour. 1ml of TC100 supplemented with 7% FCS was added in each well. Infected cells were harvested after 48 to 72 hour post infection and the expression of the protein in each well was analysed by SDS PAGE followed by western blotting.

IV.2.4 Infection of Sf9 with recombinant baculovirus

500 ml (10^6 cell per ml) of SF9 in a spinner culture were collected by centrifugation at 400 g for 5 minutes and resuspended in the presence of the virus stock. Cells were incubated 1 to 3 hours at 28°C with the virus before adding TC100 with 7% FCS to 500 ml and incubated in a spinner culture at 28°C for 48 to 72 hours.

IV.3 Purification of GST-NFI fusion protein

IV.3.1 Purification of GST-NFI expressed in *E. coli*

Induced cells were resuspended in 5 ml (per litre of culture) of 50mM Hepes-NaOH pH 7.5, 0.5 M NaCl, 1mM PMSF, 5mM DTT. Triton X-100 was added to a concentration of 1% and the extract was ultrasound sonicated with 6 burst for 30 seconds. The extract was centrifuged for 30 minutes at 30 000 g. The supernatant was then used to purify GST-NFI on a immobilised glutathione column.

IV.3.2 Purification of GST-NFI expressed in Sf9

500 ml of a spinner culture of Sf9 cells infected for 48 hours at 28°C, were harvested by centrifugation at 400 g for 5 minutes. From then on all steps were carried out either on ice or in the cold room. Buffers and centrifuges were cooled to 4°C. The cells were washed once with PBS and resuspended in 5 ml of 25mM Hepes-NaOH pH 7.5, 5mM DTT, 0.5 M NaCl, 1mM sodium metabisulfate, 1mM PMSF and 10 µg/ml pepstatin and incubated for 20 minutes. This cellular extract was centrifuged at 45 000 g for 30 minutes on a Beckman TL100 ultracentrifuge. The supernatant was used to purify GST-NFI by immobilised glutathione affinity chromatography.

IV.3.3 Glutathione affinity chromatography

The column was equilibrated in 25 to 50mM Hepes-NaOH pH 7.5, 1mM DTT and 0.5 M NaCl. The protein extract was loaded and unbound proteins were removed by washing with 5 to 10 column volume of the equilibrating buffer. Bound protein was

eluted in 50mM Hepes-NaOH pH 7.5, 0.2 M NaCl, 10 % glycerol and 10mM glutathione (pH 7 with NaOH). An aliquot of each fraction was run on a 10% SDS PAGE.

IV.4 Purification of Nuclear Factor I

IV.4.1 Purification of Nuclear Factor I as inclusion bodies and large scale folding *in vitro*

The pellet from each litre of bacterial culture was resuspended in 10 ml of 50mM Hepes-NaOH pH 7.5, 0.5 M NaCl, 1mM PMSF, 5mM DTT and 0.35 mg/ml lysozyme and incubated 30 minutes at 20°C. Triton X-100 was added to a concentration of 1% (v/v) followed by ultrasound sonication of small burst of 30 seconds till the solution became homogeneous. The extract was then treated by 20 mg/l of DNase I for 1 hour at 37°C. The soluble fraction was separated from the inclusion bodies by centrifugation for 30 minutes at 30 000 g at 4°C. The pellet was washed twice with PBS, 1% Triton X-100.

Pellets were dissolved in 50mM Hepes-NaOH pH 7.5, 6 M guanidine, 25mM DTT and incubated for 2 hours. Insoluble material was precipitated by centrifugation at 300 000g for 10 minutes at 20°C. The soluble material was diluted very rapidly to a final protein concentration of 0.1 mg/ml using a syringe and a needle into 50mM Hepes-NaOH pH 7.5, 1 M NaCl, 1mM DTT, 1 M NDSB additive under vigorous agitation at 4°C in a glass beaker. Agitation was stopped 2 minutes later and the folding mixture left overnight on ice.

The solution was then diluted to a NaCl concentration of 0.3 M using 50mM Hepes-NaOH pH 7.5, 1mM DTT and 0.1 mg/ml BSA and NFI was finally purified by ion exchange and/or by DNA affinity chromatography. The purity was estimated by SDS PAGE.

IV.4.2 Purification of NFI from Spodoptera frugiperda cells infected with recombinant baculovirus

500 ml of SF9 cell culture was infected with recombinant baculovirus for 72 hours at 28°C and harvested by centrifugation at 400 g for 5 minutes. From then on all step were carried out either on ice or in the cold room. Buffers and centrifuges were cooled to 4°C. The cells were washed once with PBS and resuspended in 5 ml of 25mM Hepes-NaOH pH 7.5, 5mM KCl, 0.5mM $MgCl_2$, 1mM benzamidine, 1mM sodium metabisulfite, 1mM PMSF, 10 µg/ml antipain, 10 µg/ml pepstatin and 10 µg/ml leupeptin. The cells were disrupted with 20 strokes in a Dounce homogeniser using a type B pestle. Nuclei were collected by centrifugation at 2 000 g for 5 minutes. The nuclei were resuspended in 5ml of 25mM Hepes-NaOH pH 7.5, 5mM DTT, 0.4 M NaCl, 1mM PMSF, 10 µg/ml antipain, 10 µg/ml pepstatin and 10 µg/ml leupeptin and incubated for 20 minutes. The nuclear extract was diluted two fold in 25mM Hepes-NaOH pH 7.5 before centrifuging the extract at 45 000 g for 30 minutes at 4°C. The supernatant was used to purify NFI by ion exchange chromatography and by affinity chromatography.

IV.4.3 Ion Exchange Chromatography

Two types of ion exchange resin were used : either Biogel 70 which is a weak anion exchange column (BioRad) or a fast flow strong anion exchange column (Fast flow SP Sepharose Pharmacia). Biogel was prepared as follow : 2 ml of resin powder was resuspended in 8ml of 250mM Hepes NaOH pH 8, 1 M NaCl, 10mM EDTA for 1 hour. The suspension was left to settle and the resin was then washed twice in deionised water.

The resin was preequilibrated in 25 mM Hepes NaOH pH7.5, 0.2 M NaCl, 1 mM DTT. The protein extract was loaded onto the column, unbound protein washed by Buffer A, 1 mM DTT and bound protein eluted by 25 mM Hepes NaOH pH 7.5, 0.7 M NaCl and 1 mM DTT. An aliquot of each fraction was used for SDS PAGE.

IV.4.4 DNA Affinity Chromatography

Preparation of an DNA affinity resin

The scheme for preparing the oligonucleotide affinity resin was based on Kadonaga and Tijan (1986). To generate a NFI affinity column, 4 mg of a single stranded amino linked to the 5' end oligonucleotide (5' -GAT CTT ATT TGG ATT GAA GGC CAA TAT G) was annealed to a complementary oligonucleotide (5' - GAT CCA TAT TGG CCT TCA ATC CAA ATA A), precipitated twice in ethanol and resuspended in deionised water.

For the cross linking reaction, 3 g of cyanogen bromide activated (CNBr) Sepharose 4B (Sigma) was swollen in 15 ml of 1mM HCl for 1 hour. Beads were washed on a stintered glass funnel with 400ml of 1mM HCl, then by 500 ml of ultrapure water, followed by 400 ml of 10mM sodium phosphate pH 8. The resin was then left to react with 8 mg of double stranded oligonucleotide overnight at 20°C under gentle rotation. The resin was collected by centrifugation at 150 g for 1 minute. The supernatant discarded, and the resin was washed once with 5 ml of 10mM sodium phosphate pH 8. Unreacted CNBr sites were blocked by incubating the resin in 100 ml of 1M ethanolamine pH 8 under mild rotation for 4 hours. The resin was then washed with 100 ml of 10mM sodium phosphate pH 8, 0.1 M NaCl, followed by 100ml of 2M NaCl and finally using 100ml of TEN300 with 0.02% sodium azide. The NFI affinity resin was stored at 4°C in TEN300, 0.02% NaN3.

Purification of NFI by affinity chromatography

Prior to loading with a protein extract, the DNA-Sepharose was washed with 10 column volumes with 25mM Hepes-NaOH pH7.5, 1mM EDTA, 1mM DTT and 0.2 M NaCl. The protein extract was diluted so that the NaCl concentration was brought down to 0.25 - 0.3 M with 25mM Hepes-NaOH pH7.5, 10mM DTT. 6 μg of poly (dI-dC) or non specific double stranded oligonucleotide was added per mg total protein. The extract was then loaded onto the column. Unbound protein were eluted by 25mM Hepes-NaOH pH 7.5, 1mM EDTA, 1mM DTT and 0.2 M NaCl. Bound NFI

was eluted in 25mM Hepes-NaOH pH 7.5, 1 M NaCl, 1mM DTT, 10 % glycerol. Fraction were analysed by SDS PAGE.

V. Analysis of NFI protein

V.1 Determination of protein concentration

The protein concentration was determined using the method of Bradford (1976). A kit was purchased from Bio-rad and used following the manufacturer's instructions. Protein samples (5 to 50 µl) were mixed with 990 µl of Bradford's reagent and incubated for 5 minutes at room temperature before measuring the absorbance of the solution at 595 nm. Protein concentration of the sample was determined from a standard curve obtained using BSA as standard.

V.2 Polyacrylamide gel electrophoresis in denaturing condition and Western blot analysis.

10 to 15% polyacrylamide gels were cast using a Bio-Rad Mini Protean apparatus from a 30% (w/v) stock polyacrylamide solution (29% acrylamide : 1% bisacrylamide) in 0.375 M Tris pH 8.9 and polymerised by the addition of 0.1 % (w/v) ammonium persuphate and 1/1000 volume of TEMED (BDH). The stacking gel contained 5% acrylamide in 0.125 M Bis-Tris HCl pH 6.5. Protein sample were diluted in 4x denaturing buffer (1x : 1% SDS, 0.175 M β-mercaptoethanol, 8% v/v glycerol and 0.06% w/v bromophenol blue) and boiled for 5 minutes at 100°C before loading. The electrophoresis buffer was 0.025 M Tris, 0.19 M glycine and 0.1% SDS. Once the migration front reached the bottom of the gel, electrophoresis was stopped and the gel apparatus dismantled. The gels were either fixed in 10% acetic acid, stained in 0.1% Coomassie blue G250, 10% acetic acid and 20% ethanol and destained in 10% acetic acid and 20% ethanol, or transferred onto a PVDF membrane for immunodetection as described as follow.

Separated polypeptides were transferred unto a PVDF membrane (Millipore) in a semi dry blotter in 48mM Tris, 190mM glycine, 20% ethanol for 1 hour at 1 mA/mm^2 of gel. After transfer, the membrane was blocked with PBS containing 5% (w/v) skimmed milk powder and 0.1% Tween 20, then incubated with the appropriated antibody at a dilution ranging from 1/1000 to 1/5000 in the same buffer. The membrane was washed twice before adding an HRP antibody raised against the previous antibody in blocking buffer for 30 minutes at a dilution of 1/5000. The membrane was washed then developed by the enhanced chemiluminiscence (ECL; Amersham International) protocol according to the manufacturers recommendations.

When protein separated by SDS PAGE were to be sequenced, the proteins were transferred to a PVDF membrane as described above with the following variations : i) the separating and stacking gel were degassed prior to polymerisation. ii) The separating gel was prepared the day before its use and stored in 0.375 M Tris pH 8.8 overnight. iii) The gel was prerun with electrophoresis buffer at the anode and a buffer of 0.125 M Tris, 0.1% SDS, 60µM glutathione at the cathode. iv) The gel was run with electrophoresis buffer supplemented with 100µM β-mercaptoacetic acid at the cathode. v) Transfer was carried out in a wet blotter (Bio-Rad) for 1 hours at 300 mA in 10mM CAPS pH 11, 10% ethanol on a PVDF membrane (Perkin Elmer). vi) The membrane was stained with 0.1% Coomassie blue R250, 40% ethanol and 1% acetic acid for 1 to 2 minutes and destined in 40% ethanol, 1% acetic acid. Bands were cut out and stored in an eppendorf tube at -20 C.

V.3 Preparation of a radioactively labelled double stranded oligonucleotide

Two complementary oligonucleotides (GAT CTT ATT TTGGCT TCA AGC CAA TAT G and GAT CCA TAT TGG CTT CAA GCC AAA ATA A) which contained a NFI DNA binding site were synthesised and annealed.

The recessives ends of the double stranded oligonucleotide were labelled with ^{32}P dCMP using the *E. coli* DNA polymerase (Klenow fragment) (NEB). 500 ng of

double stranded oligonucleotide was mixed with 33 μM of dATP, dGTP, dTTP and 20 μCi [α-^{32}P] dCTP (3000 Ci/mmol) with 1 to 5 units of DNA polymerase in 10mM Tris HCl pH 7.5, 5mM $MgCl_2$, 7.5mM dithiothreitol. The reaction was incubated at 25°C for 15 minutes. 33 μM of dCTP was added and the reaction left for a further 5 minutes before stopping the reaction by the addition at a final concentration of 20mM EDTA and 10% glycerol.

The radioactively labelled DNA was purified by running on a 12% polyacrylamide gel (29 acrylamide : 1 bisacrylamide) in 1 x TBE. An X-ray film was exposed to the gel for 5 minutes in order to detect the position of the probe. The appropriate band was cut out and eluted overnight in 300 μl of TEN300 at room temperature. The buffer containing the probe was decanted and stored at 4°C

V.4 Gel Electrophoresis DNA binding assay

Prior to the gel retention assay, NFI was incubated with 20mM freshly prepared DTT. 1 to 100 ng of purified NFI or 1 to 6 μg of nuclear or cellular extract was incubated with 50mM Hepes-NaOH pH 7.5, 200mM NaCl, 5mM $MgCl_2$ and 10 ng to 1 μg of a non specific double stranded oligonucleotide (the amount depended on the purity of NFI). After 20 minutes, 1 to 5 ng of labelled double stranded DNA was added and incubated a further 30 minutes at 4°C. The NFI-DNA complex was separated from free DNA on a 8% polyacrylamide gel (55 acrylamide : 1 bisacrylamide). The gel was prerun at 150 V for 30 minutes prior to loading. Electrophoresis was carried out in 0.5xTBE for 90 minutes at 150 V. The gel was fixed in 10% acetic acid for 10 minutes, dried and exposed either to X-Ray film or to a phosphoimager screen.

V.5 Limited proteolysis of Nuclear Factor I

NFI was dialysed to 25mM Hepes-NaOH pH 7.5, 0.25 M NaCl, 1mM DTT. Prior to digestion, NFI was incubated 20 minutes on ice with 20mM DTT. Trypsin, chymotrypsin, V8 and proteinase K were added at ratio ranging from 1/1 to 1/1000 (w/w) of protease to NFI. The digestion reaction was incubated at 25°C for 1 hour.

1mM PMSF was then added to quench the proteolysis reaction and incubated on ice for 20 minutes. The digestion products obtained were analysed by SDS PAGE and the various digest product were N terminally sequence (Perkin Elmer procise protein sequencer, University of St. Andrews) and their molecular weight determined by laser desorption mass spectrometry (University of Aberdeen).

To analyse the susceptibility of DNA bound NFI to proteolytic attach was studied by adding a 2 fold molar excess of a double stranded DNA containing a NFI binding and incubating for 30 minutes on ice.

V.6 Iodoacetic labelling of Nuclear Factor I

100 to 200 µg of NFI was dialysed in 25mM Hepes-NaOH pH 7.5, 0.2 M NaCl, 1mM DTT and 1mM EDTA. NFI was incubated at 4°C with 200µl of NFI affinity resin for 1 hour. The resin was collected in a eppendorf by a brief centrifugation at 200 g and the supernatant removed. The resin was then washed twice with 1 ml of buffer A. NFIDBD bound to the column was treated with 10mM iodoacetic acid (Sigma) in buffer A for 30 minutes at 20°C (the iodoacetic acid was equilibrated to pH7 with NaOH). The affinity resin was then washed twice with 1 ml of buffer A and 2mM DTT. NFI was eluted in 25mM Hepes-NaOH pH 7.5, 1 M NaCl and 2mM DTT. NFI was then dialysed for 4 hours buffer A and 1mM DTT before concentration in a microconcentrater (Centricon-30) with a membrane with a molecular cut off of 30 kDa by centrifugation for 1 hour. NFI was then treated with a neutral solution of ^{14}C iodoacetic acid (ICN) for 30 minutes at 20°C. The reaction was quenched by the addition of 50mM DTT and left on ice for 15 minutes.

V.7 Isolation, identification of radiolabelled cysteine containing peptide

NFI labelled by ^{14}C was precipitated by the adding to a final concentration 10% trichoroacetic acid and left on ice for 1 hour prior to centrifugation at 100 000 g at 4°C for 20 minutes. The supernatant was discarded and the pellet washed twice with ice

cold acetone before being dried in air. The pellet was resuspended in 8 M deionised urea, 100mM Tris HCl pH 8 and the urea concentration lowered to 2 M by addition of water. NFI was digested by trypsin at a ratio of 1/50 (w/w) for 20 hours at 37°C. The peptides were separated by reverse phase chromatography using a Waters HPLC system with a Delta C18 column. The peptides were eluted in a gradient of 0 to 80% acetonitrile in 0.06% (v/v) trifluoroacetic acid at a flow rate of 0.5 ml/min. 0.5 ml fractions were collected and a 10 µl aliquot of each fraction was added to 5 ml Ecoscint A and scintillation counted. Peak radioactivity fractions were sent for N terminal amino acid sequencing (University of St. Andrews).

V.8 Small Scale folding of Nuclear Factor I

Folding *in vitro* of NFI was performed by quick dilution. Purified inclusion body was resuspended in 6 M guanidine, 50mM Hepes NaOH pH 7.5, 25mM DTT (or β-mercaptoethanol) for 1 hour at 20°C. Insoluble material was precipitated by centrifugation at 100 000g for 10 minutes at 20°C. Unless stated, the folding was performed in 50mM Hepes-NaOH pH 7.5, 1 M guanidine, 5mM DTT and 1 M NDSB 256t (synthesised by Dr Vuillard) in a final volume of 0.5 ml. The folding buffer and the denatured NFI were added into the same eppendorf tube taking care not to mix them. The amount of stock NFI was chosen to achieve a final concentration of 0.1 mg/ml. Mixing was carried out by vigorous vortexing for 10 seconds at 20°C. The folding reaction was left overnight at 4°C before testing the activity of NFI by gel retention assay.

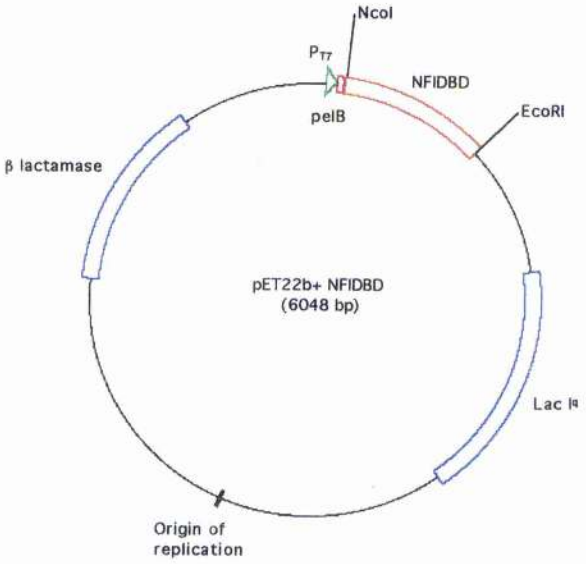
Chapter 1 : Cloning, expression and purification of Nuclear Factor I DNA Binding Domain (NFIDBD)

Understanding the precise details of the interaction between protein and DNA requires large amounts of pure and active protein. In mammalian cells, several isoforms of Nuclear Factor I are produced in small amounts due to variable splicing and post translational modification. Various means of overexpressing the protein were investigated in order to obtain large amounts of an active fragment of NFI sufficient for DNA binding (NFIDBD). Although a recombinant baculovirus expressing NFIDBD already existed, this protein had a higher molecular weight than expected (determined by mass spectrometry, University of Aberdeen). Furthermore NFIDBD migrated as a doublet and contained a blocked amino terminal, making the study of the regions involved in DNA binding difficult. One of my first objectives was to obtained a NFIDBD with the proper molecular weight and with a free amino terminus. Therefore in this chapter, two different protein expression systems (bacteria and baculovirus) were investigated.

I. Cloning and expression in bacteria

Two different *E. coli* expression vectors were chosen for the expression of NFIDBD: the pET22b+ and the pGEX2T cPk (figure 1.1). Both vectors contain a bacterial origin of replication, an ampicillin resistance gene (β lactamase) and a gene encoding for the lactose repressor protein (Lac I^q).

In pET 22b+, the recombinant protein is expressed at the C terminus of a periplasmic localisation signal (pelB) under the control of a T7 promoter and a lactose operator. Recombinant proteins should then be exported to the periplasmic region of the bacteria.



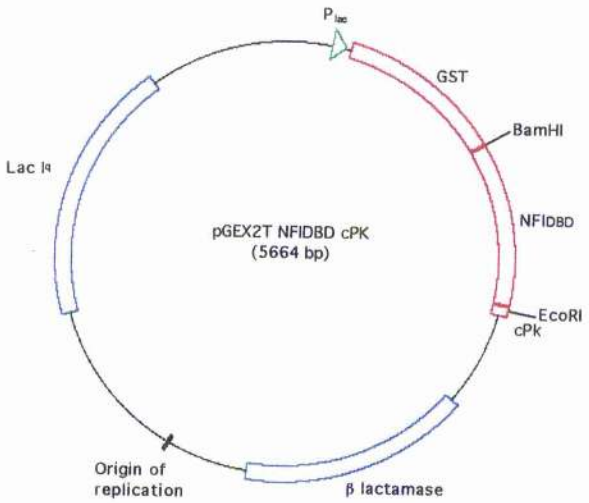


Figure 1.1 Schematic representation of the vectors pET22b+ and pGEX2T cPk used for the expression of Nuclear Factor I in bacteria.

The cDNA encoding for the 222 amino terminal of Nuclear Factor I was inserted in the restriction sites NcoI and EcoRI in the vector pET22b+ and BamHI and EcoRI in the pGEX2TcPk. The green arrow represents the promoter involved in the expression of the inserted gene. Red and orange boxes represent the protein expressed after induction with IPTG and the blue boxes represent other proteins expressed by the vector in the bacteria.

During the process of export of the protein, the localisation signal is cleaved off from the protein by a specific peptidase.

The vector pGEX2T cPk was derived from the vector pGEX2T (Smith and Johnson 1988). This vector had the coding region of a portion of the P protein of the Simian Virus (SV5) for which a monoclonal antibody is available (Pk). The coding region of the antibody binding site was inserted at the 3' end of the multiple cloning site. In this vector, Nuclear Factor I is expressed as a fusion protein with the glutathione-S-Transferase at the N terminus which allows purification by affinity chromatography on immobilised glutathione and the Pk site at the C terminus. Protein expression is under the control of the lactose promoter which is inducible by IPTG.

I.1 Construction of the bacterial expression vectors.

The coding region for NFIDBD was amplified using the polymerase chain reaction (PCR). The template was the original CTF1 plasmid obtained from Dr. R. Tjian (Santoro *et al*, 1988). Primers used in the PCR for the cloning into the pET22 system had an NcoI site at the 5' end of the gene which kept the coding sequence of NFIDBD in the same reading frame as the localisation signal. The reverse primer had a stop codon followed by a Eco RI restriction site.

Primers used in the PCR for the cloning of NFIDBD into pGEX2T contained a BamHI site at the 5' end and the coding sequences of the first six residues of IkBα inserted after the thrombin coding sequence. The expressed GST-IκBα was found to be efficiently cleaved by thrombin. It was expected that by adding these six residues from the N terminal region of IκBα between NFIDBD and the thrombin cleavage site, the fusion protein GST-NFIDBD would be more sensitive to proteolysis by thrombin. The reverse primer possessed an EcoRI site and the reading frame for the Pk antibody recognition site was not altered. Therefore a codon encoding for a glycine residue was added between the NFIDBD coding region and the EcoRI restriction site.

Analysis of the PCR products revealed a single band of DNA migrating with the expected mobility in an agarose gel. Vectors and PCR products were digested using

appropriate restriction enzymes: NcoI and EcoRI for cloning into pET22b+, BamHI and EcoRI for cloning into pGEX2T. Ligation products were used to transform competent *E. coli* XL1 Blue strain (chemically prepared competent *E. coli*). Colonies resistant to ampicillin were individually picked and grown to obtain a small scale preparation of plasmid DNA. The presence of the cDNA encoding for NFIDBD in the vector was checked in each plasmid.

I.2 Protein expression and purification in bacteria

I.2.1 Expression and purification of NFIDBD

For protein expression using the pET22b+ vector, the plasmid had to be transformed into a special strain of bacteria which expressed the T7 RNA polymerase. The bacterial strain BL21 (DE3) is an *E. coli* strain in which the T7 RNA polymerase, inserted into the chromosomal DNA, is under the control of an inducible lactose promoter. By adding IPTG to the media, the T7 RNA polymerase is expressed which then specifically transcribes genes which are under the control of a T7 promoter.

NFIDBD expression was obtained by the addition of 1 mM IPTG into the media of the transformed bacterial culture when the culture reached an OD of 0.6 at 600 nm. The culture was incubated for a further 4 hours at 37°C. Total bacterial lysates were analysed by SDS PAGE followed by staining with Coomassie Brilliant Blue (figure 1.2). In the presence of IPTG, a polypeptide was detected with an apparent molecular weight of about 30 kDa.

To lyse the cells, bacteria were treated by lysozyme (0.7 mg/ml) and their membranes disrupted by Triton X-100 and sonication. The lysate was then centrifuged to separate insoluble and soluble material. The insoluble fraction was resuspended in a buffer containing a strong detergent (1% SDS in PBS). Each fraction was analysed by SDS PAGE.

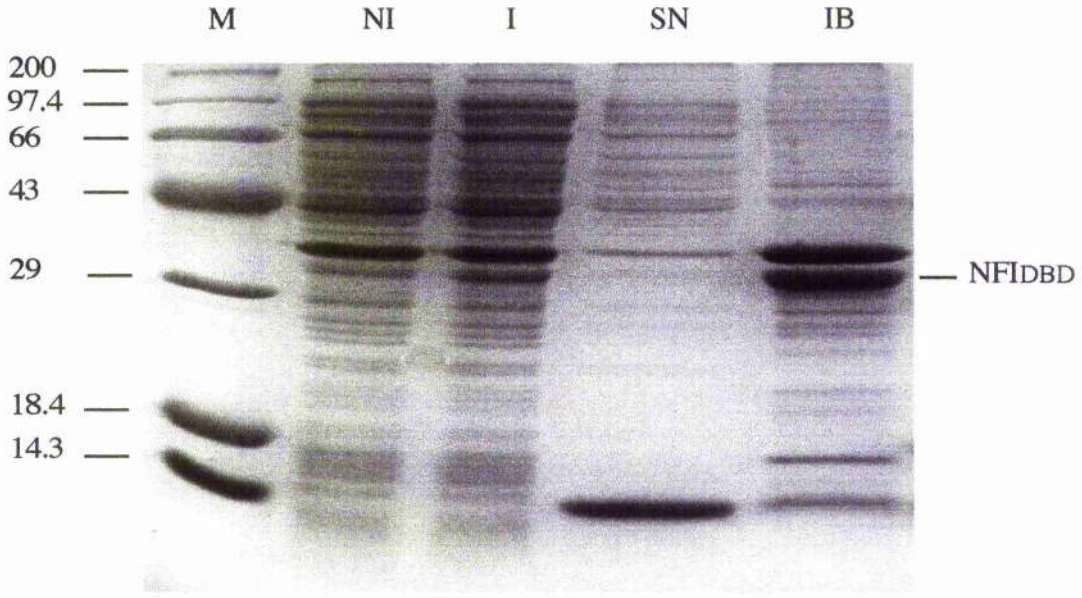


Figure 1.2 Induction and purification of NFIDBD expressed from the vector pET22b+.

Transformed BL21DE3 with pET22b+ NFIDBD was grown and induced for 4 hours at 37°C with 1 mM IPTG. Proteins in an aliquot of the non induced (NI) and of the induced (I) bacterial culture were separated on a 12 % SDS PAGE and the proteins stained with Coomassie blue. The bacterial culture was harvested and bacteria lysed in gentle condition. The bacterial extract was centrifuged to separate the soluble material (SN) and the insoluble material (IB) which was resuspended in PBS suplemented with 1% SDS. An aliquot of the soluble (SN) and insoluble material (IB) was also loaded in the gel. M are the protein markers and the molecular weight of each marker is indicated in kDa.

Almost all of NFIDBD was detected in the insoluble fraction. Furthermore the protein migrated with an apparent molecular weight of 30 kDa while NFIDBD alone migrated with an apparent molecular weight of 27 kDa, suggesting that the pelB export signal was not cleaved and that NFIDBD was not exported to the periplasmic region of the bacteria.

In other experiments, NFIDBD could not be detected in a periplasmic fraction (data not shown). Induction of protein expression at lower temperature (20°C) or with lower amounts of IPTG did not affect significantly the solubility of the protein.

I.2.2 Expression and purification of GST-NFIDBD

Protein expression of GST-NFIDBD was induced in *E. coli* XL1 Blue strain at 20°C with 0.4mM IPTG. Analysis by SDS PAGE, a lysate from IPTG induced bacteria showed the appearance of a polypeptide with an apparent molecular weight of 55 kDa (figure 1.3A).

In a similar fashion as described above, proteins in the soluble and insoluble fraction were separated on a SDS PAGE (figure 1.3A). Although most of the GST-NFIDBD was detected in the insoluble fraction, some protein was also present in the soluble fraction. Therefore the soluble GST-NFIDBD was purified by affinity chromatography on immobilised glutathione. Specifically bound protein was eluted by competing by free glutathione (10mM). An aliquot of each fraction was analysed by SDS PAGE (figure 1.3B).

The protein GST-NFIDBD could be detected in all the various fractions but the amounts of protein varied. Most of the fusion protein was unable to bind glutathione and was found in the flow through fraction (lane FT). DNA binding activity of NFIDBD could also be detected in this fraction. A very small fraction of the GST-NFIDBD bound to the column.

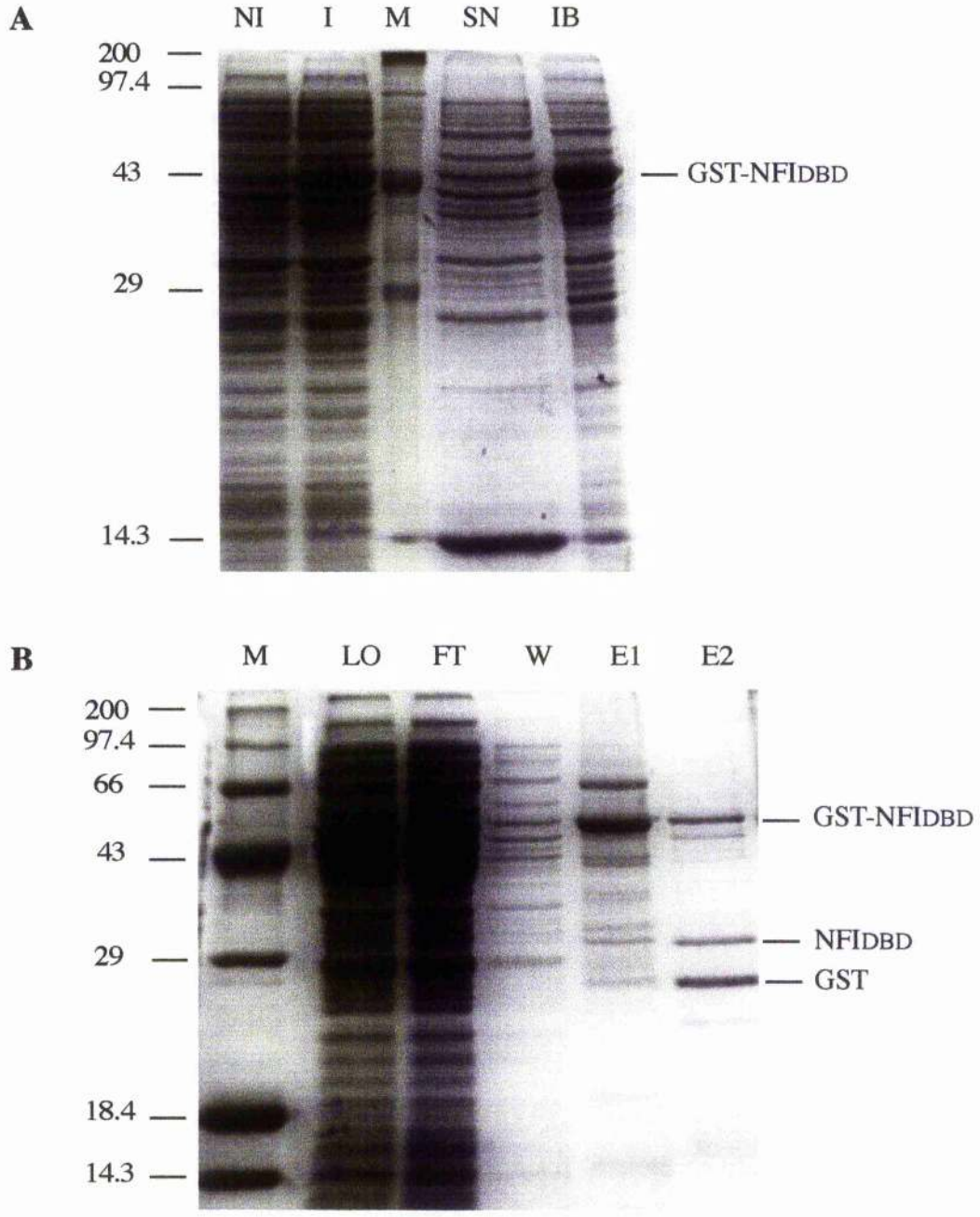


Figure 1.3 Expression and purification of GSTNFIDBD expressed in E.coli.

Expression of GST-NFIDBD was induced with 1 mM IPTG and incubated for 4 hours before harvesting the cells. An aliquot of the bacteria were lysed with 1%SDS in PBS from an extract before (NI) and after 4 hours induction (I) and proteins were separated on a 12% SDS PAGE which were stained using Coomassie blue. Bacteria were harvested and lysed. Insoluble (IB) and soluble (SN) material were separated by centrifugation (figure A).
The GST-NFIDBD in the soluble fraction was purified onto a glutathione agarose column (figure B, LO and FT), the column was washed (W) and the bound protein eluted by adding 10 mM glutathione in the buffer (E1). The eluted GST-NFIDBD was treated for 2 hours with thrombin (E2). M are the protein markers and the molecular weight of each marker is indicated in kDa.

The DNA binding activity of the fusion protein GST-NFIDBD in the elution fraction was measured by a gel retardation assay (figure 1.4). Various amounts of protein were incubated with a dsDNA containing a specific DNA binding site for NFI. Free DNA and GST-NFIDBD - DNA complex were separated in a native polyacrylamide gel. Between 2 μg and 200 ng of GST-NFIDBD, the DNA was fully saturated by GST-NFIDBD. At 20 ng most of the probe was migrating as the free DNA. However several bands could be detected but they did not correspond to any band found at higher concentrations, suggesting these were due to non specific binding. 10 ng of active NFIDBD is sufficient to saturate the probe. Therefore the amount of DNA binding activity of GST-NFIDBD was estimated between 1 and 5 % of the total protein.

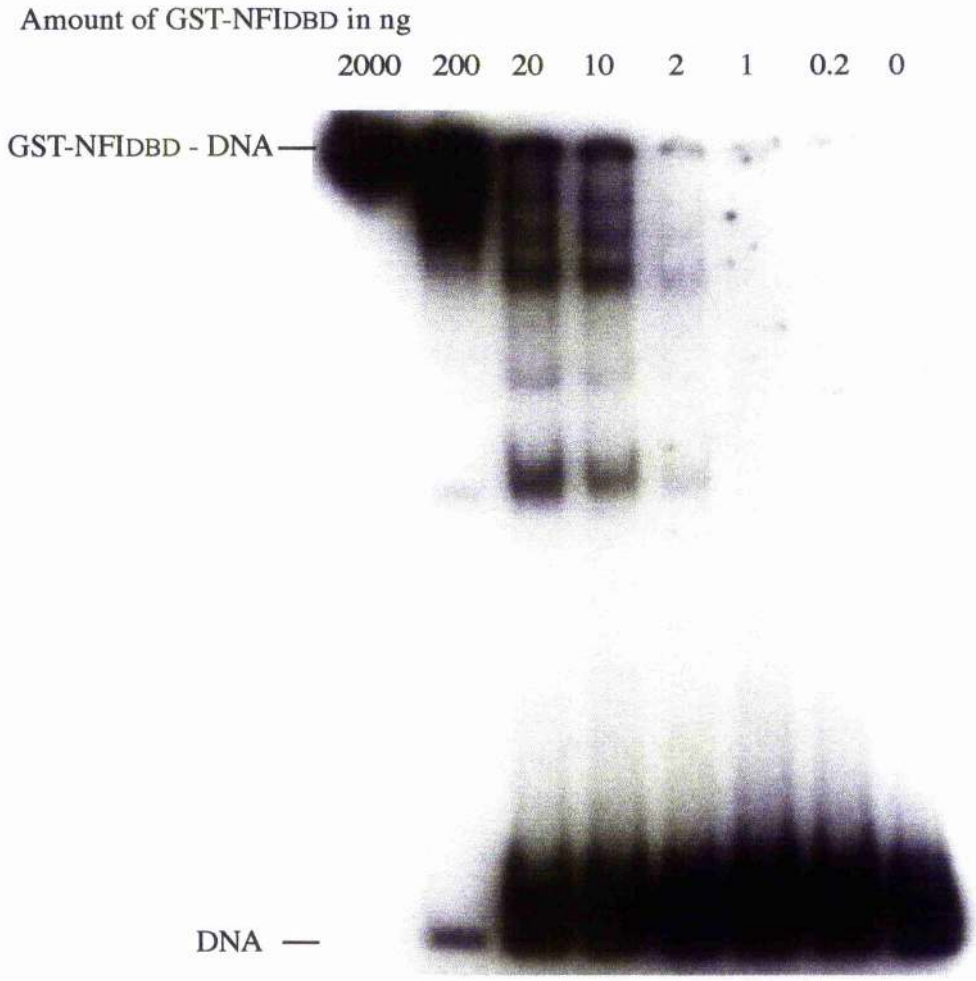


Figure 1.4 DNA binding activity of the GST-NFIDBD.

The DNA binding activity of the GST-NFIDBD bound to the glutathione agarose column, was measured by gel retardation assay. The protein concentration was estimated by Bradford and the equivalent of 2μg to 0.2ng of protein was used per gel retardation assay. DNA protein complex and freeDNA were separated in a 8% polyacrylamide gel with 0.5xTBE for 90 minutes at 150V. The gel was dried and exposed on a phosphoimager plate. The amount of protein used in each assay is indicated on top of the gel in nanograms.

II. Cloning, Expression and purification of NFIDBD in baculovirus infected cells.

Unlike bacterial expression, previous studies showed that expression of NFIDBD using recombinant baculovirus (Bosher *et al*, 1990) or vaccinia virus (Gounari *et al*, 1989) produced an active and soluble protein. As the DNA binding activity of NFIDBD did not seem to be affected by GST in the bacterial expression system (figure 1.4), we used therefore the baculovirus system to try to obtain soluble and active GST-NFIDBD.

II.1 Cloning of NFIDBD in pAcG2T and pVL1393.

NFIDBD was cloned into the transfer vectors pAcG2T and pVL1393 in order to obtain recombinant baculovirus (figure 1.5). Both vectors contained a bacterial origin of replication, an ampicillin resistant gene and a multiple cloning site to insert the required gene under the control of the polyhedrin promoter flanked by baculovirus sequences required for homologous recombination. In the vector pAcG2T, the cDNA of NFIDBD is inserted at the 3'end of the GST gene.

The coding region of NFIDBD was amplified by PCR using the pGEX2T NFIDBD cPk as template constructed in section "1.1". The same 5' end oligonucleotide was used but the 3' end primer recognised the Pk region and contained a BamHI restriction site and the sequence encoding a stop codon. Vectors and PCR were digested with BamHI and ligated into one another. Ligation mix was transformed into *E. coli* XL1 Blue strain. The presence of inserts in the plasmid was checked by digestion using the restriction enzyme BamHI and its orientation using EcoRI.

II.2 Production of recombinant baculovirus

The resulting plasmids pAcG2T NFIDBDcPk and pVL1393 NFIDBDcPk were co transfected into SF9 cells as described by Kitts *et al* (1990) with linearised viral DNA purified from the recombinant baculovirus AcRK6 (obtained from Prof. Hay).

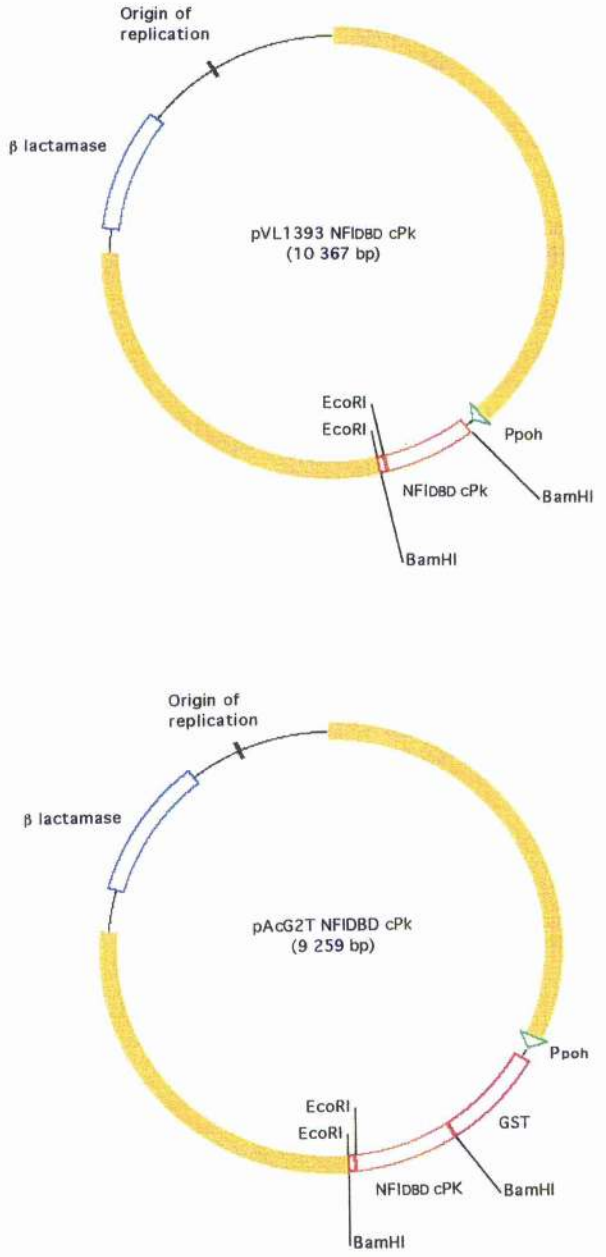


Figure 1.5 Schematic representation of the vectors pVL1393 and pAcG2T used for the production of recombinant baculovirus

The cDNA encoding for the 222 amino terminal of Nuclear Factor I was inserted in the restriction site BamHI of the vectors pVL1393 and pAcG2T. The restriction sites EcoRI are also represented which were used for checking the orientation of the insert in the vector. In yellow is represented the region of the baculovirus genome required for recombination. The green arrow represents the polyhedrin promoter involved in the expressionof the inserted gene. Red and orange boxes represent the protein expressed and the blue boxes represent the bacterial selection marker.

This recombinant virus contained the gene encoding for the β galactosidase under the control of a polyhedrin promoter. Furthermore a Bsu36I restriction site was engineered in the ORF 1629 which is an essential gene for viral infectivity. Therefore by subjecting the DNA purified from this recombinant virus to digestion using the endonuclease Bsu36I, the viral DNA was cleaved twice: once in the ORF1629 and once in the β galactosidase gene. By cotransfection, two main recombinant viruses were produced: either a AcRK6 type virus which would produce blue plaques in the presence of X-gal or the required recombinant baculovirus (white plaques). The virus was isolated from the supernatant of these cotransfections by plaque assay in the presence of X-gal. Clearly separated white plaques on a SF9 monolayer were selected and the recombinant virus was used for subsequent infections.

II.3 Purification of NFIDBD

The purification of NFIDBD from recombinant baculovirus infected cells has already been described by Bosher *et al* (1990). NFIDBD was purified from a Sf9 cells culture infected for 72 hours using a three steps strategy: in the first step, a nuclear extract of these Sf9 cells was prepared. In the second step, NFIDBD was purified by ion exchange chromatography and then, by affinity chromatography on DNA immobilised in the final step.

NFIDBD cPk was also purified in this fashion. An aliquot of each fraction was analysed by western blotting using the Pk antibody (figure 1.6). This revealed that NFIDBD bound efficiently to the ion exchange column and to the DNA affinity column. A very limited amount of protein was either found in the flow through fractions or in the wash fraction of each column. The purification resulted in a pure fraction of NFIDBD (figure 1.8C). The addition of these 21 residues from the IkBα at the N terminus and the Pk at the C terminus did not affect the DNA binding activity, suggesting that NFIDBD cPk was folded in a similar way than in the original NFIDBD recombinant baculovirus.

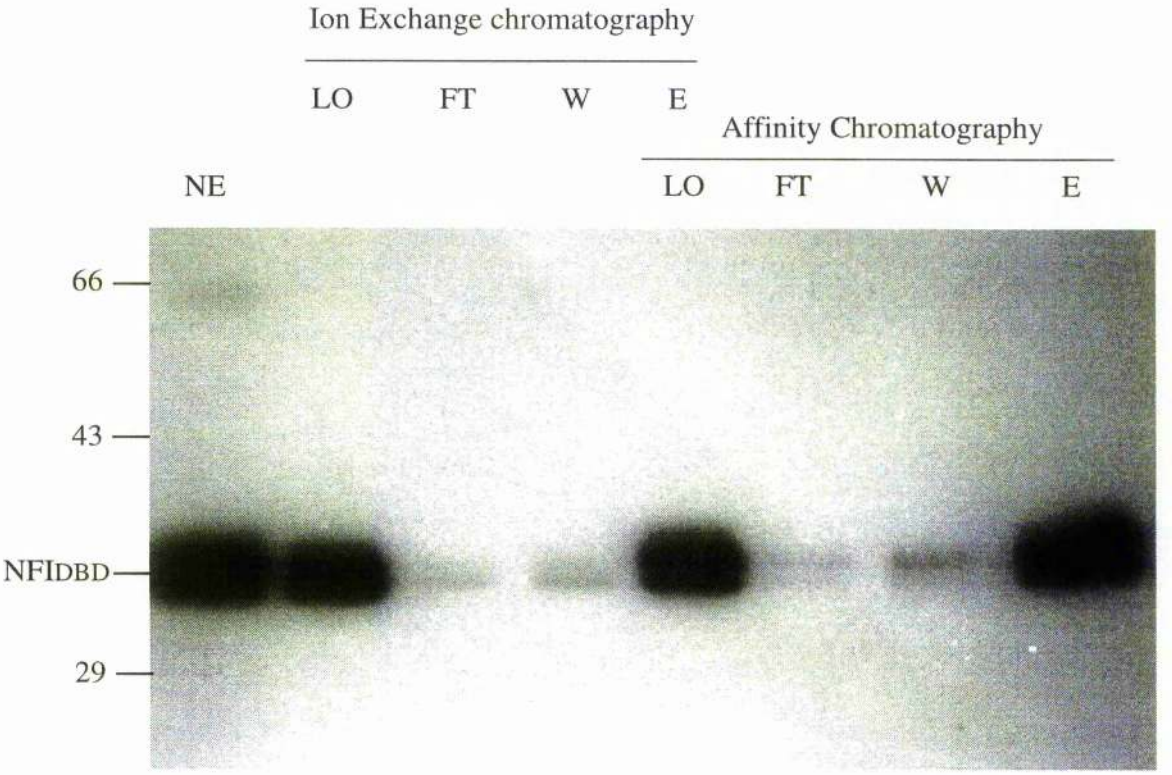


Figure 1.6 Purification of NFIDBD cPk from infected Sf9 cells infected with a recombinant baculovirus expressing NFIDBD cPk.

A nuclear extract (NE) was obtained from infected insect cells with a recombinant baculovirus expressing NFIDBD. NFIDBD was purified on an anion exchange column and by affinity chromatography using a DNA immobilised Sepharose column. Proteins of the loaded (LO), the flow through (FT), the wash (W) and the elution (E) fractions were separated on a 12% SDS PAGE and blotted onto a PVDF membrane. Proteins on the membrane were probed using the Pk antibody and bands were revealed using the enhanced chemiluminescence procedure.

II.4 Purification of the GST-NFIDBD

II.4.1 Purification by glutathione affinity chromatography

For large scale protein purification, Sf9 cells were infected for 48 to 60 hours with the recombinant virus expressing GST-NFIDBD. Each stock of virus was tested individually for optimal protein expression (see material and methods). An infected cell extract was clarified by centrifugation prior to purification of the fusion protein by affinity chromatography on glutathione agarose. The column was washed and specifically bound protein eluted by competition with free glutathione (10 mM) added to the buffer. An aliquot of each fraction was analysed on a 10% SDS PAGE stained with Coomassie Brilliant Blue (figure 1.7A). A 55 kDa polypeptide corresponding to GST-NFIDBD was not detected in the glutathione elution. However examination of the stained gel revealed two polypeptide species with apparent molecular weights of 26 kDa., suggesting that they could represent GST protein.

To identify full length GST-NFIDBD and follow its purification, western blotting was performed on each fraction using the Pk antibody which recognises the Tag at the C-terminus of the protein. The western blot revealed the existence of a polypeptide species specifically recognised by the antibody at an apparent molecular weight similar to GST-NFIDBD produced in bacteria (figure 1.7B), which did not appear to bind to the column, but was present in the flow through of the glutathione affinity column. A small amount of immunoreactive material was however detected in the fraction eluted with glutathione which was below the level of detection in the Coomassie stained gel.

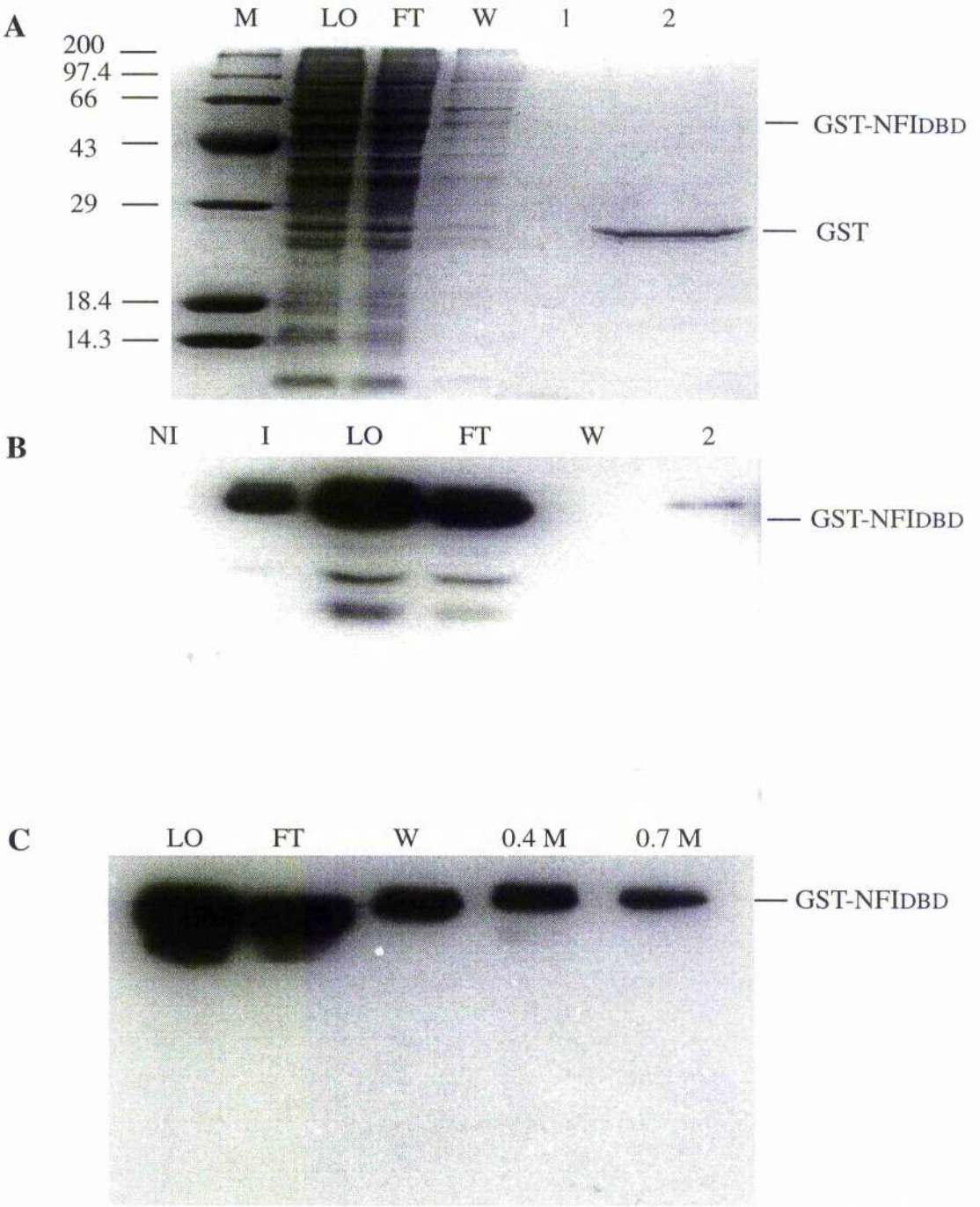


Figure 1.7 Purification of GST-NFIDBD expressed in infected Sf9 cells using a recombinant baculovirus.

The GST-NFIDBD fusion protein was purified by affinity chromatography on a glutathione immobilised column. The GST-NFIDBD content in every fraction during purification was followed by separating proteins on a 12% SDSPAGE and analysing the various fractions either by Coomassie blue staining (A) or by western blot using the Pk antibody (B). GST-NFIDBD was also loaded on an ion exchange column. The GST-NFIDBD content of the various fractions were analysed by western blot using the Pk antibody. LO is the load on to the column fraction, FT the flow through. NI and I are fractions of non induced and induced *E.coli* transformed with pGEX2T NFIDBD. W is the wash fraction. 1 and 2 are elution fractions of the glutathione immobilised column while 0.4 M and 0.7M represent the salt concentration used in the fraction to elute the ion exchange column. M are the protein markers and the molecular weight of each is indicated in kDa.

It seems that GST was active but most of the fusion protein was unable to bind glutathione. The same purification was performed in the presence of either additives (1% Triton X100, 8 mM CHAPS, 1 M NDSB 201) or in variable NaCl concentration (0.2, 0.4 and 1 M). None of these conditions enabled the purification of GST-NFIDBD by glutathione immobilised affinity chromatography. Thus GST-NFIDBD could not be purified by affinity chromatography on immobilised glutathione.

II.4.2 Purification by anion exchange chromatography

An alternative strategy employing anion exchange chromatography was used. The infected cell extract was loaded onto an ion exchange column at 0.2 M NaCl. The resin was washed and bound proteins were eluted by raising the NaCl concentration first to 0.4 M NaCl followed by 0.7 M. An aliquot of each fraction was analysed by Western blotting using the Pk tag monoclonal antibody (Figure 1.7C). The vast majority of the fusion protein did not bind to the column and was present in both the flow through and the wash fraction. A small fraction of protein eluted at 0.4 M NaCl and also some in 0.7 M NaCl.

II.4.3 Purification of NFIDBD after treatment of thrombin

NFIDBD could be purified only after treatment of the fusion protein with thrombin. The GST-NFIDBD was dialysed to remove any protease inhibitors and to lower the salt concentration to 0.2M. The GST-NFIDBD was treated with thrombin for 3 hours. NFIDBD could then be purified as described earlier on in section 2.3 (figure 1.8). A clear band eluted from the anion exchange column. On the affinity column, about half of the protein did not bind to the column and was detected either in the flow through fraction or the wash fraction. Although other proteins were also present, NFIDBD represented a large proportion. NFIDBD migrated as a single band with an apparent molecular weight of 30 kDa while NFIDBD purified directly from the insect cells migrated as a doublet with an apparent molecular weight between 30 to 32 kDa.

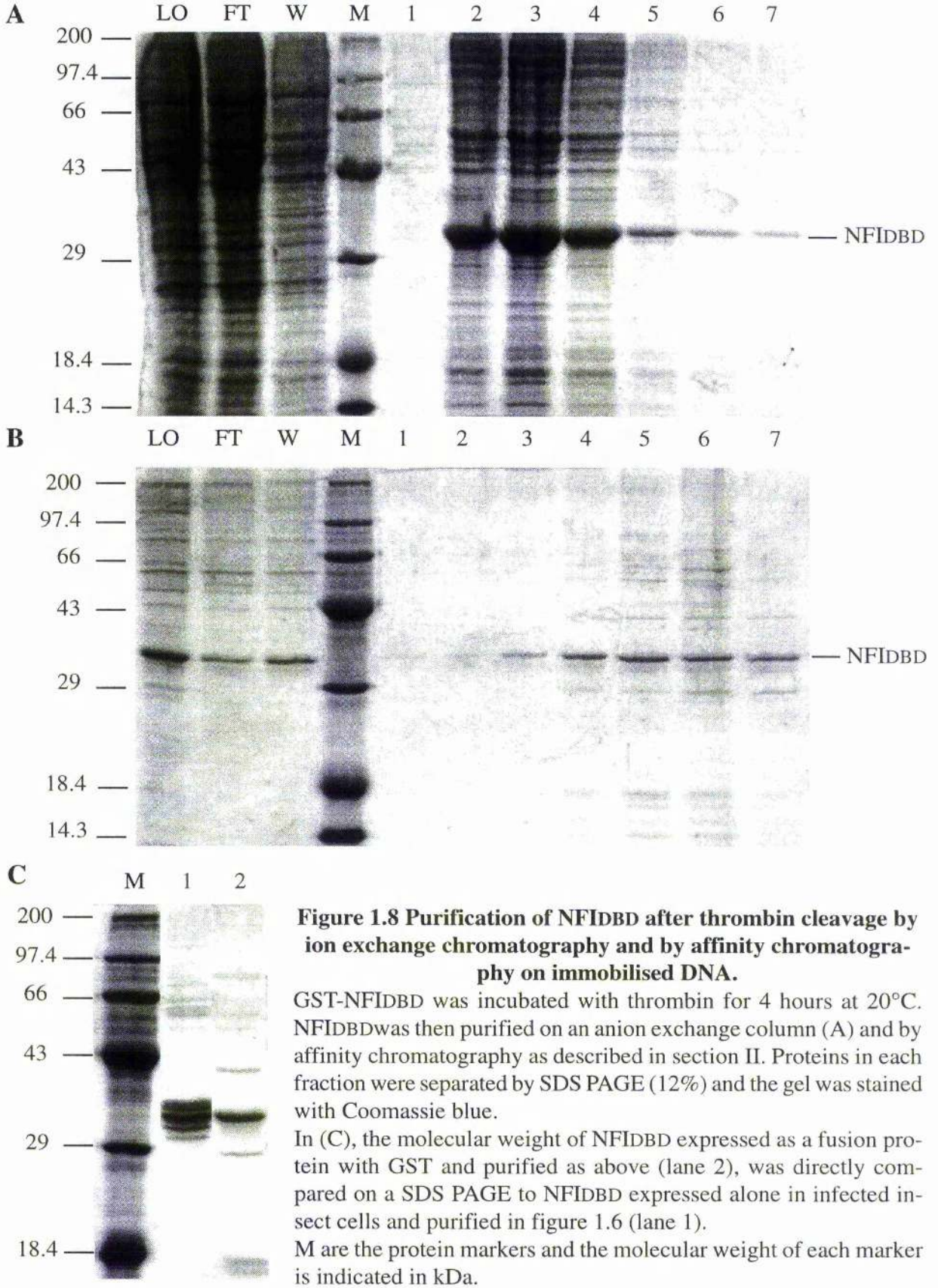


Figure 1.8 Purification of NFIDBD after thrombin cleavage by ion exchange chromatography and by affinity chromatography on immobilised DNA.

GST-NFIDBD was incubated with thrombin for 4 hours at 20°C. NFIDBDwas then purified on an anion exchange column (A) and by affinity chromatography as described in section II. Proteins in each fraction were separated by SDS PAGE (12%) and the gel was stained with Coomassie blue.

In (C), the molecular weight of NFIDBD expressed as a fusion protein with GST and purified as above (lane 2), was directly compared on a SDS PAGE to NFIDBD expressed alone in infected insect cells and purified in figure 1.6 (lane 1).

M are the protein markers and the molecular weight of each marker is indicated in kDa.

III. Summary and conclusion

In an attempt to obtain active protein, we tried two expression systems. In bacteria the protein was either insoluble or the amount of active NFIDBD was very low. Using baculovirus we could purify active NFIDBD. However the NFIDBD produced as a fusion protein was difficult to purify and involved the removing of the GST. If produced alone, NFIDBD produced a variety of polypeptides with different molecular weights (mainly a doublet). The apparent molecular weight was slightly higher than expected. The reason for heterogeneity in the protein preparation is still unknown. No post translational modification has yet been found for the DNA binding region of NFI. It was suspected that the doublet was due to proteolysis. However this method remains the only available and convenient source of pure and active NFIDBD.

Chapter 2 : The folding *in vitro* of Nuclear Factor I DNA Binding Domain expressed in bacteria

Expression of proteins in bacteria is inexpensive and convenient compared to using the baculovirus expression system. As previously shown, NFIDBD produced in bacteria produced homogenous protein which could be easily purified. Therefore, we investigated if this protein could be folded into an active protein. Earlier research did show that CTF 1 could be renatured from bacterial inclusion bodies by dialysis however this work only described the production of very limited amount of protein (at most micrograms, Amemiya *et al*, 1994). Similarly NFIDBD was renatured by dialysis but this was carried out using a fusion protein with GST (Coenjaert *et al*, 1994). The work presented in this chapter is an investigation of the expression and the optimal condition for renaturation of NFIDBD.

I. Construction of a vector for the expression of NFIDBD.

The pET LS NFIDBD expression vector described in chapter 1, expresses a fusion protein of NFIDBD fused to a periplasmic localisation signal at its N terminus. To obtain a vector expressing only NFIDBD, the periplasmic coding sequence was digested out of the vector by using the restriction enzymes Nco I and Nde I (figure 2.1). The cohesive ends were filled in using the DNA polymerase (Klenow fragment). This construct was religated with T4 ligase which enabled us to express NFIDBD proteins (pET NFIDBD).

The pET 22b+ vector also contains a sequence coding for a six histidines residues at the C terminus of a protein. Adding these residues could allow protein purification under native and denaturing condition, in addition to providing an epitope tag. There

were two major problems in the construction of a vector expressing NFIDBD fused to the C terminus of 6 histidines (pET NFIDBD Tag) : first the nucleotides GA in the Eco RI restriction site was part of the stop codon of the NFIDBD gene and had to be removed, secondly the sequence coding for the 6 histidines had to be fused to NFIDBD in the same reading frame.

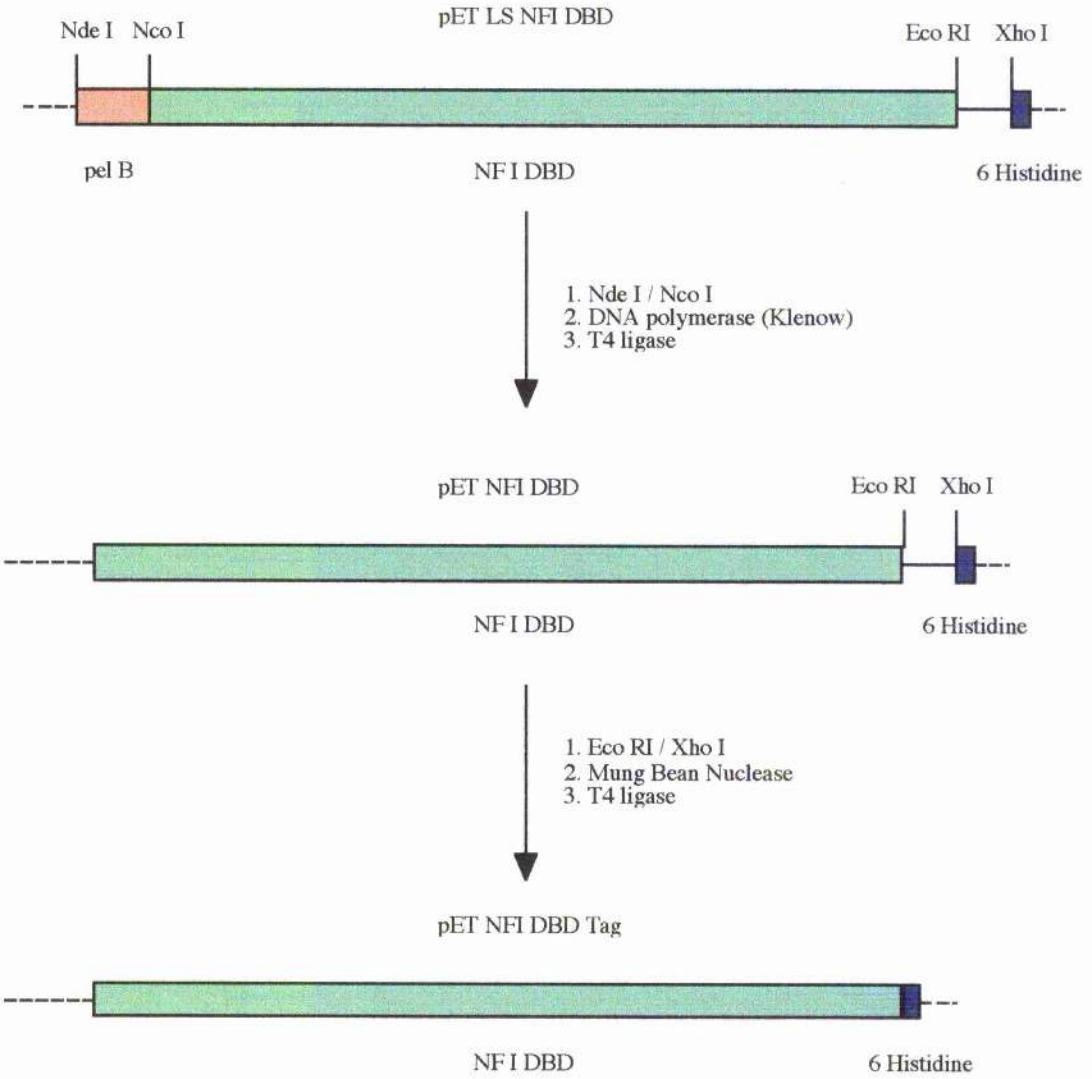


Figure 2.1 Construction of expression vector pET NFI DBD and pET NFI DBD Tag.

The coding sequences are represented by color boxes, the localisation signal (pelB) in pink, the NFI DBD in green and the 6 histidines in blue. The first cloning step was to construct pET NFI DBD to express NFI DBD. The second cloning step was to construct pET NFI DBD cTag which expresses NFI DBD fused to 6 histidines residues at the C terminus of the protein.

Hence the easiest way to obtain the new construct was to digest the pET NFIDBD using Eco RI and Xho I restriction enzymes and the cohesive ends digested using mung bean nuclease to produce a blunt ended vector. This procedure changed the stop codon (TGA) into a TGG sequence coding for a tryptophan while putting in frame the six histidines sequence. The vector was religated with T4 ligase.

To identify positive clones expressing NFIDBD - 6 histidines, competent BL 21 (DE3) were transformed. Bacteria from individual colonies were screened by western blotting using a monoclonal antibody recognising the 6 histidine sequence at the C terminus of the protein (Invitrogen).

II. Expression and purification of NFIDBD (and NFIDBD Tag)

NFIDBD was expressed in *E. coli* BL21 (DE3) strain. The expression was induced by 1mM IPTG as described in chapter 1. Growth of bacteria and induction of protein production were performed at 37°C. When total cellular proteins were analysed by Coomassie blue staining of SDS PAGE, an induced species with an estimated molecular mass of 26 kDa was detected. As previously shown, the protein was found in the insoluble fraction. To purify the inclusion bodies, bacteria were treated with lysozyme and their membrane disrupted by Triton X-100 and sonication. DNA was digested by the addition of DNase and $MgCl_2$. Soluble material and inclusion bodies were separated by centrifugation and the pellet washed with PBS supplemented with 1% Triton X-100 (v/v). To solubilize the inclusion bodies, various concentrations of urea and guanidine were tried (2, 4, 6 and 8M). Soluble and insoluble fractions were separated by centrifugation at 50 000 g.

NFIDBD remained insoluble in urea even when using a 8 M concentration (figure 2.2). Instead in 6 M guanidine, NFIDBD was soluble, while most contaminant proteins remained insoluble. At this stage, NFIDBD was thus the major protein in this fraction.

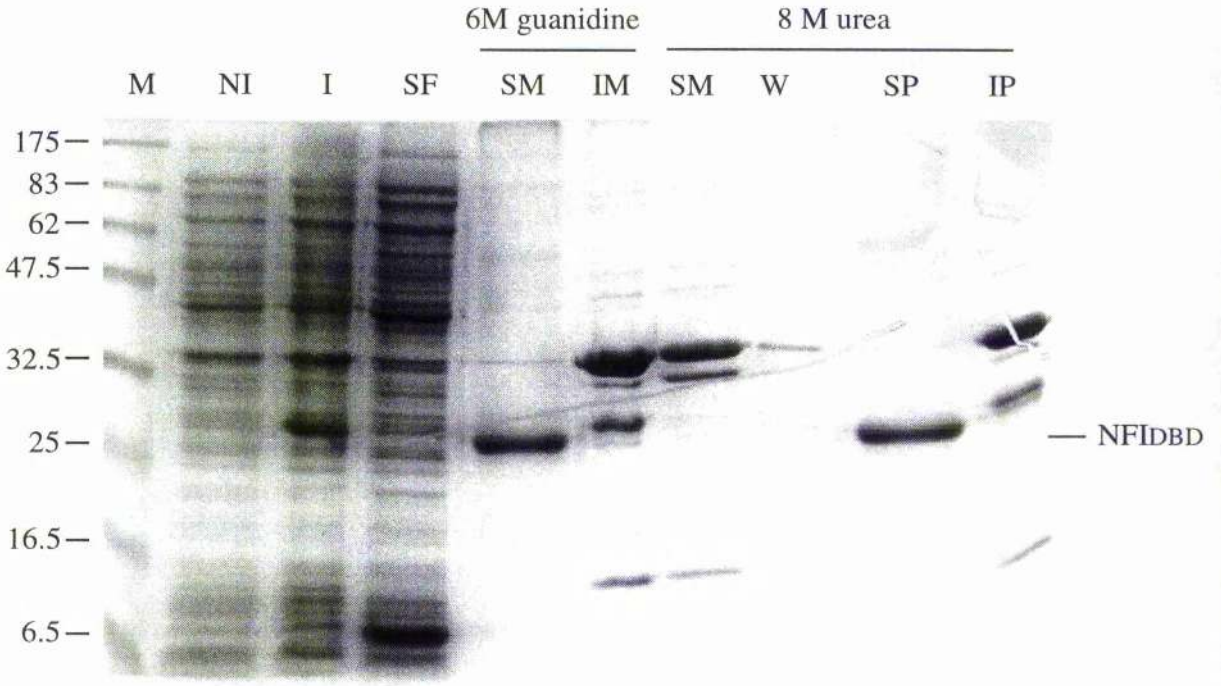


Figure 2.2 Expression and purification of Nuclear Factor I from BL21 (DE3).

NFIDBD was expressed in E.coli transformed with pET22b NFIDBD. NFIDBD was expressed by adding 1 mM IPTG to the bacterial culture. Total protein content in an aliquot of the bacterial culture before (NI) and after 3 hours addition of IPTG (I), was analysed on a 12%SDS PAGE. The bacterial culture was harvested by centrifugation and resuspended in extraction buffer. Soluble(SF) and insoluble proteins were separated by centrifugation. As NFIDBD was in the insoluble fraction, solubility in the presence of 6M guanidine or 8M urea was tested. The inclusion bodies were resupended in the denaturant for 2 hours, and insoluble (IM) and soluble material (SM) were separated by centrifugation. Each fraction was analysed by SDS PAGE. Proteins were stained with Coomassie blue. As NFI was not soluble in urea, the pellet was washed (W) with water and the pellet solubilised with 6M guanidine. Insoluble proteins were separated from the soluble proteins (SP) by centrifugation. The proteins in this pellet were resuspended by PBS supplemented with 1% SDS (IP). Lane M are the protein markers and the molecular weight of each marker is indicated in kDa.

III. Renaturation assay and interpretation of results.

This section describes the renaturation assay used in this study and the method to quantify yield of folded NFIDBD.

Figure 2.3 shows the main steps of the renaturation procedure. The renaturation solution was added to the bottom of an eppendorf tube. A small volume of NFIDBD solution in denaturing buffer was laid carefully on the side of a tube. It was then mixed to the renaturation solution by vortexing vigorously for 10 seconds. The folding mix was then left overnight at 4 °C without stirring.

A gel retention assay was performed to determine the proportion of functional NFI in each renaturation test. NFIDBD was incubated with a radioactively labelled oligonucleotide which contained a specific binding site for NFI. (To quench any non specific DNA binding which may occur with improperly folded NFI, an excess of non specific oligonucleotide was present).

NFI-DNA complex was separated from free DNA in a native polyacrylamide gel. Gels were analysed quantitatively using a phosphoimager. The more slowly migrating species represented the DNA-protein complex while the free DNA migrated close to the dye front. The yield of correctly folded, and hence active, NFI in each renaturation assay was determined by estimating the radioactivity present in the DNA - protein complex compared to the total radioactivity in each assay.

Our preliminary results showed that the folding yield was critically dependent on experimental parameters. We therefore decided to investigate these parameters in order to optimise the folding yield.

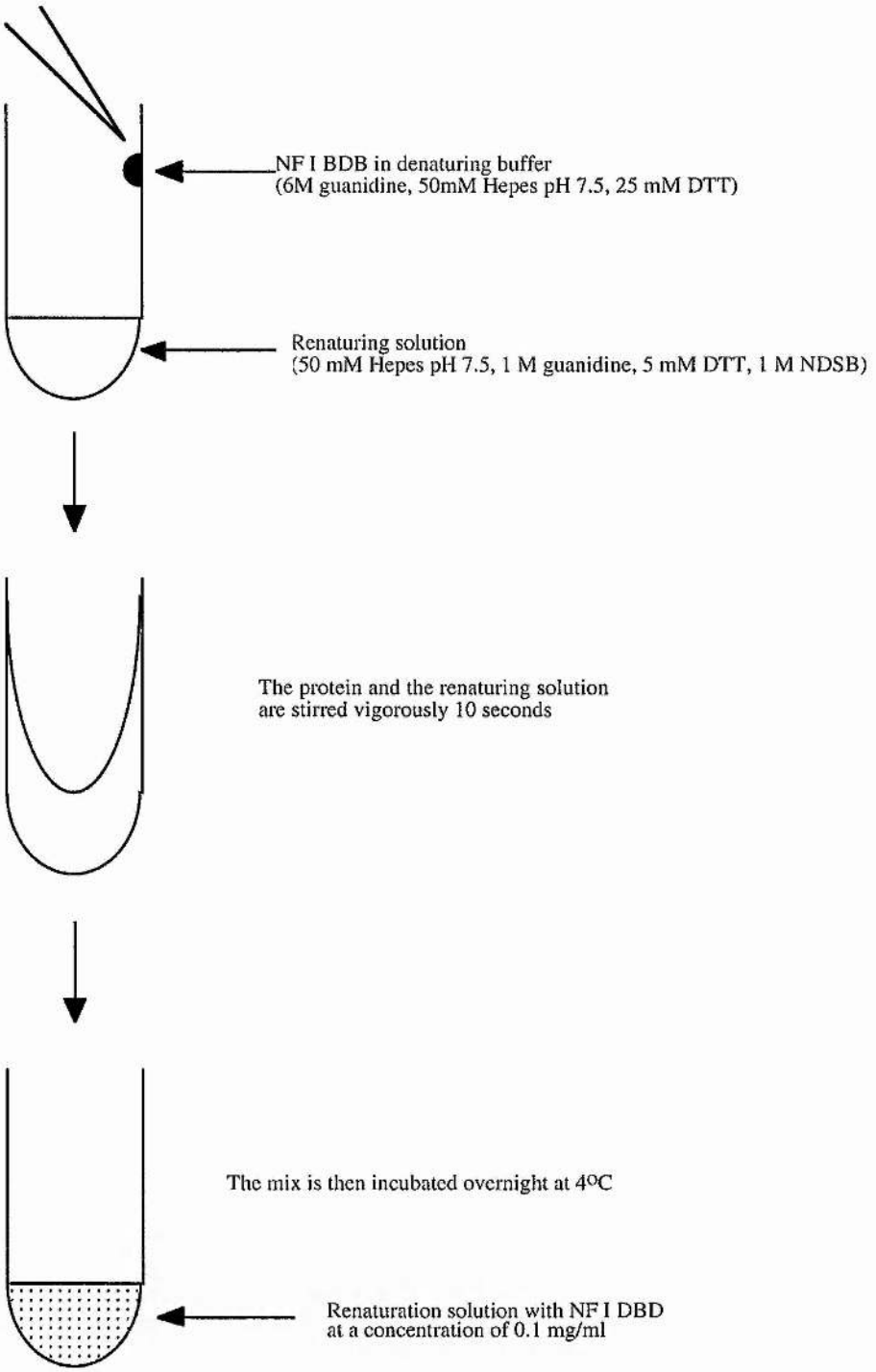


Figure 2.3 Schematic representation of the renaturation assay.

The folding of NFI DBD was performed by rapid dilution. In step 1, both solutions were added to an eppendorf tube but not mixed. In step 2, the protein was mixed to the renaturation buffer by vigorous vortexing. The protein was left to incubate overnight in in the renaturation buffer (step 3).

IV. Renaturation of NFIDBD using non detergent sulphobetaine

A new family of chemicals was found to increase folding yield of protein (Vuillard *et al*, 1998, Goldberg *et al*, 1996). These chemicals known as non detergent sulphobetaines (NDSB), were made by reacting propane sultone to an amine which produced different types of NDSB depending on the amine used. However all NDSB share similar properties which are : i) water soluble (up to 60 %), ii) improve yield in protein renaturation and purification by increasing protein solubility and by reducing aggregation, iii) zwitterionic molecule that could be used at pH between 3 and 12, and iv) compared to detergent, they are non denaturing and easily removed by dialysis.

The chemical structure of the various NDSB family members (figure 2.4) shows two distinct regions :

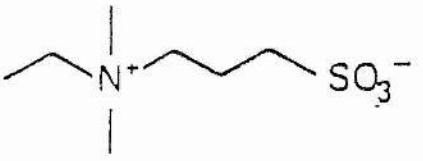
A hydrophilic region similar in all NDSB containing a sulphonate group separated by a three carbon arm from a quaternary ammonium. In solution, it is supposed that the negatively charged oxygen from the sulphonic group might interact with the positive ammonium thus forming a six membered ring. The other region is hydrophobic and is variable between NDSB. A whole range of NDSB exists with a small hydrophobic region such as in NDSB 195 to a larger one such as seen in NDSB 256.

The effect of five different NDSB were tested in NFIDBD renaturation : with NFIDBD, NDSB 195 decreased the amount of renatured NFIDBD by 50% while NDSB 201 and 223 did not improve the folding yield. Both forms of NDSB 256 increase renaturation up to 300% compared to the control (Figure 2.5).

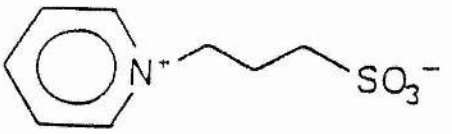
With NFIDBD Tag, the effect was different. NDSB 195 had no effect, while 201 and 223 increase renaturation 200% and both forms of NDSB 256 by 400%. However the yield in the control experiment was 2 times lower using NFIDBD Tag compared to NFIDBD.

NDSB 256 increased significantly the folding yield of NFIDBD and NFIDBD Tag. In further experiments, NDSB 256 tertio butyl was chosen to do the renaturation assay.

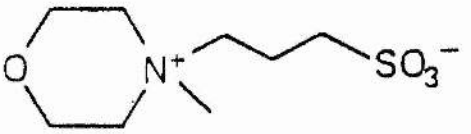
NDSB 195



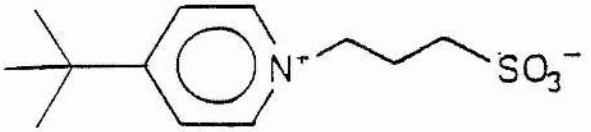
NDSB 201



NDSB 223



NDSB 256 tertio butyl (NDSB 256t)



NDSB 256 phenyl (NDSB 256p)

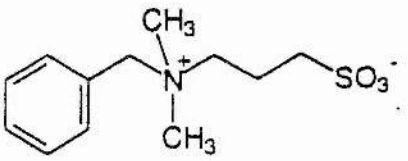


Figure 2.4 Schematic representation of the chemical structure of the various Non Detergent Sulphobetaine used in this study.

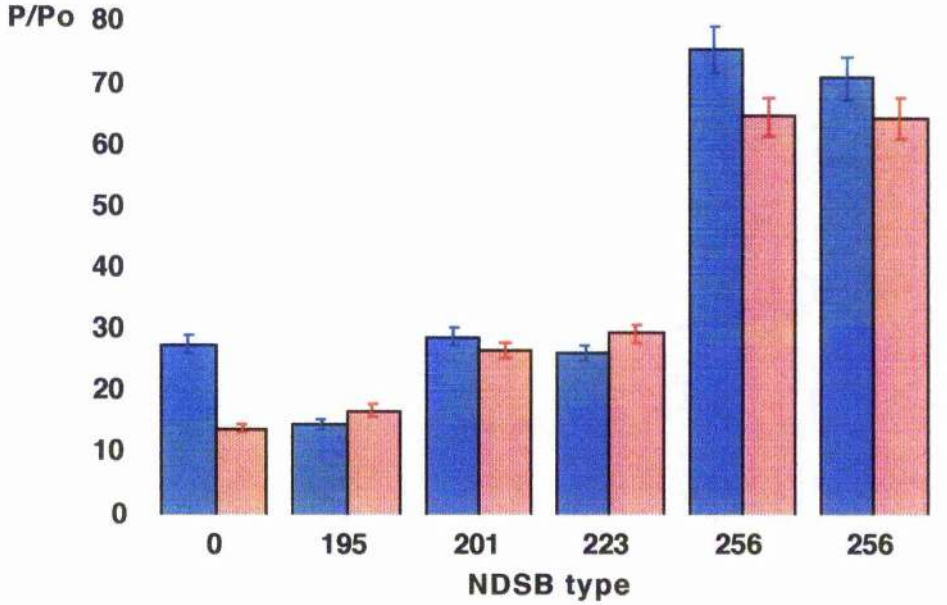


Figure 2.5 **Effect of various Non Detergent Sulphobetaines on the folding yield of NFI DBD and NFI DBD Tag.**
The folding yields were determined by gel retention assay of the renaturation assay in the presence of different NDSB (195, 201, 223, 256t, 256p) at a concentration of 1M as described in the text. Blue columns represent the renaturation of NFI DBD while red columns represent NFI DBD Tag. The yield was an average of three independent renaturation assays. P/Po is the fraction of bound DNA to the total amount of DNA added in each assay

V. Determination of the optimal condition for NFIDBD renaturation.

Various parameters in the renaturation assay were tested in order to determine which condition allowed an optimal folding yield. Substitution of NDSB by related chemical was also investigated.

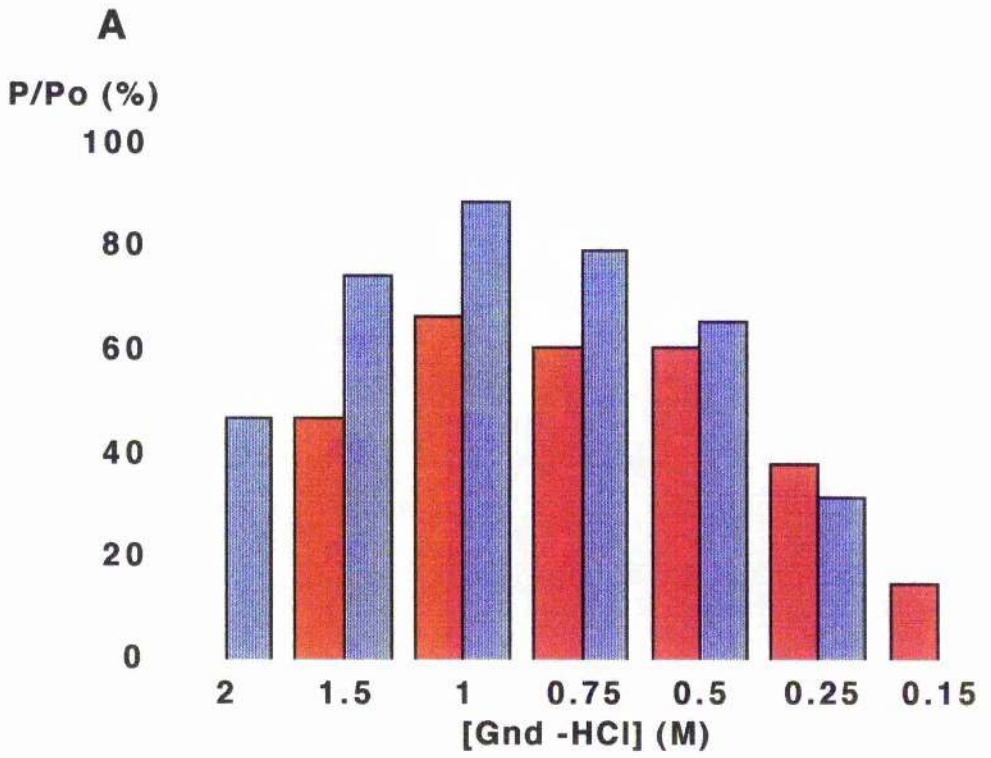
V.1 Ionic strength

Ionic strength is critical for NFIDBD stability. Purified NFIDBD precipitates at an ionic strength of 0.1 M. To determine the effect of guanidine on refolding *in vitro*, NFIDBD was renatured in concentration ranging from 2 to 0.25 M guanidine.

At 1 M guanidine, the renaturation of NFIDBD was optimal but remained high at concentration of 0.75 M guanidine (figure 2.6). When folding was performed in 0.25 M guanidine, there was a sharp decrease in yield (up to 3 times) compared to 1 M. As expected, increasing the guanidine concentration was also detrimental for the folding of NFIDBD. The guanidine affects both the ionic strength but is also a denaturant and probably affects the structure of NFIDBD at higher concentration.

An experiment was also performed where the guanidine was replaced by NaCl. Similar results were obtained, however the folding reached a plateau (instead of a peak) at concentration of NaCl between 0.75 and 2 M.

Although no difference in refolding yield was observed using NaCl instead of guanidine, one has to bear in mind that guanidine is a chaotropic salt and affects the solution.



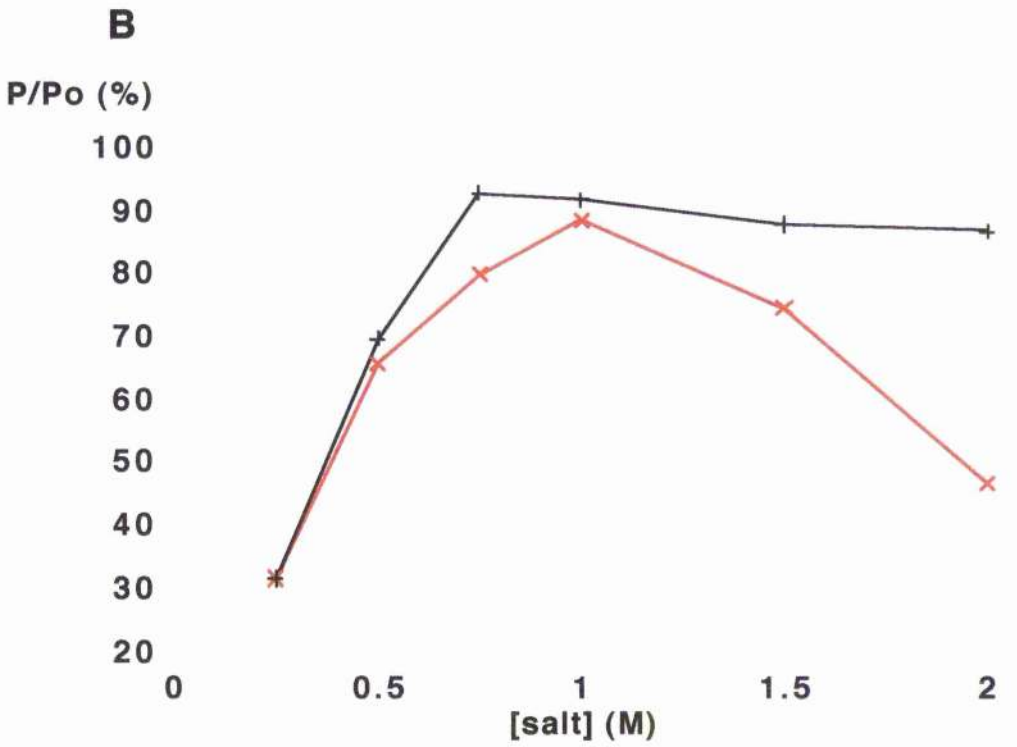


Figure 2.6 Effect of the guanidine concentration on the folding of NFI DBD and NFI DBD Tag (A) and comparison of NaCl and guanidine concentration on folding of NFI DBD cTag (B).

(A) The renaturation was done in concentration of guanidine varying between 2M and 0.15M. Blue column represent the renaturation of NFIDBD while red column represents NFIDBD Tag (previous page). This experiment was repeated twice.
(B) The folding yield of NFIDBD Tag was compared using either NaCl or guanidine in the renaturation buffer used at concentration between 2 and 0.25 M. The black line represents the renaturation in the presence of NaCl and in red, in the presence of guanidine. This experiment was done only once.

V.2 Effect of the protein concentration in the renaturation assay

It must be pointed out that protein concentration was determined using the "Bradford's method" with BSA as standard. This is valid for comparison purpose but may not reflect on the real protein concentration.

The final protein concentration was varied between 0.5 and 0.01 mg/ml in the renaturation assay. Both NFIDBD and NFIDBD Tag gave similar results (figure 2.7). The yield of renatured protein peaked at a concentration of 0.1 mg/ml. At a concentration of 0.25 mg/ml, the folding yield decreased to a half and by 4 fold at a concentration of 0.5 mg/ml.

In more diluted assay, the renaturation also decreased although not as dramatically. However if the protein was very dilute (0.01 mg/ml), virtually no refolded protein was detected.

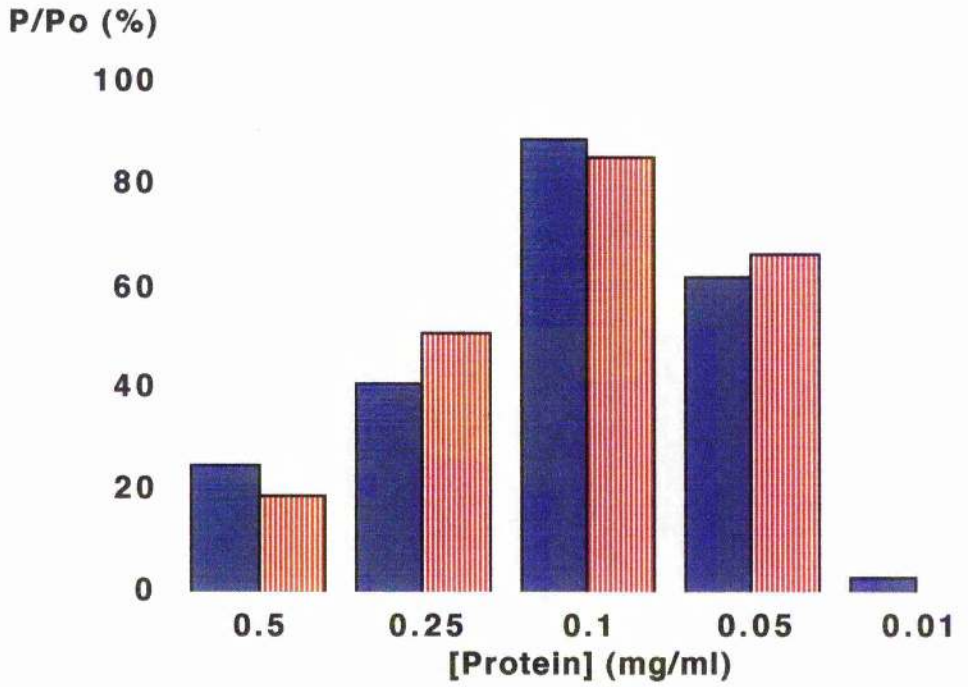


Figure 2.7 Effect of the protein concentration on the folding yields of NFI DBD and NFI DBD Tag.

NFI DBD and NFI DBD Tag were folded at a final protein concentration ranging from 0.5 to 0.01 mg/ml in the renaturation solution. The folding yield was determined by using 100 ng of protein per gel retention assay. Blue column represents the folding yield of NFI DBD and in red the folding yield of NFI DBD Tag. The results are from one experiment.

V.3 Effect of pH in NFIDBD folding

The pH mainly affects the charges of acidic and basic residues in the protein. Renaturation assays were performed in pH ranging from 4.7 to 10.4. Our choice of buffer used for this experiment dictated the pH studied. The pH was set at the pKa of the buffer to ensure maximum buffering capacity and that the amount of acid or base added did not vary between buffers (and the ionic strength). There was however one exception : at pH 7, Bis Tris propane was used (pKa2 @ pH 6.8). Maximum

renaturation was observed at pH 7 (figure 2.8), but remained high at pH 7.5, 8.1 and 9. There was a four fold decrease in renaturation yield if the pH was lowered from 7 to 6.1 and even stronger between pH 9 and 10.4.

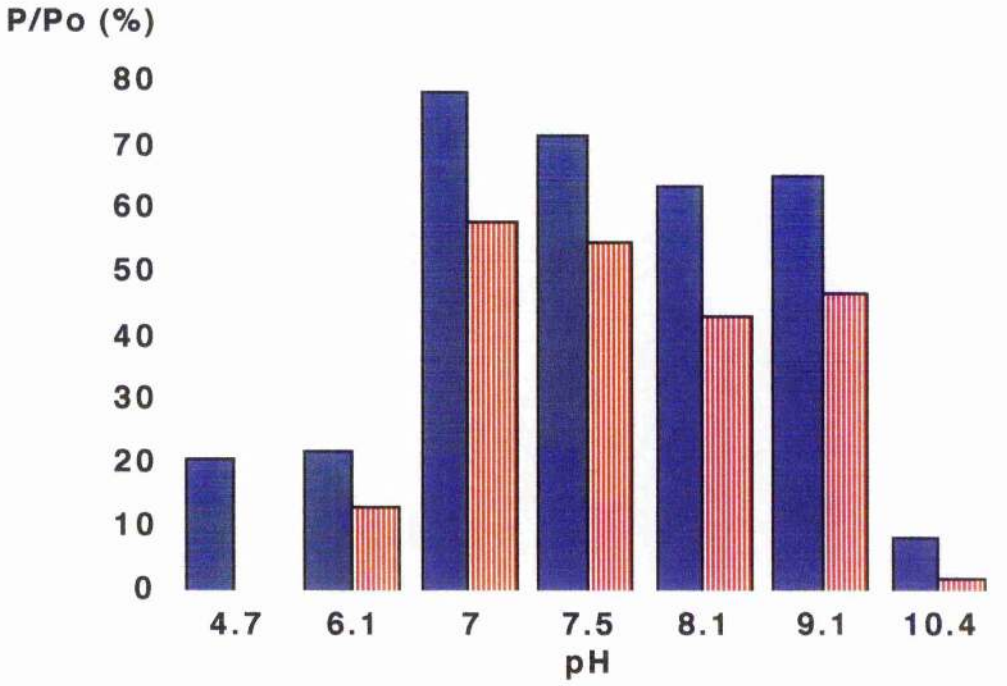


Figure 2.8 Effect of pH on NFI DBD and NFI DBD cTag folding

The renaturation were performed at the stated pH on the graph as described in the text. Each renaturation assay were performed using the following buffer at the concentration of 50 mM :
pH 4.7 with acetic acid/NaOH
pH 6.1 with MES/NaOH
pH 7 with Bis Tris Propane/HCl
pH 7.5 with Hepes/NaOH
pH 8.1 with Tris/HCl
pH 9.1 with Bis Tris Propane/HCl
pH 10.4 with CAPS/NaOH
The gel retention assay was performed in 50 mM Hepes pH 7.5 as described in material and methods. Blue column represent the folding of NFI DBD and in red, the folding of NFI DBD Tag. The results are from one experiment.

To ensure that an effect of pH was not masked by the high ionic strength in the renaturation buffer, the same experiment was performed by substituting 1M guanidine by 0.25 M NaCl. The folding yield in 0.25M NaCl was three times lower than in 1M guanidine. Except for pH 7.5, the same pH dependence (figure 2.9) was observed.

However at each pH, the refolding yield was still 3 times lower for each pH tested compared to the refolding done in the presence of 1 M guanidine. This is unfortunate as it means that the protein solution would need to be diluted significantly before the next step in the purification of NFIDBD.

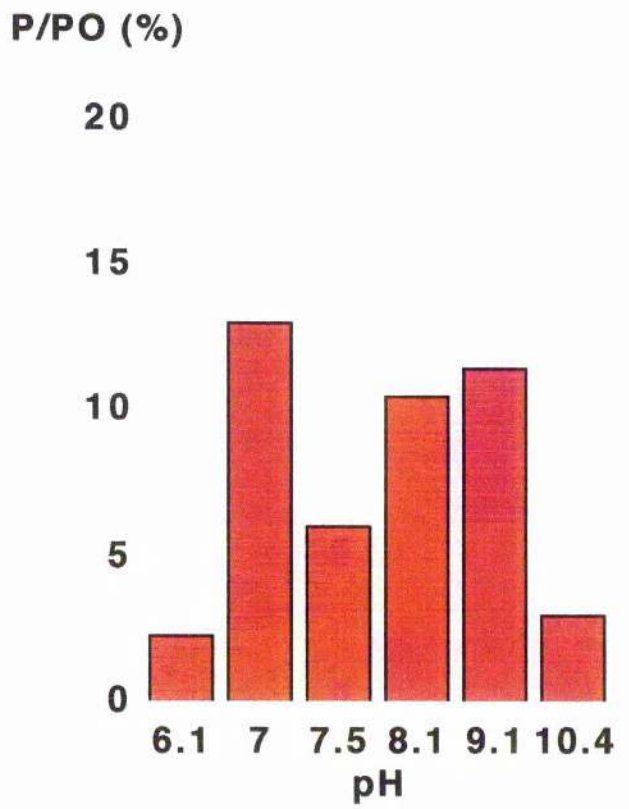


Figure 2.9 Effect of pH on NFI DBD folding at 0.25M NaCl.

NFI DBD folding was performed in the presence of 0.25M NaCl instead of 1M guanidine. Renaturation was performed as described in figure 2.8. Folding yields were determined by gel retention assay as described earlier. Each assay was performed in only once.

V.4 Influence of Temperature

Two slightly different approaches were chosen to study the influence of temperature on the renaturation. In both experiments the renaturation buffer was incubated at the stated temperature. Following mixing, in the first experiment, the protein was immediately incubated at 4°C. In the second experiment the renaturation assay was incubated for a further 1 hour at the stated temperature.

The folding of the protein by quick dilution can schematically be divided in two stages. Initially the protein is unfolded as a strand and collapses forming a molten globule by mixing NFIDBD in the renaturation solution. In a second stage, the protein rearranges itself to form active protein. the first event occurs in milliseconds while the next step can take up a few minutes to several hours. The first experiment is to study the influence of temperature on the collapse. In the second experiment the influence of temperature is tested in the rearrangement process.

The temperature studied were -10, 0, 10, 20 and 37°C. Temperature did not influence significantly the renaturation yield of NFIDBD (figure 2.10). Folding yield remained the same thought out the temperatures tested except when the renaturation assay was further incubated at 37°C, which resulted in a three fold decrease in yield (figure 2.10).

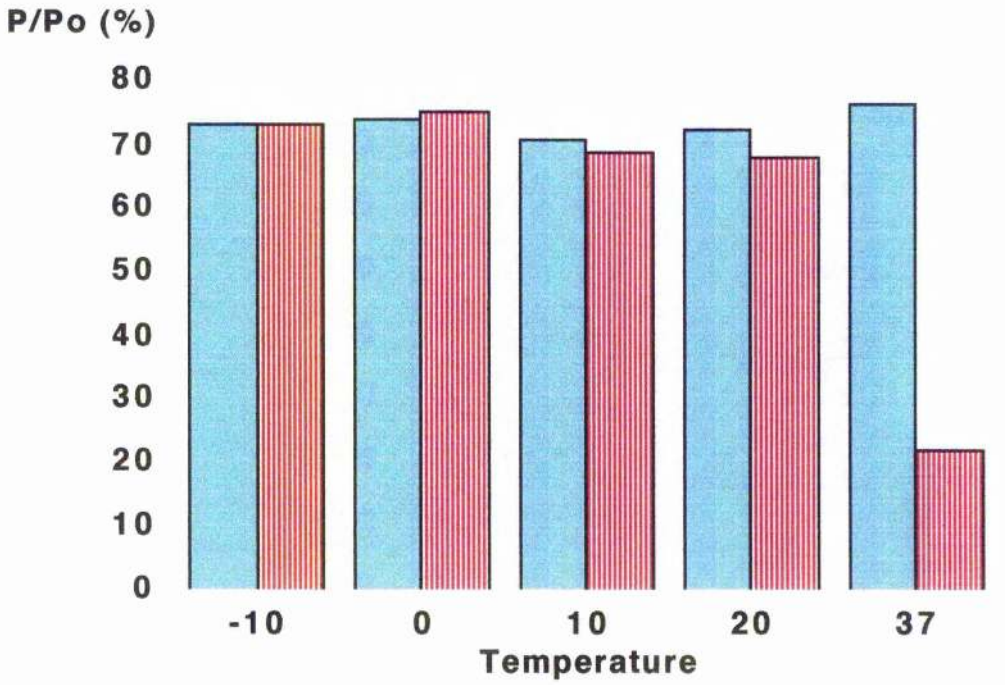


Figure 2.10 Influence of temperature on NF I DBD Folding

Renaturation was performed at the stated temperature under standard conditions. The renaturation solution was either left overnight at 4°C (green columns) or incubated for a further hour at the stated temperature (red columns).The results are from only one assay.

V.5 Effect of reducing agent

NFIDBD is very sensitive to red-ox conditions and requires the presence of a reducing agent in order to be active. The effect of the type of reducing agent and its optimal concentration in the folding procedure were investigated.

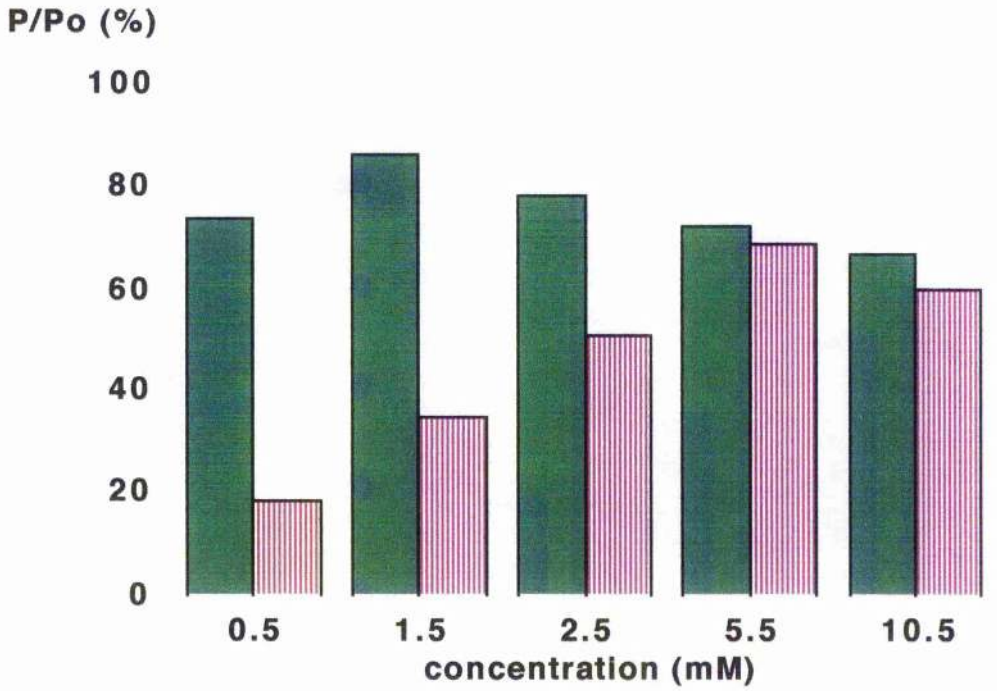


Figure 2.11 Effect of reducing agent in the folding of NFI DBD.

Inclusion bodies were resuspended with 25 mM of the studied reducing agent for 2 hour prior to renaturation. Renaturation were performed as described before supplemented with DTT (green columns) or β–Mercaptoethanol (purple column) at a concentration ranging between 0.5 to 10.5 mM. The results are from one assays at each concentration.

In this experiment, NFIDBD Tag was used as it may be used for purification purpose on a Nickel column. NFIDBD Tag was dissolved in denaturing condition with either 25 mM DTT or 25 mM β-mercaptoethanol. The protein was then renatured in the presence of varying concentration of reducing agent. The final concentration of the reducing agent ranged between 0.5 and 10.5 mM. Figure 2.11 shows that the optimal concentration of DTT was 1.5 mM and did not fluctuate significantly between the different concentrations used. With β–mercaptoethanol, the yield of folding was

concentration dependent. The yield increased with higher concentration of β–mercapoethanol to reach a peak at 5.5 mM. Nevertheless, in β-mercaptoethanol, folding was not as effective as in DTT.

V.6 Dilution versus dialysis

Dialysis is often suggested as a way to renature protein but the intermediate form of the protein tends to aggregate with long residence times in renaturation buffer. Dialysis has successfully been used with NDSB 201 to fold the protein complex HDH/Ku (Ochem *et al*, 1998). The principle of this method is that guanidine, being a significantly smaller molecule than NDSB, would diffuse faster in the dialysis buffer and the protein is left without guanidine but with NDSB.

NFIDBD was diluted before dialysis to 4 M guanidine, 1 M NDSB 201, 50 mM Hepes pH 7.5 and 25 mM DTT. Dialysis was tried at 0.5 mg/ml and 0.25 mg/ml NFIDBD. The dialysis buffer was the same as the renaturation buffer. At the same time, a quick dilution renaturation assay was performed using the same protein concentration to compare both methods.

As shown in figure 2.12, both methods gave very different results. The yield achieved by dialysis at both protein concentrations tested was 4 times lower than when using the quick dilution procedure (folding yields were 16 % by dialysis and 64 % by quick dilution).

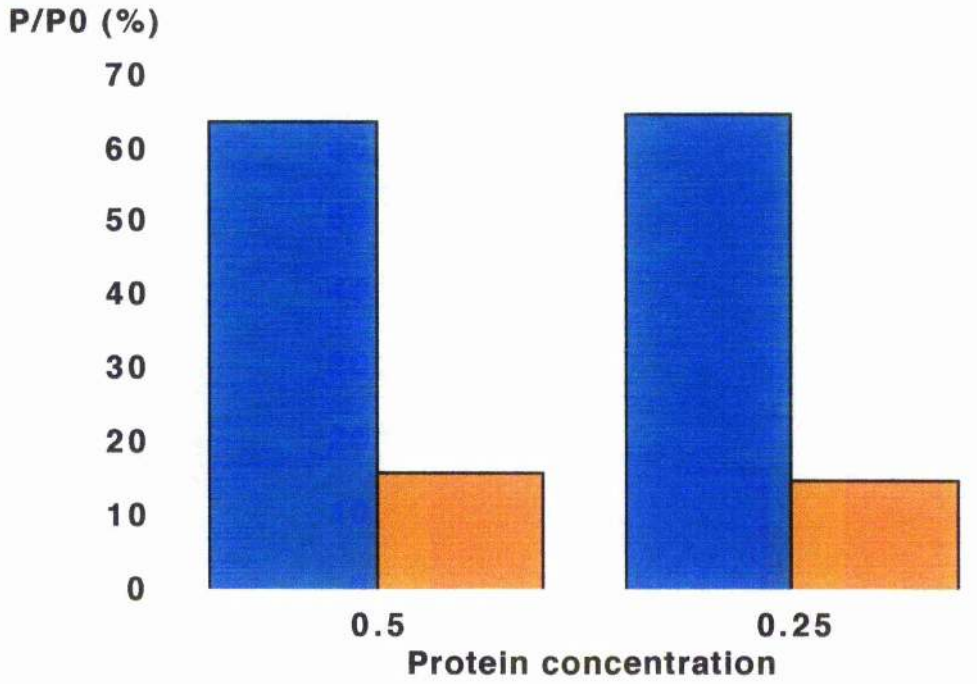


Figure 2.12

Renaturation by quick dilution versus dialysis

NFI DBD was renatured either by quick dilution (blue columns) or by dialysis (orange columns) at a final concentration of 0.5 mg/ml or 0.25 mg/ml. The folding yields were determined by using 1 microliter of each renaturation into a gel retention assay and analysed as described previously. This experiment was performed twice with very similar results

VI. Renaturation using detergent

Two different types of detergent were studied for their ability to influence the folding yield of NFIDBD. CHAPS is a zwitterionic detergent, chemically similar to NDSB, but with a much larger hydrophobic region. A non ionic detergent was also tested, n-octylglucoside (OPG), which is a pyranose with a 8 carbon chain on C1 of the sugar.

The concentration of each detergent used in the renaturation buffer was dictated by the critical micelle concentration (CMC). CMC is the concentration at which detergent forms micelles. CHAPS has a CMC of 4 to 8 mM and therefore was tested at 8 and 80

mM, whereas OPG having a CMC of 14.5 was tried at 15 and 150 mM. The renaturation assay was done in parallel with NDSB, so folding yield could be compared.

In the presence of CHAPS, folding yields were respectively 77 % and 63 % (figure 2.13). These figures were very similar to the yield obtained with NDSB 256 which gave a folding yield of 75 %. In the presence of OPG in the renaturation buffer, the yields were much lower (2 and 19%) even compared to the control which had no additive (20%).

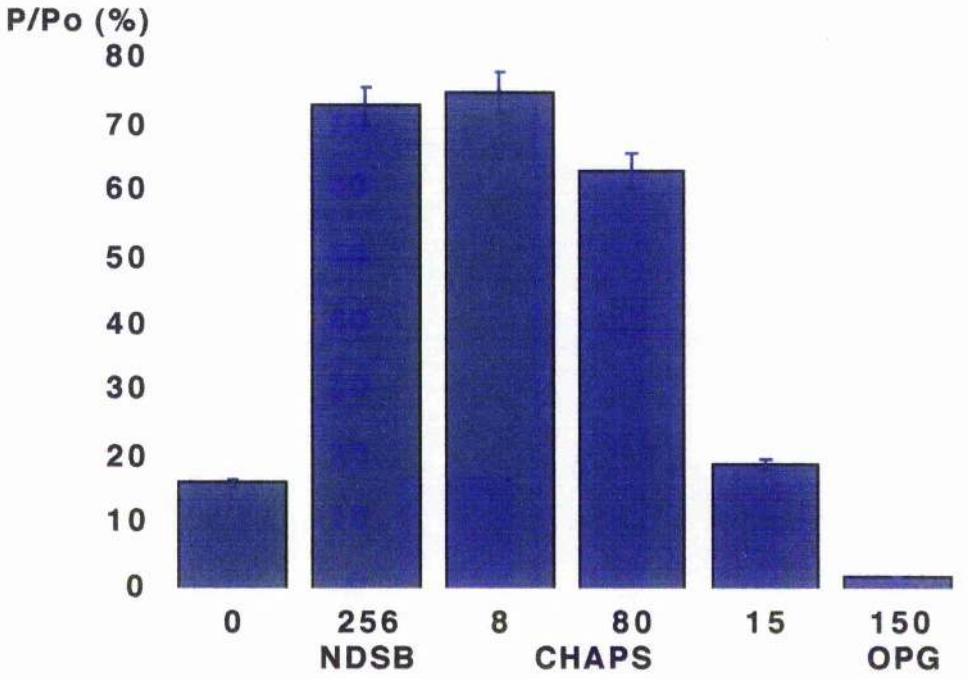


Figure 2.13 Folding yield using detergents CHAPS and OPG

Renaturations were performed as in figure 2.5 except that NDSB was replaced by either CHAPS (8 and 80 mM) or OPG (15 and 150 mM). For comparison purpose, renaturation in the presence of no additive and 1 M NDSB 256 were also performed. The yields are an average of three renaturation assays.

VII. *In vitro* folding of NFI : an overview

The optimal conditions did not vary significantly between NFIDBD and NFIDBD Tag. However with the latter form of the protein, the yields were never as good. The optimal conditions for folding NFIDBD were very sensitive to specific conditions. These were the ionic strength which should be kept at about 1 M guanidine or NaCl, a final protein concentration of 0.1 mg/ml, a pH of 7, DTT at 1 mM and either 1 M NDSB 256 or 8mM CHAPS. Folding should be performed using the quick dilution procedure.

VIII. Renaturation and purification of large amount of NFIDBD

Large scale folding and purification scheme of NFIDBD is shown in figure 2.14. Purified inclusion bodies from induced bacteria were resolubilised in denaturation conditions and NFIDBD was folded in the buffer described. Before purification, the guanidine concentration had to be lowered to 0.3M guanidine so that the protein could either be purified by affinity or by ion exchange chromatography. The renaturation buffer was diluted by stirring the renatured protein solution and pouring in a buffer without salt. More gentle methods were investigated to reduce the salt concentration : dialysis, concentration of protein by ultra centrifugation before dilution and purification on a nickel column. None of these methods improved the yield (data not shown). These other methods had major inconveniences (volume, slow and presence of DTT) and were left out.

To purify NFIDBD, the protein was loaded onto an affinity chromatography column as described in chapter 1 at a flow rate of 1 ml per min. The column was washed and NFIDBD was eluted in 1M NaCl. Most of the NFIDBD was in the flow through fraction (compare LO and FT in the SDS PAGE). The wash removed all of the contaminant protein remaining and a homogenous fraction of NFIDBD was recovered in the

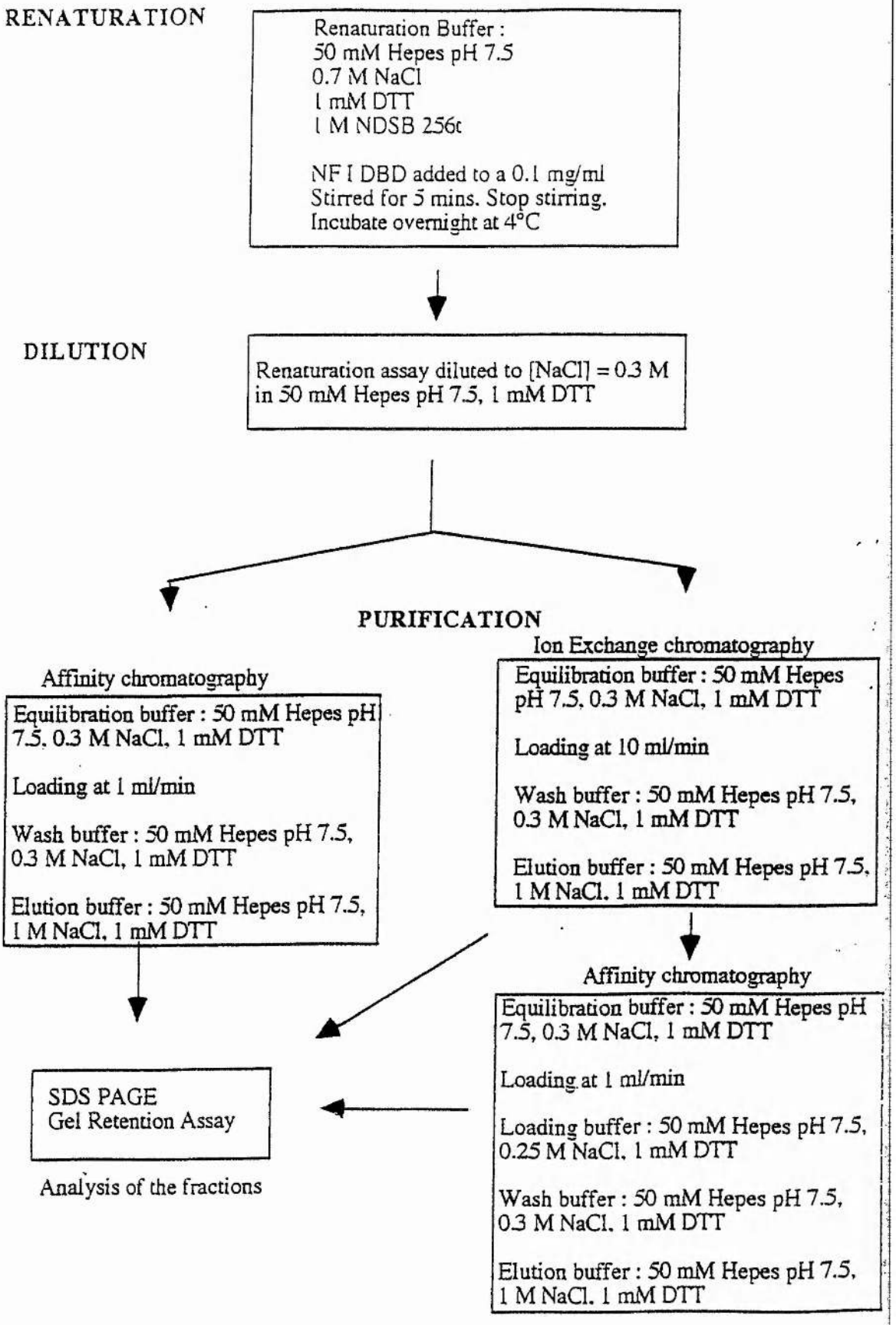
fractions eluting with 1M NaCl. All of the NFIDBD with DNA binding activity bound to the column as no NFIDBD was detected in the flow through or the wash fraction (figure 2.15). The DNA binding site on the column was therefore not saturated. It is important to note that as low as 0.5 ng per microliter can be detected by a gel retention assay. Thus the DNA affinity column is an excellent step to separate active and inactive NFIDBD.

Using the affinity chromatography is in fact time-consuming as the renaturation is diluted and large volumes are being handled. Therefore other means of purification were investigated. The most obvious one was to use ion exchange chromatography. A strong anion exchange resin was chosen which can withstand fast loading capacity (750 cm/hr) (Pharmacia SP Sepharose Fast Flow). The column (1.5 cm2) was loaded at 10 ml/min at 0.3M NaCl. The column was washed and eluted fractions were analysed both by SDS PAGE and gel retention assay. The majority of the protein did not bind on the resin (figure 2.16A and B). On the gel retention assay, the active protein bound to the column (> 95%) and eluted in the 1M NaCl buffer.

NFIDBD was further purified by affinity chromatography. Suprisingly all of NFIDBD bound to the affinity column, suggesting that only properly folded NFIDBD bound to the ion exchange resin. Although the fraction is not as pure as in the previous purification, the protein extract was in this case not perfect. This problem could be solved by using a more homogenous extract of NFIDBD.

From 150 mg of protein extract, 1.9 mg of pure protein was obtained in renaturation of NFIDBD Tag in the presence of NDSB 201. It is expected that the final folding yield in 1M guanidine with no added additive should about 2 % while in the presence of NDSB 256 it reached at least 5 % for NFIDBD.

Figure 2.14 procedure used to purify *in vitro* folded NFI with DNA binding activity after renaturation



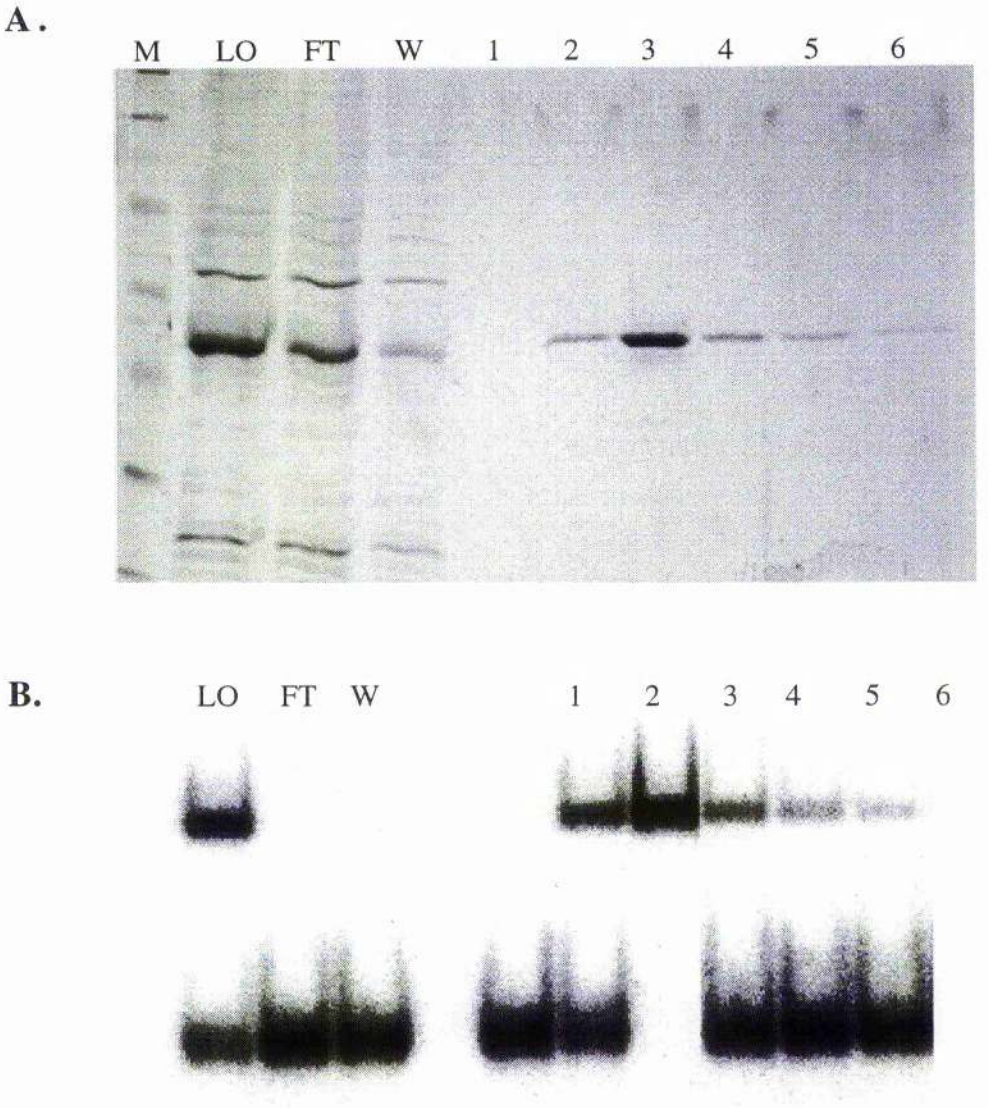


Figure 2.15 :
Purification of *in vitro* folded NFI DBD by affinity chromatography.

NFI DBD was renatured and diluted to a salt concentration of 0.3M. The protein was then loaded (fraction LO) onto a DNA affinity column containing a double stranded oligonucleotide with an NFI DBD site. The column was washed (fraction W) and fractions (fraction 1 to 6) were eluted in 1M NaCl. 10 microliters aliquots of each fraction were analysed on a 12% SDS PAGE stained with Coomassie blue (A) and 1 microliter by gel retention assay (B). FT is the flow through fraction of the column during loading of the affinity chromatography. M is the protein standards (MW are from top to bottom 175, 83, 62, 47.5, 32.5 25, 16.5, 6.5 kDa)

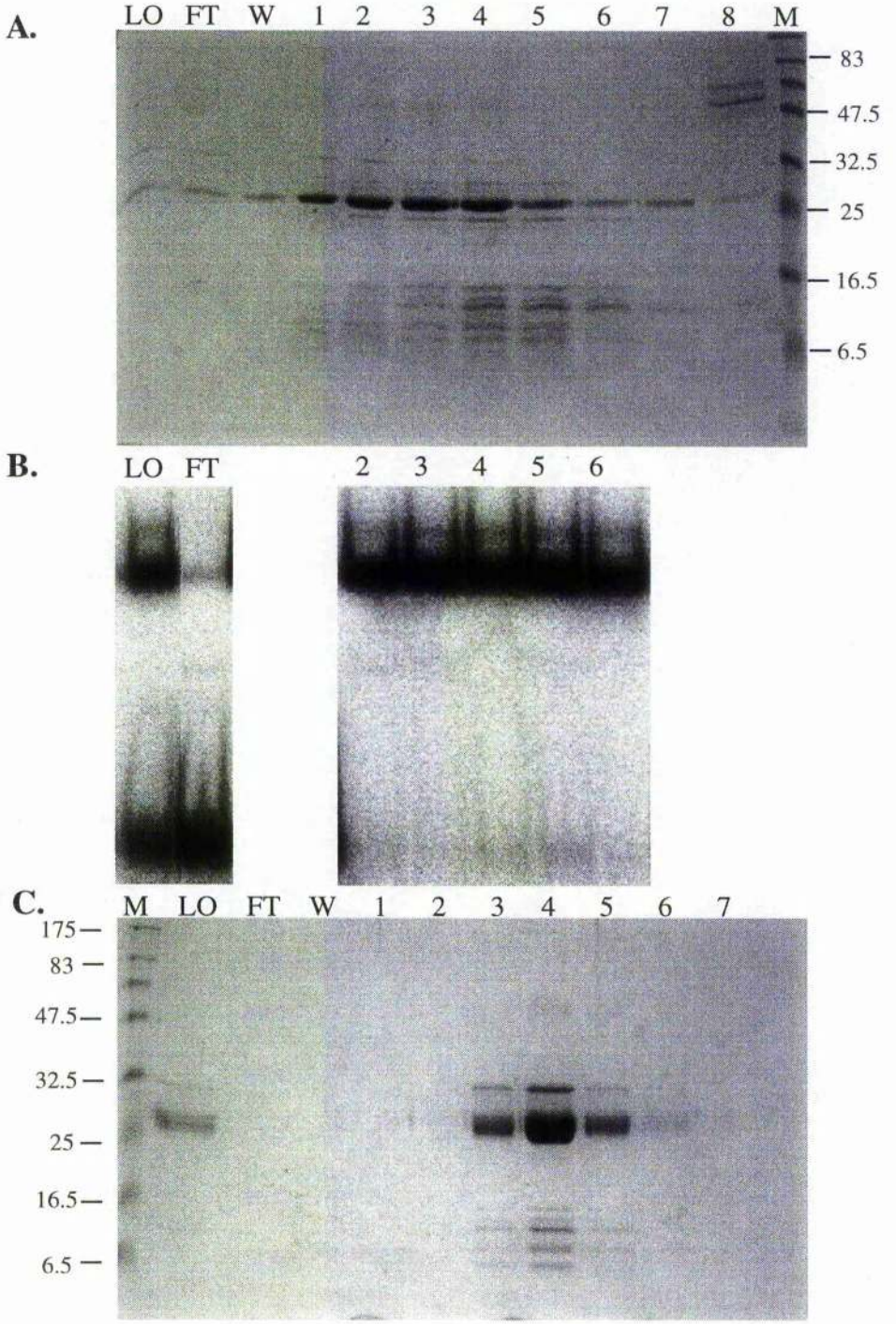


Figure 2.16 :

Purification of in vitro folded NFI DBD by ion exchange (A and B) and affinity chromatography (C)

NFI DBD was renatured and diluted to 0.3 M salt concentration. The protein was loaded on an ion exchange column, wash and eluted in 1 M NaCl. The protein was rediluted to 0.3 M NaCl before purifying it on a affinity column as described in figure 2.15. Fraction LO is the fraction loaded on the column, FT is the flow through of the column, W is the wash fraction and eluted fraction are numbed. M are protein markers. Gels A and C are Coomassie stained 12 % SDS PAGE and B is a gel retention assay.

IX. Comparison of the activity and structure of renatured NFIDBD with baculovirus purified NFIDBD.

In this section, the DNA binding affinity and the structure of the renatured NFIDBD was compared to baculovirus expressed NFIDBD purified as described in chapter 1.

IX.1 Comparison of the dissociation rates

To compare the affinity for both NFIDBD, a DNA competition assay was performed. Both proteins bound to DNA in a gel retention, however, this did not give an idea of the protein affinity for DNA. This could only be tested by determining the equilibrium constant from the association rate and the dissociation rate. We chose just to determine the dissociation rate. This is experimentally more straight forward and gave a better idea of the stability of the DNA-Protein complex.

NFIDBD was incubated first with a molar excess of radioactively labelled DNA containing an NFI binding site. Once equilibrium was reached, the preformed DNA - NFIDBD was challenged with a 100 fold molar excess of the same oligonucleotide but unlabelled. At different time points, an 20 microliter aliquot was taken and loaded immediately onto a running non denaturing polyacrylamide gel where bound and free DNA were separated. All experiments were performed at 20°C.

As it is difficult to estimate the time for the DNA-protein complex to be effectively separated from the free oligonucleotide, the radioactivity in the band corresponding to the complex DNA-NFI was compared at different time points with the first time point loaded after the addition of the cold oligonucleotide. The percentage of the remaining complex DNA-NFIDBD was plotted versus time. The dissociation of the complex followed an exponential decay which was very similar for baculovirus purified and refolded protein (figure 2.17B).

To determine the dissociation rate, the time points between 1 and 30 minutes were used, as at 60 minutes, the reaction had reached an equilibrium with the competitor DNA.

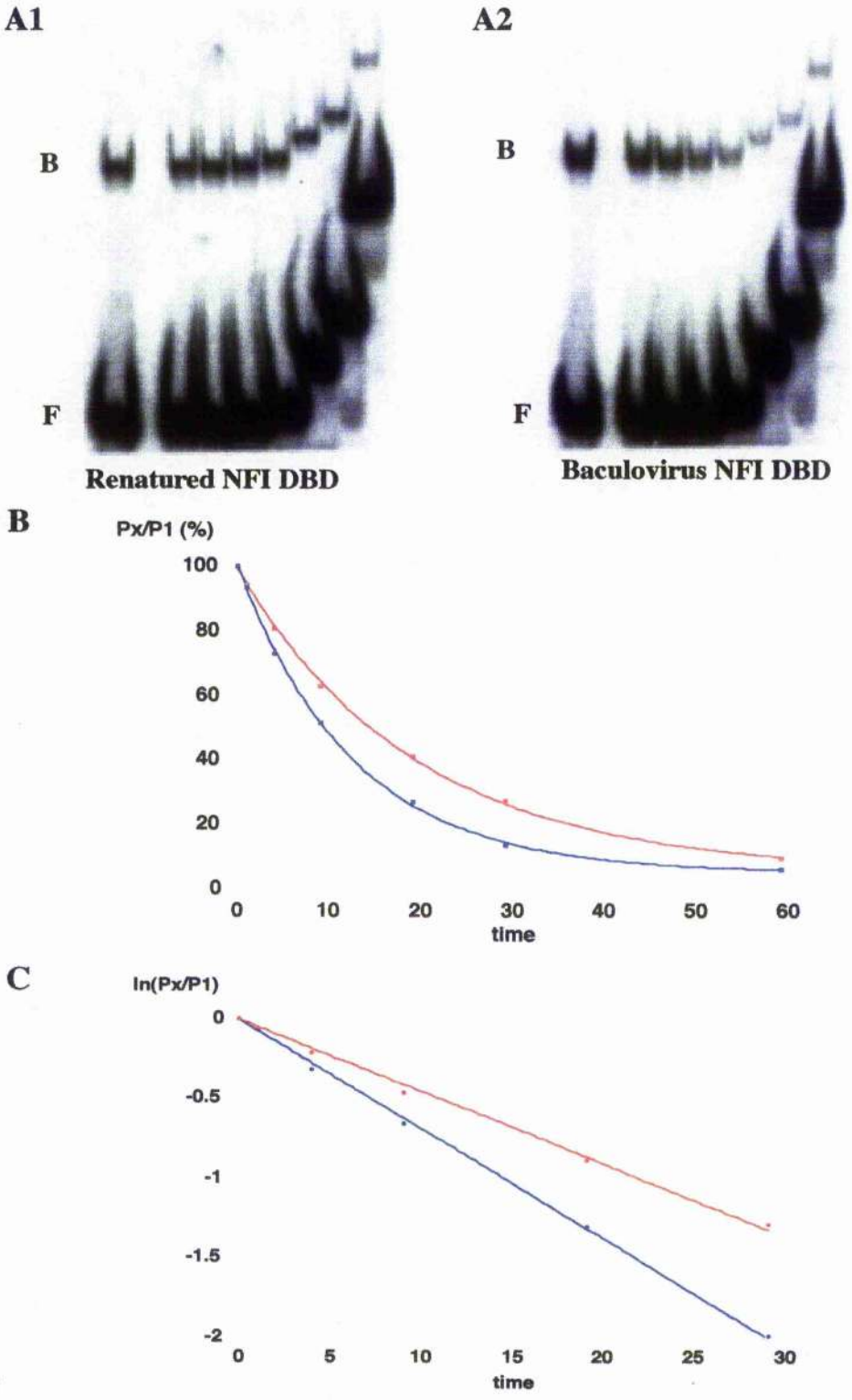


Figure 2.17

Comparison of the dissociation rate of renatured NFI DBD(A1) and baculovirus NFI DBD (A2)

(A1) and (A2) are the gel of the competition assay run. At various time points (0,1, 2, 5, 10, 20, 30, 60 minutes) an aliquot of the competition assay was loaded on a 8%polyacrylamide gel. The gel was dried and analysed by a phosphoimager.

(B) is the percentage of the remaining complex DNA-NFI at various times compared to the complex at time 1 minute.

(C) is the plot to determine the dissociation rate

The logarithm of the remaining probe at time x compared to time 1 minute was plotted versus time. The slope showed a dissociation rate of 1.2 10^{-3} per second for the baculovirus purified NFIDBD and 7.6 10^{-4} per second for *E. coli* NFIDBD, which is rather similar given the very different expression and purification of the two proteins.

IX.2 Comparison of their structure studied by limited proteolysis

An alternative approach was used to examine the structure by studying the digestion pattern of NFIDBD subjected to limited proteolysis in native conditions. NFIDBD cPk purified from baculovirus infected cells and renatured NFIDBD were incubated with chymotrypsin at a ratio from 1:10 to 1 :1000 (w/w) chymotrypsin to NFIDBD for 1 hour at 20 °C. Each digest was performed on the free protein and also complexed to the DNA. Digestion products were analysed on a SDS PAGE.

Figure 2.18 shows the digestion of each NFIDBD. Both proteins are very resistant to the degradation effect of chymotrypsin especially when the protein is bound to a specific recognition site on the DNA. The renatured NFIDBD and baculovirus purified NFIDBD gave similar patterns in the digestion. A first digestion product was observed at about 19 kDa. This digest product was protected from cleavage by DNA. A further digest product, which was not protected by DNA was observed around 15 kD at a ratio of 1:10 NFIDBD to chymotrypsin. The digest product was not only similar for renatured NFIDBD but no other bands were identified than those already mentioned. This signifies that the purified NFIDBD from the renaturation was homogenous concerning its structure.

Both methods suggest that the renatured NFIDBD has a structure that is indistinguishable from that of the baculovirus purified NFIDBD. Furthermore renatured NFIDBD binds DNA with an affinity comparable to that of the baculovirus NFIDBD.

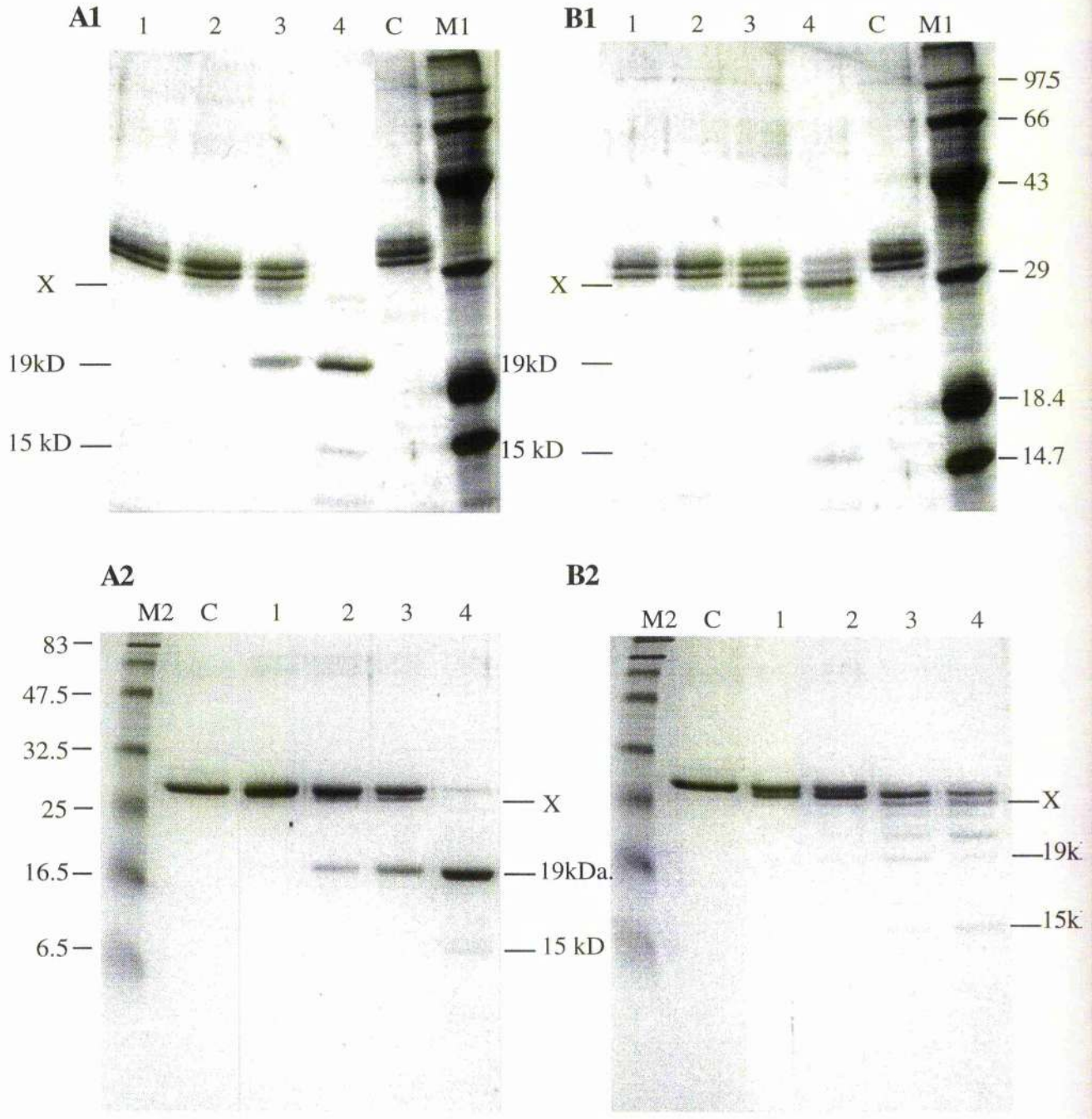


Figure 2.18 comparison of the structure of folded in vitro NFIDBD and NFIDBD purified from baculovirus by limited proteolysis.

Prior to proteolysis, NFIDBD was incubated with DTT and in (B) with DNA for 30 minutes at 20°C. Chymotrypsin was added in varying ratio from 1/100 (w/w) of protease to NFIDBD (lane 1), 1/50 (lane 2), 1/10 (lane 3) to 1/5 (lane 4). The proteolytic reaction was incubated for 1 hour at 20°C. The digestion was stopped by the addition at a final concentration of 1mM PMSF. In the top gel (A1 and B1), the reaction was done on NFIDBD purified from infected insect cells using a recombinant baculovirus and in (A2) and (B2), in vitro folded NFIDBD was used. The various proteolytic digest products were separated in a 12% SDS PAGE and proteins stained by Coomassie blue. Lane C represent NFI incubated with no protease. Lane M1 are the proteins markers (200, 97.5, 66, 43, 29, 18.4, 14.7 kDa.) while M2 are (175, 83, 62, 47.5, 32.5, 25, 16.5, 6.5 kDa)

Chapter 3 : Characterisation of the NFI DNA binding domain by limited proteolysis

In the absence of any high resolution structural data on the DNA binding domain of NFI, alternative methods have been used to define the functional domains of this protein. Site directed mutagenesis experiments did not yield a great deal of information. Most mutants produced inactive monomeric proteins even when expressed in HeLa cells using recombinant vaccinia viruses (Armentero *et al*, 1994).

For these reasons, limited proteolysis was used to study the domain organisation of NFI. The active protein was subjected to limited proteolysis using various proteases, so that the distinct structural domains of NFI could be defined. The protease is thought to preferentially recognise residues available in the protein sequence that are in unstructured region of the protein such as flexible loops. By performing the proteolysis in the absence or the presence of DNA, residues protected by DNA may be identified.

The protease digestion products were analysed by SDS PAGE and the various fragment obtained were characterised by N-terminal sequencing and mass spectrometry. By combining both sets of data, it was possible to precisely determine the position of the cleavage sites.

Digests were initially performed on NFIDBD purified from insect cells infected by recombinant baculovirus. However the identification of the digestion sites was difficult as the protein had a higher molecular weight than expected and the amino terminus of the protein was modified. None of the digestion products identified could be sequenced.

As seen in chapter 2, both NFIDBD purified from infected insect cells and NFIDBD folded *in vitro* following expression in *E. coli*, had a similar chymotrypsin digestion pattern. The digestion patterns obtained with trypsin, endoproteinase Glu-C or

proteinase K were also similar. Therefore the *in vitro* folded NFIDBD was used in this work as its N terminal group was not blocked.

In this chapter, we used the following conventions: ratios are indicated as weight to weight of protease to NFIDBD. Fragments V, T, C and K are proteolytic products of the Endoproteinase Glu-C, Trypsin, Chymotrypsin and proteinase K respectively (Mr is the relative molecular weight).

I. Limited proteolysis of NFIDBD using endoproteinase Glu C

Endoproteinase Glu-C or protease V8, cleaves either after glutamic acid residues in ammonium bicarbonate buffer or after acidic residues (aspartic acid and glutamic acid) in other buffer conditions. As the digestions were performed in Hepes, it was expected that any accessible acidic residues would be susceptible to cleavage.

In the presence of endoproteinase Glu-C, NFIDBD is degraded into fragments of molecular weight of 19.3 and 19.5 kDa at a V8 to NFIDBD ratio of 1/100 (V1, figure 3.1A). At a ratio of 1/10, the initial product (V0) was completely degraded into the V1 product. However at higher ratios, a band was detected with the same Mr as V0. This was real as the amount of protease increased from lane 4 to 7 (figure 3.1A). This band does not appear to be a by product of the protease as it is not seen in the digestion of NFIDBD done in the presence of DNA (figure 3.1B lanes 6 and 7). The band is probably NFIDBD which is difficult to explain as the NFIDBD was totally degraded at a lower amount of protease.

When NFIDBD is complexed to DNA, the degradation into the product V1 is protected from the protease V8 at ratio (w/w) of 1/100 and 1/50 (a weak band of similar apparent molecular weight could be detected from a ratio of 1/10 to 1/1). Instead following a similar cleavage pattern, NFIDBD was degraded into a V2 protein fragment of a

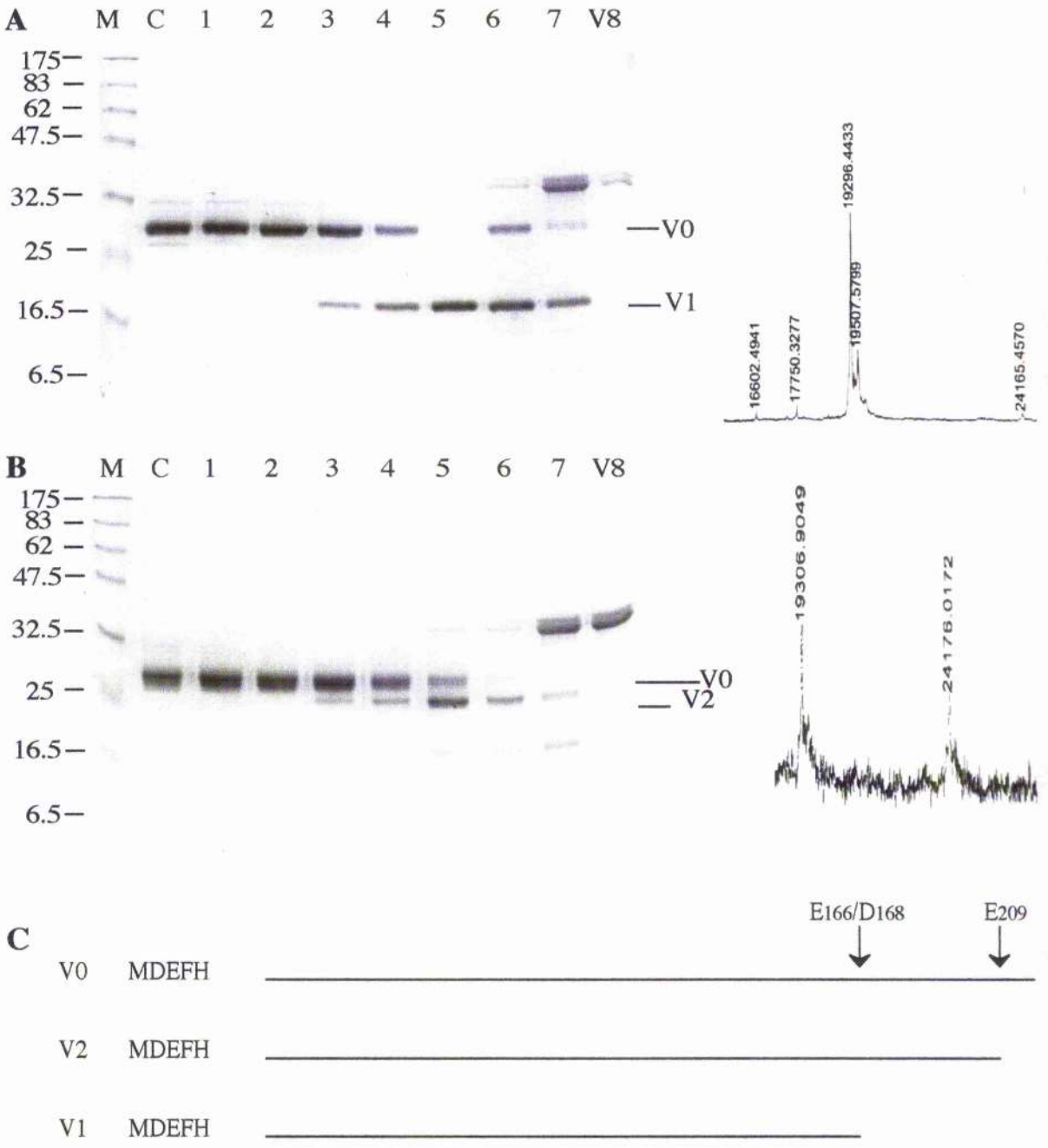


Figure 3.1: Partial proteolysis with endoproteinase Glu-N of NFI DNA Binding Domain(A) and when complexed to a double stranded oligonucleotides with a specific DNA binding site for NFI(B).

Prior to proteolysis NFI was incubated with 5mM DTT and in (B) with DNA for 30 minutes at 20°C. Endoproteinase Glu-N was added in varying ratio from 1/1000 (w/w) protease to NFI (lane 1), 1/500 (lane 2), 1/100 (lane 3), 1/50 (lane 4), 1/10 (lane 5), 1/5 (lane 6) to 1/1 (lane 7). The protease reaction was incubated for 1 hour at 20°C. The protease was inactivated by incubation for 15 mins on ice with 1 mM PMSF. An aliquot (2 μg) of protein was separated on a 12% SDS PAGE and proteins bands revealed by staining with Coomassie blue. The molecular weight of each band was determined by mass spectrometry. The mass spectrum is shown to the right of each gel. Lane C represent NFI non treated with protease and lane V8 represent the protease alone. Lane M are the protein markers and the molecular weight of each marker is indicated in kDa. (C) is a schematic representation of the different cleavage sites of V8 in NFIDBD.

molecular weight of 24.2 kDa. (although the signal was poor in the mass spectrometry, no other peak could be detected between 20 and 25 kDa.).

Protein sequencing of the band V1 and V2 yielded the original N terminal sequence of NFIDBD and thus the V8 protease digested at the C terminus of NFIDBD (figure 3.1C). The digestion site was after the Glu 166 and after the Asp in position 168 in the absence of DNA, and after the Glu in position 209 in the presence of DNA.

Interestingly in the absence of DNA, no V2 product was observed in the gel, suggesting that the cleaving rate of the V8 protease for the acidic residues in position 166 and 168 was higher than the one in position 209.

In both assays, the protein products (V1 and V2) were optimal at a V8 to NFIDBD ratio of 1/10 (w/w). Increasing V8 proportion, the intensity of the band representing the V1 and V2 protein fragment decreased suggesting further proteolysis on other acidic residues in NFIDBD. Some of these products were detected in the SDS PAGE. However the amount of protein was insufficient to determine their amino terminal sequence and hence to determine the cleavage site.

II. Limited proteolysis using chymotrypsin

Chymotrypsin cleaves protein preferentially after aromatic, leucine, methionine and alanine residues. The digestion of NFIDBD using chymotrypsin followed a simple pattern. NFIDBD was degraded from a ratio of 1/100 of chymotrypsin to NFIDBD in to a single band (figure 3.2A). However the laser desorption mass spectrometry showed that the protein is in fact cleaved at several sites producing several fragments of a molecular weight ranging from 19.8 to 20.4 kDa (C1) while N terminal sequencing revealed that the fragments had common N-termini. Chymotrypsin cleaves after several different residues between the tyrosine residue in position 170 to the valine residue in position 175. In fact this region contains an alanine, a leucine, another tyrosine and a phenylalanine, all residues susceptible to cleavage. As the concentration of chymotrypsin increases, C1 appears to be of a slightly smaller molecular weight. It is

most likely that the chymotrypsin cleaves after the valine 175 first and then down to the tyrosine in position 170. Several bands of different molecular weight were observed in various mass spectrophotometry charts.

At a ratio of 1/1, two further proteolytic products were detected in the presence and the absence of DNA. Sequencing of each band showed that the lower band (C3) contained the N terminal sequence of NFIDBD (MW = 4 and 4.15kDa). The upper band (C2) was the cleavage product of C1 at the tyrosine and at the phenylalanine residue in position 32 and 33 respectively. However from the gel, the molecular weight of the C2 product clearly indicated that this fragment migrated between the 6 and the 16.5 kDa protein marker. If the C2 and C3 were digestion products simply at position 32 and 33 from the C1 fragment, the molecular weight of the C2 product would be 15.5 kDa. No peak at this molecular weight was observed on the mass spectrometry charts. Instead peaks in the mass spectrometry chart were seen at 10 to 10.3 kDa. Therefore the C2 product contained the NFIDBD sequence from the Y32/F33 to either a leucine or to a tryptophane residue at position 119. However other peaks were observed in the mass spectrometry (i.e. a peak at 12 kDa.) but their molecular weight determination was not as accurate as for the 10.3 kDa fragment. The apparent molecular weight was calculated and was 11 kDa.

When complexed to DNA, the digestion of NFIDBD by chymotrypsin produced a complex patterns of products. The digestion of the protein to residues 170 and 173 (to produce C1) was barely detectable at ratio of 1/10 to 1/1. Both sites were protected from cleavage in the presence of DNA (figure 3.2B). However the C2 and C3 proteolytic product appeared with the same pattern, although the intensity of the bands was slightly lower. At a ratio of 1/100 an uncharacterised proteolytic product of NFIDBD was detected with an apparent molecular mass of 25 kDa (C4) and a further product at a chymotrypsin to NFIDBD ratio from 1/10 with an apparent molecular weight of 24.5 kDa (C5).

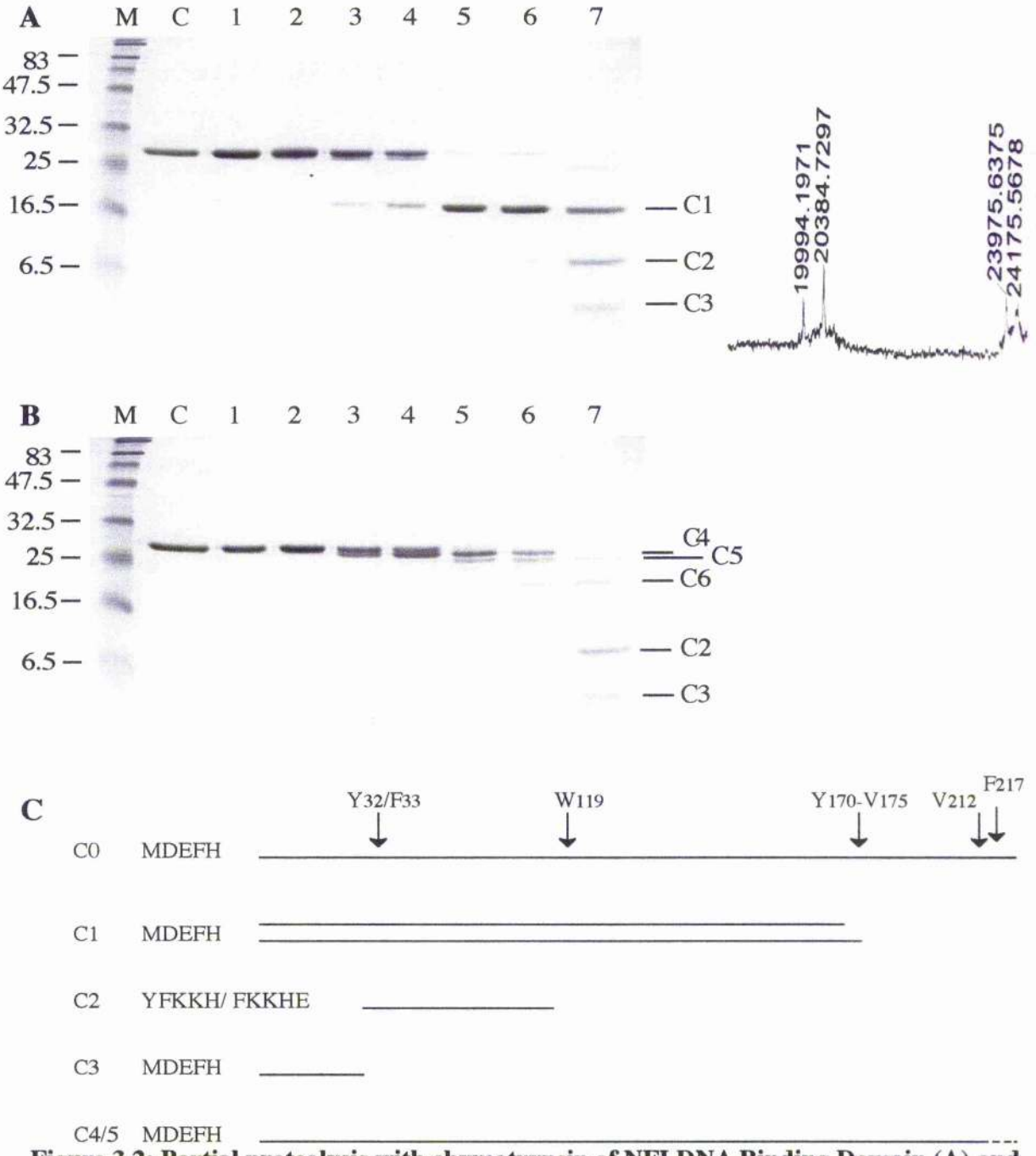


Figure 3.2: Partial proteolysis with chymotrypsin of NFI DNA Binding Domain (A) and when complexed to a double stranded oligonucleotides with a specific DNA binding site for NFI(B).

Prior to proteolysis NFI was incubated with 5mM DTT and in (B) with DNA for 30 minutes at 20°C. Chymotrypsin was added in varying ratio from 1/1000 (w/w) protease to NFI (lane 1), 1/500 (lane 2), 1/100 (lane 3), 1/50 (lane 4), 1/10 (lane 5), 1/5 (lane 6) to 1/1 (lane 7). The protease reaction was incubated for 1 hour at 20°C. The protease was inactivated by incubation for 15 mins on ice with 1 mM PMSF. An aliquot (2 µg) of protein was separated on a 15% SDS PAGE and proteins bands revealed by staining with Coomassie blue. The molecular weight of bands C1, C2 and C3 were determined by mass spectrometry. The mass spectrum is shown to the right of the gel (A). Lane C represent NFI non treated with protease. Lane M are the protein markers and the molecular weight of some markers are indicated in kDa. In (C) is a schematic representation of the different cleavage sites of Chymotrypsin in NFIDBD.

N terminal sequencing of the C4/C5 protein fragment showed that the cleavage took place at the C terminal end of the protein but as no precise molecular weight was obtained, the digestion site could only be mapped approximately, after the phenylalanine 217 and valine 212. Further bands could be detected just under the C5 protein fragment such as C6. Unfortunately there were not enough materials for sequencing and no molecular weight of the C6 fragment was obtained. A band was detected just under the C5 product but a similar band was also seen in the digest performed on free NFIDBD (Mr= 24 kDa). This band was probably a fragment of the chymotrypsin protein.

III. Limited proteolysis using Trypsin

Trypsin cleaves protein specifically after positively charged residues (lysine and arginine). As shown in figure 3.3, NFIDBD is very sensitive to trypsin.

The various digestion products obtained are shown in figure 3.3. NFIDBD is cleaved in three products of molecular weight 21.6 (T1) 20.5 and 20.6 kDa (T2) (figure 3.3A and B). At a ratio of 1/100, NFIDBD is degraded into the T1 and T2 product. The amount of T1 and T2 were similar at that ratio. At higher ratio, NFIDBD was degraded into T2 to be totally degraded and could not be detected at a ratio of 1/5 of trypsin to NFIDBD (w/w). These digestions take place at the C terminus of NFIDBD as the N termini of the product was identical to the original NFIDBD. The T2 protein product is composed of 2 different proteolytic products. As the concentration of trypsin is increased, NFIDBD seems to first be degraded at the arginine residue in position 178 followed by the arginine in position 176 as suggested by mass spectrometry analysis.

When complexed to DNA, NFIDBD was cleaved into the same products. However in the presence of DNA, the proteolytic pattern of NFI was different. At a ratio of 1/100 the main product was the T1 protein fragment and the degradation from T1 to T2 proceeded at a reduced speed. Proteolysis did not run to completion as T2 could still be detected at a trypsin to NFIDBD ratio of 1/1.

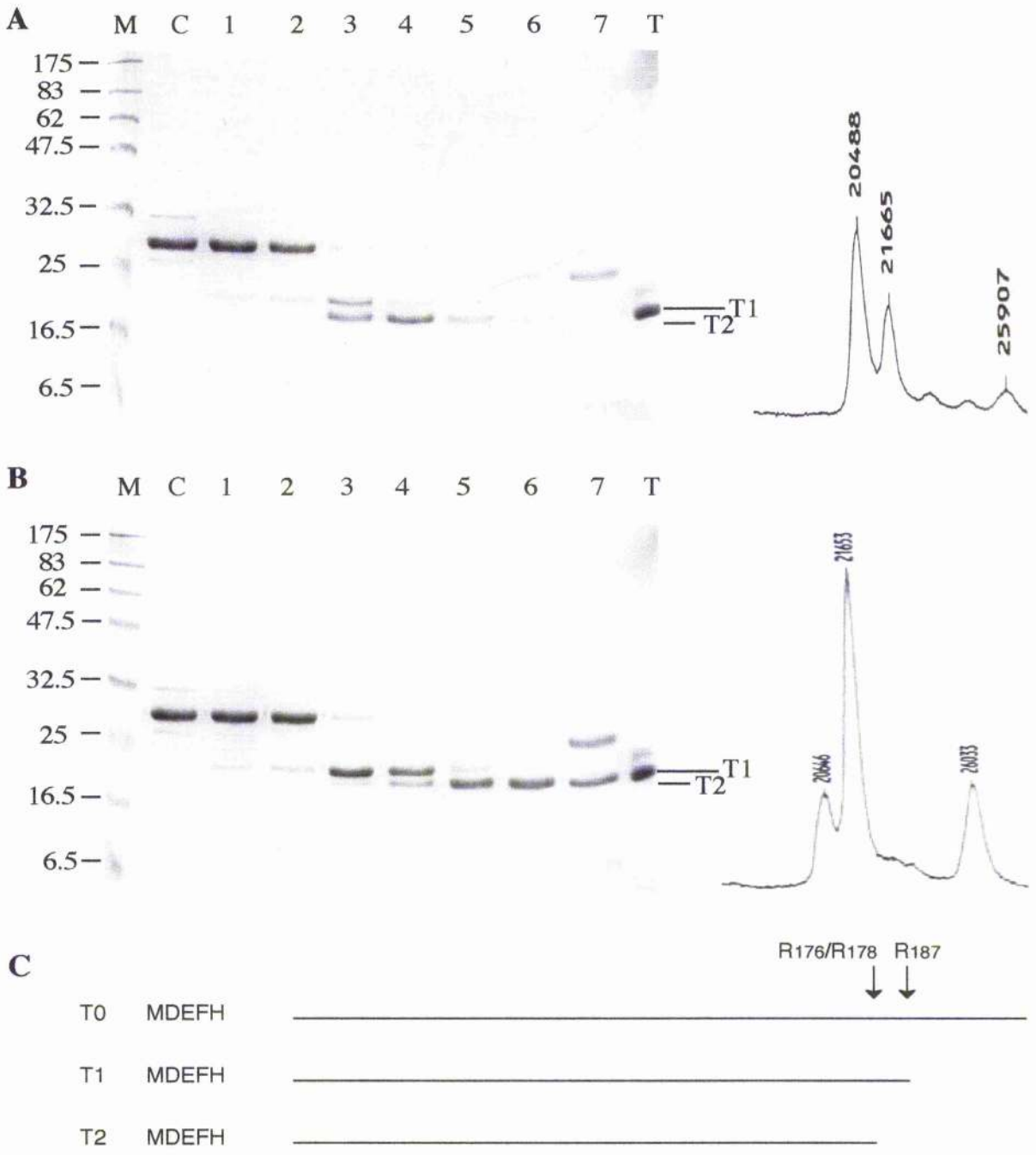


Figure 3.3: Partial proteolysis using trypsin of NFI DNA Binding Domain (A) and when complexed of a double stranded oligonucleotides with a specific DNA binding site for NFI(B).

Prior to proteolysis NFI was incubated with 5mM DTT and in (B) with DNA for 30 minutes at 20°C. Trypsin was added in varying ratio from 1/1000 (w/w) protease to NFI (lane 1), 1/500 (lane 2), 1/100 (lane 3), 1/50 (lane 4), 1/10 (lane 5), 1/5 (lane 6) to 1/1 (lane 7). The protease reaction was incubated for 1 hour at 20°C. The protease was inactivated by incubation for 15 mins on ice with 1 mM PMSF. An aliquot (2μg) of protein was separated by migration in a 12% SDS PAGE and proteins bands revealed by staining with Coomassie blue. The molecular weight of each band was determined by mass spectrometry. The mass spectrum is shown on the right of the gel. Lane C represent NFI non treated with protease and lane T represent the protease alone. Lane M are the protein markers and the molecular weight of each marker is indicated in kDa. (C) is a schematic representation of the different cleavage sites of trypsin in NFIDBD.

In a further experiment performed at a higher temperature, NFIDBD was totally degraded by trypsin even in the presence of DNA. At a trypsin to NFIDBD ratio of 1/1, other forms of the protein migrating faster in the gel were detected: these protein products migrated between the protein markers of 16.5 kDa and 6 kDa. None of these polypeptides species were identified.

IV. Limited proteolysis using proteinase K.

Proteinase K is a non specific protease and was therefore expected to cleave after any residue. From a proteinase K to NFIDBD ratio of 1/100, free NFIDBD is cleaved into a major product K1. In the presence of DNA, this K1 product was protected from cleavage, and instead two major protein fragment K2 and K3 were obtained, they had a higher molecular weight than K1.

The N terminal sequencing showed that K1 and K2 were cleaved at their C terminal end. Although K3 was not sequenced, it was expected to have the same N terminal sequence as it is an intermediate proteolytic product of degradation to the K2 fragment. This product was also apparent in the digestion of free NFIDBD (albeit weaker). The molecular weight of these protein fragments were not obtained by mass spectrometry, but by calculating their apparent molecular weight in SDS PAGE using the V1 and T1 and T2 as molecular weight markers (figure 3.4C). The K1 fragment migrated just between the T2 and V1 while the K2 fragment migrated between T1 and T2. The apparent molecular weight for K1 was estimated at 19.6 kDa while K2 and K3 were 21.1 kDa and 24 kDa respectively. This meant that K1 was cleaved after the leucine in position 169, K2 after the glutamic acid 181 and K3 after the glutamine in position 208.

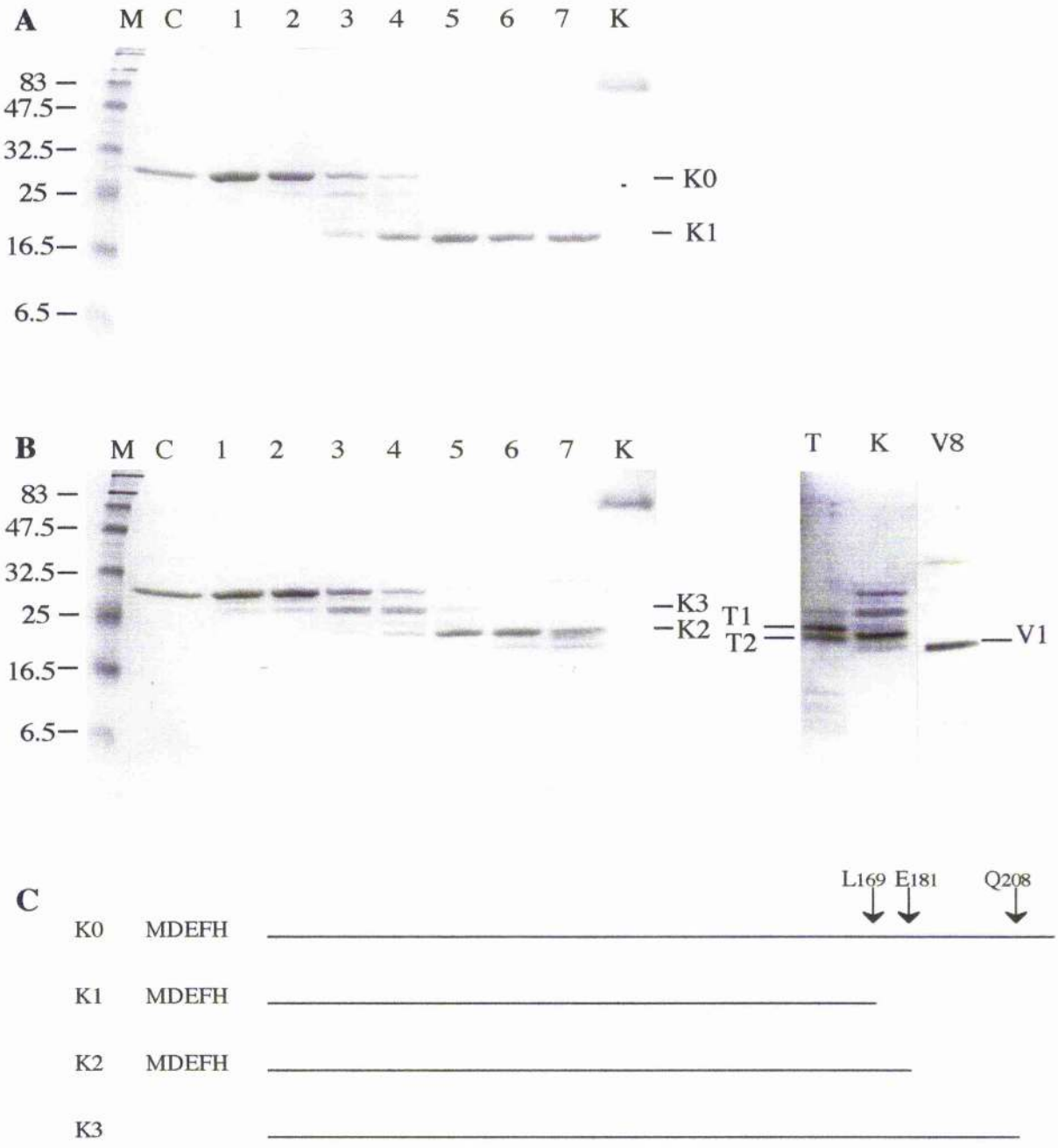


Figure 3.4: Partial proteolysis using proteinase K of NFI DNA Binding Domain (A) and when complexed of a double stranded oligonucleotides with a specific DNA binding site for NFI(B).

Prior to proteolysis NFI was incubated with 5mM DTT and in (B) with DNA for 30 minutes at 20°C. Proteinase K was added in varying ratio from 1/1000 (w/w) protease to NFI (lane 1), 1/500 (lane 2), 1/100 (lane 3), 1/50 (lane 4), 1/10 (lane 5), 1/5 (lane 6) to 1/1 (lane 7). The protease reaction was incubated for 1 hour at 20°C. The protease was inactivated by incubation for 15 mins on ice with 1 mM PMSF. An aliquot (2μg) of protein was separated by migration in a 15% SDS PAGE and proteins bands revealed by staining with Coomassie blue. The apparent molecular weight of K1, K2 and K3 was determined by comparing on a SDS PAGE with other NFIDBD proteolytic products of known mass (T1, T2 and V1). Lane C represent NFI non treated with protease and lane K represents the protease. Lane M are the protein markers and the molecular weight of some of the markers are indicated in kDa. In (C), the cleavage site are schematicaly represented.

V. Summary

By performing these partial proteolysis experiment on NFIDBD in the presence and absence of DNA, different patterns of digestion were obtained. The susceptibility to degradation by proteases of the residues in NFIDBD enabled us to determined three different groups of residues (figure 3.5): i) those which were degraded in the presence DNA (R187, Y32 and F33, Q208, E181), ii) those whose degradation was slowed down in the presence of DNA (R176, R178) and iii) those which were protected from degradation by DNA (E166, Y170 to V175 and L169). Other cleavage sites are likely since some of them were detected by SDS PAGE but the cleavage site has not yet been determined.

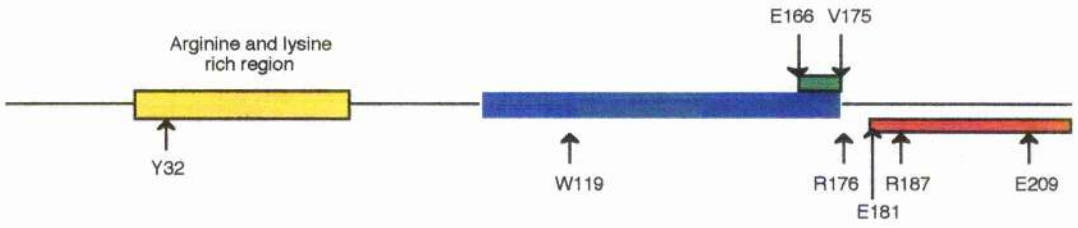


Figure 3.5 Representation of the various domains susceptible to proteolytic degradation.

The various cleavages site identified in the partial proteolysis experiments are represented. Cleavage sites identified in the absence of DNA, are represented over the diagram, while those which were detected in the presence of DNA are shown under it. The domain sensitive to cleavage by proteases, is shown in a green box while the domain not required for DNA binding is in a red box. The others region previously identified as being involved in DNA binding are the lysine-arginine rich domain (yellow box). The conserved region between residues 100 to 175 is also shown by a blue box.

Chapter 4 : Identification of the cysteines residues of Nuclear Factor I involved in the DNA binding activity.

In vivo, NFI DNA binding activity seems to be regulated through a red ox mechanism involving cysteine residues which are required for DNA binding (Bandyopadhyay *et al*, 1998). However, which of the four conserved cysteine residues is involved in DNA binding is still unknown. Site directed mutagenesis showed that all of them, except cysteine 111 (positioning in human), were sensitive to mutations which led to an inactive protein (probably due to misfolding of the protein, Novak *et al*, 1992, Armentero *et al*, 1994, Bandyopadhyay *et al*, 1994). To identify the cysteine residues involved in the DNA binding activity, a direct labelling method with iodoacetic acid was used (prior to use, iodoacetic acid was neutralised by titration with NaOH to form iodoacetate). Unlike site directed mutagenesis, this method has the advantage that iodoacetate labelled all the different cysteine residues involved in the active site of a native protein and ensures that the overall fold is conserved. Iodoacetate labelling is a method that was already successfully used to label cysteine residue in position 62 involved in the DNA binding of the p50 subunit of NFkB (Matthews *et al*, 1992).

At a pH lower than 8, iodoacetate specifically reacts with the SH group of cysteine via a direct nucleophilic subtitution releasing a proton and a negatively charged iodine and resulting in the addition of a carboxyl group on the S. This new carboxyl function is a negatively charged group at physiological pH.

I. Effect of iodoacetate on DNA binding activity of NFI

It had been shown on cellular extract fractions, that the DNA binding activity of NFI was abolished by treatment with a number of cysteine modifying chemicals (such as N ethylmaleinede, diamide, Novak *et al*, 1992, Bandyopadhyay *et al*, 1994). Furthermore it was noticed that the addition of fresh DTT to fractions containing purified NFIDBD enhanced the DNA binding activity of this protein. This suggested that cysteine residues in NFIDBD were directly involved in the DNA binding activity.

This was investigated further in the following manner: cysteine residues were first fully reduced by incubation in 20 mM DTT for 30 minutes (this step ensured that the chemical modification could take place on all available sulphydryl groups). NFIDBD was then diluted into the gel retardation buffer and treated with iodoacetate at a concentration varying from 2 to 15 mM for 30 minutes at 20°C. The iodoacetate modification reaction was stopped by the addition of 50 mM DTT for 15 minutes. The remaining DNA binding activity of NFI in each assay was determined by a gel retention assay in the presence of an excess of probe. This enabled to detect any inhibitory effect of iodoacetate on the DNA binding activity of NFI. Quantification of the amount of radioactivity in each NFI band (Figure 4.1A) was done by scanning the film followed by analysis with Biorad analyser software. The intensity of each band in the presence of iodoacetate was compared to the control (incubated without any iodoacetate). The percentage of the radioactivity remaining was plotted versus the concentration of iodoacetate. In all samples containing iodoacetate, the DNA binding activity of free NFIDBD was inhibited and the remaining radioactivity varied between 5 and 10%. Addition of 2 mM iodoacetate was sufficient to inhibit the DNA binding activity of NFI (Figure 4.1A). However some DNA binding activity still remained even after treatment at higher concentration of iodoacetate (10 or 15 mM) which can not be explained by the residual DTT present in the assay buffer (less than 0.5mM).

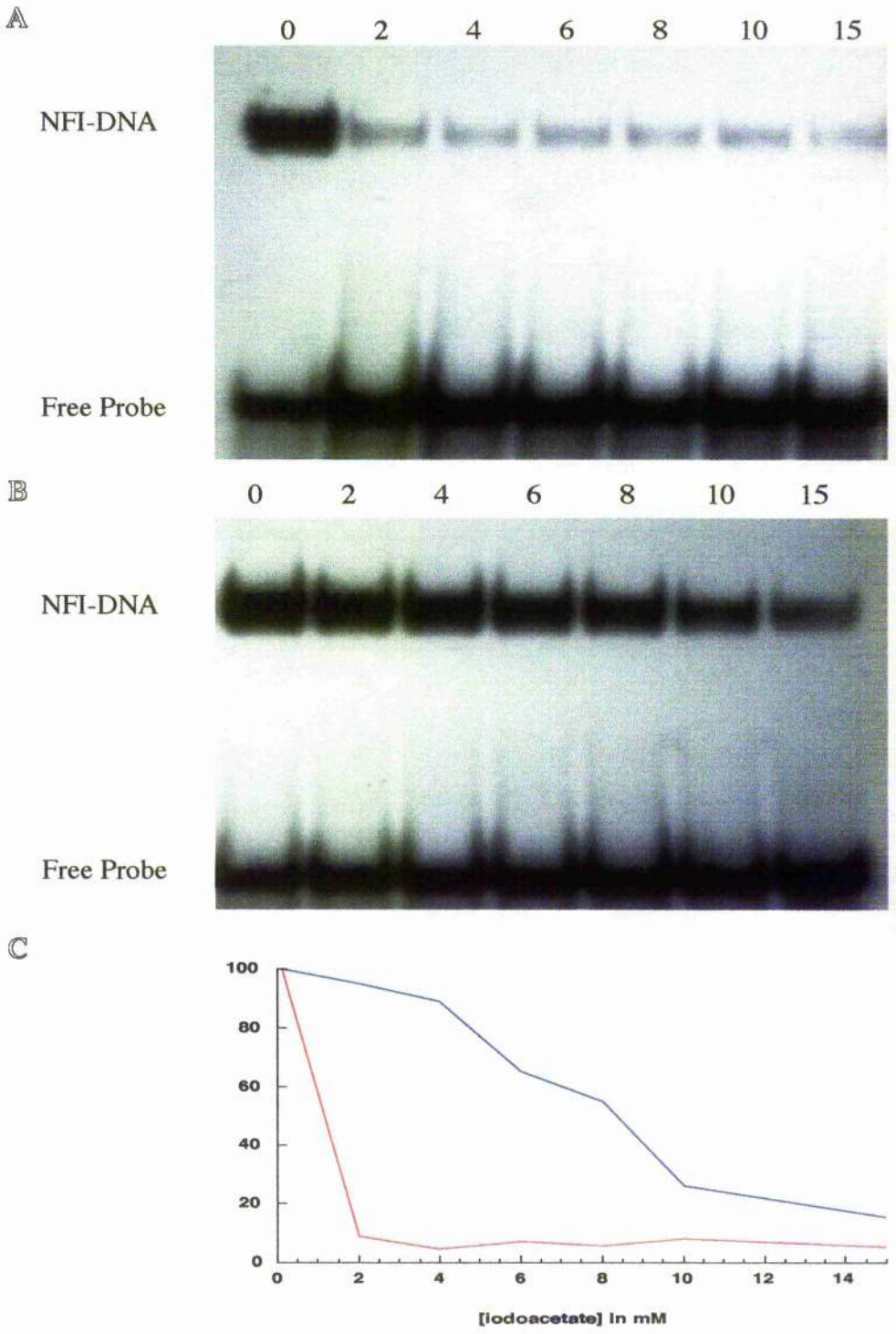


Figure 4.1 Effect of various concentration of iodoacetic acid on the DNA binding activity of NFI. Iodoacetic acid was added either to NFI (A) or NFI complexed to DNA (B) and incubated for 30 minutes at 20°C. The reaction was quenched by addition of DTT to a final concentration of 50mM. The complex DNA-NFI was separated from the free probe on a 8% native polyacrylamide gel at 150 V for 90 minutes. The gel was dried and exposed with a film overnight. The positioning of the complex NFI-DNA and of the free DNA are indicated. The concentration of the iodoacetic acid used in each assay is also indicated on top of each lane in mM. In (C) is a representation of the percentage of DNA binding activity of NFIDBD remaining after treatment with iodoacetic acid either free (red) or complexed to DNA (blue).

As opposed to what has been described above where DNA was added after modification, NFIDBD was also treated with the same chemical but when complexed to a radioactively labelled oligonucleotide containing an NFI binding site (figure 4.1B). This was performed as described above except for the fact that the probe (1.6 ng) was added before treating NFIDBD with iodoacetate. NFIDBD, when complexed to the DNA, was resistant to inactivation by iodoacetate at concentration of 2 to 15 mM. However at concentration of 15 mM of iodoactetic acid, the modifying agent had an inhibitory effect on the DNA binding activity of NFI as the quantity of complex probe-NFI decreased. When NFIDBD was complexed to DNA, the percentage of remaining DNA binding activity of NFIDBD followed a sigmoid curve. At concentration under 4mM of iodoacetate, the DNA binding activity of NFIDBD was over 90% of the control. Above 4mM of iodoacetate, the DNA binding activity of NFIDBD decrease dramatically falling to 15% at 15mM iodoacetate.

These experiments clearly demonstrated the involvement of cysteines residues in the DNA binding activity of NFI.

II. Labelling of cysteines using ^{14}C iodoacetate.

To identify the cysteine residue(s) modified by iodoacetate in DNA binding experiments, the differential labelling procedure develop by Koshland was used (1959). This two step procedure involved the treatment of NFI when complexed to DNA with "cold" iodoacetate in order to modify exposed and sensitive residues to iodoacetate but which are not involved in the DNA binding activity. Elution from the DNA column then exposed any cysteines residues, in intimate association with DNA, which can then be labelled with ^{14}C iodoacetate. To avoid any alteration to the protein state, ^{14}C iodoacetate labelling was performed under identical conditions to the initial cold iodoacetate treatment.

The procedure used is described in figure 4.2. NFI was first complexed to DNA using an NFI binding site immobilised on Sepharose beads. Bound NFIDBD was extensively

washed and then exposed to unlabelled 10 mM iodoacetate (in the same buffer) for 30 minutes at 20°C. The concentration of iodoacetate used was kept to 10mM to ensure that the reaction ran to completion. Beads were then washed by the equilibration buffer supplemented by 2mM DTT (to quench any residual iodoacetate) followed by elution of NFI from the beads by raising the salt concentration to 1M.

NFI was then dialysed to reduce the NaCl concentration to 0.25M and the protein was concentrated by centrifugation using a Centricon membrane with a molecular weight cut off of 30 kDa. NFI was labelled with 10mM ^{14}C iodoacetate (ICN) under the same conditions as described above. The labelling reaction was terminated by the addition of 50 mM DTT and incubation continued for a further 15 minutes.

III. Fractionation of the peptides from the tryptic digestion and identification of the labelled peptides.

^{14}C labelled Nuclear Factor I was precipitated by addition of 10% (w/v) trichloroacetic acid and the pellet washed twice with acetone. This step separated the unincorporated iodoacetate from the protein. The NFIDBD precipitate was resuspended in 8 M deionised urea, 100 mM Tris pH 7.5 and boiled for 10 minutes. The urea concentration was lowered to 2 M by addition of water, and NFI was then digested using trypsin at a ratio of 1/25 (w/w) for 16 hours at 37°C.

The various peptides were purified by reverse phase HPLC using a Waters HPLC system and a "delta pak" C18 column. Elution was achieved using a gradient of 0 to 98% acetonitrile in water in the presence of 0.06% trifluoroacetic acid. Fractions were collected and their radioactivity was determined by liquid scintillation counting. The elution profile of the HPLC column at 210 nm and scintillation counting are represented in figure 4.3. Six main radioactive peaks were obtained. Fraction 7 corresponded to unreacted iodoacetate. Fraction 49, 58, 60, 61 and 73 were dried and peptides sequenced by automated Edman degradation.

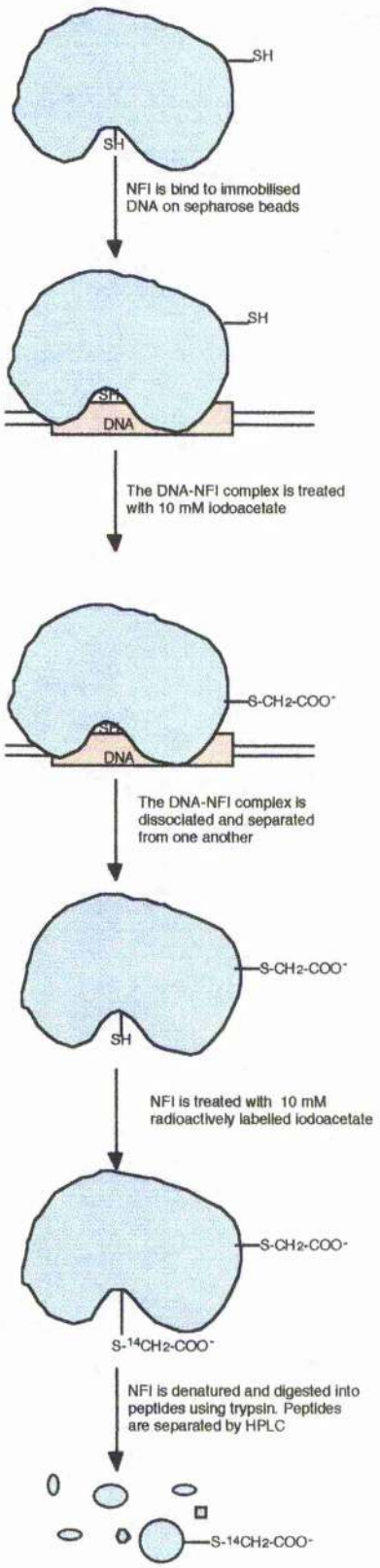


Figure 4.2 Procedure used for the differential labelling with iodoacetate of cysteine residues in Nuclear Factor I

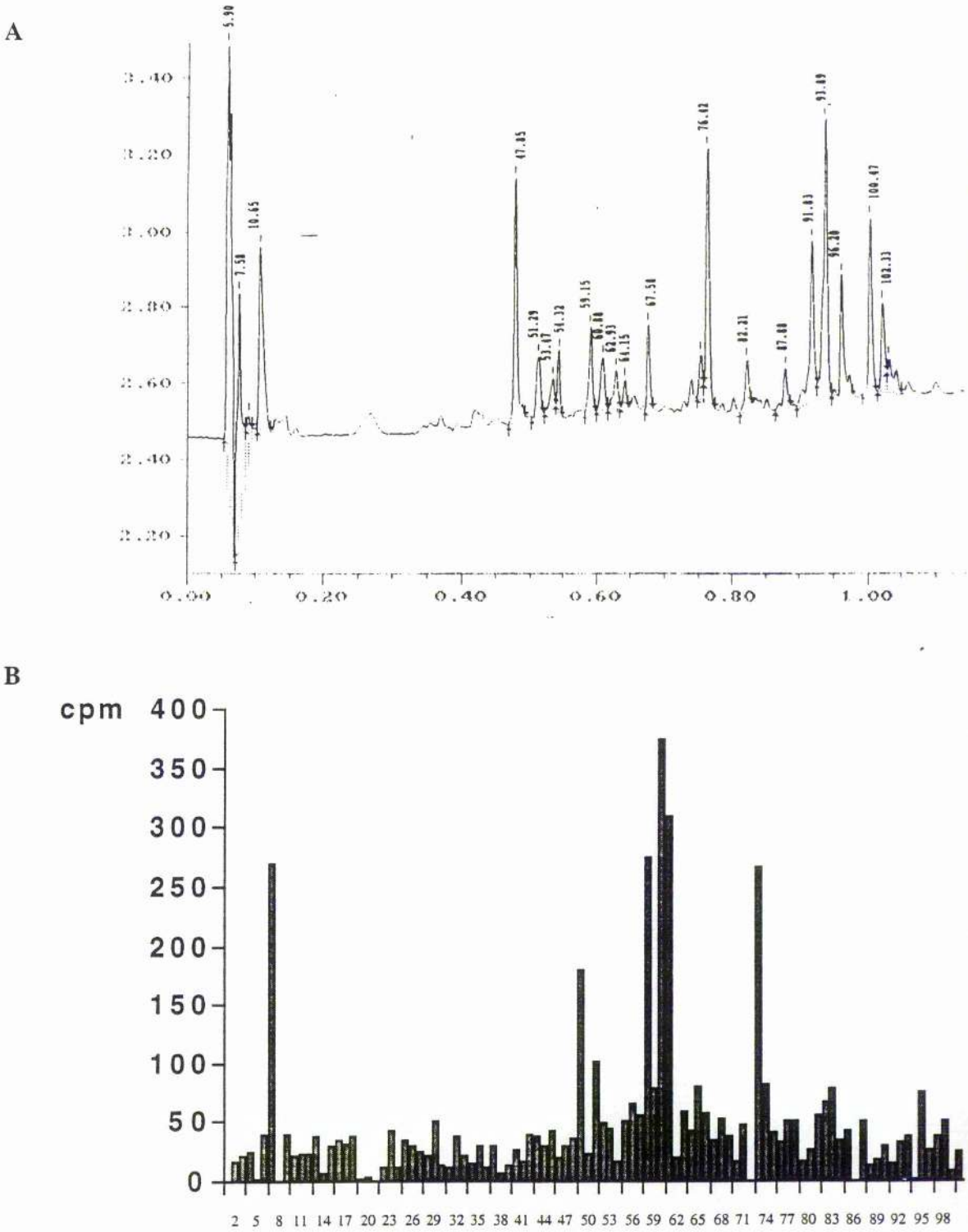


Figure 4.3 Fractionation of the various peptides obtained from the trypsin digestion of NFI.

The elution profile of the reverse phase HPLC is represented in (A). Identification of C14 labelled peptides (B) was done by scintillation counting for 5 minutes and plotted. The scintillation counting of each fraction is represented in counts per minute (cpm).

IV. Sequencing of the various fractions and identification of the residues labelled with ^{14}C iodoacetate.

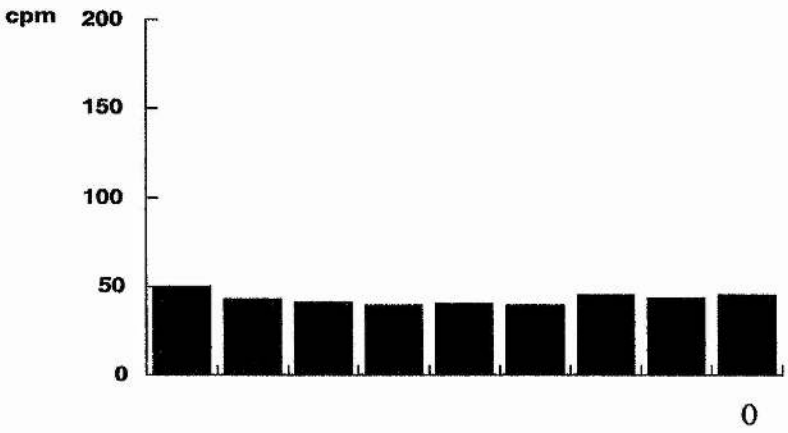
Each fraction was sequenced for eight cycles. 80% of the material released at each cleavage step was loaded onto a HPLC in order to identify the amino acid residue. The remaining 20% was used to determine the radioactivity of the cycle by liquid scintillation counting. Figure 4.4 shows both sequences and the scintillation counting profile of each fraction.

No clear sequence could be determined for fraction 49 nor were any radioactive peaks detected. An aliquot of the resuspended fraction was also monitored by scintillation counting. No radioactivity was found in this fraction (data not shown). It seems that the peptide was either lost during the drying procedure or remained bound to the tube.

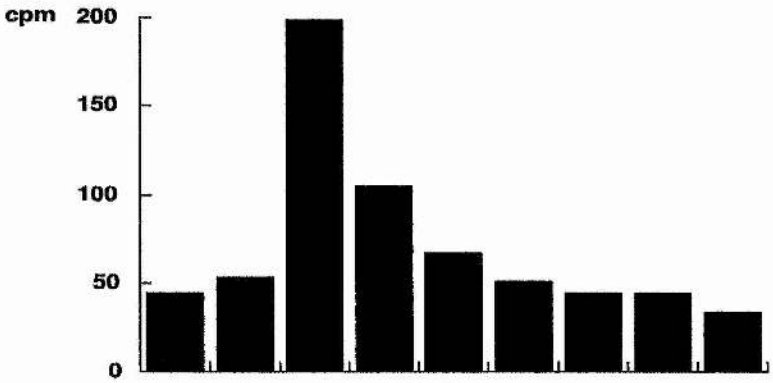
With fraction 73, a sequence could barely be determined, the scintillation counting also revealed a small peak coinciding with a cysteine residue in position 148 in the human NFI.

Using the three other fractions (58, 60 and 61), the sequences could be easily be determined. In fraction 58, the peptide which was sequenced, corresponded to the sequence between isoleucine 93 and arginine 97 in the human NFI. Furthermore another sequence was also obtained which span from the lysine 107 to leucine 113. The radioactivity associated with the third sequencing cycle increased dramatically coinciding with a residue which eluted from the HPLC as a carboxymethyl cysteine residue (cysteine residues are broken down in the Edman reaction). This corresponds to the labelling of the cysteine residue in position 95. In the other sequence, a cysteine residue was expected in the fifth cycle. However, the slight increased in the radioactivity could not account for the precise labelling of this cysteine residue.

Fraction 49



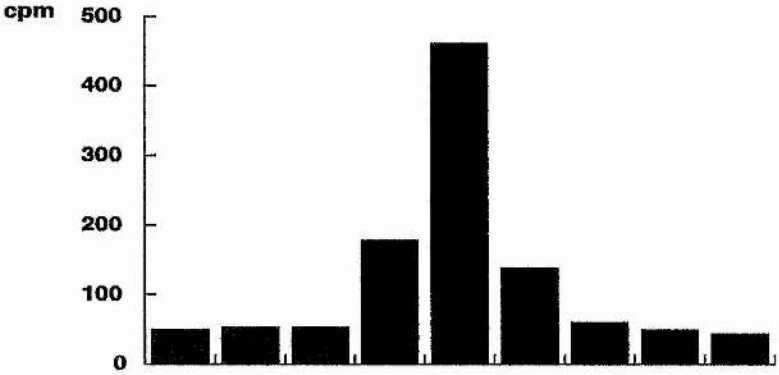
Fraction 58



Sequence 1 : I D Cmc L R 0

Sequence 2 : K A P G Cmc V L

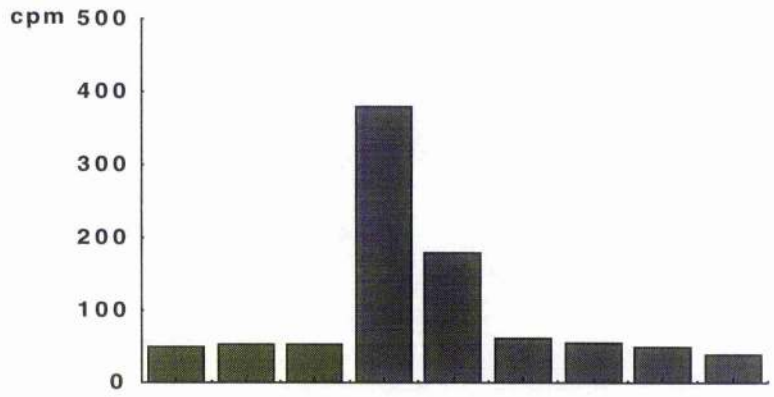
Fraction 60



Sequence 1 : K A P G Cmc V L S 0

Sequence 2 : R I D Cmc L R

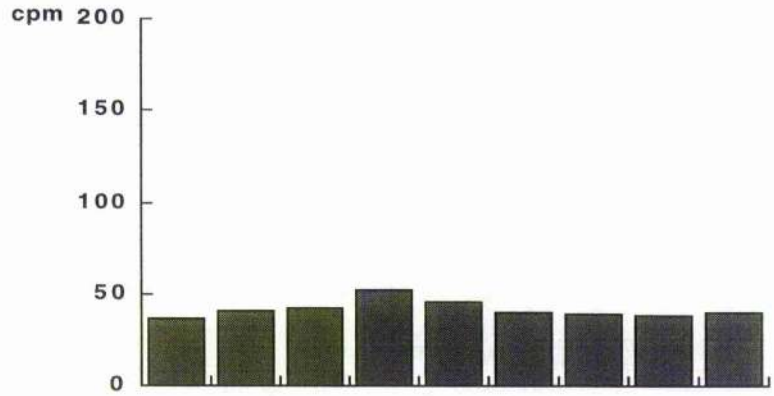
Fraction 61



Sequence 1 : R I D Cmc L R 0

Sequence 2 : K A P G Cmc V L S

Fraction 73



Sequence 1 : A A Q C G H P 0

Figure 4.4 Sequence of the peptides in fraction 49, 58, 60, 61 and 73 purified by reverse phase HPLC.

The fractions were sequenced and each flow through of each sequencing cycle was recovered. The radioactivity of each cycle was determined by scintillation counting for 5 minutes and plotted as above. The result of the sequencing is represented under the plot. The sequence (Sequence 1) is the major sequence whereas the sequence 2 is the minor sequence obtained. 0 represents the background of the scintillation counter

Sequencing of fractions 60 and 61 also revealed in both of them the existence of two different peptides: from lysine 97 to a serine 114 and from an arginine 92 to arginine 97. The first peptide was predominant in fraction 60 while the second one was the main peptide of fraction 61. In fraction 60, two peptides labelled by ^{14}C were detected by scintillation counting: the cysteine modified residue in the minor sequence eluted in the fourth cycle which coincided with an increase of radioactivity. In the fifth cycle, the radioactive signal was further increased due to the presence of a labelled cysteine residue present at that stage in the major sequence. In fraction 61, the radioactivity peaks in each fraction also coincided with carboxymethyl cysteine residues. The cysteine residue in position 95 was clearly labelled. The radioactive labelling of the cysteine residue in position 111 is less obvious as the radioactive peak could be also interpreted as the carry over from the previous fraction. However comparing the fifth and sixth cycle of fraction 60 and 61 does suggest that the cysteine residue in the fifth cycle of fraction 61 is ^{14}C labelled.

Surprisingly this method shows the existence not of one but of two cysteine residues protected from modification by iodoacetate by DNA: cysteine residue in position 95 and 111. The cysteine residue in position 95 did not seem to be labelled more preferentially than the cysteine residue in position 111. These labelled residues accounted for two thirds of the radioactivity present in the modified NFIDBD. Although other cysteine residues seemed to be labelled such as cysteine in position 148, these could only account for about 10% of the total incorporation of radioactivity, hence showing the reactivity of cysteine residues 95 and 111.

Discussion

I. Nuclear factor I: a family of proteins with two distinct functional domains.

Nuclear Factor I is divided into two distinct functional domains: a DNA binding region between residue 1 and 222 and a region at the C terminus involved in the regulation of transcription. This N terminal region alone can bind DNA and stimulate DNA replication of the genome of adenovirus 2 and 5. This is achieved by enhancing the formation of a nucleoprotein preinitiation complex at the origin of replication through a direct interaction with the Ad pol. Although the role of NFI in this replication process is now well established, structural information about the protein is still scarce. Using deletion analysis, Chen *et al*, (1990) found the interaction site between the Adenovirus polymerase and NFI lies between residues 68 and 150. Armentero and co workers (1994) used site directed mutagenesis, to pinpoint the residues: two double mutants at residues cysteine 79/arginine 80 and isoleucine 87/threonine 88 were unable to interact with the Ad pol. However as most mutants did not form active protein, it is possible that other regions are also important for the interaction of NFI with Ad pol.

Two subdomains in the NFI DNA Binding Domain possess different DNA binding activity. A region between residues 1 to 78 binds DNA non specifically. Another region between residues 78 to 222 recognises specifically the NFI DNA biding site (Dekker *et al*, 1996). Further site directed mutagenesis showed that cysteines are involved in the DNA binding (Novak *et al*, 1992). In this work, we attempted to define these regions more precisely and in particular find out which cysteines are important. This was achieved using two techniques: limited proteolysis and differential labelling. However to study NFIDBD by limited proteolysis, it is important to have a protein with a correct molecular weight and a free N terminus.

II. Expression and purification of Nuclear Factor I DNA Binding Domain.

II.1 Expression of Nuclear Factor I

Both expression systems (bacterial and baculovirus) produced very different results. Production of Nuclear Factor I in bacteria enabled us to obtain large quantity of protein (20 mg/l of induced cells) but here NFIDBD formed insoluble and inactive material. The reason why protein forms inclusion bodies is still not understood. It has been found that the tendency of inclusion bodies formation does not correlate with the size or the hydrophobicity of the expressed recombinant protein (Wilkinson *et al*, 1989). Overexpression of a naturally occurring bacterial protein can even lead to aggregates (Gribskov *et al*, 1983). A kinetic model has been proposed where the formation of inclusion bodies would depend on the rate of protein synthesis, the rate of aggregation and the rate of protein folding of the protein being expressed (Kiefhaber *et al*, 1991). This interpretation correlates with the experimental finding that induction carried out under reduced growth condition (extreme pH values, low temperature) and limited level of induction resulted in an increased recovery of soluble material. The solubility of the protein GST-NFIDBD was increased by inducing protein production at a lower temperature, but the effect was not reproducible for NFIDBD. NFIDBD solubility was increased by inducing protein production at low cell density in the media (Dekker *et al*, 1996). Maybe a way to interpret the formation of NFIDBD inclusion bodies in bacteria would be to relate them with results obtained during folding *in vitro*. We showed that low salt concentration, high temperature and high protein concentration are the worst conditions for NFIDBD folding *in vitro*. These conditions are precisely those which prevail in the bacterial cells when protein production is induced. Another possibility which may explain the insolubility of NFIDBD expressed in bacteria is that this protein might be toxic for bacteria. NFIDBD binds tightly to DNA and may cause a problem for

the bacteria to grow or even survive. Therefore only bacteria producing insoluble protein survived after induction by IPTG. However using NFIDBD fused to a periplasmic localisation signal should have solved this problem. The protein with the localisation signal was still insoluble and not exported to the periplasmic region. NFIDBD is probably insoluble as soon as it is synthesised.

Baculovirus system allowed the overproduction of NFIDBD in an active form. The construct which was described in chapter 1, contained added sequences which were not related to NFIDBD sequence: at the N terminus 5 residues from IkB and at the C terminus a sequence from the viral protein P of SV5 were added. This modified NFIDBD (named NFIDBD cPk) construct was soluble, imported to the nucleus of infected insect cells and could be purified in a similar fashion as NFIDBD from other recombinant virus construct (Bosher *et al*, 1990). Therefore we could expect this protein to fold like the native one. Using this system of expression, the yields obtained were generally low (1 to 1.5 mg of purified protein per litre of infected cell). This was much lower than the 10 mg of protein typically expected.

II.2 Expression of Nuclear Factor I as a fusion protein with GST.

The NFIDBD was produced in baculovirus as a fusion protein for a main reason: to avoid post translational modification of the N terminus. NFIDBD cPk was then produced as a fusion protein with GST at its N terminus in infected insect cells. The main difficulty was that the GST-NFIDBD cPk could not be purified by affinity or ion exchange chromatography, despite the fact that several different conditions were used to attempt to destabilise the putative interaction between GST and NFIDBD. Purifications were tried in the presence of various additives (Triton, CHAPS, NDSB) or various NaCl concentration (0.2, 0.4 and 1.5 M). None of these conditions enabled the purification of GST-NFIDBD cPk by affinity chromatography on immobilised glutathione. This suggests that no weak hydrophobic or ionic interaction occurred

between both proteins. Dialysis of the cellular extract also ensured that GST binding to glutathione - agarose was not hindered by the presence of free glutathione present in the insect cell. We can also rule out that GST (and NFIDBD) were mutated as both proteins could be easily purified once separated. This further supported the hypothesis that no strong interaction existed between the two proteins but steric hindrance. Also no frame shift occurred as a monoclonal recognition site at the C terminus was still recognised by the antibody Pk.

The methods set up for purification of NFIDBD were awkward and hard to reproduce. The critical step was the cleavage of the GST-NFIDBD with thrombin in a crude cellular extract. Although the cellular extract was dialysed to remove protease inhibitors, the proteolysis reaction using thrombin was not always successful. One of the critical conditions which was identified, was the salt concentration in the extract. This needed to be at 0.2 M NaCl (data not shown). Other unidentified factors also could affect this step.

II.3 Comparison of NFIDBD proteins produced under different conditions

NFIDBD purified from baculovirus, migrated as a doublet in a SDS PAGE and had an apparent higher molecular weight than NFIDBD purified when fused to GST (figure 1.8). The nature of the post translational modification of this protein is still unclear. However this doublet disappears after treatment with protease (figure 2.18) which suggests that the modification took place at the extreme C terminus. As phosphorylation and glycosylation seem to occur outside the DNA binding domain (Bosher, personal data), this doublet is most likely due to proteolytic degradation. The high apparent molecular weight of this protein compared to the other protein is unclear. It may be just a result of the modification of the N terminus which affects binding of the SDS to the protein. The exact molecular weight would need to be determined by other methods such as mass spectrometry.

III. *In vitro* folding of Nuclear Factor I

III.1 Advantages

Besides the high quantity of protein produced, the formation of inclusion bodies has several advantages: this enabled the production of a potentially toxic protein, protected the protein from proteolytic degradation and made the protein easy to purify. By disrupting the bacterial membrane using lysozyme, Triton and sonication, and after digestion of the DNA, inclusion bodies are easily separated by simple centrifugation from the rest of the cellular proteins. This produces an extract with 4 major proteins. None of the other proteins found in the inclusion bodies were solubilised in 6M guanidine (figure 2.2) except for NFIDBD. No attempts were made to purify further the protein. NFIDBD could have been further purified by gel filtration or for the six histidine Tag protein, by metal immobilised affinity chromatography.

III.2 Optimal conditions for folding

Refolding experiments have described several stages in the folding process. In denatured condition, the protein can be considered as a long flexible chain, however some structural element may still exist. During the folding process, the protein first adopts a loose tertiary structure known as the molten globule. Then secondary structure is acquired. In the final stage the protein compacts and rearranges to obtain the final structure.

As the primary structure of proteins dictates their tertiary structure, the folding pathway may differ between proteins. In chapter 2, we attempted to determine the optimal conditions for NFI folding *in vitro*. These depended on the additive used, the concentration of salt, the concentration of protein and the choice of method used for the renaturing. Other factors were also investigated such as the effect of temperature, pH

and the redox state. The yields obtained were rather low and varied between 2% for folding in the presence of guanidine alone to 8% in the presence of NDSB 256.

III.3 Role of hydrophobic residues

The driving force in protein folding are the hydrophobic interactions. To modulate these interactions, additives were devised. How Non Detergent Sulfobetaine (NDSB) acts in the folding process was studied (Vuillard *et al*, 1998). These molecules have a small hydrophobic region with a hydrophilic tail which prevents NDSB remaining in the hydrophobic core of the folded protein and avoids protein-protein aggregation. The effect of NDSB is to bind to hydrophobic regions transiently thus allowing the protein to fold correctly and avoid the formation of aggregates.

The effect of the various NDSB (of variable hydrophobicity) in folding could be classified into three groups: i) NDSB 195 had a negative influence in the folding yield. ii) NDSB 201 and 223 did not affect the folding yields while iii) NDSB 256 improved the folding yield by 3 to 4 fold. The more hydrophobic NDSB were the most effective. The folding yield increased up to concentration of 0.8 M NDSB.

In vitro folding experiments suggest that influencing the hydrophobic interaction alone could not result in a folding yield over 10%. Using detergents such as CHAPS and dimethyl octyl ammonium propane sulfonate (SB 3-8) at concentrations above and below the CMC (critical micelle concentration) instead of NDSB, the yield could not be increased over 8%. The structure of these additives are similar to NDSB. Their hydrophobic region is much larger causing this compound to form micelles. At the concentration used, these detergents did not affect significantly the folding yield of NFI. Using other detergents which contain a similar hydrophobic region but lack this zwitterionic region such as N-laurylsarcosine, the folding yields were close to zero.

In the optimal conditions set up, the folding yield did not exceed 8%. This suggests that to increase the folding yield further, the formation of the native structure is dependent on interactions other than hydrophobic. This was further supported by the

observation in these optimal conditions, that the unfolded protein remained in solution and did not form any obvious aggregates as generally observed for proteins where hydrophobic patches are exposed.

III.4 The role of charged residues in folding

The charged residues have a significant effect in NFIDBD folding. NFIDBD contains at least 30% of charged residues (38 basic residues and 34 acidic residues). Some observations showed the role of the charged residues in the folding of NFIDBD.

The folding was successful with only protonated basic residues (compare folding yields at pH 9.1 and 10.4). However we did not study the effect of deprotonated acidic residues on the folding yields. Furthermore the folding yields were increased by a high ionic strength which seemed to indicate that the ions in the solution should screen the interaction between charged residues bearing opposite charges. Folding yields in the presence of a non ionic detergent were poor.

The folding is pH dependent hence depends on ionic interaction. A striking fact is that at pH 7.5, NFIDBD can bind to an anion exchange column at a high salt concentration (0.3M NaCl) although the pI is 8.6. The incorrectly folded NFIDBD does not bind the resin at this ionic strength (figure 2.16). The positively charged residues are clustered in a region at the N terminus of NFIDBD and it is most likely that this region interacts with DNA with no sequence preference(Dekker *et al*, 1996). To improve folding yields, one should examine ways to get this region to fold properly as it seems that this region is needed to bind to the ion exchange column. A factor which may help and was shown to increase yield in the folding of a family of DNA binding factor, the AraC (Egan *et al*, 1994), is DNA. DNA would screen and lock in position positive charge residues which salt or guanidine cannot do. Furthermore it may help dimerisation.

III.5 The role of protein concentration in folding of NFIDBD *in vitro*

The final concentration of protein was a major factor in obtaining the optimal folding yield. In our studies the optimal yield was obtained between 0.25 and 0.05 mg/ml. In other proteins, optimal folding yield is obtained at a lower concentration usually between 10 to 50 mg/l (Clark, 1998). For NFIDBD, the effect of the protein concentration on folding seems to be a consequence of two competing reactions. On the one hand, folding at high protein concentration increases the chance of protein aggregation either via hydrophobic or ionic interactions. This effect will be minimised at low protein concentration. On the other hand, it is clear that NFIDBD is only active as a dimer (data not shown, Kruse *et al*, 1994, Armentero *et al*, 1994) and as this is a bimolecular reaction, it will be favoured at high concentration.

III.6 The effect of the methodology, of pH and temperature

Another critical step in the folding *in vitro* of NFIDBD is the way in which guanidine is removed from the solution. Two conventional methods were tried (dialysis and rapid dilution). For some proteins, precipitation is observed when the denatured form is placed in native solvent conditions and therefore gradual removal of the denaturant may be the best way to obtain high yields (Clark, 1998). However by dialysis, the protein is exposed to intermediate denaturant concentration for an extended period of time. At these intermediate concentrations, the protein is likely to exist as a folding intermediate in which exposed hydrophobic regions lead to aggregation. In the case of NFIDBD, folding yields were much higher by rapid dilution than using dialysis (figure 2.12).

The optimal pH and temperature required for the folding *in vitro* of NFIDBD are broad (pH between 7 and 9 when folding was performed at a temperature below 20°C, figures 2.8 and 2.10).

III.7 the role of cysteine residues in the folding

Reduced cysteine residues are important for the DNA binding activity of NFIDBD. Two different reducing agents yielded rather different results. Using β mercaptoethanol, the optimal concentration was 10.5mM while the optimum for DTT was 1.5 mM. The difference in concentration is probably due to the difference in reducing capacity of these two chemicals.

Disulphide bridges and thus a proper oxidative state are important for protein folding i.e. for a secretory protein. Although the cell is in a reducing environment, some protein in cells may contain disulphides bonds important for their activity (such as the Adenovirus protease). Although folding yield of NFI is improved by the presence of a reducing agent, the mutation of any of 3 (out of the 4 conserved cysteine) residues affects proper folding of NFIDBD, suggesting that these residues are important for folding (Novak *et al*, 1992). It would therefore be of interest to attempt folding in an oxidised environment and to test the activity of NFIDBD after reduction of cysteine residues.

III.8 Role of the histidine residues in the folding of NFIDBD

Some evidence suggest that histidines residues are important in the folding of NFIDBD. When NFIDBD contains six histidines at the C terminus, folding yields in guanidine are half those obtained using NFIDBD (figure 2.). Furthermore, the folding yield of NFIDBD at pH 7 was increased 400% compared to pH 6.1 (figure 2.). The pKa of a histidine residue is around 6.1 but may vary depending on the environment indicating that successful refolding requires deprotonated histidine residues. A metal ion could also be involved in the structure of NFIDBD. A putative zinc motif could be determined in the NFIDBD sequence between residues cysteine 148 and histidine 166. Folding in the presence of various metals did not improve the folding yield, however the solutions did not contain EDTA.

III.9 Folding of GST-NFIDBD

Attempts have been made by us and other groups to renature the fusion protein as the interaction of GST with glutathione can be carried out at up to 1.5 M NaCl. The folding yields using GST-NFI were much lower than using NFIDBD. This could partly be due to the fact that the yield were estimated by gel retardation and that GST-NFIDBD from baculovirus binds poorly to DNA.

IV. Analysis of the DNA binding domain of NFI

Direct methods exist which may be employed to define the mechanism by which NFIDBD recognises DNA. This can be done by solving the structure of the DNA protein complex by X-ray crystallography (Nuclear Magnetic resonance is impractical due to the high molecular weight of NFIDBD). However crystallography requires a large amount of pure and homogenous protein (at least 10 mg). Expression of NFI in baculovirus produced several isoforms and it is only by using homogenous *in vitro* folded NFI that we may generate NFI of sufficient purity and quantity to crystallise the protein.

The DNA binding domain of NFI has mainly been studied using site directed mutagenesis. However the problem with introducing mutations in NFI is that mutations could alter the overall structure of the protein. As this seems to happen with most mutants of NFI, a different approach using correctly folded NFI was employed. Experiments were performed *in vitro*, on purified and active NFIDBD to determine which region and residues were important for DNA binding activity.

The first objective was to identify cysteines residues involved in DNA binding, as initial experiments indicated that one, or several cysteine residues were involved. DNA binding of NFI was stimulated in the presence of reducing agents and inhibited by

cysteine modifying reagents. To identify the residues responsible for this effect, we chose iodoacetate labelling. Iodoacetate inhibited DNA binding activity of NFI at a low concentration and was specific for cysteines residues at pH under 8. Furthermore radioactively labelled iodoacetic acid was commercially available.

Using the differential labelling procedure, two cysteine residues were identified (cysteines 95 and 111) in which bound DNA blocks modification by iodoacetate. In other DNA binding proteins, cysteine residues have been shown to play multiple roles in DNA protein interactions: in interacting directly with the DNA structure (with the backbone as seen in the crystal structure of p50 (Ghosh et al, 1995), or interacting with the nucleoside as seen in the crystal structure of the transcription factor AP1 (Glover and Harisson, 1995)). Alternatively cysteine residues can participate in the formation of a structural motif such as the zinc finger or the zinc ribbon. In the zinc finger, four residues (histidines or cysteines) residues coordinate an atom of zinc (Klug and Rhodes, 1987).

In NFI, site directed mutagenesis demonstrated that cysteine 111 did not perturb its DNA binding activity but mutation of this residue resulted in the loss of the redox regulation of NFIDBD. Thus it appears that cysteine 111 is important for the redox regulation of NFI DNA binding activity.

This result is very similar to what was observed with the transcription factor AP1 (Abate et al 1990). The subunits of AP1 contain a cysteine in the region involved in DNA binding activity. When mutated, AP1 could still bind DNA but could not be regulated by changes in the redox state. In fact the extreme sensitivity of this residue to oxidation necessitated the use of a cysteine to serine mutant in the crystallographic analysis (Glover et al, 1995). The serine residue makes a direct contact with the base in the major groove of DNA. In a similar fashion, cysteine 111 in NFI may be susceptible to redox regulation and make direct contact with DNA. Interestingly, mutation of the residues immediately following this cysteine (leucine 112 and arginine 113) generated a protein that could form dimers but could not bind DNA (Armentero et al 1994). These experiments suggest that cysteine 111 is involved in direct contact with DNA.

Another means to regulate the DNA binding activity of these transcriptions factor could be by direct regulation with NO. NO generating compounds such as sodium nitroprusside inactivate the DNA binding activity of transcription factors such as AP1 (Tabuchi *et al*, 1994) and NF kB (Matthews *et al*, 1996). The C62S mutant of p50 was more resistant to inactivation indicating that this cysteine residue was the principal site of NO modification. In preliminary experiments, these NO compounds were also shown to inactivate the DNA binding activity of NFI.

Although the cysteine residue in position 95 is not in any known DNA binding motif of NFI, it is not uncommon to have additional residues making contact with the DNA. This was clearly seen in the crystal structure of NFkB. However this residue is important in the correct folding of the protein as the mutation of this residue results in a misfolded protein.

Dekker et al (1996) identified two domains involved in the interaction of NFIDBD with DNA: a region binding DNA without specificity (residues 1 to 78) and a region which specifically recognised the DNA binding site of NFI (residues 78 to 236). To investigate further the role of these different functional domains, the method of partial proteolysis was employed.

These experiments were performed on functionally active NFIDBD, in the absence or the presence of DNA. By comparing the various proteolytic products, residues which were protected by the presence of DNA could be identified.

The digest products obtained were clearly different due to the addition of DNA. This suggested that the region whose residues were cleaved in the absence of DNA, but not the site of interaction with DNA, were either directly involved in DNA binding or close to the DNA binding site. Following determination of the cleavage sites for most proteolytic fragments, two regions of different proteolytic susceptibility were identified. A region between residues 166 and 175 was degraded in the absence of DNA but protected in the presence of DNA, a region between 181 and 222 was cleaved

in the presence, but not the absence of DNA (figure 3.5). A few other cleavages site were mapped in NFI using chymotrypsin but the protein remained largely resistant to further degradation even when using substantial amounts of protease.

Our results may cast light on previously unexplained observations and may explain the role of this region in DNA binding. Most of the DNA binding region, from residue 2 to 178, is encoded from a single exon. It was thought that this exon encoded for the whole active DNA binding domain. However experimental data from us and other groups showed that overexpression of only the 178 residues of the N terminal of CTF1 produced a protein with a dramatically reduced DNA binding activity compared to the full length NFIDBD (222 residues) or CTF1 (499 residues). One could interpret this result as an incorrectly folded protein.

However, in vivo, the protein NFI B3 was identified which contains only the N terminal region. NFI B3 is the result of an alternative splicing of the intron between exon 2 and 3. For the production of CTF1, this intron is normally spliced (figure 5.1), but for the production of NFI B3, this intron remains in the mRNA. The first codon of this intron encodes for a stop codon. Therefore NFI B3 has the equivalent DNA binding region of the construct presented before, that is the 177 residues of exon 2 and an additional 9 residues at the N terminus. The key point about NFI B3 is that this isoform of NFI can not bind DNA while a fragment of NFIDBD cut after residue 181 still retains DNA binding activity.

Co expression of NFI B3 with CTF1 produced two types of DNA binding proteins: the CTF1 homodimer and the heterodimer CTF1-NFI B3. However CTF1 homodimer had a higher DNA binding affinity than the heterodimer CTF1-NFI B3. Therefore NFI B3 seems to contain all the required elements for proper folding and dimerisation but lacks an essential element for DNA binding. This was surprising as the arginine and lysine rich region and the specific DNA recognition region were both present in this protein. Our result therefore showed that this region between residue 166 and 181 is critical for DNA binding.

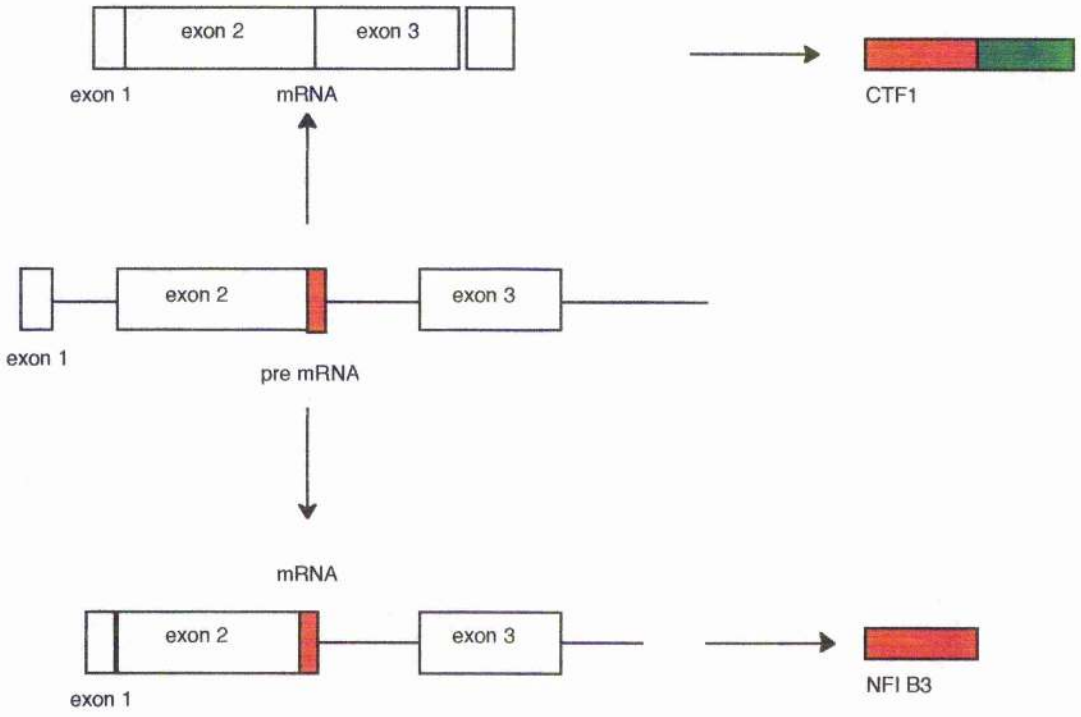


Figure 5.1 Schematic representation of the alternative splicing events occuring in the production of NFI B3 and CTF1 at the 5'end of the pre mRNA encoding for NFI proteins.

The stop codon in the intron between exon 2 and 3 is represented by a red box. Due to the non splicing of the intron between the exon 2 and 3, the mRNA producing for NFI B3, encodes for a protein with the N terminal 186 residues (red box) and does not contain any transcription activation domain. The protein produced can not bind DNA. The removal of this intron in the pre mRNA encodes for a protein such as CTF1 with a transcription activation domain (green box) with DNA binding activity.

This region is very sensitive to proteolytic degradation. This observation suggests that this domain is very exposed in the molecule and is less compact than the rest of the protein. The amino acid sequence also showed that it was composed of two regions: a highly conserved region between residue 166 to 175. Closer to the C terminus, only an acidic residue (usually an aspartic acid residue) in position 179 was conserved among NFI species. The digest pattern also suggests that this region probably exists into two subdomains: a region very well protected by DNA which may interact directly with DNA and a region between 176 and 181 required for the DNA binding but nevertheless sensitive to the protease (i.e. the arginine residues in position 176 and 178). Cleavage of these arginine residues in position 176 and 178 did inactivate the DNA binding of NFI.

In an attempt to determine the role of this region in DNA binding, the amino acid sequence of this region (166 to 181) was tested for homologies with other proteins in the "genbank" database. NFI was the only DNA binding protein possessing this type of sequence. Therefore it is difficult to establish the exact role of this region in DNA binding until structural data is obtained.

One could suggest that the difference of the digestion pattern of NFI in the presence or absence of DNA could be due to a conformational change in the protein. If NFI displayed large conformational changes when binding to DNA all digestion patterns would differ significantly in the presence or absence of DNA. This is obviously not the case, as the digest pattern using trypsin in the presence or absence of DNA, is the same at the C terminus. Furthermore, the digest products C2 and C3 are also present in both sets of condition suggesting that no N terminal conformational change could be detected. Therefore no large scale change of conformation was evident either at the C or at the N terminus of the protein. To explain the difference in the digest pattern with V8 in the presence or absence of DNA (the V2 product is not observed in the digest without DNA), the protease reactivity for both cleavage sites may vary. Once exposed, the glutamic acid in position 166 might be more reactive than the residue in position 209.

V. Conclusion

By using new refolding techniques, we were able to set up a method for obtaining homogenous functional NFI. This was helpful to understand which residues are involved in DNA binding. How Nuclear Factor I binds to DNA and how the domains bind still remains unclear. However our results from the purification, labelling with iodoacetate and partial proteolysis shows that NFI uses several regions to bind DNA. A region rich in positively charged residues was previously shown to interact with DNA, most likely via the phosphate backbone. We showed that another region, located between residue 166 and 181, is crucial in DNA binding. Labelling with iodoacetic acid showed that cysteines 95 and 111 are in close contact with the DNA. This confirms results from site directed mutagenesis showing that the region close to 111 is involved in recognition of the DNA binding site and that the role of cysteine 111 is crucial for redox regulation. It is now becoming apparent the DNA binding activity is located in the whole length of the DNA binding domain of the protein. However without any spatial structure, it is as yet impossible to have any understanding of how this protein folds and therefore the relation between these various regions in DNA binding.

VI. Future Work

To characterise further the residues of NFI involved in DNA binding activity, several footprinting methods have been recently developed which employ the use of an irreversible and reversible modifying chemical. This has the advantage that the protein has to be functional only for the time of the first modifying reaction. Also a hydroxyl

radical cleavage reaction has been developed. These may help to better characterise the various regions involved in the DNA binding.

Obviously the next step would be to solve the structure of NFI. Although this has already been attempted, the production of NFI has mainly relied on using infected cell with a recombinant virus. However the protein in these cells is often modified and results in a heterogeneous population. Therefore the folding in vitro is a very attractive method as the protein is most unlikely to contain post translational modification. Still the folding yield remains low. This seems to be either due to the lack of a factor required for the correct folding such as DNA or a metal. Further research should therefore focus on optimising further the folding yield.

Another method would to refold with the addition of a cytoplasmic extract from HeLa cells. These may contain a factor (such as a chaperonine) required for optimal folding of NFI. If successful, one could therefore attempt to purify this factor and characterise it.

References

R Abate, C., L. Patel, F. J. d. Rauscher and T. Curran (1990).
Redox regulation of fos and jun DNA-binding activity in vitro. Science *249*, p1157-1161

Adhya, S., P. S. Shneidman and J. Hurwitz (1986).
Reconstruction of adenovirus replication origins with a human nuclear factor I binding site. J Biol Chem *261*, p3339-3346

Alonso, C. R., C. G. Pesce and A. R. Kornblihtt (1996).
The CCAAT-binding proteins CP1 and NF-I cooperate with ATF-2 in the transcription of the fibronectin gene. J Biol Chem *271*, p22271-22279

Altmann, H., W. Wendler and E. L. Winnacker (1994).
Transcriptional activation by CTF proteins is mediated by a bipartite low-proline domain. Proc Natl Acad Sci U S A *91*, p3901-3905

Ambinder, R. F., M. A. Mullen, Y. N. Chang, G. S. Hayward and S. D. Hayward (1991).
Functional domains of Epstein-Barr virus nuclear antigen EBNA-1. J Virol *65*, p1466-1478

Amemiya, K., L. Durham and E. O. Major (1994).
Overexpression and reactivation of binding activity of the recombinant CTF/NF-1 transcription factor. Protein Expr Purif *5*, p57-64

Anderson, K. P. and D. F. Klessig (1983).
Posttranscriptional block to synthesis of a human adenovirus capsid protein in abortively infected monkey cells. J Mol Appl Genet *2*, p31-43

Anderson, C. W. (1993).
DNA damage and the DNA-activated protein kinase. Trends Biochem Sci *18*, p433-437

Apt, D., Y. Liu and H. U. Bernard (1994).
Cloning and functional analysis of spliced isoforms of human nuclear factor I-X: interference with transcriptional activation by NFI/CTF in a cell-type specific manner. Nucleic Acids Res *22*, p3825-3833

Armentero, M. T., M. Horwitz and N. Mermod (1994).
Targeting of DNA polymerase to the adenovirus origin of DNA replication by interaction with nuclear factor I. Proc Natl Acad Sci U S A *91*, p11537-11541

Auble, D. T., K. E. Hansen, C. G. Mueller, W. S. Lane, J. Thorner and S. Hahn (1994).
Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. Genes Dev *8*, p1920-1934

Bagby, S., S. Kim, E. Maldonado, K. I. Tong, D. Reinberg and M. Ikura (1995).
Solution structure of the C-terminal core domain of human TFIIB: similarity to cyclin A and interaction with TATA-binding protein. Cell *82*, p857-867

Bandyopadhyay, S. and R. M. Gronostajski (1994).
Identification of a conserved oxidation-sensitive cysteine residue in the NFI family of DNA-binding proteins. J Biol Chem *269*, p29949-29955

Bandyopadhyay, S., D. W. Starke, J. J. Mieyal and R. M. Gronostajski (1998). Thioltransferase (glutaredoxin) reactivates the DNA-binding activity of oxidation-inactivated nuclear factor I. J Biol Chem *273*, p392-397

Bangur, C. S., T. S. Pardee and A. S. Ponticelli (1997).
Mutational analysis of the D1/E1 core helices and the conserved N- terminal region of yeast transcription factor IIB (TFIIB): identification of an N-terminal mutant that stabilizes TATA-binding protein-TFIIB-DNA complexes. Mol Cell Biol *17*, p6784-6793

Bernad, A., L. Blanco, J. M. Lazaro, G. Martin and M. Salas (1989).
A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. Cell *59*, p219-228

Bernard, H. U. and D. Apt (1994).
Transcriptional control and cell type specificity of HPV gene expression. Arch Dermatol *130*, p210-215

Birnboim, H. C. and J. Doly (1979).
A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res *7*, p1513-1523

Blau, J., H. Xiao, S. McCracken, P. O'Hare, J. Greenblatt and D. Bentley (1996).
Three functional classes of transcriptional activation domain. Mol Cell Biol *16*, p2044-2055

Blomquist, P., Q. Li and O. Wrange (1996).
The affinity of nuclear factor 1 for its DNA site is drastically reduced by nucleosome organization irrespective of its rotational or translational position. J Biol Chem *271*, p153-159

Bochkarev, A., J. A. Barwell, R. A. Pfuetzner, E. Bochkareva, L. Frappier and A. M. Edwards (1996).
Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA. Cell *84*, p791-800

Borgmeyer, U., J. Nowock and A. E. Sippel (1984).
The TGGCA-binding protein: a eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. Nucleic Acids Res *12*, p4295-4311

Bosher, J., E. C. Robinson and R. T. Hay (1990).
Interactions between the adenovirus type 2 DNA polymerase and the DNA binding domain of nuclear factor I. New Biol *2*, p1083-1090

Bosher, J., I. R. Leith, S. M. Temperley, M. Wells and R. T. Hay (1991).
The DNA-binding domain of nuclear factor I is sufficient to cooperate with the adenovirus type 2 DNA-binding protein in viral DNA replication. J Gen Virol *72*, p2975-2980

Bosher, J., A. Dawson and R. T. Hay (1992).
Nuclear factor I is specifically targeted to discrete subnuclear sites in adenovirus type 2-infected cells. J Virol *66*, p3140-3150

Bradford, M. M. (1976).
A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, p248-254

Breg, J. N., J. H. van Opheusden, M. J. Burgering, R. Boelens and R. Kaptein (1990).
Structure of Arc repressor in solution: evidence for a family of beta- sheet DNA-binding proteins. Nature *346*, p586-589

Bruggemeier, U., M. Kalff, S. Franke, C. Scheidereit and M. Beato (1991).
Ubiquitous transcription factor OTF-1 mediates induction of the MMTV promoter through synergistic interaction with hormone receptors. Cell *64*, p565-572

Buetti, E. and B. Kuhnel (1986).
Distinct sequence elements involved in the glucocorticoid regulation of the mouse mammary tumor virus promoter identified by linker scanning mutagenesis. J Mol Biol *190*, p379-389

Buratowski, S., S. Hahn, L. Guarente and P. A. Sharp (1989).
Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, p549-561

Burke, T. W. and J. T. Kadonaga (1996).
Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev *10*, p711-724

Candeliere, G. A., P. W. Jurutka, M. R. Haussler and R. St-Arnaud (1996).
A composite element binding the vitamin D receptor, retinoid X receptor alpha, and a member of the CTF/NF-1 family of transcription factors mediates the vitamin D responsiveness of the c-fos promoter. Mol Cell Biol *16*, p584-592

Carlson, M. (1997).
Genetics of transcriptional regulation in yeast: connections to the RNA polymerase II CTD. Annu Rev Cell Dev Biol *13*, p1-23

Cato, A. C., P. Skroch, J. Weinmann, P. Butkeraitis and H. Ponta (1988).
DNA sequences outside the receptor-binding sites differently modulate the responsiveness of the mouse mammary tumour virus promoter to various steroid hormones. Embo J *7*, p1403-1410

Chakraborty, T. and G. C. Das (1991).
Proteins of the nuclear factor-1 family act as an activator of the late promoter in human polyomavirus BK in vitro. J Gen Virol *72*, p1935-1942

Chalepakis, G., J. Arnemann, E. Slater, H. J. Bruller, B. Gross and M. Beato (1988).
Differential gene activation by glucocorticoids and progestins through the hormone regulatory element of mouse mammary tumor virus. Cell *53*, p371-382

Challberg, M. D. and T. J. Kelly, Jr. (1979).
Adenovirus DNA replication in vitro. Proc Natl Acad Sci U S A *76*, p655-659

Challberg, M. D. and T. J. Kelly, Jr. (1981).
Processing of the adenovirus terminal protein. J Virol *38*, p272-277

Challberg, M. D. and D. R. Rawlins (1984).
Template requirements for the initiation of adenovirus DNA replication. Proc Natl Acad Sci U S A *81*, p100-104

Chase, J. W. and K. R. Williams (1986).
Single-stranded DNA binding proteins required for DNA replication. Annu Rev Biochem *55*, p103-136

Chaudhry, A. Z., A. D. Vitullo and R. M. Gronostajski (1998).
Nuclear factor I (NFI) isoforms differentially activate simple versus complex NFI-responsive promoters. J Biol Chem *273*, p18538-18546

Chavez, S. and M. Beato (1997).
Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. Proc Natl Acad Sci U S A *94*, p2885-2890

Chen, M. and M. S. Horwitz (1989).
Dissection of functional domains of adenovirus DNA polymerase by linker- insertion mutagenesis. Proc Natl Acad Sci U S A *86*, p6116-6120

Chen, M., N. Mermod and M. S. Horwitz (1990).
Protein-protein interactions between adenovirus DNA polymerase and nuclear factor I mediate formation of the DNA replication preinitiation complex. J Biol Chem *265*, p18634-18642

Chen, H., M. Ramachandra and R. Padmanabhan (1994).
Biochemical characterization of a temperature-sensitive adenovirus DNA polymerase [published erratum appears in Virology 1995 Feb 20;207(1):342]. Virology *205*, p364-370

Chiang, C. M. and R. G. Roeder (1995).
Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science *267*, p531-536

Choder, M. and R. A. Young (1993).
A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. Mol Cell Biol *13*, p6984-6991

Choy , B. and M. R. Green (1993).
Eukaryotic activators function during multiple steps of preinitiation complex assembly. Nature 366, p531-536

Clark, E. D. B. (1998).
Refolding of recombinant proteins. Curr Opin Biotechnol *9*, p157-163

Cleat, P. H. and R. T. Hay (1989).
Co-operative interactions between NFI and the adenovirus DNA binding protein at the adenovirus origin of replication. Embo J *8*, p1841-1848

Cleghon, V. and D. F. Klessig (1992).
Characterization of the nucleic acid binding region of adenovirus DNA binding protein by partial proteolysis and photochemical cross-linking. J Biol Chem *267*, p17872-17881

Cleghon, V., A. Piderit, D. E. Brough and D. F. Klessig (1993).
Phosphorylation of the adenovirus DNA-binding protein and epitope mapping of monoclonal antibodies against it. Virology *197*, p564-575

Coenjaerts, F. E., E. De Vries, G. J. Pruijn, W. Van Driel, S. M. Bloemers, N. M. Van der Lugt and P. C. Van der Vliet (1991).
Enhancement of DNA replication by transcription factors NFI and NFIII/Oct-1 depends critically on the positions of their binding sites in the adenovirus origin of replication. Biochim Biophys Acta *1090*, p61-69

Coenjaerts, F. E. and P. C. van der Vliet (1994).

Early dissociation of nuclear factor I from the origin during initiation of adenovirus DNA replication studied by origin immobilization. Nucleic Acids Res 22, p5235-5240

Coenjaerts, F. E., J. A. van Oosterhout and P. C. van der Vliet (1994).
The Oct-1 POU domain stimulates adenovirus DNA replication by a direct interaction between the viral precursor terminal protein-DNA polymerase complex and the POU homeodomain. Embo J *13*, p5401-5409

Coleman, R. A., A. K. Taggart, L. R. Benjamin and B. F. Pugh (1995).
Dimerization of the TATA binding protein. J Biol Chem *270*, p13842-13849

Coleman, R. A. and B. F. Pugh (1997).
Slow dimer dissociation of the TATA binding protein dictates the kinetics of DNA binding. Proc Natl Acad Sci U S A *94*, p7221-7226

Conaway, R. C. and J. W. Conaway (1990).
Transcription initiated by RNA polymerase II and purified transcription factors from liver. Transcription factors alpha, beta gamma, and delta promote formation of intermediates in assembly of the functional preinitiation complex. J Biol Chem *265*, p7559-7563

Cordingley, M. G., A. T. Riegel and G. L. Hager (1987).
Steroid-dependent interaction of transcription factors with the inducible promoter of mouse mammary tumor virus in vivo. Cell *48*, p261-270

Cordingley, M. G. and G. L. Hager (1988).
Binding of multiple factors to the MMTV promoter in crude and fractionated nuclear extracts. Nucleic Acids Res *16*, p609-628

Cortes, P., O. Flores and D. Reinberg (1992).
Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. Mol Cell Biol *12*, p413-421

Crawford, D. R., P. Leahy, C. Y. Hu, A. Chaudhry, R. Gronostajski, G. Grossman, J. Woods, P. Hakimi, W. J. Roesler and R. W. Hanson (1998).
Nuclear factor I regulates expression of the gene for phosphoenolpyruvate carboxykinase (GTP). J Biol Chem *273*, p13387-13390

Dahmus, M. E. (1995).
Phosphorylation of the C-terminal domain of RNA polymerase II. Biochim Biophys Acta *1261*, p171-182

Daniell, E. (1976).
Genome structure of incomplete particles of adenovirus. J Virol *19*, p685-708

Davies, A. H., J. B. Jowett and I. M. Jones (1993).
Recombinant baculovirus vectors expressing glutathione-S-transferase fusion proteins. Biotechnology (N Y) *11*, p933-936

de Vries, E., W. van Driel, S. J. van den Heuvel and P. C. van der Vliet (1987).
Contactpoint analysis of the HeLa nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix. Embo J *6*, p161-168

Dekker, N., M. Cox, R. Boelens, C. P. Verrijzer, P. C. van der Vliet and R. Kaptein (1993).
Solution structure of the POU-specific DNA-binding domain of Oct-1. Nature *362*, p852-855

Dekker, J., J. A. van Oosterhout and P. C. van der Vliet (1996).
Two regions within the DNA binding domain of nuclear factor I interact with DNA and stimulate adenovirus DNA replication independently. Mol Cell Biol *16*, p4073-4080

Dekker, J., P. N. Kanellopoulos, A. K. Loonstra, J. A. van Oosterhout, K. Leonard, P. A. Tucker and P. C. van der Vliet (1997).
Multimerization of the adenovirus DNA-binding protein is the driving force for ATP-independent DNA unwinding during strand displacement synthesis. Embo J *16*, p1455-1463

Dekker, J., P. N. Kanellopoulos, J. A. van Oosterhout, G. Stier, P. A. Tucker and P. C. van der Vliet (1998).
ATP-independent DNA unwinding by the adenovirus single-stranded DNA binding protein requires a flexible DNA binding loop. J Mol Biol *277*, p825-838

Dikstein, R., S. Ruppert and R. Tjian (1996).
TAFII250 is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. Cell *84*, p781-790

Drapkin, R. and D. Reinberg (1994).
The multifunctional TFIIH complex and transcriptional control. Trends Biochem Sci *19*, p504-508

Drapkin, R., J. T. Reardon, A. Ansari, J. C. Huang, L. Zawel, K. Ahn, A. Sancar and D. Reinberg (1994).
Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. Nature *368*, p769-772

Drapkin, R., G. Le Roy, H. Cho, S. Akoulitchev and D. Reinberg (1996).
Human cyclin-dependent kinase-activating kinase exists in three distinct complexes. Proc Natl Acad Sci U S A *93*, p6488-6493

Dunsworth-Browne, M., R. E. Schell and A. J. Berk (1980).
Adenovirus terminal protein protects single stranded DNA from digestion by a cellular exonuclease. Nucleic Acids Res *8*, p543-554

Edwards, A. M., C. M. Kane, R. A. Young and R. D. Kornberg (1991).
Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J Biol Chem *266*, p71-75

Egan, S. M. and Schleif R. F. (1994).
DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at rhaBAD. J. Mol. Biol. 243, p821-829

Enomoto, T., J. H. Lichy, J. E. Ikeda and J. Hurwitz (1981).
Adenovirus DNA replication in vitro: purification of the terminal protein in a functional form. Proc Natl Acad Sci U S A *78*, p6779-6783

Fang, S. M. and Z. F. Burton (1996).
RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. J Biol Chem *271*, p11703-11709

Farrell, S., N. Simkovich, Y. Wu, A. Barberis and M. Ptashne (1996).
Gene activation by recruitment of the RNA polymerase II holoenzyme. Genes Dev *10*, p2359-2367

Feaver, W. J., J. Q. Svejstrup, N. L. Henry and R. D. Kornberg (1994).

Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, p1103-1109

Field, J., R. M. Gronostajski and J. Hurwitz (1984).
Properties of the adenovirus DNA polymerase. J Biol Chem *259*, p9487-9495

Flores, O., E. Maldonado and D. Reinberg (1989).
Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. J Biol Chem *264*, p8913-8921

Fowlkes, D. M., S. T. Lord, U. Petterson and L. Philipson (1979).
Interaction between the adenovirus DNA binding protein and double stranded DNA. J. Mol. Biol. 129, p163-180.

Frappier, L. and M. O'Donnell (1991).
Epstein-Barr nuclear antigen 1 mediates a DNA loop within the latent replication origin of Epstein-Barr virus. Proc Natl Acad Sci U S A *88*, p10875-10879

Fredman, J. N. and J. A. Engler (1993).
Adenovirus precursor to terminal protein interacts with the nuclear matrix in vivo and in vitro. J Virol *67*, p3384-3395

Frenkel, G. D. (1978).
Adenovirus DNA synthesis in vitro in an isolated complex. J Virol *25*, p459-463

Friefeld, B. R., M. D. Krevolin and M. S. Horwitz (1983).
Effects of the adenovirus H5ts125 and H5ts107 DNA binding proteins on DNA replication in vitro. Virology *124*, p380-389

Furlong, E. E. M., N. K. Keon, F. D. Thornton, T. Rein and F. Martin (1996).
Expression of a 74-kDa nuclear factor 1 (NF1) protein is induced in mouse mammary gland involution. Involution-enhanced occupation of a twin NF1 binding element in the testosterone-repressed prostate message- 2/clusterin promoter. J Biol Chem *271*, p29688-29697

Gao, B., L. Jiang and G. Kunos (1996).
Transcriptional regulation of alpha(1b) adrenergic receptors (alpha(1b)AR) by nuclear factor 1 (NF1): a decline in the concentration of NF1 correlates with the downregulation of alpha(1b)AR gene expression in regenerating liver. Mol Cell Biol *16*, p5997-6008

Garlatti, M., M. Aggerbeck, J. Bouguet and R. Barouki (1996).
Contribution of a nuclear factor 1 binding site to the glucocorticoid regulation of the cytosolic aspartate aminotransferase gene promoter. J Biol Chem *271*, p32629-32634

Garon, C. F., K. W. Berry and J. A. Rose (1972).
A unique form of terminal redundancy in adenovirus DNA molecules. Proc Natl Acad Sci U S A *69*, p2391-2395

Garon, C. F., R. P. Parr, R. Padmanabhan, I. Roninson, J. W. Garrison and J. A. Rose (1982).
Structural characterization of the adenovirus 18 inverted terminal repetition. Virology *121*, p230-239

Geiger, J. H., S. Hahn, S. Lee and P. B. Sigler (1996).
Crystal structure of the yeast TFIIA/TBP/DNA complex [see comments]. Science *272*, p830-836

Ghosh, G., G. Van Duyne , S. Ghosh and P. B. Sigler (1995).
Structure of NF-kappa B p50 homodimer bound to a kappa B site. Nature 373, p303-310.

Gibson, T. J., J. D. Thompson, A. Blocker and T. Kouzarides (1994).
Evidence for a protein domain superfamily shared by the cyclins, TFIIB and RB/p107. Nucleic Acids Res *22*, p946-952

Gil, G., J. R. Smith, J. L. Goldstein, C. A. Slaughter, K. Orth, M. S. Brown and T. F. Osborne (1988).
Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Proc Natl Acad Sci U S A *85*, p8963-8967

Glover, J. N. and S. C. Harrison (1995).
Crystal structure of the heterodimeric bZIP transcription factor cFos-cJun bound to DNA. Nature 373, p257-261.

Goldberg, M. E., N. Expert-Bezancon, L. Vuillard and T. Rabilloud (1996).
Non-detergent sulphobetaines: a new class of molecules that facilitate in vitro protein renaturation. Fold Des *1*, p21-27

Goodrich, J. A., T. Hoey, C. J. Thut, A. Admon and R. Tjian (1993).
Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. Cell *75*, p519-530

Goodrich, J. A. and R. Tjian (1994).
Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell *77*, p145-156

Gounari, F., R. De Francesco, J. Schmitt, P. van der Vliet, R. Cortese and H. Stunnenberg (1990).
Amino-terminal domain of NF1 binds to DNA as a dimer and activates adenovirus DNA replication. Embo J *9*, p559-566

Goyal, N., J. Knox and R. M. Gronostajski (1990).
Analysis of multiple forms of nuclear factor I in human and murine cell lines. Mol Cell Biol *10*, p1041-1048

Greenblatt, J. (1997).
RNA polymerase II holoenzyme and transcriptional regulation. Curr Opin Cell Biol *9*, p310-319

Gribskov, M. and R. R. Burgess (1983).
Overexpression and purification of the sigma subunit of Escherichia coli RNA polymerase. Gene 26, p109-118.

Gronostajski, R. M., S. Adhya, K. Nagata, R. A. Guggenheimer and J. Hurwitz (1985).
Site-specific DNA binding of nuclear factor I: analyses of cellular binding sites. Mol Cell Biol *5*, p964-971

Gronostajski, R. M. (1986).
Analysis of nuclear factor I binding to DNA using degenerate oligonucleotides. Nucleic Acids Res *14*, p9117-9132

Gronostajski, R. M. (1987).

Site-specific DNA binding of nuclear factor I: effect of the spacer region. Nucleic Acids Res *15*, p5545-5559

Gronostajski, R. M., J. Knox, D. Berry and N. G. Miyamoto (1988).
Stimulation of transcription in vitro by binding sites for nuclear factor I. Nucleic Acids Res *16*, p2087-2098

Gunzburg, W. H. and B. Salmons (1992).
Factors controlling the expression of mouse mammary tumour virus. Biochem J *283*, p625-632

Ha, I., W. S. Lane and D. Reinberg (1991).
Cloning of a human gene encoding the general transcription initiation factor IIB. Nature *352*, p689-695

Ha, I., S. Roberts, E. Maldonado, X. Sun, L. U. Kim, M. Green and D. Reinberg (1993).
Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. Genes Dev *7*, p1021-1032

Hampsey, M. (1998).
Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev *62*, p465-503

Hanke, T., C. Botting, E. A. Green, P. W. Szawlowski, E. Rud and R. E. Randall (1994).
Expression and purification of nonglycosylated SIV proteins, and their use in induction and detection of SIV-specific immune responses. AIDS Res Hum Retroviruses *10*, p665-674

Harrison, S. C. and A. K. Aggarwal (1990).
DNA recognition by proteins with the helix-turn-helix motif. Annu Rev Biochem *59*, p933-969

Hatfield, L. and P. Hearing (1993).
The NFIII/OCT-1 binding site stimulates adenovirus DNA replication in vivo and is functionally redundant with adjacent sequences. J Virol *67*, p3931-3939

Hay, R. T., N. D. Stow and I. M. McDougall (1984).
eplication of adenovirus mini-chromosomes. J Mol Biol *175*, p493-510

Hay, R. T. (1985).
Origin of adenovirus DNA replication. Role of the nuclear factor I binding site in vivo. J Mol Biol *186*, p129-136

Hay, R. T. (1985).
The origin of adenovirus DNA replication: minimal DNA sequence requirement in vivo. Embo J *4*, p421-426

Hay, R. T. and I. M. McDougall (1986).
Viable viruses with deletions in the left inverted terminal repeat define the adenovirus origin of DNA replication. J Gen Virol *67*, p321-332

Hegde, R. S., S. R. Grossman, L. A. Laimins and P. B. Sigler (1992).
Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA- binding domain bound to its DNA target [see comments]. Nature *359*, p505-512

Hennighausen, L., U. Siebenlist, D. Danner, P. Leder, D. Rawlins, P. Rosenfeld and

T. Kelly, Jr. (1985).
High-affinity binding site for a specific nuclear protein in the human IgM gene. Nature *314*, p289-292

Hernandez, N. (1993).
TBP, a universal eukaryotic transcription factor? Genes Dev *7*, p1291-1308

Herr, W., R. A. Sturm, R. G. Clerc, L. M. Corcoran, D. Baltimore, P. A. Sharp, H. A. Ingraham, M. G. Rosenfeld, M. Finney, G. Ruvkun and et al. (1988).
The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes Dev *2*, p1513-1516

Herr, W. and M. A. Cleary (1995).
The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. Genes Dev *9*, p1679-1693

Hilleman, M. R., and J. R. Werner (1954).
Recovery of a new agent from patients with acute respiratory illness. Proc. Soc. Exp. Biol. Med. 85, p183-188.

Hoey, T., B. D. Dynlacht, M. G. Peterson, B. F. Pugh and R. Tjian (1990).
Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. Cell *61*, p1179-1186

Hoffmann, A., C. M. Chiang, T. Oelgeschlager, X. Xie, S. K. Burley, Y. Nakatani and R. G. Roeder (1996).
A histone octamer-like structure within TFIID [see comments]. Nature *380*, p356-359

Hoffmann, A., T. Oelgeschlager and R. G. Roeder (1997).
Considerations of transcriptional control mechanisms: do TFIID-core promoter complexes recapitulate nucleosome-like functions? Proc Natl Acad Sci U S A *94*, p8928-8935

Horwitz, M. S. (1990).
Adenoviridae and their replication. In: Fields, B. N. and D. M. Knipe. Ed. Virology, volume 2, 2nd ed. New York: Raven Press Ltd; 1679-1722.

Ikeda, J. E., T. Enomoto and J. Hurwitz (1982).
Adenoviral protein-primed initiation of DNA chains in vitro. Proc Natl Acad Sci U S A *79*, p2442-2446

Imbalzano, A. N., H. Kwon, M. R. Green and R. E. Kingston (1994).
Facilitated binding of TATA-binding protein to nucleosomal DNA [see comments]. Nature *370*, p481-485

Imbalzano, A. N., K. S. Zaret and R. E. Kingston (1994).
Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. J Biol Chem *269*, p8280-8286

Inostroza, J., O. Flores and D. Reinberg (1991).
Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. J Biol Chem *266*, p9304-9308

Inoue, T., T. Tamura, T. Furuichi and K. Mikoshiba (1990).
Isolation of complementary DNAs encoding a cerebellum-enriched nuclear factor I family that activates transcription from the mouse myelin basic protein promoter. J Biol Chem *265*, p19065-19070

Jackson, S. P. and R. Tjian (1988).
O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell *55*, p125-133

Jackson, S. P. and R. Tjian (1989).
Purification and analysis of RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. Proc Natl Acad Sci U S A *86*, p1781-1785

Jones, K. A., K. R. Yamamoto and R. Tjian (1985).
Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell *42*, p559-572

Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly and R. Tjian (1987).
A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell *48*, p79-89

Joung, I., M. S. Horwitz and J. A. Engler (1991).
Mutagenesis of conserved region I in the DNA polymerase from human adenovirus serotype 2. Virology *184*, p235-241

Joung, I. and J. A. Engler (1992).
Mutations in two cysteine-histidine-rich clusters in adenovirus type 2 DNA polymerase affect DNA binding. J Virol *66*, p5788-5796

Kadonaga, J. T. and R. Tjian (1986).
Affinity purification of sequence-specific DNA binding proteins. Proc Natl Acad Sci U S A *83*, p5889-5893

Kalff, M., B. Gross and M. Beato (1990).
Progesterone receptor stimulates transcription of mouse mammary tumour virus in a cell-free system. Nature *344*, p360-362

Kanellopoulos, P. N., H. van der Zandt, D. Tsernoglou, P. C. van der Vliet and P. A. Tucker (1995).
Crystallization and preliminary X-ray crystallographic studies on the adenovirus ssDNA binding protein in complex with ssDNA. J Struct Biol *115*, p113-116

Kanellopoulos, P. N., D. Tsernoglou, P. C. van der Vliet and P. A. Tucker (1996).
Alternative arrangements of the protein chain are possible for the adenovirus single-stranded DNA binding protein. J Mol Biol *257*, p1-8

Kang, M. E. and M. E. Dahmus (1995).
The unique C-terminal domain of RNA polymerase II and its role in transcription. Adv Enzymol Relat Areas Mol Biol *71*, p41-77

Kaplan, L. M., R. E. Kelinman and M. S. Horwitz (1977).
Replication of adenovirus type 2 DNA in vitro. Proc Natl Acad Sci U S A *74*, p4425-4429

Kedinger, C., O. Brison, F. Perrin and J. Wilhelm (1978).
Structural analysis of viral replicative intermediates isolated from adenovirus type 2-infected HeLa cell nuclei. J Virol *26*, p364-379

Kenny, M. K., L. A. Balogh and J. Hurwitz (1988).
Initiation of adenovirus DNA replication. I. Mechanism of action of a host protein required for replication of adenovirus DNA templates devoid of the terminal protein. J Biol Chem *263*, p9801-9808

Kiefhaber, T., R. Rudolph, H. H. Kohler and J. Buchner (1991).
Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. Biotechnology 9, p825-829.

Killeen, M. T. and J. F. Greenblatt (1992).
The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. Mol Cell Biol *12*, p30-37

Kim, T. K. and R. G. Roeder (1993).
Transcriptional activation in yeast by the proline-rich activation domain of human CTF1. J Biol Chem *268*, p20866-20869

Kim, J. L., D. B. Nikolov and S. K. Burley (1993).
Co-crystal structure of TBP recognizing the minor groove of a TATA element [see comments]. Nature *365*, p520-527

Kim, Y., J. H. Geiger, S. Hahn and P. B. Sigler (1993).
Crystal structure of a yeast TBP/TATA-box complex [see comments]. Nature *365*, p512-520

Kim, T. K. and R. G. Roeder (1994).
Proline-rich activator CTF1 targets the TFIIB assembly step during transcriptional activation. Proc Natl Acad Sci U S A *91*, p4170-4174

King, A. J. and P. C. van der Vliet (1994).
A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends in vitro by a jumping back mechanism. Embo J *13*, p5786-5792

King, A. J., W. R. Teertstra and P. C. van der Vliet (1997).
Dissociation of the protein primer and DNA polymerase after initiation of adenovirus DNA replication. J Biol Chem *272*, p24617-24623

King, A. J., W. R. Teertstra, L. Blanco, M. Salas and P. C. van der Vliet (1997).
Processive proofreading by the adenovirus DNA polymerase. Association with the priming protein reduces exonucleolytic degradation. Nucleic Acids Res *25*, p1745-1752

Kitchingman, G. R. (1985).
Sequence of the DNA-binding protein of a human subgroup E adenovirus (type 4): comparisons with subgroup A (type 12), subgroup B (type 7), and subgroup C (type 5). Virology *146*, p90-101

Kitts, P. A., M. D. Ayres and R. D. Possee (1990).
Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vectors. Nucleic Acids Res *18*, p5667-5672

Klein, C. and K. Struhl (1994).
Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. Science *266*, p280-282

Klemm, J. D., M. A. Rould, R. Aurora, W. Herr and C. O. Pabo (1994).
Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, p21-32

Klug, A. and D. Rhodes (1987).
Zinc fingers: a novel protein fold for nucleic acid recognition. Cold Spring Harb Symp

Quant Biol *52*, p473-482

Knezetic, J. A. and G. Felsenfeld (1993).
Mechanism of developmental regulation of alpha pi, the chicken embryonic alpha-globin gene. Mol Cell Biol *13*, p4632-4639

Krebs, C. J., B. Dey and G. Kumar (1996).
The cerebellum-enriched form of nuclear factor I is functionally different from ubiquitous nuclear factor I in glial-specific promoter regulation. J Neurochem *66*, p1354-1361

Kruijer, W., F. M. van Schaik and J. S. Sussenbach (1981).
Structure and organization of the gene coding for the DNA binding protein of adenovirus type 5. Nucleic Acids Res *9*, p4439-4457

Kruse, U. and A. E. Sippel (1994).
Transcription factor nuclear factor I proteins form stable homo- and heterodimers. FEBS Lett *348*, p46-50

Kumar, K. U., A. Pater and M. M. Pater (1993).
Human JC virus perfect palindromic nuclear factor 1-binding sequences important for glial cell-specific expression in differentiating embryonal carcinoma cells. J Virol *67*, p572-576

Kumar, K. U., L. R. Devireddy, S. C. Tang, A. Pater and M. M. Pater (1996).
Human JC virus nuclear factor 1 binding motifs and large tumor antigen region rquired for transactivation of late promoter. J Neurochem *67*, p473-481

Kusukawa, J., M. Ramachandra, R. Nakano and R. Padmanabhan (1994).
Phosphorylation-dependent interaction of adenovirus preterminal protein with the viral origin of DNA replication. J Biol Chem *269*, p2189-2196

Lechner, R. L. and T. J. Kelly, Jr. (1977).
The structure of replicating adenovirus 2 DNA molecules. Cell *12*, p1007-1020

Leegwater, P. A., W. van Driel and P. C. van der Vliet (1985).
Recognition site of nuclear factor I, a sequence-specific DNA-binding protein from HeLa cells that stimulates adenovirus DNA replication. Embo J *4*, p1515-1521

Leegwater, P. A., P. C. van der Vliet, R. A. Rupp, J. Nowock and A. E. Sippel (1986).
Functional homology between the sequence-specific DNA-binding proteins nuclear factor I from HeLa cells and the TGGCA protein from chicken liver. Embo J *5*, p381-386

Li, S. and J. M. Rosen (1995).
Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. Mol. Cell. Biol. 15, p2063-70

Li, Y. and R. D. Kornberg (1994).
Interplay of positive and negative effectors in function of the C- terminal repeat domain of RNA polymerase II. Proc Natl Acad Sci U S A *91*, p2362-2366

Liao, S. M., J. Zhang, D. A. Jeffery, A. J. Koleske, C. M. Thompson, D. M. Chao, M. Viljoen, H. J. van Vuuren and R. A. Young (1995).
A kinase-cyclin pair in the RNA polymerase II holoenzyme [see comments]. Nature *374*, p193-196

Lichy, J. H., M. S. Horwitz and J. Hurwitz (1981).
Formation of a covalent complex between the 80,000-dalton adenovirus terminal protein and 5'-dCMP in vitro. Proc Natl Acad Sci U S A *78*, p2678-2682

Lichy, J. H., J. Field, M. S. Horwitz and J. Hurwitz (1982).
Separation of the adenovirus terminal protein precursor from its associated DNA polymerase: role of both proteins in the initiation of adenovirus DNA replication. Proc Natl Acad Sci U S A *79*, p5225-5229

Lindenbaum, J. O., J. Field and J. Hurwitz (1986).
The adenovirus DNA binding protein and adenovirus DNA polymerase interact to catalyze elongation of primed DNA templates. J Biol Chem *261*, p10218-10227

Liu, Y., H. U. Bernard and D. Apt (1997).
NFI-B3, a novel transcriptional repressor of the nuclear factor I family, is generated by alternative RNA processing. J Biol Chem *272*, p10739-10745

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997).
Crystal structure of the nucleosome core particle at 2.8 A resolution [see comments]. Nature *389*, p251-260

Ma, D., I. Olave, A. Merino and D. Reinberg (1996).
Separation of the transcriptional coactivator and antirepression functions of transcription factor IIA. Proc Natl Acad Sci U S A *93*, p6583-6588

Makela, T. P., J. D. Parvin, J. Kim, L. J. Huber, P. A. Sharp and R. A. Weinberg (1995).
A kinase-deficient transcription factor TFIIH is functional in basal and activated transcription. Proc Natl Acad Sci U S A *92*, p5174-5178

Martinez, E., Y. Dusserre, W. Wahli and N. Mermod (1991).
Synergistic transcriptional activation by CTF/NF-I and the estrogen receptor involves stabilized interactions with a limiting target factor. Mol Cell Biol *11*, p2937-2945

Matthews, J. R., N. Wakasugi, J. L. Virelizier, J. Yodoi and R. T. Hay (1992).
Thioredoxin regulates the DNA binding activity of NF-kappa B by reduction of a disulphide bond involving cysteine 62. Nucleic Acids Res *20*, p3821-3830

Matthews, J. R., W. Kaszubska, G. Turcatti, T. N. Wells and R. T. Hay (1993).
Role of cysteine62 in DNA recognition by the P50 subunit of NF-kappa B. Nucleic Acids Res *21*, p1727-1734

Matthews, J. R., C. H. Botting, M. Panico, H. R. Morris and R. T. Hay (1996).
Inhibition of NF-kappaB DNA binding by nitric oxide. Nucleic Acids Res *24*, p2236-2242

Maxon, M. E. and R. Tjian (1994).
Transcriptional activity of transcription factor IIE is dependent on zinc binding. Proc Natl Acad Sci U S A *91*, p9529-9533

McCracken, S. and J. Greenblatt (1991).
Related RNA polymerase-binding regions in human RAP30/74 and Escherichia coli sigma 70 [published erratum appears in Science 1992 Mar 6;255(5049):1195]. Science *253*, p900-902

McQuillan, J. J., G. D. Rosen, T. M. Birkenmeier and D. C. Dean (1991).
Identification of a protein that interacts with the nuclear factor-1 (NF-1) binding site in

cells that do not express NF-1: comparison to NF- 1, cellular distribution, and effect on transcription. Nucleic Acids Res *19*, p6627-6631

Meisterernst, M., I. Gander, L. Rogge and E. L. Winnacker (1988).
A quantitative analysis of nuclear factor I/DNA interactions. Nucleic Acids Res *16*, p4419-4435

Meisterernst, M., L. Rogge, R. Foeckler, M. Karaghiosoff and E. L. Winnacker (1989).
Structural and functional organization of a porcine gene coding for nuclear factor I. Biochemistry *28*, p8191-8200

Mellon, I. and P. C. Hanawalt (1989).
Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. Nature *342*, p95-98

Mermod, N., E. A. O'Neill, T. J. Kelly and R. Tjian (1989).
The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell *58*, p741-753

Mink, S., E. Hartig, P. Jennewein, W. Doppler and A. C. Cato (1992).
A mammary cell-specific enhancer in mouse mammary tumor virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor. Mol Cell Biol *12*, p4906-4918

Mittal, V. and N. Hernandez (1997).
Role for the amino-terminal region of human TBP in U6 snRNA transcription. Science *275*, p1136-1140

Mizzen, C. A., X. J. Yang, T. Kokubo, J. E. Brownell, A. J. Bannister, T. Owen-Hughes, J. Workman, L. Wang, S. L. Berger, T. Kouzarides, Y. Nakatani and C. D. Allis (1996).
The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, p1261-1270

Monaghan, A., A. Webster and R. T. Hay (1994).
Adenovirus DNA binding protein: helix destabilising properties. Nucleic Acids Res *22*, p742-748

Monaghan, A. and R. T. Hay (1996).
Pyridoxal 5'-phosphate inhibition of adenovirus DNA polymerase. J Biol Chem *271*, p24242-24248

Moqtaderi, Z., Y. Bai, D. Poon, P. A. Weil and K. Struhl (1996).
TBP-associated factors are not generally required for transcriptional activation in yeast. Nature *383*, p188-191

Morgan, D. O. (1995).
Principles of CDK regulation. Nature *374*, p131-134

Morin, N., C. Delsert and D. F. Klessig (1989).
Nuclear localization of the adenovirus DNA-binding protein: requirement for two signals and complementation during viral infection. Mol Cell Biol *9*, p4372-4380

Mows, C. C., T. Preiss, E. P. Slater, X. Cao, C. P. Verrijzer, P. C. van Der Vliet and M. Beato (1994).
Two independent pathways for transcription from the MMTV promoter. J Steroid Biochem Mol Biol *51*, p21-32

Mul, Y. M., R. T. van Miltenburg, E. De Clercq and P. C. van der Vliet (1989).
Mechanism of inhibition of adenovirus DNA replication by the acyclic nucleoside triphosphate analogue (S)-HPMPApp: influence of the adenovirus DNA binding protein. Nucleic Acids Res *17*, p8917-8929

Mul, Y. M., C. P. Verrijzer and P. C. van der Vliet (1990).
Transcription factors NFI and NFIII/oct-1 function independently, employing different mechanisms to enhance adenovirus DNA replication. J Virol *64*, p5510-5518

Mul, Y. M. and P. C. Van der Vliet (1992).
Nuclear factor I enhances adenovirus DNA replication by increasing the stability of a preinitiation complex. Embo J *11*, p751-760

Mul, Y. M. and P. C. van der Vliet (1993).
The adenovirus DNA binding protein effects the kinetics of DNA replication by a mechanism distinct from NFI or Oct-1. Nucleic Acids Res *21*, p641-647

Nagata, K., R. A. Guggenheimer, T. Enomoto, J. H. Lichy and J. Hurwitz (1982).
Adenovirus DNA replication in vitro: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc Natl Acad Sci U S A *79*, p6438-6442

Nagata, K., R. A. Guggenheimer and J. Hurwitz (1983).
Specific binding of a cellular DNA replication protein to the origin of replication of adenovirus DNA. Proc Natl Acad Sci U S A *80*, p6177-6181

Nagata, K., R. A. Guggenheimer and J. Hurwitz (1983).
Adenovirus DNA replication in vitro: synthesis of full-length DNA with purified proteins. Proc Natl Acad Sci U S A *80*, p4266-4270

Nakajima, N., M. Horikoshi and R. G. Roeder (1988).
Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. Mol Cell Biol *8*, p4028-4040

Nakatani, Y., S. Bagby and M. Ikura (1996).
The histone folds in transcription factor TFIID. J Biol Chem *271*, p6575-6578

Nebl, G., N. Mermod and A. C. Cato (1994).
Post-transcriptional down-regulation of expression of transcription factor NF1 by Ha-ras oncogene. J Biol Chem *269*, p7371-7378

Nikolov, D. B., S. H. Hu, J. Lin, A. Gasch, A. Hoffmann, M. Horikoshi, N. H. Chua, R. G. Roeder and S. K. Burley (1992).
Crystal structure of TFIID TATA-box binding protein [see comments]. Nature *360*, p40-46

Nikolov, D. B., H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder and S. K. Burley (1995).
Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature *377*, p119-128

Nonet, M. L. and R. A. Young (1989).
Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of Saccharomyces cerevisiae RNA polymerase II. Genetics *123*, p715-724

Novak, A., N. Goyal and R. M. Gronostajski (1992).

Four conserved cysteine residues are required for the DNA binding activity of nuclear factor I [published erratum appears in J Biol Chem 1993 Dec 5;268(34):26032]. J Biol Chem *267*, p12986-12990

Nowock, J. and A. E. Sippel (1982).
Specific protein-DNA interaction at four sites flanking the chicken lysozyme gene. Cell *30*, p607-615

Nowock, J., U. Borgmeyer, A. W. Puschel, R. A. Rupp and A. E. Sippel (1985).
The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer. Nucleic Acids Res *13*, p2045-2061

O'Connor, M. and H. U. Bernard (1995).
Oct-1 activates the epithelial-specific enhancer of human papillomavirus type 16 via a synergistic interaction with NFI at a conserved composite regulatory element. Virology *207*, p77-88

O'Neill, E. A., C. Fletcher, C. R. Burrow, N. Heintz, R. G. Roeder and T. J. Kelly (1988).
Transcription factor OTF-1 is functionally identical to the DNA replication factor NF-III. Science *241*, p1210-1213

O'Neill, E. A. and T. J. Kelly (1988).
Purification and characterization of nuclear factor III (origin recognition protein C), a sequence-specific DNA binding protein required for efficient initiation of adenovirus DNA replication. J Biol Chem *263*, p931-937

O'Reilly, D. R., L. K Miller and V. A. Luckow (1992).
Baculovirus Expression Vectors: a laboratory manuel. In: W. H. Freeman and Co. New York.

Ochem, A. E., D. Skopac, M. Costa, T. Rabilloud, L. Vuillard, A. Simoncsits, M. Giacca and A. Falaschi (1997).
Functional properties of the separate subunits of human DNA helicase II/Ku autoantigen. J. Biol. Chem. 272, p29919-29926.

Ohkuma, Y., H. Sumimoto, M. Horikoshi and R. G. Roeder (1990).
Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE. Proc Natl Acad Sci U S A *87*, p9163-9167

Oikarinen, J., A. Hatamochi and B. de Crombrugghe (1987).
Separate binding sites for nuclear factor 1 and a CCAAT DNA binding factor in the mouse alpha 2(I) collagen promoter. J Biol Chem *262*, p11064-11070

Ori, A., D. Atzmony, I. Haviv and Y. Shaul (1994).
An NF1 motif plays a central role in hepatitis B virus enhancer. Virology *204*, p600-608

Orphanides, G., T. Lagrange and D. Reinberg (1996).
The general transcription factors of RNA polymerase II. Genes Dev *10*, p2657-2683

Osada, S., S. Daimon, T. Nishihara and M. Imagawa (1996).
Identification of DNA binding-site preferences for nuclear factor I-A. FEBS Lett *390*, p44-46

Ozer, J., P. A. Moore, A. H. Bolden, A. Lee, C. A. Rosen and P. M. Lieberman (1994).

Molecular cloning of the small (gamma) subunit of human TFIIA reveals functions critical for activated transcription. Genes Dev *8*, p2324-2335

Paonessa, G., F. Gounari, R. Frank and R. Cortese (1988).
Purification of a NF1-like DNA-binding protein from rat liver and cloning of the corresponding cDNA. Embo J *7*, p3115-3123

Parker, C. S. and J. Topol (1984).
A Drosophila RNA polymerase II transcription factor contains a promoter- region-specific DNA-binding activity. Cell *36*, p357-369

Parker, E. J., C. H. Botting, A. Webster and R. T. Hay (1998).
Adenovirus DNA polymerase: domain organisation and interaction with preterminal protein. Nucleic Acids Res *26*, p1240-1247

Pearson, G. D. and P. C. Hanawalt (1971).
Isolation of DNA replication complexes from uninfected and adenovirus- infected HeLa cells. J Mol Biol *62*, p65-80

Pearson, G. D. (1975).
Intermediate in adenovirus type 2 replication. J Virol *16*, p17-26

Pettit, S. C., M. S. Horwitz and J. A. Engler (1988).
Adenovirus preterminal protein synthesized in COS cells from cloned DNA is active in DNA replication in vitro. J Virol *62*, p496-500

Pettit, S. C., M. S. Horwitz and J. A. Engler (1989).
Mutations of the precursor to the terminal protein of adenovirus serotypes 2 and 5. J Virol *63*, p5244-5250

Price, D. H., A. E. Sluder and A. L. Greenleaf (1989).
Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. Mol Cell Biol *9*, p1465-1475

Pronk, R., M. H. Stuiver and P. C. van der Vliet (1992).
Adenovirus DNA replication: the function of the covalently bound terminal protein. Chromosoma *102*, pS39-45

Pronk, R. and P. C. van der Vliet (1993).
The adenovirus terminal protein influences binding of replication proteins and changes the origin structure. Nucleic Acids Res *21*, p2293-2300

Pruijn, G. J., W. van Driel and P. C. van der Vliet (1986).
Nuclear factor III, a novel sequence-specific DNA-binding protein from HeLa cells stimulating adenovirus DNA replication. Nature *322*, p656-659

Pruijn, G. J., H. G. Kusters, F. H. Gmelig Meyling and P. C. van der Vliet (1986).
Inhibition of adenovirus DNA replication in vitro by autoimmune sera. Eur J Biochem *154*, p363-370

Pruijn, G. J., W. van Driel, R. T. van Miltenburg and P. C. van der Vliet (1987).
Promoter and enhancer elements containing a conserved sequence motif are recognized by nuclear factor III, a protein stimulating adenovirus DNA replication. Embo J *6*, p3771-3778

Pruijn, G. J., R. T. van Miltenburg, J. A. Claessens and P. C. van der Vliet (1988).
Interaction between the octamer-binding protein nuclear factor III and the adenovirus origin of DNA replication. J Virol *62*, p3092-3102

Pruijn, J. M., P. C. van der Vliet, N. A. Dathan and I. W. Mattaj (1989).
Anti-OTF-1 antibodies inhibit NFIII stimulation of in vitro adenovirus DNA replication. Nucleic Acids Res *17*, p1845-1863

Puzianowska-Kuznicka, M. and Y. B. Shi (1996).
Nuclear factor I as a potential regulator during postembryonic organ development. J Biol Chem *271*, p6273-6282

Qian, X., C. Jeon, H. Yoon, K. Agarwal and M. A. Weiss (1993).
Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS [published erratum appears in Nature 1995 Jul 20;376(6537):279]. Nature *365*, p277-279

Qian, F., U. Kruse, P. Lichter and A. E. Sippel (1995).
Chromosomal localization of the four genes (NFIA, B, C, and X) for the human transcription factor nuclear factor I by FISH. Genomics *28*, p66-73

Rajas, F., M. Delhase, M. De La Hoya, P. Verdood, J. L. Castrillo and E. L. Hooghe-Peters (1998).
Nuclear factor 1 regulates the distal silencer of the human PIT1/GHF1 gene. Biochem J *333*, p77-84

Ranish, J. A., W. S. Lane and S. Hahn (1992).
Isolation of two genes that encode subunits of the yeast transcription factor IIA. Science *255*, p1127-1129

Rawlins, D. R., P. J. Rosenfeld, R. J. Wides, M. D. Challberg and T. J. Kelly, Jr. (1984).
Structure and function of the adenovirus origin of replication. Cell *37*, p309-319

Rijnders, A. W., B. G. Van Bergen, P. C. Van der Vliet and J. S. Sussenbach (1983).
Immunological characterization of the role of adenovirus terminal protein in viral DNA replication. Virology *131*, p287-295

Rijnders, A. W., B. G. van Bergen, P. C. van der Vliet and J. S. Sussenbach (1983).
Specific binding of the adenovirus terminal protein precursor-DNA polymerase complex to the origin of DNA replication. Nucleic Acids Res *11*, p8777-8789

Robert, F., D. Forget, J. Li, J. Greenblatt and B. Coulombe (1996).
Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. J Biol Chem *271*, p8517-8520

Roberts, M. M., J. L. White, M. G. Grutter and R. M. Burnett (1986).
Three-dimensional structure of the adenovirus major coat protein hexon. Science *232*, p1148-1151

Robinson, A. J., H. B. Younghusband and A. J. Bellett (1973).
A circula DNA-protein complex from adenoviruses. Virology *56*, p54-69

Roovers, D. J., P. F. Overman, X. Q. Chen and J. S. Sussenbach (1991).
Linker mutation scanning of the genes encoding the adenovirus type 5 terminal protein precursor and DNA polymerase. Virology *180*, p273-284

Roovers, D. J., F. M. van der Lee, J. van der Wees and J. S. Sussenbach (1993).
Analysis of the adenovirus type 5 terminal protein precursor and DNA polymerase by

linker insertion mutagenesis. J Virol *67*, p265-276

Rosenfeld, P. J. and T. J. Kelly (1986).
Purification of nuclear factor I by DNA recognition site affinity chromatography. J Biol Chem *261*, p1398-1408

Roth, S. Y. (1995)
Chromatin mediated transcriptional repression in yeast. Curr. Opin. Gen. Dev. 5, p168-173.

Roulet, E., M. T. Armentero, G. Krey, B. Corthesy, C. Dreyer, N. Mermod and W. Wahli (1995).
Regulation of the DNA-binding and transcriptional activities of Xenopus laevis NFI-X by a novel C-terminal domain. Mol Cell Biol *15*, p5552-5562

Rowe, W. P., R. J. Huebner, L. K. Gilmore, R. H. Parrott and T. G. Ward (1953).
Isolation of a cyto pathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc. Soc. Exp. Biol. Med. 84, p570-573.

Rupp, R. A., U. Kruse, G. Multhaup, U. Gobel, K. Beyreuther and A. E. Sippel (1990).
Chicken NFI/TGGCA proteins are encoded by at least three independent genes: NFI-A, NFI-B and NFI-C with homologues in mammalian genomes. Nucleic Acids Res *18*, p2607-2616

Santoro, C., N. Mermod, P. C. Andrews and R. Tjian (1988).
A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature *334*, p218-224

Schaack, J., W. Y. Ho, P. Freimuth and T. Shenk (1990).
Adenovirus terminal protein mediates both nuclear matrix association and efficient transcription of adenovirus DNA. Genes Dev *4*, p1197-1208

Schechter, N. M., W. Davies and C. W. Anderson (1980).
Adenovirus coded deoxyribonucleic acid binding protein. Isolation, physical properties, and effects of proteolytic digestion. Biochemistry *19*, p2802-2810

Scheidereit, C., S. Geisse, H. M. Westphal and M. Beato (1983).
The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus. Nature *304*, p749-752

Schild, C., F. X. Claret, W. Wahli and A. P. Wolffe (1993).
A nucleosome-dependent static loop potentiates estrogen-regulated transcription from the Xenopus vitellogenin B1 promoter in vitro. Embo J *12*, p423-433

Schuur, E. R., U. Kruse, J. S. Iacovoni and P. K. Vogt (1995).
Nuclear factor I interferes with transformation induced by nuclear oncogenes. Cell Growth Differ *6*, p219-227

Sharp, P. A., C. Moore and J. L. Haverty (1976).
The infectivity of adenovirus 5 DNA-protein complex. Virology *75*, p442-456

Shenk, T. (1996).
Adenoviridae: the viruses and their replication. In: Fields, B. N., D. M. Knipe and P. M. Howley, et al. Ed. Virology, 3nd ed. Lippincott: Raven Press Ltd; 979-1016.

Shu, L. M., M. S. Horwitz and J. A. Engler (1987).
Expression of enzymatically active adenovirus DNA polymerase from cloned DNA

requires sequences upstream of the main open reading frame. Virology *161*, p520-526

Siebenlist, U., L. Hennighausen, J. Battey and P. Leder (1984).
Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in Burkitt lymphoma. Cell *37*, p381-391

Smart, J. E. and B. W. Stillman (1982).
Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA. J Biol Chem *257*, p13499-13506

Smith, D. B. and K. S. Johnson (1988).
Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene *67*, p31-40

Sock, E., M. Wegner, E. A. Fortunato and F. Grummt (1993).
Large T-antigen and sequences within the regulatory region of JC virus both contribute to the features of JC virus DNA replication. Virology *197*, p537-548

Spangenberg, C., K. Eisfeld, W. Stunkel, K. Luger, A. Flaus, T. J. Richmond, M. Truss and M. Beato (1998).
The mouse mammary tumour virus promoter positioned on a tetramer of histones H3 and H4 binds nuclear factor 1 and OTF1. J Mol Biol *278*, p725-739

Sprengel, J., B. Schmitz, D. H.-Neitzel, C. Zock and W. Doerfler (1984).
Nucleotide sequence of human type 12 DNA: comparative functional analysis. J. Virol. 68, p379-389.

Stillman, B. W., J. B. Lewis, L. T. Chow, M. B. Mathews and J. E. Smart (1981).
Identification of the gene and mRNA for the adenovirus terminal protein precursor. Cell *23*, p497-508

Stillman, B. W., F. Tamanoi and M. B. Mathews (1982).
Purification of an adenovirus-coded DNA polymerase that is required for initiation of DNA replication. Cell *31*, p613-623

Stow, N. D. (1982).
The infectivity of adenovirus genomes lacking DNA sequences from their left-hand termini. Nucleic Acids Res *10*, p5105-5119

Stunnenberg, H. G., H. Lange, L. Philipson, R. T. van Miltenburg and P. C. van der Vliet (1988).
High expression of functional adenovirus DNA polymerase and precursor terminal protein using recombinant vaccinia virus. Nucleic Acids Res *16*, p2431-2444

Sun, X., D. Ma, M. Sheldon, K. Yeung and D. Reinberg (1994).
Reconstitution of human TFIIA activity from recombinant polypeptides: a role in TFIID-mediated transcription. Genes Dev *8*, p2336-2348

Sun, Z. W. and M. Hampsey (1995).
Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *92*, p3127-3131

Suzuki, M. (1990).
The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA. Nature *344*, p562-565

Suzuki, M. (1993).

Common features in DNA recognition helices of eukaryotic transcription factors [published erratum appears in EMBO J 1993 Oct;12(10):4042]. Embo J *12*, p3221-3226

Tabuchi, A., K. Sano, E. Oh, T. Tsuchiya and M. Tsuda (1994).
Modulation of AP-1 activity by nitric oxide (NO) in vitro: NO-mediated modulation of AP-1. FEBS Lett *351*, p123-127

Taggart, A. K. and B. F. Pugh (1996).
Dimerization of TFIID when not bound to DNA. Science *272*, p1331-1333

Takada, R., Y. Nakatani, A. Hoffmann, T. Kokubo, S. Hasegawa, R. G. Roeder and M. Horikoshi (1992).
Identification of human TFIID components and direct interaction between a 250-kDa polypeptide and the TATA box-binding protein (TFIID tau). Proc Natl Acad Sci U S A *89*, p11809-11813

Tamanoi, F. and B. W. Stillman (1983).
Initiation of adenovirus DNA replication in vitro requires a specific DNA sequence. Proc Natl Acad Sci U S A *80*, p6446-6450

Tan, S., K. P. Garrett, R. C. Conaway and J. W. Conaway (1994).
Cryptic DNA-binding domain in the C terminus of RNA polymerase II general transcription factor RAP30. Proc Natl Acad Sci U S A *91*, p9808-9812

Tan, S., Y. Hunziker, D. F. Sargent and T. J. Richmond (1996).
Crystal structure of a yeast TFIIA/TBP/DNA complex [see comments]. Nature *381*, p127-151

Tanese, N., B. F. Pugh and R. Tjian (1991).
Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev *5*, p2212-2224

Temperley, S. M. and R. T. Hay (1992).
Recognition of the adenovirus type 2 origin of DNA replication by the virally encoded DNA polymerase and preterminal proteins. Embo J *11*, p761-768

Tjian, R. and T. Maniatis (1994).
Transcriptional activation: a complex puzzle with few easy pieces. Cell *77*, p5-8

Tolun, A., P. Alestrom and U. Pettersson (1979).
Sequence of inverted terminal repetitions from different adenoviruses: demonstration of conserved sequences and homology between SA7 termini and SV40 DNA. Cell *17*, p705-713

Trentin , J. J., Y. Yabe and G. Taylor (1962).
The quest for human cancer viruses. Science 137, p835-849.

Tsernoglou, D., A. Tsugita, A. D. Tucker and P. C. van der Vliet (1985).
Characterization of the chymotryptic core of the adenovirus DNA-binding protein. FEBS Lett *188*, p248-252

Tucker, P. A., D. Tsernoglou, A. D. Tucker, F. E. Coenjaerts, H. Leenders and P. C. van der Vliet (1994).
Crystal structure of the adenovirus DNA binding protein reveals a hook- on model for cooperative DNA binding. Embo J *13*, p2994-3002

van Amerongen, H., R. van Grondelle and P. C. van der Vliet (1987).

Interaction between adenovirus DNA-binding protein and single-stranded polynucleotides studied by circular dichroism and ultraviolet absorption. Biochemistry *26*, p4646-4652

van Bergen, B. G. and P. C. van der Vliet (1983).
Temperature-sensitive initiation and elongation of adenovirus DNA replication in vitro with nuclear extracts from H5ts36-, H5ts149-, and H5ts125-infected HeLa cells. J Virol *46*, p642-648

van Bergen, B. G., P. A. van der Ley, W. van Driel, A. D. van Mansfeld and P. C. van der Vliet (1983).
Replication of origin containing adenovirus DNA fragments that do not carry the terminal protein. Nucleic Acids Res *11*, p1975-1989

Van der Vliet, P. C., J. Zandberg and H. S. Jansz (1977).
Evidence for a function of the adenovirus DNA-binding protein in initiation in DNA synthesis as well as in elongation of nascent DNA chains. Virology *80*, p98-110

van der Vliet, P. C. and M. M. Kwant (1978).
Role of DNA polymerase gamma in adenovirus DNA replication. Nature *276*, p532-534

van der Vliet, P. C., W. Keegstra and H. S. Jansz (1978).
Complex formation between the adenovirus type 5 DNA-binding protein and single-stranded DNA. Eur J Biochem *86*, p389-398

Van der Vliet, P. C. (1995).
Adenovirus DNA replication. Curr Top Microbiol Immunol *199*, p1-30

Verrijzer, C. P., A. J. Kal and P. C. Van der Vliet (1990).
The DNA binding domain (POU domain) of transcription factor oct-1 suffices for stimulation of DNA replication. Embo J *9*, p1883-1888

Verrijzer, C. P., J. A. van Oosterhout, W. W. van Weperen and P. C. van der Vliet (1991).
POU proteins bend DNA via the POU-specific domain. Embo J *10*, p3007-3014

Verrijzer, C. P., M. Strating, Y. M. Mul and P. C. van der Vliet (1992).
POU domain transcription factors from different subclasses stimulate adenovirus DNA replication. Nucleic Acids Res *20*, p6369-6375

Verrijzer, C. P., K. Yokomori, J. L. Chen and R. Tjian (1994).
Drosophila TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. Science *264*, p933-941

Vinson, C. R., P. B. Sigler and S. L. McKnight (1989).
Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science *246*, p911-916

Vliet, P. C. v. d. and A. J. Levine (1973).
DNA-binding proteins specific for cells infected by adenovirus. Nat New Biol *246*, p170-174

Vuillard, L., T. Rabilloud and M. E. Goldberg (1998).
Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. Eur J Biochem *256*, p128-135

Wang, T. S., S. W. Wong and D. Korn (1989).

Human DNA polymerase alpha: predicted functional domains and relationships with viral DNA polymerases. Faseb J *3*, p14-21

Wang, Z., S. Buratowski, J. Q. Svejstrup, W. J. Feaver, X. Wu, R. D. Kornberg, T. F. Donahue and E. C. Friedberg (1995).
The yeast TFB1 and SSL1 genes, which encode subunits of transcription factor IIH, are required for nucleotide excision repair and RNA polymerase II transcription. Mol Cell Biol *15*, p2288-2293

Wang, J., A. K. Sattar, C. C. Wang, J. D. Karam, W. H. Konigsberg and T. A. Steitz (1997).
Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. Cell *89*, p1087-1099

Watson, C. J. and R. T. Hay (1990).
Expression of adenovirus type 2 DNA polymerase in insect cells infected with a recombinant baculovirus. Nucleic Acids Res *18*, p1167-1173

Webster, A., R. T. Hay and G. Kemp (1993).
The adenovirus protease is activated by a virus-coded disulphide-linked peptide. Cell *72*, p97-104

Webster, A., I. R. Leith and R. T. Hay (1994).
Activation of adenovirus-coded protease and processing of preterminal protein. J Virol *68*, p7292-7300

Webster, A., I. R. Leith, J. Nicholson, J. Hounsell and R. T. Hay (1997).
Role of preterminal protein processing in adenovirus replication. J Virol *71*, p6381-6389

Webster, A., I. R. Leith and R. T. Hay (1997).
Domain organization of the adenovirus preterminal protein. J Virol *71*, p539-547

Wendler, W., H. Altmann and E. Ludwig-Winnacker (1994).
Transcriptional activation of NFI/CTF1 depends on a sequence motif strongly related to the carboxyterminal domain of RNA polymerase II. Nucleic Acids Res *22*, p2601-2603

Wendler, W. M. F., E. Kremmer, R. Forster and E. L. Winnacker (1997).
Identification of pirin, a novel highly conserved nuclear protein. J Biol Chem *272*, p8482-8489

Wenzelides, S., H. Altmann, W. Wendler and E. L. Winnacker (1996).
CTF5--a new transcriptional activator of the NFI/CTF family. Nucleic Acids Res *24*, p2416-2421

Wolffe, A. P. (1994).
Nucleosome positioning and modification: chromatin structures that potentiate transcription. Trends Biochem Sci *19*, p240-244

Wolfson, J. and D. Dressler (1972).
Adenovirus-2 DNA contains an inverted terminal repetition. Proc Natl Acad Sci U S A *69*, p3054-3057

Wong, M. L. and M. T. Hsu (1988).
Psoralen-cross-linking study of the organization of intracellular adenovirus nucleoprotein complexes. J Virol *62*, p1227-1234

Woychik, N. A. and R. A. Young (1994).
Exploring RNA polymerase II structure and function In: R. C. Conaway and J. W. Conaway, Transcription: mechanisms and regulation. Raven Press, New York, p227-242.

Xiao, H., J. T. Lis, J. Greenblatt and J. D. Friesen (1994).
The upstream activator CTF/NF1 and RNA polymerase II share a common element involved in transcriptional activation. Nucleic Acids Res *22*, p1966-1973

Xie, X., T. Kokubo, S. L. Cohen, U. A. Mirza, A. Hoffmann, B. T. Chait, R. G. Roeder, Y. Nakatani and S. K. Burley (1996).
Structural similarity between TAFs and the heterotetrameric core of the histone octamer [see comments]. Nature *380*, p316-322

Yamada, K., T. Tanaka and T. Noguchi (1997).
Members of the nuclear factor 1 family and hepatocyte nuclear factor 4 bind to overlapping sequences of the L-II element on the rat pyruvate kinase L gene promoter and regulate its expression. Biochem J *324*, p917-925

Yokomori, K., A. Admon, J. A. Goodrich, J. L. Chen and R. Tjian (1993).
Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex. Genes Dev *7*, p2235-2245

Young, R. A. (1991).
RNA polymerase II. Annu Rev Biochem *60*, p689-715

Zhao, L. J. and R. Padmanabhan (1988).
Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. Cell *55*, p1005-1015

Zhao, L. J. and R. Padmanabhan (1991).
Three basic regions in adenovirus DNA polymerase interact differentially depending on the protein context to function as bipartite nuclear localization signals. New Biol *3*, p1074-1088

Zhou, Q., T. G. Boyer and A. J. Berk (1993).
Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. Genes Dev *7*, p180-187

Zhu, W., Q. Zeng, C. M. Colangelo, M. Lewis, M. F. Summers and R. A. Scott (1996).
The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc ribbon [letter]. Nat Struct Biol *3*, p122-124

Zijderveld, D. C., M. H. Stuiver and P. C. van der Vliet (1993).
The adenovirus DNA binding protein enhances intermolecular DNA renaturation but inhibits intramolecular DNA renaturation. Nucleic Acids Res *21*, p2591-2598

Zijderveld, D. C. and P. C. van der Vliet (1994).
Helix-destabilizing properties of the adenovirus DNA-binding protein. J Virol *68*, p1158-1164

Zorbas, H., L. Rogge, M. Meisterernst and E. L. Winnacker (1989).
Hydroxyl radical footprints reveal novel structural features around the NF I binding site in adenovirus DNA. Nucleic Acids Res *17*, p7735-7748

Zorbas, H., T. Rein, A. Krause, K. Hoffmann and E. L. Winnacker (1992).
Nuclear factor I (NF I) binds to an NF I-type site but not to the CCAAT site in the

human alpha-globin gene promoter. J Biol Chem *267*, p8478-8484