CLONING, EXPRESSION, PURIFICATION AND CRYSTALLISATION OF THE HUMAN Ets-1/USF1/DNA COMPLEX
&
CRYSTAL STRUCTURE OF *THERMOTOGA MARITIMA* HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9)

A Dissertation Presented

by

FRANCISCO J FERNÁNDEZ-PÉREZ

Submitted to the Graduate School of the
Universitat Autònoma de Barcelona in partial fulfilment
of the requirements for the degree of

PHILEAS DOCTOR

January 1998 – January 2002

UAB Biochemistry and Molecular Biology Official Graduate Programme
EMBL International PhD Graduate Programme

I herein declare that the dissertation entitled 'Cloning, Expression, Purification and Crystallisation of Human Ets-1/USF1/DNA Complex & Crystal Structure of *Thermotoga maritima* Histidinol-Phosphate Aminotransferase (EC 2.6.1.19)' has been written with only the help stated therein and has not, and will not, be presented to any other faculty.

Francisco J Fernández-Pérez
Hamburg, at February 9, 2002

Para Cristina, con todo mi amor

ACKNOWLEDGMENTS

My most sincere acknowledgements to Matthias Wilmanns, for his support and confidence at all times; to Michael Sieweke, Paul Tucker and Hans Schöler, for their commitment as an excellent thesis advisory committee; to Luis Serrano, who encouraged me to join the EMBL PhD programme; to Enric Querol, for his unshakeable braveness as university tutor; to Xavier Avilés for his assistance and advice; to Josep Biosca and Mari Carmen Martínez, for lots of time and patience; to the people of the EMBL Hamburg Outstation and EMBL Heidelberg (predocs, postdocs, staff, group leaders), for a splendid time together; and, last but not least, to Cristina Vega-Fernández, whose help throughout this thesis was truly invaluable and who generously lent me her time along with plenty of wonderful ideas.

ABSTRACT ON 'CLONING, EXPRESSION, PURIFICATION AND CRYSTALLISATION OF THE HUMAN Ets-1/USF1/DNA COMPLEX'

Ets-1 and USF1 are transcription factors involved in many key processes in development and differentiation. Defects in their regulation have been associated with pathological disorders and viral infections. Furthermore, Ets-1 is a paradigm of a regulation mechanism in transcriptional control whereupon its DNA-binding activity and, thereby, its transactivation properties, are constitutively silenced (autoinhibition). Derepression of Ets-1 can be achieved by protein-protein interactions, which unmask Ets-1's DNA binding domain and lead to the formation of a high affinity ternary complex onto DNA between Ets-1, the partner protein and the combined site for both transcription factors.

In this work, we undertook the structural elucidation of the autoinhibitory mechanism of Ets-1. We started off by cloning, expressing and purifying Ets-1 and the partner protein USF1. As new challenges were met in this part of the project, we devised methods and approaches directed towards the efficient production of DNA-binding proteins for structural biology. A key concept turn out to be the necessity to provide proteins with an environment closest to their biological contexts in order to stabilise them. For example, Ets-1 stabilisation could not be attained until a pre-formed USF1/DNA complex was added to it, resembling the in vivo situation, where USF1, bound to their composite binding site, recruits Ets-1 onto DNA.

We successfully produced a number of Ets-1/USF1/DNA ternary complexes and have partially screened the vast space of crystallisation conditions. Crystals have been obtained, and they have been proved to contain both protein partners plus DNA. Refinement of these preliminary crystallisation conditions will ultimately render crystals of sufficient quality for diffraction experiments. We are excited about the possibilities open up by these crystal for deeper insight into a not well understood regulatory mechanism, but which underlies such important diseases as the human immunodeficiency syndrome and many different types of leukaemia.

RESUMEN DE 'CLONING, EXPRESSION, PURIFICATION AND CRYSTALLISATION OF THE HUMAN Ets-1/USF1/DNA COMPLEX'

Ets-1 y USF1 son factores de transcripción implicados en muchos procesos fundamentales del desarrollo y la diferenciación celular. Defectos en su regulación se han asociado con desórdenes patológicos e infecciones virales. Ets-1, en particular, es un paradigma del mecanismo de regulación transcripcional en virtud del cual la capacidad de Ets-1 para unir DNA y, por tanto, de transactivación, están silenciadas constitutivamente (autoinhibición). La desrepresión de Ets-1 se puede lograr por medio de interacciones proteína-proteína, que desenmascaran el dominio de unión a DNA de Ets-1 y conducen a la formación de un complejo ternario de gran afinidad entre Ets-1, una proteína compañera y el sitio de unión combinado para ambos factores de transcripción.

En este trabajo, emprendimos la elucidación estructural del mecanismo de autoinhibición de Ets-1. Comenzamos con la clonación, expresión y purificación de Ets-1 y la proteína compañera USF1. A medida que íbamos enfrentando nuevos desafíos, diseñamos métodos y aproximaciones que facilitaran la producción eficaz de proteínas de unión a DNA para biología estructural. Uno de los conceptos fundamentales que ha reaparecido en el presente estudio es la necesidad de proporcionar a cada proteína un entorno lo mas parecido posible al que disfrutan en su contexto biológico. Por ejemplo, la estabilización de Ets-1 no fue posible hasta que se añadió el complejo USF1/DNA , que semeja las condiciones fisiológicas en las que USF1, unido a DNA, recluta Ets-1 sobre su sitio de unión.

Hemos producido con éxito un conjunto de complejos ternarios Ets-1/USF1/DNA para los cuales hemos buscado condiciones de cristalización. Hemos obtenido cristales, que hemos probado contienen Ets-1, USF1 y DNA. El refinamiento (optimización) de esas condiciones de cristalización preliminares debería conducir en definitiva a la obtención de cristales de suficiente calidad para difracción. Estos cristales ofrecen la posibilidad de profundizar nuestro conocimiento en un mecanismo de regulación que tiene una importancia capital en enfermedades como el síndrome de inmunodeficiencia adquirida o muchos tipos de leucemias.

ABSTRACT ON 'CRYSTAL STRUCTURE OF *THERMOTOGA MARITIMA* HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9)'

THE WORK EXPOSED HERE IS A RESULT OF A COLLABORATION WITH MC VEGA.

The structure of *Thermotoga maritima* (*tma*) Histidinol-phosphate aminotransferase (HisC) (EC 2.6.1.9) has been solved by X-ray crystallography at 2.85 Å resolution by the multiple anomalous dispersion (MAD) method, as part of an ongoing effort to elucidate the structures and detailed reaction mechanisms of the enzymes of the entire histidine biosynthesis pathway. HisC catalyses the transfer of an amino group from a donor (glutamic acid) to an acceptor (imidazole acetol phosphate) to yield two products, 2-oxo-glutarate and l-histidinol phosphate. In subsequent reactions, l-histidinol phosphate is dephosphorylated and oxidised to l-histidinol, l-histidinal, and, finally, l-histidine.

RESUMEN DE 'CRYSTAL STRUCTURE OF *THERMOTOGA MARITIMA* HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9)'

EL TRABAJO AQUÍ EXPUESTO ES EL RESULTADO DE UNA COLABORACIÓN CON MC VEGA.

La estructura del enzima Histidinol fosfato aminotransferasa (HisC) (EC 2.6.1.9) de *Thermotoga maritima* (*tma*) ha sido resuelta por cristalografía de rayos X a 2.85 Å de resolución por el método de dispersión anómala múltiple (MAD), como parte de un esfuerzo continuado por elucidar la estructura y los mecanismos de reacción detallados de todos los enzimas de la ruta biosintética de la histidina. HisC cataliza la transferencia de un grupo amino de un donante (una molécula de ácido glutámico) a un aceptor (imidazol acetol fosfato) generando 2 productos, 2-oxo-glutarato y fosfato de l-histidinol. En las siguientes reacciones, el fosfato de histidinol es fosforilado y oxidado a l-histidinol, l-histidinal y, finalmente, l-histidina.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	ampere
A	angstrom (10^{-10} m)
ACN	acetonitrile
AI	autoinhibitory
AIEX	anion exchange chromatography
amp	ampicillin
A_x	absorbance at x nm
asu	asymmetric unit
ATP	adenosine triphosphate
βME	β–mercapto ethanol
bp	base pairs
BPB	bromophenol blue
BSA	bovine serum albumin
BT	Bis-Tris, bis(2-hydroxyethyl)imino-tris(hydroxymethyl)methane
CC	correlation coefficient
cDNA	complementary DNA
CHAPS	3-(3-Cholamidopropyl)dimethylamino-1-propanesulphonate
CIEX	cation exchange chromatography
CMC	critical micellar concentration
CoM	centre of mass
crb	carbenicillin
ch	chloramphenicol
*dd*H2O	milliQ water, 18.0 mOhm resistance
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotides triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
E244	Ets-1 244-441
E244ΔS	Ets-1 244-441, C261S, C350S, C416S
E280	Ets-1 280-441
E280ΔS	Ets-1 280-441, C350S, C416S
E301	Ets-1 301-441
E301ΔS	Ets-1 301-441, C350S, C416S
E335	Ets-1 335-441
E335ΔS	Ets-1 335-441, C350S, C416S
EBS	Ets binding site
EDTA	ethylendiaminetetraacetate
eg	latin exampli gratia, for example
EMSA	electrophoretic mobility shift assay
Ets	E-twenty six transcription factor
Fig	figure
fom	figure of merit
FPLC	fast performance liquid chromatography
g	gram
g	acceleration of gravity

GF	gel filtration chromatography
GST	glutathione-S-transferase
GuHCl	guanidinium hydrochloride
h	hour
HEPES	N-2-Hydroxyethyl-piperazine-N-2-ethanesulfonic acid
HPLC	high performance/pressure liquid chromatography
HisC	l-histidinol-phosphate aminotransferase
His6 tag	$(polyhistide)_6$-tag
Hsp	l-histidinol-phosphate
ie	latin id est, that is to say
IEX	ion exchange chromatography
IMAC	immobilised metal chelating chromatography
IPTG	isopropyl-D-beta-galactopyranoside
kan	kanamycin
kb	kilobase
kd	dissociation constant
kDa	kilodalton
krpm	1000x revolutions per minute
laqI	gene encoding the lac repressor
l	litre
LB	Luria-Bertani
LDS	lithium dodecyl sulphate
loc	lack of closure
LU	USF1 194-310
LUΔS	USF1 194-310, C229S, C248S
m	micro (10^{-6})
M	molar
m	milli (10^{-3})
MALDI	matrix assisted laser desorption ionisation
MCS	multicloning site
MES	2-(N-morpholino) ethane sulphonic acid
min	minute
MPD	2-methyl-pentane-2,4-diol
MPE	mean phase error
mRNA	messenger RNA
MS	mass spectrometry
MW	relative molecular mass
MWapp	approximate MW
MWacc	accurate MW
mwco	molecular weight cutoff
NMR	nuclear magnetic resonance spectroscopy
n	nano (10^{-9})
NCS	non-crystallographic symmetry
OD_x	optical density at x nm
orf	open reading frames
ox	oxidised
p	page
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer saline
PCR	polymerase chain reaction

PDB	protein data bank
PEG	polyethyleneglycol
Pfu	*Pyrococcus furiosus*
pI	isoelectric point
PLP	pyridoxal phosphate
Pol	polymerase
PPOW	phasing power
PVDF	polyvinylidene fluoride
PX	protein/macromolecular crystallography
red	reduced
rms	root mean square
rmsd	root mean square deviation
RNA	ribonucleic acid
Rnase	ribonuclease
RPC	reverse phase chromatography
rpm	revolutions per minute
RT	room temperature
s	seconds
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
SeMet	seleno-L-methionine
ssDNA	single-stranded DNA
SU	USF1 194-260
SUΔS	USF1 194-260, C229S, C248S
Taq	*Thermus aquaticus*
TAE	Tris-acetate-EDTA buffer
TB	Tris buffer
TBE	Tris-borate-EDTA buffer
TCA	trichloroacetic acid
TCEP	Tris-chloride-ethanol-phosphine
TE	Tris-EDTA buffer
TEMED	N,N,N,N-tetramethylendiamine
TEN	Tris-EDTA-sodium chloride
TEV	tobacco edge virus
TFE	trifluoroacetic acid
TOF	time of flight
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
USF1	upstream stimulatory factor
UV	ultraviolet
V	volt
vol	volumes
(v/v)	volume per volume
(w/v)	weight per volume
wt	wild-type

TABLE OF CONTENTS

PART ONE – CLONING, EXPRESSION, PURIFICATION AND CRYSTALLISATION OF THE HUMAN Ets-1/USF1/DNA COMPLEX

Goals of Part One

1. Background Information

2. Materials and Methods

3. Results and Discussion

PART TWO – CRYSTAL STRUCTURE OF *THERMOTOGA MARITIMA* HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9)

Goals of Part Two

4. Introduction

5. Molecular Biology Materials and Methods

6. Results and Discussion

Appendix. Vector maps

Part One

Cloning, expression, purification and crystallisation of the human Ets-1/USF1/DNA complex

1. Background Information

1.1 Gene regulation

The fundamental dogma of molecular biology states that DNA produces RNA that in turn translates into proteins. Therefore, if the information contained in the genome (the genotype) is to be converted into proteins, which confer the specific characteristics to an individual (the phenotype), it must first be transcribed into an RNA message. Then the process of transcription, whereby the coding region of a gene serves as template for messenger RNA (mRNA) synthesis, becomes an essential element in gene expression. In the event that transcription failed to occur all the ensuing regulatory steps that affect processing of the RNA message such as RNA splicing (Smith & Valcarcel, 2000), RNA editing (Gott & Emerson, 2000), transport to the cytoplasm (Jansen, 2001), and translation (Sonenberg et al, 2000), would be rendered redundant.

This central role of transcription in the process of gene expression renders it an attractive control point for regulating the expression of genes in particular cell types or in response to specific stimuli. Indeed, it is now clear that in the vast majority of the cases where a protein is expressed in specific patterns or in response to particular stimuli, the underlying control mechanism occurs on the transcriptional level. This ensures that the corresponding gene is transcribed according to that specific pattern or following such stimuli. For example, the genes encoding the immunoglobulin heavy and light chains are transcribed at high levels only in the antibody producing B cells (Garcia et al, 1986), whilst myeloblast differentiation will only occur upon transcription of the muscle specific transcription factor *myoD* (Wei & Paterson, 2001).

Early studies of eukaryotic gene regulation identified specific, short DNA sequences that were present in all genes activated in response to a common signal but absent from genes that were not regulated by that signal (reviewed in Latchman, 1998). Subsequently, it was shown that these sequences (eg the heat shock element or the glucocorticoid response element) stimulated gene transcription by recruiting positively acting (activating) transcription factors (reviewed in Hanna-Rose & Hansen, 1996).

These early studies ushered the idea that transcription factors regulated gene expression in a positive manner in response to specific stimuli or in particular cell types. Subsequently, however, it became clear that many transcription factors could act by inhibiting, as opposed to activating, gene expression. Such repression can occur either via a negatively acting factor (a repressor) interfering with a positively acting factor, or by a direct interaction between the negative factor and the basal transcriptional complex that reduces or fully impairs the activity of the complex (reviewed in Latchman, 1998, and Hanna-Rose & Hansen, 1996). In turn, these studies led to the proposal that the transcription rate of a particular gene, either in a specific cell type or following exposure to a particular stimulus, was regulated by the relative balance of activating and inhibiting transcription factors.

Although this idea is basically correct, the situation has become more complex as more experiments have been conducted. In particular, it is now becoming increasingly accepted that although some transcription factors are pure activators or pure repressors, many can both activate and repress transcription in a manner that is dependent on the particular situation or context (Lefstin & Yamamoto, 1998). Both protein-protein interactions and protein-DNA interactions play an essential role in deciding whether the binding of a transcription factor to its recognition element will lead to the activation or the repression of the downstream promoter. Furthermore, this decision is often made not by one single transcription factor, but by a combination of different transcription factors bound to DNA and to other transcriptional regulators.

It is precisely the vast array of distinct combinations of transcriptional regulators that can bind to enhancers and promoters in a cell-specific fashion that provides an explanation for the complexity of gene expression programmes observed in Nature. This realisation is commonly cited as the 'combinatorial control' model for gene expression and constitutes one of the cornerstones of the complex world of gene regulation.

1.2 Transcriptional control

Some 30 years ago, following the elucidation of transcriptional control in prokaryotes, attention turned to the corresponding problem in eukaryotes: How are so many genes transcribed in a cell type-specific, developmentally regulated manner? The answer lies in two modes of regulation, one involving chromatin, and the other the enzyme that transcribes DNA into messenger RNA, RNA polymerase II. Both modes of regulation have their greatest impact on transcription initiation,

although other control points are possible, such as elongation and termination. Although basic features of the prokaryotic mechanism have been preserved, the demands of eukaryotic control are met by a huge increase in complexity and by the addition of new layers to the transcription apparatus. The term combinatorial control encompasses both modes of regulation under a common theme, that to successfully orchestrate a tightly regulated but extremely versatile network of protein-protein and protein-DNA interactions cells need only provide a limited number of interaction partners and scenarios.

Transcription activation during development or in response to extracellular stimuli is a very complex process dominated by multiprotein complexes that assemble on promoter and enhancer sequences through orderly recruitment. Many of the proteins and protein complexes involved in eukaryotic transcription have been identified, including nucleosomes, polymerases, general transcription factors (such as TBP, the TATA-binding protein), and the regulatory apparatus (transcriptional activators and repressors and co-activators that link the basal factors to the activators and repressors).

1.2.1 Repression by nucleosomes

Coiling of DNA around a histone octamer in the nucleosome is now recognised as an essential mechanism of transcriptional control (Kornberg & Lorch, 1999; Struhl, 1999). Nucleosomes repress all genes except those whose transcription is brought about by specific positive regulatory mechanisms, setting a threshold above which transcriptional control can be observed. In other words, most eukaryotic genes remain silenced by default, and they are transcribed by active mechanisms only at precise times and locations. This general repression mechanism provides a good basis for cell type-specific regulation of the large eukaryote genomes. Furthermore, this mechanism impedes the exponential growth of repressors necessary to extinguish the expression of most genes in the cell.

Nucleosomes repress transcription in at least three different ways. First, they hinder sites of protein binding to DNA, thereby interfering with the interaction of activator and repressor proteins, DNA modifying enzymes and chromatin remodelling factors (Workman & Kingston, 1998). Second, chains of nucleosomes can become further coiled or folded (supercoiled), and thus this higher-order coiling represses transcription of entire chromosomal domains (Ramakrishnan, 1997; Bell & Felsenfeld, 1999). Finally, interactions of nucleosomes with additional chromosomal proteins in heterochromatin suppress gene expression in a hereditary manner (Grunstein, 1998).

The structural basis of repression by nucleosomes lies in the architecture of the histone octamer. Each histone is organised in two domains, a characteristic 'histone fold', and an unstructured amino-terminal 'tail' (Luger et al, 1997). The histone-fold domains constrain the DNA in a central core particle and, thereby, restrict access of DNA-binding proteins. By contrast, the tail extends from the surface of the core particle, providing points of interaction for higher-order coiling, condensation in heterochromatin and gene activation. The histone tails are the targets for a variety of post-translational modifications that includes phosphorylation, methylation, ubiquitination, ADP-ribosylation, and acetylation, the latter being a hallmark of actively transcribed chromatin (Strahl and Allis, 2000). Those modifications that, as acetylation, tend to increase the overall number of negative charges on the histones, can presumably weaken the interactions between the nucleosome and the DNA. As the histone tails lie outside the core particle of the nucleosome and do not contribute to its stability or organisation, acetylation is more likely to affect higher-order chromatin structure. Indeed, histone acetylation is not sufficient for transcriptional activation since it does not disrupt the core particle of the nucleosome, nor does it lead to a dramatic change in nucleosome stability (Hansen et al, 1998) or in the amount of DNA wrapped onto the surface of the histone octamer (Bauer et al, 1994).

The finding of histone acetyltransferase (HAT) activities in previously identified co-activators soon led to the discovery of histone deacetylase (HAD) activities in co-repressors. Subsequently, the idea that deacetylation can re-establish repression emerged from the fact that acetylation can relieve repression by chromatin. Histone deacetylation might also shed light on the long-known relationship between DNA methylation and gene inactivation: Methyl-CpG binding protein, MeCP2A, a protein that binds to methylated DNA, recruits a multiprotein deacetylase complex (Jones et al, 1998; Nan, 1998). Moreover, several studies have shown that inhibition of histone deacetylases can rescue the expression of endogenous genes or reporter constructs that DNA methylation had previously silenced (Ng & Bird, 1999).

The recruitment of chromatin remodelling complexes appears to be one of the earliest steps in transcriptional activation (Fry & Peterson, 2001). An emerging view proposes that transcriptional activators and repressors must harness two types of chromatin remodelling enzymes to the promoter region, an ATP-dependent SWI/SNF-like complex and a histone-acetyltransferase. These two enzymes appear to synergistically collaborate to establish a local chromatin structure that is permissive for the ensuing events. Their combined action would modulate the higher order structure of chromatin, facilitate or occlude the binding of transcription factors, and aid in or prevent the establishment of a transcriptional preinitiation complex.

In human cells, for example, the related co-activators CBP (cAMP-responsive-element-binding binding protein, or CREB) and p300 bind to activator proteins involved in cell cycle control, differentiation, DNA repair, and apoptosis. Both CBP and p300 are HATs, and therefore specifically acetylate the histone tails. Other functionally equivalent protein complexes are human PCAF (p300/CBP-associated factor) (Yang et al, 1996) and yeast SAGA (Spt-Ada-Gcn5-Acetyltransferase) (Grant et al, 1997) and Elongator (Otero et al, 1999).

1.2.2 Gene activator proteins

Most of the attention devoted to eukaryotic transcription regulation at the early stages of the field focused on positive regulatory elements, known as enhancers, and on proteins, termed activators, that bind to them and stimulate transcription. Thus far probing into enhancers by mutagenesis and detection of activators by DNA-binding analysis has revealed many types. Activators are specific for genes or gene families and, typically, couple transcription to the physiological needs of the cell. Activators are often latent and become functional in transcription during a physiological response. For example, the signal transducer and activator of transcription (STAT) proteins (Darnell et al, 1997) transduce signals from growth factor and cytokine receptors, nuclear factor κB (NF-κB) (Ghosh et al, 1998), which mediates immune and inflammatory responses, and nuclear hormone receptors (Mangelsdorf et al, 1995), which interact with a wide array of lipophilic signalling molecules. STATs are phosphorylated by receptor-associated Janus kinases (JAKs), whereupon they dimerise, enter the nucleus and activate transcription. NF-κB is maintained in a 'masked' state in the cytoplasm through its association with an inhibitory subunit, IκB. Phosphorylation of IκB by cell-surface receptors and its removal by proteolysis releases the NF-κB activator in a fully functional form to enter the nucleus. Finally, nuclear hormone receptors can be persistently associated with target genes and require only binding of the hormone to activate transcription.

Activator proteins are (at least to a first approximation, and with exceptions) modular, with distinct DNA binding and activation domains (Ptashne, 1998). This standpoint derived from the general success of 'domain-swap' experiments, where an activation domain or repression domain demonstrates intrinsic activity when delivered to a promoter via a heterologous DNA binding domain, such as that of the yeast GAL4 protein. The DNA binding domains (DBDs) fall into several major families, whose structures are known from crystallographic and nuclear magnetic resonance (NMR) analysis (Harrison, 1991; Luscombe et al, 2000). The affinity and specificity of DNA binding is sometimes augmented, and regulated, by interactions with accessory proteins

bound to adjacent DNA sites (thereby forming what has been term an 'enhanceosome') (Maniatis et al, 1998).

Conversely, activation (also termed transactivation) domains are poorly understood. It is thought that following the assembly of enhancer complexes they contact components of the basal transcription machinery to influence the initiation of the transcription reaction. Many such interactions have been described to date, but the chain of precise contacts leading to transcription remains to be elucidated. Several different types of activation domains have been identified and classified as acidic, glutamine rich, and proline rich. Thus far, the structural relationships and mechanisms of specificity of these different activation domains remain obscure. Indeed, it has become clear that there must exist several functionally distinct subsets of glutamine-rich and acidic activators because not all activation domains within one type aim at the same targets. Furthermore, mutagenesis experiments have shown that the particular residues that are key to activation are not necessarily the predominant residues such as glutamines or acidics (Gill et al, 1994). Instead, in some cases, bulky hydrophobic residues interspersed with glutamines or acidic amino acids seem to be central to the interaction with the transcriptional apparatus. An attractive hypothesis is that the interaction between activation domains and their targets is driven by hydrophobic interaction (much as protein folding is), while the specificity is attained by multiple contacts made by repetitive motifs of amino acids. In this way, an activation domain could establish initial contacts with multiple targets via weak interactions during different steps in the assembly of the initiation complex. These weak contacts may then consequently adapt, or 'fit,' into a more stable complex.

The notion that gene activators can refine the set of protein interactions they engage in leads quite naturally to the concept of allosterism in transcription regulation. Allosterism (Greek, *allos* = other, *stereos* = solid or space) first referred to the regulation mechanism whereupon a small effector molecule binds to an enzyme at a site (the allosteric site) other than the active site (where the catalytic activity occurs) (Monod et al, 1963). The interaction changes the conformation of the enzyme to affect the formation at the active site of the usual complex between the enzyme and its substrate. As a result, the ability of the enzyme to catalyse a reaction is modified. This is the basis of the so-called induced-fit model (Koshland et al, 1966), which states that the binding of a substrate or some other effector to an enzyme may cause conformational changes to enhance or inhibit its activity. In gene regulation, the notion of allosteric control translates in the more general statement that regulatory proteins (transcriptional regulators) perform distinct functions depending on the configuration of the macromolecular surface that they recognise. As transcriptional regulators interact with DNA and DNA-bound proteins in a direct or indirect (via co-activators)

fashion, the possibility becomes very attractive that transcriptional regulators may adapt their structures, and thereby their functions, to the particular DNA sequences they specifically target. Indeed, for those few proteins that have been studied both in the free and bound states, DNA-induced changes in secondary and tertiary structure have been described for a range of eukaryotic transcriptional regulators (Table 1 presents a number of well-documented cases of allosteric control in gene regulation).

In some cases, conformational changes are induced in response to either specific or unspecific DNA (Petersen, 1995; Meierhans et al, 1995), whereas in others they are restricted to, or observed at a greater extent in, specific DNA complexes (Weiss et al, 1990; Shuman et al, 1990; Patel et al, 1990; Epstein et al, 1994; Ikeda et al, 1996). In a few cases, a protein bound to different DNA sequences assumes distinct conformations (Ikeda et al, 1996; Tan & Richmond, 1990; Cleary et al, 1997; Scully et al, 2000; Tomilin et al, 2000). Even subtle structural changes can have long-range functional consequences. For example, the DBD of Ets-1 undergoes little or no detectable change in secondary structure in response to DNA (Werner et al, 1995; Brehm et al, 1997), but in the full-length protein DNA binding induces the dissociation and unfolding of helices far removed from the protein-DNA interface (Petersen et al, 1995; Jonsen et al, 1996). DNA binding can also affect quaternary structure by determining the protein oligomerisation state (Graves et al, 1998) or heterodimerisation patterns.

Little is known about chromatin structure beyond the nucleosome and its involvement in transcriptional regulation. It might orchestrate the coordinate regulation of multiple genes within large chromosomal domains. DNA elements responsible for such regulation include 'locus control regions' (Kornberg & Lorch, 1999), which resemble enhancers, and 'insulators' (Bell & Felsenfeld, 1999), which define the boundaries of a domain. Beyond their characterisation by molecular genetics, the mode of action of these elements remains vague. One of the many issues that call for deeper insight into the mechanisms of chromatin remodelling is the epigenetic regulation of gene expression (eg gene silencing by heterochromatin).

1.2.3 Mediator

Although interactions between activators and basal transcription factors may contribute to the process of activation, they are clearly not sufficient. This conclusion was initially based on the finding that stimulation of *in vitro* transcription by activators could be detected with partially purified TFIID, but not with purified TBP (Pugh & Tjian, 1990). Evidently, some other factors had

co-purified together with TFIID that were required for activation of transcription. Subsequently, TFIID was found to exist as a remarkably stable complex composed of TBP and at least eight TBP-associated factors (TAFs).

The identification of all eight TAFs and the demonstration that they can reconstitute activation *in vitro* led to the co-activator hypothesis (Pugh & Tjian, 1990; Dynlacht et al, 1991). This hypothesis poses that at least some of the subunits of TFIID serve to link transcription activation domains with the basal transcriptional complex. Those subunits, and the multiprotein complex that comprises them, were termed co-activators. Since there are at least eight different TAFs, with the possibility of additional substoichiometric and perhaps tissue-specific variants, it was proposed that different classes of activators might interact directly with distinct TAFs.

Following a more recent trend, we can define co-activators operationally as factors essential for the function of DNA binding activators, but not for basal (core promoter) transcription, by a minimal set of general initiation factors. They include factors intimately associated with the general transcriptional machinery factors that are recruited to the promoter via interactions with DNA bound activators. Examples of general transcription factors are the TAF components of TFIID, which on TATA-less promoters also serve as basal factors, and hSRB/MED factors associated with RNA polymerase; examples of activators include the Oct-1 associated B cell-specific OCA-B, the thyroid hormone receptor-associated TRAPs and p300/CBP, and more general co-activators isolated from the USA fraction (eg PC2 and PC4). Furthermore, negative cofactors array in a similar classification and may enhance the fold induction through selective inhibitory effects on basal transcription.

As a natural extension of the concept of co-activator, the term 'Mediator' was coined as a comprehensive description of the host of co-activators and accessory proteins that associate in large multisubunit complexes and provide the functional and structural link between activators and the transcription machinery (Flanagan et al, 1991). The first proof of Mediator's existence came from a series of elegant experiments conducted in yeast to isolate multiprotein complexes bound to or associated with the RNA Polymerase II. A 20-subunit complex could be isolated, and it was later shown to interact with Polymerase II to constitute a 'holoenzyme' complex (Thompson et al, 1993; Kim et al, 1994). The identification of two-thirds of Mediator subunits as products of genes previously identified in various screens for mutations that affect transcription control in yeast (Myers et al, 1998) corroborated the wide significance of Mediator for *in vivo* regulation. The

genetically defined subunits include the suppressor of RNA Polymerase B (SRB) proteins, one of which, when mutated, abolishes almost all transcription *in vivo* (Thompson & Young, 1995).

Protein	DNA-binding domain motif	Conformational change	Method
C/EBP	BZIP	Proteolytic resistance	Protease digestion
Ets-1	WHTH	Proteolytic sensitivity, helix destabilisation	Protease digestion, CD[b]
c-Myb	HTH	Helix induction/stabilisation, proteolytic resistance	NMR[b], CD, protease digestion
NF-κB	Rel	Proteolytic resistance, proteolytic sensitivity, altered CD spectrum	Protease digestion, CD
Steroid receptors (SR)	C4 zinc-finger	Helix induction/stabilisation, loop reorientation	NMR, X-ray[b]
Thyroid receptors (TR)	C4 zinc-finger	Proteolytic resistance, altered CD spectrum	Protease digestion, CD

Table 1.1 DNA-induced conformational changes in eukaryotic transcriptional regulators. Well-known examples of conformation changes in eukaryotic transcription factors can be found in the following references: (42, 56) for C/EBP; (39) for Ets-1; (57-59), (60, 61) for c-Myb; (62-64) for NF-κB; (65-70) for the family of steroid receptors; and (45, 71) for the family of thyroid receptors, and references thereof. [b]CD, circular dichroism, NMR, nuclear magnetic resonance, X-ray, macromolecular X-ray crystallography.

Mediator has been recently isolated from mouse and human cells, demonstrating the relevance of the Mediator mechanism in higher eukaryotes (Ito et al, 1999, and references therein). These complexes vary to some extent in subunit composition, but recent evidence emphasises their similarity. The human Mediator complex had been isolated before by virtue of its association with liganded, but not with free, nuclear receptors, providing direct evidence of activator-Mediator interaction (Fondell et al, 1996). Mediator is now recognised as a crucial interface between activators and RNA Polymerase II. It transduces information from enhancers to promoters in organisms ranging from yeast to human. However, a gamut of issues remains unanswered: What is the relative importance of activator-Mediator and activator-basal transcription factor interactions? Are different activators targeted for different subunits of Mediator, and is its subunit composition constant or vary among cell types? How do some subunits of Mediator act in both positive and negative regulation, as shown by genetic analysis in yeast? No doubt, future studies on the structure and function of Mediator will further illuminate our current understanding of transcription regulation in eukaryotes.

1.3 Blood cell differentiation

The haematopoietic system is a paradigm for the development of different specialised cell types from pluripotent stem cells. How the cell progenitors select and maintain specific gene expression programmes is key to understanding cell differentiation. Gene-inactivation studies, promoter analysis and ectopic expression of lineage-restricted factors (Cross & Enver, 1997; Orkin, 1996; Shivdasani & Orkin, 1996) have highlighted the important role of transcription factors in this complex process. However, it has not been demonstrated that a single lineage-specific master regulator exists that can activate individual patterns of gene expression. Rather, such cell-specific regulation appears to depend on a combinatorial array of transcription factors with overlapping tissue specificities, which cooperate to define individual cell types (Ness & Engel, 1994).

The complex patch of transcription factor binding sites in tissue-specific enhancer/promoter regions are generally thought to recruit subsets of binding activities from the complement of factors expressed in a particular cell type (Arnone & Davidson, 1997). Until very recently, the prevailing view considered that individual enhancer-bound regulators serve to promote cooperative assembly on the DNA (Carey, 1998, and references therein). Recent observations, however, demonstrate that major functions of typical transcription factors do not require DNA binding (Reichardt et al, 1998; Porcher et al, 1999), suggesting that the protein interaction potential of transcription factors carries information beyond the stabilisation of DNA contacts.

An alternative model, the 'cocktail party' model (Sieweke & Graf, 1998), accounts for these recent observations by proposing that combinations of transcription factors can associate independently of DNA into highly dynamic assemblies. These multiprotein complexes may exchange components and therefore exhibit distinct specificities by the selective association of factors. It is the dynamic exchange of factors in and out of these complexes that would determine cell fate decisions, or more plainly, what gene expression programme a certain group of cell progenitors would pursue. Such complexes would therefore provide a foundation for the high number of activities and specificities observed, a prerequisite for a model aiming to explain how cell differentiation routes are established. An example taken from the myeloid compartment can help illustrate the cocktail party model.

1.3.1 The myeloid compartment

The myeloid compartment offers a wealth of well-documented protein interactions. The analysis of a great number of these indicates that they are neither random nor limited to pairs, a prerequisite for

the assembly of large yet specific complexes. Moreover, the composition of transcription factors in these complexes affects not only differentiation stages but also lineage choice, through the activation of particular sets of genes or the silencing of others.

As illustrated in Fig 1.1, most of the interactions identified involve multiple partners, and yet they are highly selective. For example, the winged helix-turn-helix transcription factor Ets-1 and the bZIP factors C/EBPα and C/EBPβ are amidst the most promiscuous in the myeloid compartment (Sieweke et al, 1998; McNagny et al, 1998). Despite the apparent complexity of these interactions, there is a clear selectivity. Thus, within the bZIP family of transcription factors, Ets-1 binds to c-Jun, Maf and C/EBP proteins but not to c-Fos (McNagny et al, 1998), and within the large loop-helix-loop (HLH) family it has only been found to associate with USF-1 and TFE-3 (Sieweke et al, 1998). Likewise, the partner molecules of Ets-1 do not bind all ETS family members but a well-defined subset. In a limited number of cases, it has been possible to show that different partners bind to distinct interaction epitopes. For example, USF-1 and the runt domain protein AML-1 target different regions within the Ets-1 molecule (Kim et al, 1993; Sieweke et al, 1996). Whether such interactions can occur simultaneously within the same complex or successively in different complexes remains to be determined. To add to this complexity, myeloid transcription factors appear to interact with one another indirectly through the co-activator CBP/p300. For example, activation of the promoters for the myeloid genes mim-1 and elastase is mediated through the complex formed between c-Myb and C/EBPα (or β) through bridging by CBP/p300 (Oelgeschlager et al, 1996). Similarly, c-Myb and Ets-1 seem to interact with one another through CBP/p300 on the early myeloid promoter CD13 (Shapiro, 1995), and c-Myb and AML-1 on the myeloid peroxidase promoter (Britos-Bray et al, 1997).

1.3.2 A cocktail party scenario

As reviewed above, the cocktail party model provides an outline of the molecular mechanisms that may underpin cell differentiation. According to it, multiple but selective interactions between transcription factors allow the assembly of multicomponent complexes and dynamic shifts in subunit composition during cell differentiation. Furthermore, the cocktail party model sets a scenario where early progenitors contain an initial protein complex whose components are successively exchanged and replaced, resulting in complexes with new enhancer-binding affinities and/or functional properties that specify different cell types and differentiation stages. These complexes not only acquire the capacity to activate novel enhancers but might also repress enhancers that are no longer compatible with a given gene expression programme. As it is easier to

make the transition between complexes that share most of their subunits than between complexes that are widely different in composition, the model accommodates considerable flexibility between adjacent differentiation stages but implies a higher resistance between more distant stages. Furthermore, this model explains why a particular transcription factor can activate one type of enhancer but not another and why inducing the expression of a rate-limiting factor can activate a novel gene expression programme while silencing another. Such mechanisms may underlie, for example, the repression of erythroid specific genes by MafB (Sieweke et al, 1996), myeloid-specific genes by GATA-1 (Kulessa et al, 1995) and MEP-specific genes by C/EBP (Nerlov et al, 1998).

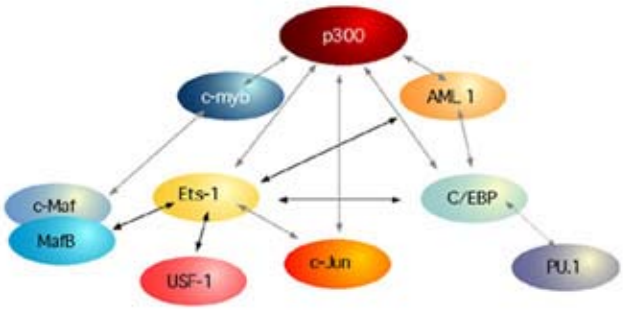

Fig 1.1 Interrelations between transcription factors in the myeloid lineage. The vast array of physical protein interactions that occur among myeloid factors relates them through multiple, but selective, partnerships. The co-activator CBP/p300 (in dark red, at the top) can mediate interactions by binding the different partners through distinct epitopes. The darker arrows point to Ets-1, USF-1, MafB and AML-1.

1.4 Retrovirus

Retroviruses constitute a family of RNA viruses that code for the enzyme reverse transcriptase, which transcribes the viral genomic RNA into a double-stranded DNA copy that ultimately integrates in the host genome. Subsequently, the provirus utilises the cell's transcriptional machinery to replicate (or transcribe) its genome, which will then be translated into viral proteins (chiefly, the structural proteins Gag, Env and Pol). A retrovirus' infective cycle ends by packaging its genome into the structural proteins of the capsid and budding off the cells in a non-lytic fashion.

The human immunodeficiency virus 1 (HIV-1) belongs to the retrovirus family. It encodes a number of regulatory and accessory proteins in addition to the characteristic *gag*, *pol* and *env* gene products common to all retroviruses. The various gene products are expressed from a set of HIV-1 specific mRNAs that are produced by alternative splicing of a single genome length transcript.

These mRNAs fall into two classes: the small multiply spliced mRNA that encode the regulatory proteins, Tat, Rev and Nef, and the singly and unspliced mRNAs that encode the structural proteins Gag, Pol and Env. This additional complexity conferred to the HIV-1 genome by the set of regulatory proteins endows it with enhanced resistance against the immune system and current therapies.

HIV-1 predominantly replicates in macrophages and helper T cells. Although this largely reflects the presence of the CD4 receptor and specific co-receptors on these cell types, viral replication requires the transcriptional activation of the integrated provirus and thus is affected by particular combinations of tissue-specific transcription factors. Several cellular transcription factors are known to regulate the promoter/enhancer located within the U3 region of the proviral 5' long terminal repeat (LTR, reviewed in (Jones & Peterlin, 1994). For example, in fully activated T cells the core promoter (TATA and Sp-1) and enhancer (NF-κB) are all that is required for high level HIV-1 gene expression. However, although these elements are sufficient for transcriptional activity, HIV-1 replication does require the integrity of regions further downstream. Thus, mutations in the LTR region -130 to -166 bp completely prevent virus replication (Kim et al, 1993). The DNA sequence in this region contains consensus binding sites for HLH proteins (E-box) and ETS proteins (EBS), as well as for the DNA bending transcription factor LEF-1 (lymphoid enhancer factor-1). Both Ets-1 and LEF-1 are highly expressed in T cells, and *in vitro* transcription assays with reconstituted chromatin revealed that these two factors in conjunction with Sp-1 could relieve nucleosomal repression of the HIV-1 LTR (Sheridan et al, 1995).

All these results underscore the leading role of cellular transcription factors in the active replication and transcription of the HIV-1 provirus, and suggest the possibility of alternative antiviral therapies directed to host proteins rather than to viral gene products. For example, a recent study (Posada et al, 2000) has demonstrated that mice carrying a dominant-negative mutant of Ets-1 are immune to HIV-1 infection, stressing two facts: First, that host proteins are indeed absolutely required for HIV-1 infection, and second, that targeting those cellular factors constitutes a promising strategy to design enhanced, more efficient antiviral drugs.

1.5 Ets-1 and USF1 as model proteins in transcription regulation

1.5.1 Ets-1 and the mechanism of autoinhibition

Ets-1 is the founder member of an expanding family of transcription factors, the ETS family, which currently encompasses nearly forty members (MacLeod et al, 1992). Ets-1 was first described based on a region of primary sequence homology with the protein product of the *v-ets* oncogene encoded by the E26 (E twenty six) avian erythroblastosis virus (Karim et al, 1990). The conserved region, termed the ETS domain, corresponds to the DNA-binding domain (DBD) (Fig 1.2). The ETS domain folds into a 'winged' helix-loop-helix of 80 amino acids, which binds as a monomer to a 10-bp DNA sequence centred about the core sequence `(C/A)GGA(A/T)` via the recognition helix H3 (Fig 1.3). Upon binding of H3 to the major groove, DNA takes up a uniform curve of 8º. The 'wing' corresponds to a β-sheet extension between β-strands β3 and β4, and collaborates in DNA binding (Table 1.2).

Fig 1.2 Domain organisation of *c-ets-1*. The human gene for Ets-1 encodes a 441 amino acid protein with two distinct domains, the ETS domain, fingerprint of the ETS family of transcription factors and DNA-binding domain, and the Pointed domain. The regions flanking the ETS domain, 301-331 and 415-441 have been shown to have a repressive effect on the DNA binding capabilities of Ets-1 and constitute the autoinhibition module (AI module) (Jonsen et al, 1996).

Protein-protein interactions with partner proteins often play major roles in targeting ETS-domain proteins to specific promoters. Several such proteins have been identified, the role of which is greatly stressed by the early discovery that some members of the ETS family are subject to negative regulation by an intramolecular mechanism that impairs the binding to DNA, including Ets-1 (Lim et al, 1992; Jonsen et al, 1996), ERM (Laget et al, 1996) and PU.1 (Crepieux et al, 1994).

According to the accepted model for the regulation of ETS proteins, Ets-1 ETS domain is kept in a repressed state by an allosteric mechanism that depends on the interaction with the two flanking autoinhibitory (AI) motifs (Sieweke et al, 1998; Wang et al, 2002). This intramolecular inhibition would be released by interaction with a partner proteins via interaction with the AI

module, the ETS domain, or both. Although this model has been extensively validated by biochemical and biophysical studies (Petersen et al, 1995; Jonsen et al, 1996; Sieweke et al, 1998; Cowley & Graves, 2000) and there is indirect structural evidence that this is the case (Petersen et al, 1995), a complex of an autoinhibited Ets-1 with DNA in the presence of a partner protein is required to fully describe the structural underpinning of the 'release of autoinhibition' mechanism (Fig 1.4).

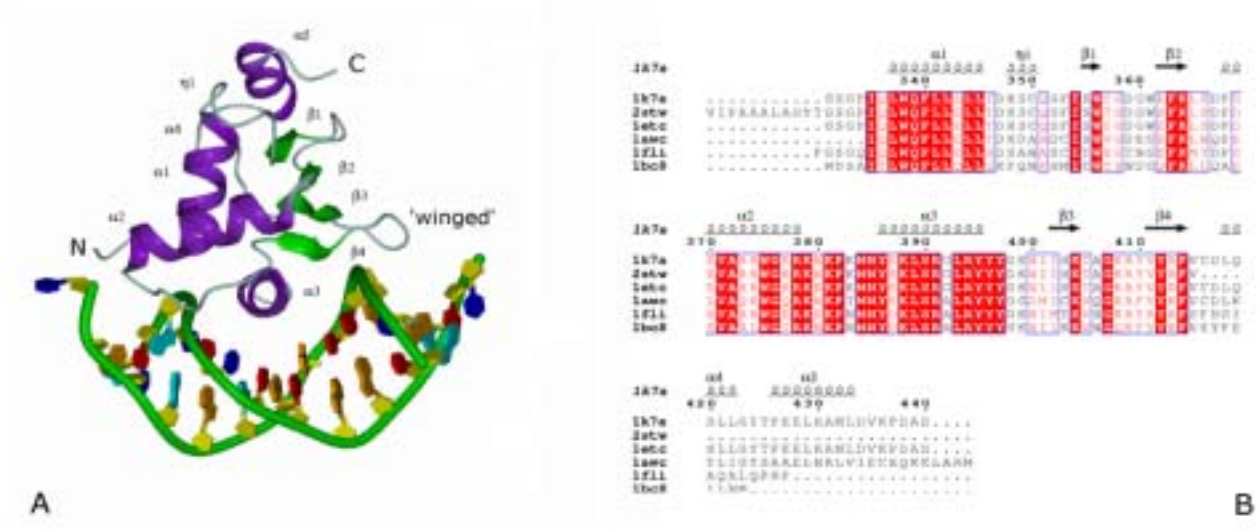

Fig 1.3 Ets-1 structure (PDB entry 1k7a) and sequence alignment of related proteins with structures in the PDB. (A) Ets-1 crystal structure 1k7a (Garvie et al, 2001) is shown docking its recognition α-helix (Á3) into the major groove of its binding site. The 'wing,' the loop connecting β-strands 3 and 4, mediates interactions with the DNA backbone. (B) Multiple sequence alignment of ETS proteins whose structure is deposited in the PDB (entry codes label each sequence). The conservation within the ETS domain is extremely high, with a considerable number of residues being strictly conserved. The secondary structure is represented on top of the alignment, and corresponds to the labels in (A).

Ets-1 is involved in the activation of HIV virus replication in humans and mice by transcriptional activation in the presence of USF1 (Sieweke et al, 1998). USF1 has been shown to recruit Ets-1 onto adjacent EBS and E-box sites, thus triggering transcription activation (Sieweke et al, 1998). Therefore, the elucidation of Ets-1 structure beyond the ETS domain can provide insights into the fundamental processes of autoinhibition in transcription regulation, and the elucidation of its complexes can shed light on pathogenic processes such as AIDS. Furthermore, a structure of the ternary complex Ets-1/USF1 bound to the HIV LTR enhancer could provide new targets for drug design.

PDB access code	ETS domain protein	Technique	Reference
1fli	Fli-1 (276-373)^{h}	NMR	(Liang et al, 1994)
2stw	Ets-1 (332-415)^{h}	NMR	(Werner et al, 1995)
1etc	Ets-1 (331-440)^{m}	NMR	(Donaldson et al, 1996)
1pue	PU.1 (171-259)^{m}	PX (2.10 Å)	(Kodandapani et al, 1996)
1bc8	SAP-1 (1-93)^{h}	PX (1.93 Å)	(Mo et al, 1998)
1awc	GABPα (311-430)^{m}	PX (2.15 Å)	(Batchelor et al, 1998)
1k79, 1k7a, 1k78	Ets-1 (331-440)^{m}	PX (2.25-2.80 Å)	(Garvie et al, 2001)

Table 1.2 PDB accession codes for proteins with ETS domain of known structure. The second column shows the span of the ETS domain and the organism it comes from (^{h} from human, ^{m} from mouse); the third column lists the experimental technique used to solve the corresponding structure (PX, protein crystallography, with resolution between round brackets; NMR, nuclear magnetic resonance); the fourth column contains the references.

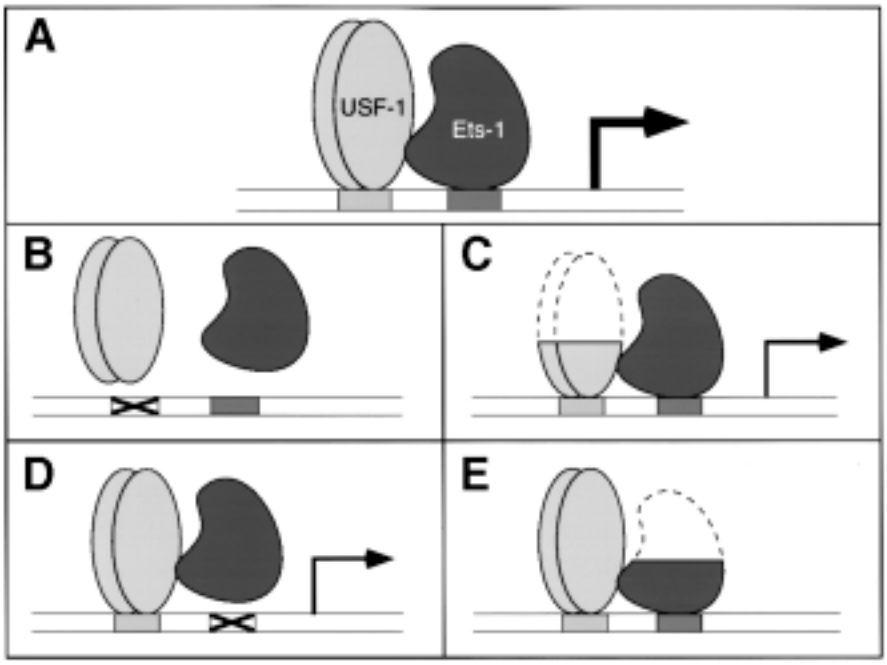

Fig 1.4 Model for the release of Ets-1 autoinhibition. Analysis of the effect of defective E-box and EBS binding sites (B, D), and USF1 and Ets-1 deletion mutants (C, E), on Ets-1-mediated transactivation (Sieweke et al, 1998). It is only when USF1 can bind to its recognition site and that Ets-1 transactivation domain is present (as in D or C) that transactivation occurs. The proposed model predicts distinct roles for USF1 and Ets-1 in their complex, whereby USF1 recruits Ets-1 onto its site which only then is able to mediate transactivation of the downstream promoter.

1.5.2 USF1 releases Ets-1 autoinhibition

USF1 (Upstream Stimulatory Factor 1, also Major Late Transcription Factor 1 or MLTF1) has been identified as a protein partner to Ets-1 by a one-yeast hybrid screen, along with another E-box protein, TFE-3 (Sieweke et al, 1998). USF1 interacts with Ets-1 in a variety of cellular and viral promoters and/or enhancers, recruiting Ets-1 via direct protein-protein interactions through their DBD and, therefore, counteracts Ets-1 autoinhibition (Fig 1.4). Ets-1 and USF1 form a stable complex on the distal enhancer region of the human immunodeficiency virus 1 (HIV-1) long terminal repeat (LTR). The distal enhancer is vital for the virus replication, and contains adjacent E-box and ETS binding sites. This interaction makes Ets-1 an enticing target for drug design, because disruption of the Ets-1/USF1 complex on the HIV-1 distal enhancer has the potential to impair viral replication. Posada et al explored the effect of an Ets-1 dominant negative mutant in T-cell infection by HIV-1 (Posada et al, 2000). Their results support the view that an intact Ets-1/USF1 is essential for HIV-1 productive infection, thereby pointing at this transcription factor complex as a novel target for anti-HIV-1 drug design.

Human USF1 is a 310 amino basic helix-loop-helix (bHLH) leucine-zipper (LZ) transcription factor, that can form both a homodimer and several heterodimers (for example, with USF-2) (Gregor et al, 1990). USF1 is expressed in all cell types (and therefore is classified as a class A HLH protein), where it binds its cognate DNA sequence, the E-box. The E-box comprises the core palindromic sequence `CACGTG`, and is found in many cellular and viral promoters and enhancers, and to which USF1 can bind as homodimer and heterodimer.

The crystal structure of the USF1 bHLH motif (as a homodimer, residues 197-260) is known at 2.9 Å resolution. It shows how each monomer folds into two amphipathic helices joined by a flexible linker that wrap around each other, forming a left-handed coiled-coil that stabilises the dimer (Ferre-D'Amare et al, 1994) (Fig 1.5). Other HLH transcription factors of known structure are listed in Table 1.3. Binding to DNA is achieved by a short stretch of positively charged residues, the basic motif, which interacts extensively with the E-box. The structure of the LZ domain of USF1 has been modelled according to sequence homology, but there is no structural information. Furthermore, dynamic light scattering (DLS) experiments have shown that the LZ motif is responsible for USF1 tetramerisation in solution presumably through the C-terminal coiled-coil interface (Ferre-D'Amare et al, 1994; Ferre-D'Amare & Burley, 1994).

PDB access code	HLH protein	Technique	Reference
1a0a	Pho4 (250-312)	PX (2.8 Å)	(Shimizu et al, 1997)
1am9	SREBP-1A (319-400)	PX (2.3 Å)	(Parraga et al, 1998)
1an4	USF1 (197-260)	PX (2.9 Å)	(Ferre-D'Amare et al, 1994)
1d7g	Hif-1 (15-73)	Model	(Michel et al, 2000)
1hlh	E47 (29-61)	Model	(Gibson et al, 1993)
1hlo	Max (4-92)	PX (2.8 Å)	(Brownlie et al, 1997)
1mdy	MyoD (105-166)	PX (2.8 Å)	(Ma et al, 1994)

Table 1.3 PDB accession codes for proteins with HLH proteins of known structure. The second column shows the name of the HLH-containing proteins and its span; the third column shows whether the structure comes from experimental X-ray diffraction data (PX, protein crystallography, with resolution between round brackets) or is a theoretical model; the fourth column lists the references.

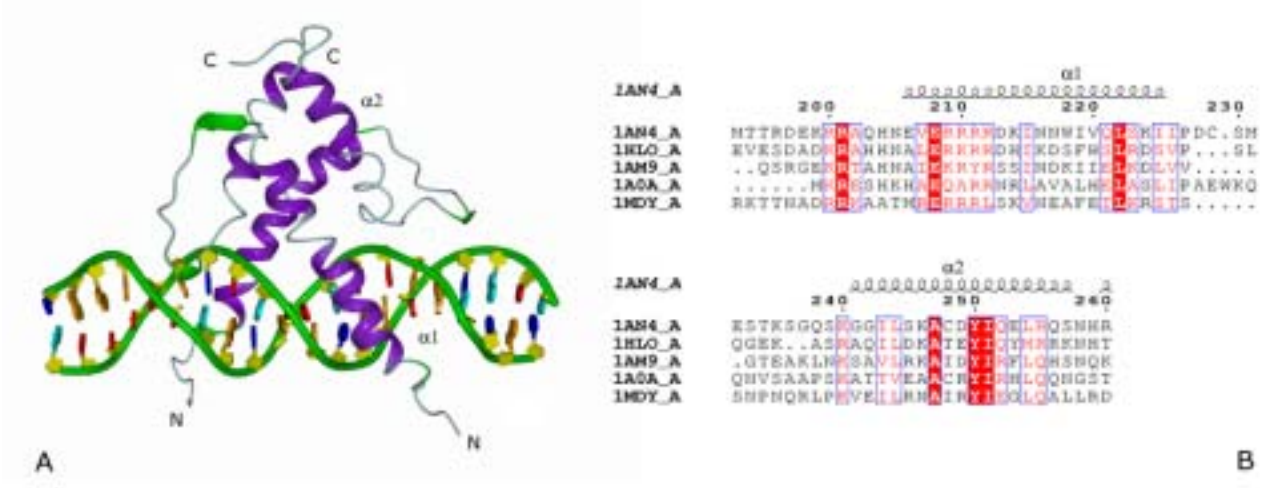

Fig 1.5 Structure of USF1 bound to its cognate DNA binding site (PDB accession code 1AN4). USF1 belongs to the basic-region (b) helix-loop-helix (HLH) leucine-zipper (ZIP) family of proteins. The basic region spans amino acids 200-212, the HLH domain residues 213-255 and the leucine-zipper residues 271-292. (A) The crystal structure contains the bHLH domain, which suffice for homo and heterodimerisation and DNA-binding (Ferre-D'Amare et al, 1994). The leucine zipper is involved in homotetramerisation, a function thought to mediate DNA looping in enhancers. Contacts with DNA are made through both recognition helices and the two extended loops. (B) Multiple sequence alignment of bHLH proteins whose structure is deposited in the PDB (entry codes label each sequence).

1.6 Summary

Transcription is the process whereupon a gene serves as template for the synthesis of a molecule of RNA that in turn produces a protein. Therefore, the initiation of transcription is a crucial regulatory event that consists in an elaborate fail-safe mechanism for controlling gene expression. It involves the correct assembly of the basal transcriptional machinery and the enhanceosome in an open chromatin region. The involvement of cellular transcription factors in disease processes, like the HIV-1 replication and cancer, makes them extremely attractive for fundamental and applied science. The elucidation of highly detailed structures of the proteins and protein-DNA complexes

that orchestrate all these processes can enhance greatly our understanding of the underlying mechanisms. However, the difficulty to produce samples of transcription factors in sufficient amounts and of sufficient quality for structural biology, often extreme, poses a great obstacle. Our endeavour is to outline and further refine procedures to optimise the production of samples for X-ray crystallography and NMR, with the ultimate goal to automate protein production for structural biology.

1.7 References

Arnone, MI. and Davidson, EH (1997) The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**(10): 1851-64.

Batchelor AH, Piper DE, de la Brousse FC, McKnight SL, and Wolberger C (1998) The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA. *Science* **279**(5353): 1037-41.

Bauer, WR, et al (1994) Nucleosome structural changes due to acetylation. *J Mol Biol* **236**(3): 685-90.

Bell, AC and Felsenfeld, G (1999) Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev* **9**(2): 191-8.

Brehm, A, Ohbo, K, and Scholer, H (1997) The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. *Mol Cell Biol* **17**(1): 154-62.

Britos-Bray, M and Friedman, AD (1997) Core binding factor cannot synergistically activate the myeloperoxidase proximal enhancer in immature myeloid cells without c-Myb. *Mol Cell Biol* **17**(9): 5127-35.

Carey, M (1998) The enhanceosome and transcriptional synergy. *Cell* **92**(1): 5-8.

Brownlie P, Ceska T, Lamers M, Romier C, Stier G, Teo H, and Suck D (1997) The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* **5**(4): 509-20.

Cleary, MA, Pendergrast, PS, and Herr, W (1997) Structural flexibility in transcription complex formation revealed by protein-DNA photocrosslinking. *Proc Natl Acad Sci USA* **94**(16): 8450-5.

Cowley DO and Graves BJ (2000) Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. *Genes Dev* **14**(3): 366-76.

Crepieux P, Coll J, and Stehelin D (1994) The Ets family of proteins: weak modulators of gene expression in quest for transcriptional partners. *Crit Rev Oncog* **5**(6): 615-38.

Cross, MA and Enver, T (1997) The lineage commitment of haemopoietic progenitor cells. *Curr Opin Genet Dev* **7**(5): 609-13.

Darnell, JE, Jr (1997) STATs and gene regulation. *Science* **277**(5332): 1630-1635.

Fry, CJ and Peterson, CL (2001). Chromatin remodeling enzymes: who's on first? *Curr Biol* **11**(5): 185-197.

Donaldson LW, Petersen JM, Graves BJ, and McIntosh LP (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J* **15**(1): 125-34.

Dynlacht, BD, Hoey, T, and Tjian, R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**(3): 563-76.

Epstein, J, et al (1994) Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem* **269**(11): 8355-61.

Ferre-D'Amare AR and Burley SK (1994) Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure* **2**(5): 357-9.

Ferre-D'Amare AR, Pognonec P, Roeder RG, and Burley SK (1994) Structure and function of the b/HLH/Z domain of USF. *EMBO J* **13**(1): 180-9.

Flanagan, PM, et al (1991) A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350**(6317): 436-8.

Fondell, JD, Ge, H, and Roeder, RG (1996) Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc Natl Acad Sci USA* **93**(16): 8329-33.

Garcia, JV, et al (1986) Synergism between immunoglobulin enhancers and promoters. *Nature* **322**(6077): 383-5.

Garvie CW, Hagman J, and Wolberger C (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**(6): 1267-76.

Ghosh, S, May, MJ, and Kopp, EB (1998) NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu Rev Immunol* **16**: 225-260.

Gibson TJ, Thompson JD, and Abagyan RA (1993) Proposed structure for the DNA-binding domain of the helix-loop-helix family of eukaryotic gene regulatory proteins. *Protein Eng* **6**(1): 41-50.

Gill, G, et al (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci USA* **91**(1): 192-6.

Gott, JM and Emeson, RB (2000) Functions and mechanisms of RNA editing. *Annu Rev Genet* **34**: 499-531.

Grant, PA, et al (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* **11**(13): 1640-50.

Graves, BJ (1998) Inner workings of a transcription factor partnership. *Science* **279**(5353): 1000-2.

Gregor PD, Sawadogo M, and Roeder RG (1990) The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev* **4**(10): 1730-40.

Grunstein, M (1998) Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* **93**(3): 325-8.

Hanna-Rose, W and Hansen, U (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet* **12**(6): 229-34.

Hansen, JC, Tse, C, and Wolffe, AP (1998) Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry* **37**(51): 17637-41.

Harrison, SC (1991) A structural taxonomy of DNA-binding domains. *Nature* **353**(6346): 715-9.

Ikeda, M, Wilcox, EC, and Chin, WW (1996) Different DNA elements can modulate the conformation of thyroid hormone receptor heterodimer and its transcriptional activity. *J Biol Chem* **271**(38): 23096-104.

Ito, M, et al (1999) Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* **3**(3): 361-70.

Jansen, RP (2001) mRNA localization: message on the move. *Nat Rev Mol Cell Biol* **2**(4): 247-56.

Jonsen MD, Petersen JM, Xu QP, and Graves BJ (1996) Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol Cell Biol* **16**(5): 2065-73.

Jones, PL, et al (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**(2): 187-191.

Jones, KA and Peterlin, BM (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem* **63**: 717-43.

Karim FD, Urness LD, Thummel CS, Klemsz MJ, McKercher SR, Celada A, Van Beveren C, Maki RA, Gunther CV, Nye JA, et al (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes Dev* **4**(9): 1451-3.

Kim, JY, et al (1993) Replication of type 1 human immunodeficiency viruses containing linker substitution mutations in the -201 to -130 region of the long terminal repeat. *J Virol* **67**(3): 1658-62.

Kim, YJ, et al (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**(4): 599-608.

Kodandapani R, Pio F, Ni CZ, Piccialli G, Klemsz M, McKercher S, Maki RA, and Ely KR (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **380**(6573): 456-60.

Kornberg, RD and Lorch, Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**(3): 285-94.

Kornberg, RD and Lorch, Y (1999) Chromatin-modifying and -remodeling complexes. *Curr Opin Genet Dev* **9**(2): 148-51.

Koshland, DE, Jr, Nemethy, G, and Filmer, D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**(1): 365-85.

Kulessa, H, Frampton, J, and Graf, T (1995) GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* **9**(10): 1250-62.

Laget MP, Defossez PA, Albagli O, Baert JL, Dewitte F, Stehelin D, and de Launoit Y (1996) Two functionally distinct domains responsible for transactivation by the Ets family member ERM. *Oncogene* **12**(6): 1325-36.

Latchman, DS (1998) Gene Regulation - a eukaryotic perspective. 3rd ed, Stansley Thorne Publishers.

Latchman, DS (1998) Eukaryotic transcription factors. 3rd ed, Academic Press.

Lefstin, JA and Yamamoto, KR (1998) Allosteric effects of DNA on transcriptional regulators. *Nature* **392**(6679): 885-8.

Liang H, Mao X, Olejniczak ET, Nettesheim DG, Yu L, Meadows RP, Thompson CB, and Fesik SW (1994) Solution structure of the ets domain of Fli-1 when bound to DNA. *Nat Struct Biol* **1**(12): 871-5.

Lim F, Kraut N, Framptom J, and Graf T (1992) DNA binding by c-Ets-1, but not v-Ets, is repressed by an intramolecular mechanism. *EMBO J* **11**(2): 643-52.

Luger, K, et al (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): 251-60.

Luscombe, NM, et al (2000) An overview of the structures of protein-DNA complexes. *Genome Biol* **1**(1).

Ma PC, Rould MA, Weintraub H, and Pabo CO (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**(3): 451-9.

Macleod K, Leprince D, and Stehelin D (1992) The ets gene family. *Trends Biochem Sci* **17**(7): 251-6.

Mangelsdorf, DJ, et al (1995) The nuclear receptor superfamily: the second decade. *Cell* **83**(6): 835-839.

Maniatis, T, et al (1998) Structure and function of the interferon-beta enhanceosome. *Cold Spring Harb Symp Quant Biol* **63**: 609-20.

McNagny, KM, et al (1998) Regulation of eosinophil-specific gene expression by a C/EBP-Ets complex and GATA-1. EMBO J **17**(13): 3669-80.

Meierhans, D, et al (1995) DNA binding specificity of the basic-helix-loop-helix protein MASH-1. *Biochemistry* **34**: 11026-36.

Michel G, Minet E, Ernest I, Roland I, Durant F, Remacle J, Michiels C (2000) A model for the complex between the hypoxia-inducible factor-1 (HIF-1) and its consensus DNA sequence. *J Biomol Struct Dyn* **18**(2): 169-79.

Mo Y, Vaessen B, Johnston K, and Marmorstein R (1998) Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by Ets proteins. *Mol Cell* **2**(2): 201-12.

Monod, J, Changeoux, J-P, and Jacob, F (1963) *J Molec Biol* **6**(306).

Myers, LC, et al (1998) The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev* **12**(1): 45-54.

Ness, SA and Engel, JD (1994) Vintage reds and whites: combinatorial transcription factor utilization in hematopoietic differentiation. *Curr Opin Genet Dev* **4**(5): 718-24.

Ng, HH and Bird, A (1999) DNA methylation and chromatin modification. *Curr Opin Genet Dev* **9**(2): 158-63.

Nan, X, et al (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**(6683): 386-389.

Nerlov, C, et al (1998) Distinct C/EBP functions are required for eosinophil lineage commitment and maturation. *Genes Dev* **12**(15): 2413-23.

Oelgeschlager, M, et al (1996) Interaction of the co-activator CBP with Myb proteins: effects on Myb- specific transactivation and on the cooperativity with NF-M. EMBO J **15**(11): 2771-80.

Orkin, SH (1996) Development of the hematopoietic system. *Curr Opin Genet Dev* **6**(5): 597-602.

Otero, G, et al (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol Cell* **3**(1): 109-118.

Parraga A, Bellsolell L, Ferre-D'Amare AR, and Burley SK (1998) Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* **6**(5): 661-72.

Patel, L, Abate, C, and Curran, T (1990) Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**(6293): 572-5.

Petersen JM, Skalicky JJ, Donaldson LW, McIntosh LP, Alber T, and Graves BJ (1995) Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science* **269**(5232): 1866-9.

Porcher, C, et al (1999) Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **126**(20): 4603-15.

Posada, R, et al (2000) Suppression of HIV type 1 replication by a dominant-negative Ets-1 mutant. *AIDS Res Hum Retroviruses* **16**(18): 1981-9.

Ptashne, M (1988) How eukaryotic transcriptional activators work. *Nature* **335**(6192): 683-9.

Pugh, BF and Tjian, R (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**(7): 1187-97.

Ramakrishnan, V (1997) Histone H1 and chromatin higher-order structure. *Crit Rev Eukaryot Gene Expr* **7**(3): 215-30.

Reichardt, HM, et al (1998) Analysis of glucocorticoid signalling by gene targeting. *J Steroid Biochem Mol Biol* **65**(1-6): 111-5.

Scully, KM, et al (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290**(5494): 1127-31.

Shapiro, LH (1995) Myb and Ets proteins cooperate to transactivate an early myeloid gene. *J Biol Chem* **270**(15): 8763-71.

Sheridan, PL, et al (1995) Activation of the HIV-1 enhancer by the LEF-1 HMG protein on nucleosome- assembled DNA in vitro. *Genes Dev* **9**(17): 2090-104.

Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, and Hakoshima T (1997) Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J* **16**(15): 4689-97.

Shivdasani, RA and Orkin, SH (1996) The transcriptional control of hematopoiesis. *Blood* **87**(10): 4025-39.

Shuman, JD, Vinson, CR, and McKnight, SL (1990) Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science* **249**(4970): 771-4.

Sieweke, MH, et al (1996) MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* **85**(1): 49-60.

Sieweke, MH and Graf, T (1998) A transcription factor party during blood cell differentiation. *Curr Opin Genet Dev* **8**(5): 545-51.

Sieweke, MH, et al (1998) Cooperative interaction of ets-1 with USF-1 required for HIV-1 enhancer activity in T cells. *EMBO J* **17**(6): 1728-39.

Smith, CW and Valcarcel, J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **25**(8): 381-8.

Sonenberg, N, Hershey, JWB, Matthews, MB (2000) Translational Control of Gene Expression. Cold Spring Harbor Laboratory Press. 1020.

Strahl, BD and Allis, CD (2000) The language of covalent histone modifications. *Nature* **403**(6765): 41-45.

Struhl, K (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**(1): 1-4.

Tan, S and Richmond, TJ (1990) DNA binding-induced conformational change of the yeast transcriptional activator PRTF. *Cell* **62**(2): 367-77.

Thompson, CM, et al (1993) A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**(7): 1361-75.

Thompson, CM and Young, RA (1995) General requirement for RNA polymerase II holoenzymes in vivo. *Proc Natl Acad Sci USA* **92**(10): 4587-90.

Tomilin, A, et al (2000) Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell* **103**(6): 853-64.

Wei, Q and Paterson, BM (2001) Regulation of MyoD function in the dividing myoblast. *FEBS Lett* **490**(3): 171-8.

Wang H, McIntosh LP, and Graves BJ (2002) Inhibitory module of Ets-1 allosterically regulates DNA binding through a dipole-facilitated phosphate contact. *J Biol Chem* **277**(3): 2225-33.

Weiss, MA, et al (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**(6293): 575-8.

Werner MH, Clore M, Fisher CL, Fisher RJ, Trinh L, Shiloach J, and Gronenborn AM (1995) The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **83**(5): 761-71.

Workman, JL and Kingston, RE (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* **67**: 545-79.

Yang, XJ, et al (1996) A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**(6589): 319-324.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals used were of the highest purity available (ACS grade) and were purchased either from Merck, Fluka, Sigma, Hampton Research, Emerald Biosciences or BioRad. Those chemicals not purchased from these companies or which deserve a comment are listed below:

Hampton Research and Emerald Biosciences: These companies provided most of the chemicals for crystallisation as preformulated screens or separate reagents (MPD, PEG).
ICM: Isopropyl-D-β-galactopyranoside (IPTG).
Pierce: Tris-chloride-ethanol-phosphine (TCEP).

2.1.2 Proteins and enzymes

Boehringer-Mannheim/La Roche: T4 DNA ligase, DNaseI and Rnase A.
Clontech: T4 DNA polymerase, T4 DNA ligase and T4 polynucleotide kinase.
MBI Fermentas: T4 DNA ligase.
Merck: Benzonase.
New England Biolabs: All restriction enzymes (including *NcoI*, *HindIII*, *KpnI*, *NdeI*, *BamHI*, *SphI*, *BglII* and *DpnI*) and BSA.
Stratagene: Shrimp alkaline phosphatase (SAP), cloned Pfu polymerase, Pfu Turbo polymerase and Taq polymerase.

2.1.3 Bacterial Strains

Table 2.1 shows the bacterial strains used in this study; JMB109 and XL1-Blue were used as cloning hosts (2.2.8), the set of B21(DE3) strains and B834 for protein expression (2.3.10); DH5a was used as either a cloning host or for protein expression; and BHS 71-18 mutS was used in site-directed mutagenesis experiments (2.2.10).

Cell strain	Genotype	Remarks
JMB109 (Yanisch-Perron et al, 1985)	*e14-(McrA$^-$) recA1 endA1 gyrA96 thi hsdR17(rK$^-$ mK$^+$) supE44 relA1 Δ(lac-proAB) [F' tra Δ36 proAB lacI^{q}ZΔM15]*	Host for cloning and for plasmid propagation; from Stratagene.
XL1-Blue (Bullock, 1987)	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F¢proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]*	Host for cloning and for plasmid propagation; from Stratagene.
DH5α (Hanahan, 1983)	*F Δ80d lac ZΔM15 Δ(lacZYA-argF) U169 deoR recA1end A1hsdR17 (rk$^-$, mk$^+$) supE44Δ$^-$ thi-1 gyrA96 relA1*	Used both for cloning and protein expression; from Life Technologies.
BL21(DE3) (Studier et al, 1986; Studier et al, 1990)	*hsdS gal [cI ts857 cndI hsdR17 recA1 endA1 gyrA96 thi-1 relA1]*	Host for high-yield expression T7 promoter-containing plasmids; T7 RNA polymerase is carried on bacteriophage λ(DE3), which is integrated in the genome of BL21; from Novagen.
B834(DE3) (Wood, 1966; Studier et al, 1986)	*E coli B dcm ompT hsdS (rB$^-$ mB$^-$) gal$^-$ met$^-$ Δ(DE3)*	Host for high-level expression T7 promotor-containing plasmids; methionine-auxotroph; from Novagen.
BL21-CodonPlus(DE3)-RIL (Carstens et al, 1999)	*E coli B F$^-$ ompT*	High-yield expression in T7 promoter-containing plasmids; T7 RNA polymerase is carried on bacteriophage λ(DE3), which is integrated in the genome of BL21; it provides the tRNA for the following rare codons: AGA and AGG (R), AUA (I) and CUA (L); from Stratagene.
BL21-(DE3)CodonPlus-RP (Carstens et al, 1999)	*E coli B F$^-$ ompT hsdS(rB$^-$ mB$^-$) dcm$^+$ Tet^r gal D(DE3) endA Hte [argU proL Camr]*	Host for high-yield expression in T7 promoter-containing plasmids; T7 RNA polymerase is carried on bacteriophage λ(DE3), which is integrated in the genome of BL21; it provides the tRNA for the following rare codons: AGA and AGG (R), AUA (I) and CUA (L); from Stratagene.
BMH 71-18 mutS (Hanahan, 1985)	*thi supE D(lac-proAB) [mutS::Tn10] [F' proAB lacI^{q}ZDM15]*	Recombination deficient strain, propagates mismatch mutations; from Promega.

Table 2.1 Cell strains and feature list. The first column of the table lists the names of the *E coli* cell strains used for cloning, vector propagation and protein expression; below each name, a suitable reference is appended; the second column contains the genotype of the strain; and the third contains some remarks about purpose, features and availability.

2.1.4 Plasmids

2.1.4.1 Cloning vectors

All inserts used during this study were amplified by PCR (2.2.9) and ligated (2.2.8) into the pCR-Blunt vector for sequencing (2.2.13), plasmid propagation of the insert (2.2.4) and as a source of template in subsequent PCR reactions. The procedure followed the specifications of the manufacturer (Invitrogen). The cDNA for the *c-ets-1* gene products was cloned into pBluescript II SK(+), and the cDNA for the *usf1* gene product in pET15b (an expression vector).

See Appendix 1 for the maps and multicloning site (MCS) of pCR-Blunt, pBluescript II SK(+) and pET15b plasmids.

2.1.4.2 Expression vectors

The pET series of vectors for high-yield expression places cloned inserts under the control of the T7 RNA promoter (Studier et al, 1990). Thus, only *E coli* strains engineered to express the T7 RNA polymerase upon IPTG induction can be used for expression, eg BL21(DE3) (2.1.3). The plasmids pET15b and pET24d contain the f1 *ori* origin of replication and the T7*lac* gene encoding the lacI repressor. pET15b is ampicillin resistant, and harbours an N-terminal His6 tag cleavable with thrombin; pET24d is kanamycin resistant, and contains an N-terminal T7 tag and a C-terminal His6 tag. Expression is induced by addition of 0.2-1 mM IPTG (2.1.1).

The pQE70 vector belongs to the pDS family of expression vectors (Brunner et al, 1987). It contains a T5 promoter, which is recognised by *E coli* RNA polymerase, and two lac operator sequences that enhance the lacI repressor binding. Additionally, it has two strong terminators, lambda *t0* and the terminator of the *E coli rrnB* operon. Finally, pQE70 bears the *bla* gene for ampicillin resistance and the colE1 origin of replication.

The pACYC9d vectors derive from the pET9d by removal of the f1 *ori* for the pACYC *ori*. There are two variants, one with His6 tag and another without it. The former contains a TEV cleavage site for convenient tag removal.

The pProExHTb expression vector contains the colE1 *ori* origin of replication, the *bla* gene for ampicillin resistance, an N-terminal His6 tag and a TEV cleavage site. Expression is under the control of the *trc* promoter, which is recognised by *E coli* RNA polymerase. A convenience of this plasmid is that cloning and expression can be carried out in a single host, eg DH5α (2.1.3).

See Appendix 1 for the maps and MCS of pET24d, pETM11, pETM30, pQE70, pProExHTb, pACYC9d and pBAT4.

Table 2.2 lists all plasmids used in this thesis except for those that I constructed.

Vector Derivatives	Resistance	Vector	Gene	Remarks
-	Kanamycin	PCR-Blunt	-	From Invitrogen
-	Ampicillin	pBluescript II SK(+)	-	From Stratagene
pEts	Ampicillin	pBluescript II SK(+)	hEts-1 1-441	Sieweke et al, 1998
-	Ampicillin	pET15b	-	From Novagen
pETS377	Ampicillin	pET15b	cEts-1 329-520	Sieweke et al, 1998
pU194	Ampicillin	pET15b	USF1 194-310	Sieweke et al, 1998
-	Ampicillin	pBAT4	-	MC Vega
-	Kanamycin	pET24d	-	From Novagen
pETM30	Kanamycin	pET24d	DCOH	G Stier
pETM11	Kanamycin	pET24d	DCOH	G Stier
pACYC9d	Kanamycin	pET9d	-	G Stier
-	Ampicillin	pProExHTb	-	From Clontech
-	Chloramphenicol	pLysS	-	From Novagen

Table 2.2 Plasmids. From left to right, construct and/or vector name, its antibiotic resistance, name of the vector from which it is derived (if applicable; otherwise '-'), gene encoded by the construct (if applicable; otherwise '-'), remarks concerning the corresponding entry (commercial availability, provider).

2.1.5 Oligodeoxyribonucleotides

The following list includes the oligodeoxyribonucleotides used for PCR amplification (2.2.9), site-directed mutagenesis (2.2.10), sequencing (2.2.13), electrophoretic mobility shift assays (EMSA, 2.3.7) and crystallisation (2.3.16). Restriction sites introduced by the PCR primers are highlighted in bold; substituted base pairs in mutagenic oligonucleotides are in lower-case letters; transcription factor binding sites in oligonucleotides for EMSA (2.3.7) and crystallisation (2.3.17) are in bold. Sequences are depicted in all cases from 5' to 3'. Synthesis of the oligodeoxyribonucleotides was done by any of the following companies or services: Genaxis, Metabion, MWG-Biotech and the EMBL oligonucleotide synthesis service.

2.1.5.1 Oligonucleotides for PCR amplification:

OE234: 5'-CGT**CATATG**GACAACATGTGTATGGGGAGG-3'

OE238: 5'-CGT**CATATG**GGGAGGACCAGTCGT-3'

OE244: 5'-CGT**CATATG**GGTAAACTCGGGGGCCAG-3'

OE441: 5'-GG**GGATCC**TCACTCGTCGGCATCTGG-3'

NcoI-OE1: 5'-CATG**CCATGG**AGGCGGCCGTCGATCTC-3'

NcoI-OE51: 5'-CATG**CCATGG**CTACTTTCAGTGGTTTC-3'

NcoI-OE73: 5'-CATG**CCATGG**AAACCCATGTTCGGGAC-3'

NcoI-OE234: 5'-CATG**CCATGG**ACAACATGTGTATGGGGAGG-3'

NcoI-OE238: 5'-CATG**CCATGG**GGGAGGACCAGTCGT-3'

NcoI-OE244: 5'-CATG**CCATGG**TAAACTCGGGGGCCAG-3'

NcoI-OE280: 5'-CATG**CCATGG**TTCCCTCCTATGACACG-3'

NcoI-OE301: 5'-CATG**CCATGG**AGGGCACCTTCAAGGAC-3'

NcoI-OE1335(A2): 5'-CATG**CCATGG**GCTTCCAGCTATGGCAG-3'

NcoI-OE335(E2): 5'-CATG**CCATGG** GAGTCCAGCTATGGCAG-3'

KpnI-OE441: 5'-GG**GGTACC**TCACTCGTCGGCATCTGGCTT-3'

SphI-OE1: 5'-CATG**GCATGC**AGGCGGCCGTCGATCTC-3'

SphI-OE51: 5'-CATG**GCATGC**CTACTTTCAGTGGTTTC-3'

SphI-OE73: 5'-CATG**GCATGC**AAACCCATGTTCGGGAC-3'

SphI-OE234: 5'-CATG**GCATGC**ACAACATGTGTATGGGGAGG-3'

SphI-OE238: 5'-CATG**GCATGC**GGGAGGACCAGTCGT-3'

SphI-OE244: 5'-CATG**GCATGC**TAAACTCGGGGGCCAG-3'

SphI-OE280: 5'-CATG**GCATGC**TTCCCTCCTATGACACG-3'

SphI-OE301: 5'-CATG**GCATGC**AGGGCACCTTCAAGGAC-3'

SphI-OE1335(A2): 5'-CATG**GCATGC**GCTTCCAGCTATGGCAG-3'

SphI-OE335(E2): 5'-CATG**GCATGC**GAGTCCAGCTATGGCAG-3'

BglII-OE441: 5'-ATA**AGATCT**ACTCGTCGGCATCTGGCTT-3'

HindIII-OE441: 5'-ATATT**AAGCTT**ATCACTCGTCGGCATCTGGCTT-3'

NcoI-OU194: 5'-CATG**CCATGG**ATGAGAAACGCAGGGCT-3'

KpnI-OU260: 5'-GG**GGTACC**TCATGATTACTGATCGTTACATC-3'

KpnI-OU310: 5'-GG**GGTACC**TCATTAGTTGCTGTCATTCTTGATG-3'

2.1.5.2 Oligonucleotides for sequencing:

T7 promotor: 5'-TAATACGACTCACTATAGGG-3'

T7 terminator: 5'-GCTAGTTATTGCTCAGCGG-3'

M13/pUC reverse: 5'-AGCGGATAACAATTTCACACAGG-3'

pQE-specific: 5'-CCCGAAAAGTGCCACCTG-3'

2.1.5.3 Oligonucleotides for site-directed mutagenesis:

OE-C261S: 5'-ATAGAGAGCTACGATAGTcTGATCGCCTCACCCAGTCC-3'

OE-C350S: 5'-TTACTCACTGATAAATCCTcTCAGTCTTTTATCAGCTGG-3'

OE-C416S: 5'-TACGTGTACCGCTTTGTGTcTGACCTGCAGAGCCTGCTG-3'

OU-C229S: 5'-TCCAAGATAATCCCAGACaGCTCTATGGAGAGCACCAAG-3'

OU-C248S: 5'-GTGGGATTCTATCCAAAGCTaGtGATTATATCCAGGAG-3'

2.1.5.4 Oligonucleotides for EMSA/crystallisation:

LTR33: 5'-TCAT**CACGTG**GCCCGAGAGCTG**CATCCG**GAGTA-3'

LTR32: 5'-TCAT**CACGTG**GCCCGAGAGCTG**CATCCG**GAGT-3'

2.1.6 DNA molecular size markers

DNA size markers or ladders were provided by MBI Fermentas. The following markers were used:

GeneRuler DNA 100 bp ladder, with fragments of size 1031, 900, 800, 700, 600, 500, 400, 300, 200, 100 and 80 bp.

GeneRuler DNA 100 bp ladder plus, with fragments of size 3000, 2000, 1500, 1200, 1031, 900, 800, 700, 600, 500, 400, 300, 200 and 100 bp.

GeneRuler DNA 1 kb ladder, with fragments of size 10000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 750, 500 and 250 bp.

Lambda DNA/ Eco130I (*StyI*) Marker, 16, with fragments of size 19329, 7743, 6223, 4254, 3472, 2690, 1882, 1489, 925, 421 and 74 bp.

2.1.7 Protein molecular weight markers

Several molecular weight standards for protein gels were used. Among them, Novex Wide 12 and SeeBlue protein markers. The size of the proteins in each marker is, for the Wide 12: 116.3, 97.4, 66.3, 55.4, 36.5, 31, 22, 14, 6, 3.5 and 2.5 kDa; and for the SeeBlue marker: 210, 78, 55, 45, 34, 23, 16, 7 and 4 kDa.

2.1.8 Media

Antibiotics: Ampicillin (50-200 mg/ml stock solution in *dd*H_2O, 50-200 µg/ml working concentration), kanamycin (25-50 mg/ml stock solution in *dd*H_2O, 25-50 µg/ml working concentration), chloramphenicol (35 mg/ml stock solution, in 100% [v/v] ethanol, 35 µg/ml working concentration).

Luria-Bertani (LB) medium (Sambrook et al, 1989): 1% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, 1% (w/v) NaCl; sterilised by autoclaving or filtration through a 0.22 µm filter. Purchased from Life Technologies as dried powder.

LB agar plates: LB medium and 1.5% (w/v) bacteriological agar. Sterilised by autoclaving. Plates should be poured when the LB agar temperature reaches 50 ºC.

M9 medium (mineral medium, 10x): 80 g/l Na_2HPO_4, 40 g/l KH_2PO_4, 5 g/l NaCl, 5 g/l NH_4Cl;

sterilise by filtration through a 0.22 μm filter.

Trace element solution (100x): 5 g/l EDTA, 0.83 g/l FeCl3?6$H_2$0, 84 mg/l $ZnCl_2$, 13 mg/l $CuCl_2.2H_2O$, 10 mg/l $CoCl_2.6H_2O$, 10 mg/l H_3BO_3, 1.6 mg/l $MnCl_2.6H_2O$; sterilise by filtration through a 0.22 μm filter.

Minimal medium: 100 ml/l M9 medium, 10 ml/l trace element solution (100x), 20 ml/l 20% (w/v) glucose, 1 ml/l 1 M Mg_2SO_4, 0.3 ml/l 1 M $CaCl_2$, 1 ml 1 mg/ml biotin, 1 ml 1 mg/ml thiamin; add appropriate antibiotics. Sterilise by filtration through a 0.22 μm filter.

Terrific broth (TB) medium: 1.2% (w/v) bacto-tryptone, 2.4% (w/v) bacto-yeast extract, 0.4% (v/v) glycerol. Sterilise by autoclaving.

SOB medium: 2.0% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, 0.5% (w/v) NaCl. Add 10 ml of a 250 mM solution of KCl (made by dissolving 1.86 g of KCl in 100 ml of *dd*H_2O). Adjust the pH to 7.0 with 5 N NaOH (~0.2 ml). Sterilise by autoclaving. Just before use, add 5 ml of a sterile solution of 2 M $MgCl_2$ (made by dissolving 19 g of $MgCl_2$ in 90 ml of *dd*H_2O).

SOC medium: SOB medium supplemented with 1.8% (v/v) glucose. Sterilise by filtration through a 0.22 μm filter.

Buffered media: 50 mM Tris or MOPS were used to buffer bacteriological media during expression at pH 7.5.

2.1.9 Buffers and solutions

2.1.9.1 Molecular biology buffers and solutions

2.1.9.1.1 Cell strains

Glycerol stock solution for storage of cell strains: Depending on cell strain, it consists of 10-50% (v/v) glycerol and suitable medium (eg LB, minimal medium) (2.1.8).

2.1.9.1.2 Agarose electrophoresis

TAE (Tris-acetate-EDTA, 50x): 242 g/l Tris base, 57.1 ml/l glacial acetic acid (37.5%), 18.6 g/l EDTA.

Agarose gel: 1-2% (w/v) agarose in 1x TAE, typically 300 ml. Agarose was of ultra high quality from Sigma.

Loading buffer (6x): 0.09% (w/v) BPB, 0.09% (w/v) xylene cyanol FF, 60% (v/v) glycerol, 60 mM EDTA.

Ethidium bromide: 10 mg/ml stock solution; working concentration is a 10^5 dilution.

SYBR Green: Concentrated stock in DMSO; working concentration is a 10^5 dilution. Typically prepared as a 6x stock premixed with the DNA electrophoresis sample buffer.

2.1.9.1.3 Buffers for minipreparation of DNA (Qiagen)

P1 buffer (resuspension buffer): 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 mg/ml RNase A.

P2 buffer (lysis buffer): 200 mM NaOH, 1% (w/v) SDS.

P3 buffer (neutralisation buffer): 3.0 M potassium acetate, pH 5.5.

QBF buffer (equilibration buffer): 750 mM NaCl, 50 mM MOPS, pH 7.0, 15% (v/v) 2-propanol, 0.15% (w/v) Triton X-100.

QC buffer (wash buffer): 1 M NaCl, 50 mM MOPS, pH 7.0, 15% (v/v) 2-propanol.

PE buffer (wash buffer): Composition not facilitated by the manufacturer.

QE buffer (elution buffer): 1.25 M NaCl, 50 mM Tris-HCl, pH 8.5, 15% (v/v) 2-propanol.

TB buffer: 10 mM Tris-HCl, pH 8.5.

TE buffer: 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA.

Quick lysis solution: 2.5% (w/v) Ficoll 400, 1.5% (w/v) SDS, 0.015% (w/v) BPB in 1x TBE buffer.

2.1.9.1.4 Buffers for midipreparation of DNA (Clontech)

S1: DNase-free RNase A (400 µg/ml) in TE buffer; stored at -20 ºC.

S2: 200 mM NaOH, 1% (v/v) SDS.

S3: 2.8 M potassium acetate, pH 5.2; stored at 4 ºC.

N2: 900 mM KCl in 100 mM Tris-H_3PO_3, pH 6.3, 15% (v/v) ethanol.

N3: 1150 mM KCl in 100 mM Tris-H_3PO_3, pH 6.3, 15% (v/v) ethanol.

N4: 1000 mM KCl in 100 mM Tris-H_3PO_3, pH 6.3, 15% (v/v) ethanol.

2.1.9.1.5 DNA modifications

Shrimp alkaline phosphatase buffer (10x): 0.1 M Tris-HCl, pH 7.5 at 37 ºC, 0.1 M $MgCl_2$, 1 mg/ml BSA.

T4 polynucleotide kinase buffer (10x): 500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 50 mM DTT, 10 mM ATP.

Ligase buffer (10x): 50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP; stored at -20 ºC.

2.1.9.1.6 PCR

PCR dNTP mix: 10 mM or 2 mM each dNTP (N = A, C, G, T) in *dd*H_2O; stored at -20 ºC.

2.1.9.2 Buffers for protein purification

2.1.9.2.1 Gel filtration (GF) buffers

GF buffer: 20 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM EDTA.

2.1.9.2.2 Anion exchange (AIEX) chromatography buffers

AIEX buffer A: 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM DTT, 2 mM EDTA.

AIEX buffer B: 50 mM Tris-HCl, pH 8.0, 1 M NaCl, 2 mM DTT, 2 mM EDTA.

2.1.9.2.3 Cation exchange (CIEX) chromatography buffers

CIEX buffer A: 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM DTT, 2 mM EDTA.

CIEX buffer B: 50 mM Tris-HCl, pH 8.0, 1 M NaCl, 2 mM DTT, 2 mM EDTA.

2.1.9.2.4 Reverse phase chromatography (RPC) buffers

RPC buffer A: 0.05% (v/v) TFE, 0 M ACN, pH 3.0.

RPC buffer B: 0.065% (v/v) TFE, 1 M ACN, pH 3.0.

2.1.9.2.5 GST affinity chromatography buffers

GST loading buffer: 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM EDTA.

GST wash buffer: 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM DTT, 2 mM EDTA.

GST elution buffer: 50 mM Tris-HCl, pH 7.5, 10 mM reduced glutathione.

2.1.9.2.6 Metal Chelating (Ni-NTA) chromatography buffers

Ni-NTA loading buffer: 50 mM Tris-HCl, pH 8.0, 5-10 mM imidazole 300 mM NaCl, 0.2% (w/v) CHAPS.

Ni-NTA wash buffer: 50 mM Tris-HCl, pH 8.0, 15-35 mM imidazole, 300 mM NaCl, 0.2% (w/v) CHAPS.

Ni-NTA high-salt (HS) buffer: 50 mM Tris-HCl, pH 8.0, 0-5 mM imidazole, 1 M NaCl, 0.2% (w/v) CHAPS.

Ni-NTA elution buffer: 50 mM Tris-HCl, pH 8.0, 150-300 mM imidazole, 300 mM NaCl, 0.2% (w/v) CHAPS.

2.1.9.2.7 Heparin-Sepharose (Pharmacia) chromatography buffers

Heparin buffer A: 20 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM EDTA.

Heparin buffer B: 20 mM HEPES, pH 7.5, 1 M NaCl, 2 mM DTT, 2 mM EDTA.

2.1.9.2.8 DNA Affinity chromatography buffers

DNA buffer A: 25 mM Tris-HCl, pH 8.0, 85 mM NaCl, 10 mM DTT, 2 mM EDTA.

DNA buffer B: 25 mM Tris-HCl, pH 8.0, 1 M NaCl, 10 mM DTT, 2 mM EDTA.

2.1.9.2.9 Dialysis buffers

TN: 50 mM Tris-HCl, pH 8.0, 250 mM NaCl, 2 mM DTT, 2 mM EDTA.

HN: 10 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM EDTA.

2.1.9.2.10 Solubilisation buffers

UB: 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5-8 M urea, 0.2% (w/v) CHAPS, 1 mg/ml lysozyme.

GU: 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 6 M guanidium hydrochloride (GuHCl) 0.2% (w/v) CHAPS, 1 mg/ml lysozyme.

NL: 50 mM CAPS, pH 10.0, 0.5-2% (v/v) N-lauryl sarcosine.

2.1.9.2.11 Gel electrophoresis

Gel staining solution: 0.1% (w/v) Coomassie Brilliant Blue R-250, 40% (v/v) ethanol, 10% (v/v) acetic acid.

Gel distaining solution: 7.5% (v/v) acetic acid, 10% (v/v) ethanol.

Gel loading buffer (6x): 40% (v/v) glycerol, 564 mM Tris base, 424 mM Tris-HCl, 8% LDS, 2.04 mM EDTA, 0.88 mM Coomassie Blue G-250, 0.70 mM Phenol red; LDS replaces SDS, much more insoluble.

MES running buffer (20x): 1 M MES, 1 M Tris base, 69.3 mM SDS, 20.5 mM EDTA; final pH should be 7.3.

TBE running buffer (5x): 54 g Tris base, 27.5 g boric acid, 2.9 g EDTA, *dd*H_2O one litre. The pH should be 8.3.

TBE loading buffer (6x): 6 ml 5x TBE running buffer, 3.2 ml glycerol, 0.3 ml 1% (w/v) BPB, 1% (w/v) xylene cyanol, 0.3 ml *dd*H_20 to 10 ml.

2.1.9.2.12 Western blot

Novex NuPAGE transfer buffer: 25 mM Bis-Tris, 25 mM bicine, 1 mM EDTA, pH 7.5.

2.1.9.2.13 Proteolytic removal of tags

Thrombin buffer: 20 mM Tris-HCl, pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$.

TEV buffer: 20 mM Tris-HCl, pH 8.0, 50-100 mM NaCl, 2 mM DTT, 2 mM EDTA.

2.1.9.2.14 Limited proteolysis

Trypsin buffer: 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 2 mM DTT.

2.1.10 Crystallisation

Crystallisation supplies and tools

Supplies and tools, including Linbro plates (24-well polysterene trays for hanging and sitting drop methodologies), siliconised cover slides (round and square slides of different thickness), sealing tape, forceps and single anodised aluminium handles for crystal manipulation were purchased from Hampton Research.

Crystallisation solutions

Crystallisation screens were purchased from Hampton Research (Crystal screens I and II, Cryo Screen, Lite Screen, PEG/Ion Screen, Low Ionic Strength Screen) or from Emerald Biosciences (Wizard Cryo Screens I and II) (2.1.1). Additional solutions were made with chemicals of the highest chemical grade (ACS grade) and filtered through a 0.22 μm filter to remove dust. All solutions were stored at 4 ºC. PEG solutions were protected from exposure to light with aluminium foil.

2.2 Microbiological Methods

2.2.1 Growth and storage of bacterial strains

Bacteria were grown on LB agar plates (2.1.8) or liquid LB medium (2.1.8) at 37 ºC, unless stated otherwise. For positive selection, media and plates were supplemented with appropriate antibiotics, in the following concentrations: 50-200 μg/ml ampicillin, 50 μg/ml carbenicillin, 25-50 μg/ml kanamycin and 35 μg/ml chloramphenicol. All media and manipulation tools were sterilised by autoclaving or, if heat-labile, by filtration with a 0.22 μm filter. Sterile laboratory practises were observed for manipulation of bacterial samples. For permanent storage, cell strains were flash-frozen in liquid nitrogen after well mixed with a glycerol-based cryoprotectant solution. BL21(DE3) and related strains (2.1.3) were preserved in 10-20% (v/v) glycerol in LB, and XL1-Blue (2.1.3) and DH5α (2.1.3) in 50% (v/v) glycerol in LB.

2.2.2 Transformation of *E coli* strains

2.2.2.1 Transformation by heat-shock

E coli cells were rendered competent by the $CaCl_2$ method (Morrison, 1977). Cells are grown to log phase, centrifuged and resuspended in a cold solution 0.1 M in $CaCl_2$. Exposure of the cell wall to calcium ions enables the cells to uptake DNA, or to become competent. Plasmid DNA is mixed with the cells and presumably adheres to them. The mixture of DNA and cells is heated up to 42 ºC for 45 s, or 'heat-shocked.' The cells are grown in non-selective media to allow synthesis of the antibiotic genes. Then, they are plated over selective agar Petri dishes for colony screening.

Preparation of cells by the $CaCl_2$ method: 2 ml of an overnight preculture (from a single colony) was used to inoculate 200 ml LB medium in a sterile flask. The culture was grown at 37 ºC with moderate shaking (250 rpm) to an OD_{600} of 0.3-0.4 (mid-log phase). Then, it was aliquoted in 10-ml pre-chilled tubes and left on ice for 5-10 min. (Cells were kept cold for all subsequent steps.) The cells were centrifuged for 5-10 min at 3000 rpm in a GSA rotor, the supernatant was poured off and each cell pellet resuspended in 5 ml of ice-cold 0.1 M $CaCl_2$ solution. The resuspended pellets were kept on ice for 30 min. Finally, cells were pelleted at 4 ºC (2500 rpm for 5 min), resuspended in 2 ml of ice cold 0.1 M $CaCl_2$ and frozen away in liquid nitrogen in 100 ml aliquots. Frozen competent cells retained competency over many months.

Transformation: 10-100 ng of DNA was added to an aliquot of competent cells previously thawed on ice, incubated for 5-30 min and heat-shocked at 42-43 ºC for 45 s. Then, 200 µl of LB or SOC medium without antibiotics were added and cells shaken at 37 ºC for 30 min. Finally, 100-200 ml were plated on a selective medium and incubate overnight.

2.2.2.2 Transformation by electroporation

Preparation of competent *E coli* cells for electroporation was done according to Dower et al, 1988, slightly modified. A starting culture was prepared as described above, but cells were harvested at a higher OD_{600} of 0.5-0.8, after pre-chilling the culture for 30 min. Cell pellets were washed with two successive cycles of resuspension and centrifugation in 10 mM HEPES, pH 7.0. The original method used *dd*H_2O, but then the cell pellets were difficult to handle. In the last two rounds of wash, the pellets were resuspended in *dd*H_2O water and then in 2 ml aliquots in 10% (v/v) glycerol. These aliquots were immediately frozen in liquid nitrogen and stored at -80 ºC.

Transformation: 10-20 ng of DNA were added to an aliquot of cells previously thawed on ice,

incubated for 5-30 min and placed at the bottom of a 0.2-cm electroporation cuvette. This cuvette was then placed inside a Stratagene transfection apparatus (electroporator), with settings to 25 μF, 200 Ohm and 2.4 kV. The time constant of the discharge was 3-4 ms under these conditions. After the discharge, the cells were quickly taken on ice, and 100-200 ml of LB or SOC medium were added to aid the cells to recover. Gentle shaking at 37 ºC for 30-60 min increased cell survival after transformation.

2.2.3 Determination of DNA concentration and purity

The concentration and degree of purity of double-stranded (ds) and single-stranded (ss) DNA oligonucleotides was determined based on the Lambert-Beer's law (Eq 1) by measuring the absorbance at 260 and 280 nm. The molar absorptivity coefficient at 260 nm was calculated from the sequence of the oligonucleotide.

(Eq 1) $$A_{260} = \varepsilon_{260}\, d\, c$$

A_{260} is the absorbance at 260 nm, ε_{260} is the molar absorptivity coefficient, c is molar concentration and d is the optical pathway of the cuvette. For a protein-free and RNA-free dsDNA solution, the ratio of A_{260}/A_{280} should be close to 2. Protein contaminants would decrease this ratio, whereas RNA contamination could increase it.

2.2.4 Isolation of plasmid DNA

Plasmid isolation (or preparation) was done according to different methods, all of them based on the alkaline SDS cell lysis method (Brinboim et al, 1979; Sambrook et al, 1989).

2.2.4.1 Small-scale plasmid preparation (minipreparation):

The protocol for small-scale plasmid preparation was based on the QIAquick Plasmid Miniprep kit manual and supplies (Qiagen), with slight modifications.

A 2-10 ml overnight culture was centrifuged at 5 krpm for 10 min, the pellet was resuspended in 100 μl of chilled P1 buffer (2.1.9.1) and incubated on ice for 5 min. Next, the cells were lysed by addition of 250 μl of P2 buffer (2.1.9.1) for 5 min. Then, lysis was stopped with 150 μl of chilled P3 buffer (2.1.9.1) followed by a 5 min incubation on ice. Cell debris and precipitated protein were removed by centrifugation of the lysate at 10 krpm for 30 min, and the supernatant decanted into a 800-μl column. The matrix of these columns is based on silica, and thus can bind

dsDNA molecules of widely different sizes. Binding occurred by spinning the loaded columns at 10 krpm for a maximum of 1 min. Likewise, impurities were washed off the column with 750 µl of QC wash buffer (2.1.9). Finally, elution was performed with 30-50 µl of QE buffer (2.1.9.1). In my experience, loading the lysate twice through the column increased the final yield by 30-50%.

2.2.4.2 Large-scale plasmid preparation (midi- and maxipreparation):

These methods were upscaled versions of the minipreparation protocol outlined before. Either the Qiagen Maxipreparation kit or the Clontech Nucleobond kit were used routinely.

The Qiagen Maxipreparation kit recommended a protocol identical to that outlined before for the Qiagen Minipreparation kit; the major difference was that the operations concerned with the column were done here by gravity-flow instead than by centrifugation. The volumes of each reagent were increased proportionally. Therefore, for a cell pellet coming from 100 ml culture, 10 ml of chilled P1 buffer (2.1.9.1) was used to resuspend the cells; then, 10 ml of P2 buffer was added to induce cell lysis, and the resulting reaction was incubated at RT for 5 min. Next, cell lysis was halted by adding 10 ml of the P3 neutralisation buffer, and the sample placed on ice. After 30 min, the sample was centrifuged in a Sorvall SS34 rotor at 18 krpm for 1-2 h. The DNA contained in the supernatant was precipitated by adding 2-propanol to a final concentration of 70% (v/v). Subsequently, the 2-propanol-precipitated DNA was collected by centrifugation at 18 krpm for 2 h in a Sorvall SS34 rotor, washed with ice-cold 70% ethanol and resuspended in an appropriate volume of TE buffer. Consistently higher yields were obtained by keeping the 2-propanol-precipitated DNA for 12-24 h at 4 ºC before centrifugation. Final yields in DNA ranged between 5-10 µg for very low copy number plasmids to 40 µg for high copy number plasmids.

The Nucleobond procedure was used for midipreps (expected yield for a high copy number plasmid of 10 mg). 100 ml of LB medium were inoculated with a single colony and grown overnight at 37 °C with vigourous shaking (300 rpm) until saturation. Next morning, the cells were centrifuged in a Sorvall GSA centrifuge at 5 krpm for 10 min, resuspended in 4 ml of S1 buffer and placed on ice for 5-10 min. Lysis was achieved by adding 4 ml of S2 buffer to the cells, shaking gently and incubating them for 5 min. Then, addition of 4 ml of S3 buffer stopped lysis, and the lysate was kept on ice for 5-10 min. To separate the supernantant from the precipitated protein and cell debris, the lysate was centrifuged at 18 krpm for 1 h in a Sorvall SS34 rotor. The supernatant was poured onto a Nucleobond AX100 column, previously equilibrated with 2 ml of N2 buffer, and flow-through discarded. Next, the column was washed with 8 ml of N3 buffer, and the bound DNA

eluted with 2 ml of N4 buffer. To remove salts and concentrate the sample, the DNA was precipitated with 0.7 vol of 2-propanol and subsequently centrifuged at 18 krpm for 30 min in a Sorvall SS34 rotor. The pellet was rinsed with pre-chilled 70% (v/v) ethanol, let dry for 10 min, and resolved in TE buffer.

2.2.5 DNA precipitation

To eliminate salts from ligation mixtures for electroporation, ethanol or n-butanol precipitation was carried out.

2.2.5.1 Ethanol precipitation

The volume of the ligation mixture was first brought to 20 µl. Next, 0.1 vol of 3 M sodium acetate (pH 5.5) and 3 vol of 100% (v/v) ethanol were added to the sample, and then it was centrifuged at 14 krpm for 30 min in an Eppendorf tabletop centrifuge at 4 ºC. The pellet was rinsed with 70% (v/v) ethanol and re-centrifuged for 10 min (same conditions). After complete evaporation of the remaining ethanol, DNA was dissolved in 20 µl *dd*H2O or TB buffer.

2.2.5.2 Butanol precipitation

The volume of the ligation mixture was first brought to 20 µl. Then, 200 µl of n-butanol were added to the sample and this was centrifuged for 5 min at 14 krpm in an Eppendorf tabletop centrifuge. The pellet was rinsed was 70% (v/v) and dried in a SpeedVac for 5 min. The pellet was dissolved in 20 µl of *dd*H_2O or TB buffer.

2.2.6 DNA electrophoresis

DNA fragments can be separated and purified by electrophoresis in agarose gels for analytical and preparative purposes.

1-2% (w/v) agarose gels were made first by adding the appropriate amount of agar in a 500 ml Erlenmeyer flask; then, 300 ml 1x TAE buffer was added and the flask stirred softly to soak the fine powder up in buffer. Next, the agarose suspension was heated in a microwave at 300 watt for 10-15 min, until the agarose was mostly melted; then, to complete the melting of the agarose, and to favour its subsequent cooling, it was further mixed with a magnetic stirrer. After 15 min, the suspension was poured into a plastic holder, a suitable comb (with 0.5-mm or 1-mm wells) was placed at the top end and allowed 30 min for solidification. Samples were loaded on a gel outside the electrophoresis chamber in 1x loading buffer (prepared by addition of 0.2 vol of the 6x loading

buffer [2.1.9.1]), and then electrophoresed at 300 V for 2 min until the bands corresponding to the dye had moved 0.5-1 cm into the gel. Then, the gel was covered with extra 1x TAE buffer, and electrophoresis continued at 50-70 V (equivalent to 90-140 mA) for 30-60 min.

DNA bands were visualised with either ethidium bromide or SYBR Green under the UV light. For a typical experiment, with DNA samples of 100-200 ng and a 100 ml agarose gel, 0.2-0.3 µl of the ethidium bromide stock solution (2.1.9.1) (a 30-50x10^{-3} dilution) or SYBR Green-supplemented loading buffer (2.1.9.1) sufficed to stain bright bands. SYBR Green has the advantages of being less toxic than ethidium bromide (Singer et al, 1999) and, because it is added directly to the samples in the loading buffer, has less chance to contaminate the electrophoresis buffer.

2.2.7 Purification of DNA from agarose gel

To ensure high purity of plasmid DNA and/or PCR fragments, they were purified from agarose gels with the Qiagen QIAquick Gel Extraction kit and the QIAquick PCR Purification kit, respectively. In all cases the agarose was of the highest quality possible (ultra high purity, Sigma). Plasmids were purified from 1% (w/v) agarose gels and PCR fragments from 1.5-2% (w/v) agarose gels. To avoid DNA damage, gels were visualised on a UV transilluminator for the briefest time, and suitable bands excised with a clean scalpel.

For both plasmids and PCR fragments, 3 vol were added per vol of agarose (100 µg corresponds approximately to 100 µl) in an Eppendorf tube, and then incubated at 50 ºC for 10 min with occasional vortexing. After the agarose dissolved completely, it was poured into a QIAquick column and centrifuged for a 1 min at 13 krpm in an Eppendorf tabletop centrifuge. Then, 750 µl of PE buffer was applied to the column and centrifuged as before. The DNA was eluted with 30-50 µl of buffer TB.

2.2.8 DNA engineering methods

2.2.8.1 Restriction of DNA by restriction endonucleases

Type I restriction endonucleases (Sambrook et al, 1989) cleave DNA at specific target sequences within the enzymes recognition sites. Endonucleases can produce protruding ('sticky') or blunt ends.

DNA plasmids and PCR products were digested at 37 ºC in the buffers recommended by the manufacturer (New England Biolabs). Some of the reactions (those with *NcoI* and *KpnI*) were

supplemented with 10 mM DTT to enhance their effectiveness. BSA (10 mg/ml, from NEB) was added to all reactions to avoid excessive adsorption of the enzymes to the plastic walls of the Eppendorf tubes. All enzymes and buffers were stored at -20 ºC. During this study, the following enzymes were used: *NcoI*, *HindIII*, *KpnI*, *NdeI*, *BamHI*, *SphI*, *BglII*, *NotI* and *DpnI* (2.1.2).

2.2.8.2 DNA ligation

In a ligation reactions a digested PCR fragment and a linearised vector are incubated with ATP and T4 DNA ligase (2.1.2) to render a circularised vector containing the insert, by the formation of a phosphodiester bond between the 3' hydroxyl of one molecule and the 5' phosphate of another.

During this study, most ligations involved a 1:3 molar ratio of vector to insert for cohesive compatible fragments and a 1:10 ratio for blunt-ended fragments. All ligations were performed in 10-20 µl at 16 ºC overnight, with 1-2 µl of T4 DNA ligase buffer, 0.5-1 µl of 10 mM ATP and 5 U of T4 DNA ligase (2.1.2).

2.2.8.3 Dephosphorylation of DNA 5' termini

Dephosphorylation of the 5' terminal phosphate groups of digested plasmids avoids recircularisation and therefore decreases the background of false positives during colony screening (2.2.12).

Dephosphorylation of 1-3 µg of DNA (0.5-1.5 pmoles of 5' termini for a 6 kbp linearised plasmid) was performed with 1 U shrimp alkaline phosphatase (SAP) at 37 ºC for 30 min (Sambrook et al, 2001). SAP was heat inactivated by incubation at 65 ºC for 15 min. The dephosphorylated plasmid was then purified with the QIAquick PCR Purification kit (Qiagen) or gel extracted with the QIAquick Gel Extraction kit (Qiagen).

2.2.8.4 Primer phosphorylation

Phosphorylation of mutagenic primers for use in the Transformer site-directed mutagenesis kit (Deng et al, 1992) was done in a 20 µl volume reaction, with 2 µl of T4 polynucleotide kinase buffer (2.1.9) and 1 U of T4 polynucleotide kinase (Clontech) per 1 µg primer, at 37 ºC for 60 min. The reaction was stopped by heat-shock inactivation at 65 ºC for 10 min. The phosphorylated primer was stored at -20 ºC for several months. 2 µl of phosphorylated primer was used per PCR reaction (2.2.9).

2.2.9 Polymerase Chain Reaction (PCR)

The polymerase chain reaction consists in the amplification of a specific DNA fragment from a template DNA molecule, eg a vector (Mullis et al, 1987). Modifications of the basic PCR reaction permit the introduction of mutations (deletions, insertions, site-directed mutagenesis). The targeted fragment is amplified by a thermostable DNA polymerase, which extends the fragment delimited by two specific oligonucleotides of 20-40 bp, the PCR primers, which anneal to the 5' and 3' ends of the fragment.

The DNA polymerases Taq polymerase (Stratagene), cloned Pfu polymerase (Stratagene) or Pfu Turbo (Stratagene) were employed. Pfu polymerase was preferred due to its proof-reading activity, which ensures higher fidelity; Taq polymerase was used mainly to confirm the successful cloning of a given insert into a plasmid. PCR reactions were prepared as follows (100-µl reactions): For every combination of PCR primers and template DNA, 10 µl of 10x PCR buffer, 10 µl of 2 mM dNTP, 10-100 ng template DNA, 50 pmoles of each primer and *dd*H_20 up to 99 µl were mixed well and centrifuged in 200-500 µl Eppendorf PCR tubes. Reactions were incubated at 95 ºC for 3-5 min to ensure complete denaturation of the template DNA, and then 2.5-5 U of Pfu or Taq polymerase was added, and the reactions continued. Two PCR machines were used, a Biometra Trio Thermoblock and an Eppendorf Mastercycler 384 thermocycler. The PCR machine from Biometra lacked a heated cover, therefore reactions needed to be overlayed with light, non-miscible mineral oil to reduce evaporation. The Mastercycler could maintain high levels of reproducibility with very small samples (20 µl), thus being ideal for screening of positive clones, because it controlled the temperature on a per-tube basis.

The annealing temperature varied from 55 ºC for site-directed mutagenesis with the QuickChange kit (Stratagene) (2.2.10) to 62-67 ºC for the PCR amplifications of inserts. The exact choice of annealing temperature depended on the melting temperatures of the PCR primers, being 5 ºC lower than the smallest melting temperature. Melting temperatures were provided by the oligonucleotide synthesis service and/or calculated with VectorNTI 6.0 (InforMax, Inc).

The basic program utilised with most reactions is shown in Table 2.3.

Cycle	Conditions
Initial denaturation	95 ºC, 3-5 min
Denaturation	95 ºC, 1 min
Annealing	55-67 ºC, 1 min
Extension	72 ºC, 1 min per 1 kbp
Back to step 1 for 20-25x	
Final extension	72 ºC, 10 min
Hold at 4 ºC	

Table 2.3 PCR reaction program. Typical PCR reaction scheme showing the distinct phases of the reaction; initial denaturation, denaturation (first step in a cycle), annealing (the melting temperature, here 55-67 ºC is the most variable parameter since it depends on the base composition of the PCR primers), extension (last step in a cycle); number of reaction cycles, often between 20-25x and rarely greater than 30 for regular subcloning and/or mutagenesis. The final extension ensures completion of those fragments not fully polymerised at the end of the last cycle.

2.2.10 Site-directed mutagenesis

Point mutations were introduced in the coding sequences of *c-ets-1* and *usf1* genes to substitute cysteines for serines. Three distinct methods were used, the megaprimer PCR (Sarkar et al, 1990), the Transformer kit (Clontech) and the QuickChange kit (Stratagene).

2.2.10.1 Megaprimer PCR

The megaprimer PCR strategy is based on the generation of PCR fragments which contain the desired mutation by amplifying the stretch of DNA spanning from the position of the mutation to the closest end. The change is introduced by one or more PCR primers.

The amplified fragment had to be purified and used as a PCR primer (hence the name 'megaprimer') with a third PCR primer, which annealed to the opposite end. Once the full-length fragment was amplified, it was re-amplified again using only the two end primers to obtain higher yields and validate the strategy.

2.2.10.2 Transformer kit

The Transformer kit uses two oligonucleotide primers that are simultaneously annealed to one strand of a denatured double-stranded template. The so-called mutagenic primer introduces the desired mutation and the other, the selection primer, mutates a unique restriction site in the plasmid, creating a new restriction site or eliminating the site completely. Elongation by T4 DNA polymerase, which lacks strand displacement activity, results in the incorporation of both mutations

in the same newly synthesised strand. The DNA is then digested with a restriction enzyme that cuts at the original restriction site. The uncut, mutated DNA will transform E coli more efficiently than the linear DNA with no mutations.

E coli BMH 71-18 mutS, which is mismatch repair deficient, was used to propagate the mutated plasmid. Two rounds of DNA digestion and transformation ensured that a very high frequency of transformants carried the mutated plasmid, which nearly always contained both mutations - the desired mutation and the selection mutation (Deng and Nickoloff, 1992).

2.2.10.3 QuickChange kit

Primers containing the mutagenic sites are synthesised to only one strand of the double-stranded template. Pfu Turbo DNA polymerase (Stratagene) (2.1.2) extends the mutagenic primers under non-strand displacing conditions, generating double-stranded DNA molecules with one strand bearing multiple mutations and nicks. The nicks are then sealed by the enzyme mix. The PCR reaction products are treated with the restriction endonuclease DpnI (2.1.9) to digest the parental DNA template. This mixture is then transformed into a suitable cloning host.

The basic protocol provided by the manufacturer was modified to increase the efficiency of the mutagenesis by increasing the incubation time of the *DpnI* digestion from 1 h to 4-6 h. This change reduced the number of false positives from 10% to less than 1%.

2.2.11 Gene subcloning

2.2.11.1 Subcloning of the human *c-ets-1* gene in pET15b

Fragments 234-441, 238-441 and 244-441 of human Ets-1 were amplified by standard PCR methods (2.2.9) using the cDNA of human *c-ets-1* gene (2.1.4) as template and the oligonucleotides OE234, OE238, OE244 and OE441 as PCR primers (2.1.5). The first three primers are sense primers and introduce an *NdeI* site; the fourth is a common antisense primer, which introduces a *BamHI* site. Each fragment was digested with *NdeI*/*BamHI* (2.1.2) under standard conditions (2.2.8), purified with the QIAquick PCR Purification kit (2.2.4), and ligated (2.2.8) with linearised pET15b (2.1.4, p 3). The ligation mixtures were precipitated in n-butanol (2.2.5, p 16), resuspended in 10 µl of *dd*H_2O, heat-shock transformed (2.2.2) into *E coli* XL1-Blue (Stratagene) and plated on agar plates supplemented with 50 µg/ml ampicillin. Transformants were assayed for correct

incorporation of the insert by PCR (2.2.9) and their sequence verified by sequencing (2.2.12) with the T7 promoter sequencing primer (2.1.5).

The resulting clones were termed pETSH234, pETSH238 and pETSH244, and comprised a HT and a thrombin-cleavage site at the amino terminus (see Table 2.4).

2.2.11.2 Subcloning of the human *c-ets-1* gene in pETM11, pETM30, pACYC and pProExHTb

Fragments 1-441 (full-length), 51-441, 73-441, 234-441, 238-441, 244-441, 280-441, 301-441 and 335-441 of human Ets-1 were amplified by standard PCR methods (2.2.9) using the cDNA of human *c-ets-1* gene (2.1.4) as template and the oligonucleotides NcoI-OE1, NcoI-OE51, NcoI-OE73, NcoI-OE234, NcoI-OE238, NcoI-OE244, NcoI-OE280, NcoI-OE301, NcoI-OE335(A2) and NcoI-OE335(E2) and KpnI-OE441 as PCR primers (2.1.5, p 4-5), respectively. The sense primers introduce an *NcoI* site and the antisense primer a *KpnI* site. Each fragment was digested with *NcoI*/*KpnI* (2.1.2) under standard conditions (2.2.8), purified with the QIAquick PCR Purification kit (2.2.4), and ligated (2.2.8) either with linearised pETM11 or pETM30 (2.1.4); the PCR fragments 244-441 and 280-441 were also ligated with linearised pACYC (2.1.4); the PCR fragments 335-441, 301-441 and 280-441 were also ligated with linearised pProExHTb (2.1.4). Ligation mixtures were precipitated in n-butanol (2.2.5), resuspended in 10 µl of *dd*H_2O, and transformed (2.2.2) into *E coli* XL1-Blue and plated on agar plates supplemented with 25 mg/ml kanamycin (for the pETM11, pETM30 and pACYC clones) or 50 µg/ml (for the pProExHTb clones). Transformants were assayed for correct incorporation of the insert by PCR (2.2.9) and their sequence verified by sequencing (2.2.12) with the T7 promoter (for pETM11 and pACYC), the T7 terminator (for pETM30) or the M13/pUC reverse (for pProExHTb) sequencing primers (2.1.5).

The resulting clones were termed pETM11-EFL (full-length), pETM11-E51 (51-441), pETM11-E73 (73-441), pETM11-E234 (234-441), pETM11-E238 (238-441), pETM11-E244 (244-441), pETM11-E280 (280-441), pETM11-E301 (301-441), pETM11-E335(A2) (335-441, with A2) and pETM11-E335(E2) (335-441, with E2), and similarly for the pETM30 set of clones; the pETM11 set of clones bore a His6 tag and a TEV cleavage site at the amino terminus, and the pETM30 set of clones a GST and a His6 tag, and a TEV cleavage site at the amino terminus; the pACYC clones, pACYC-E280 and pACYC-E244, had no tags; the pProExHTb clones,

pProExHTb-E335, pProExHTb-E301 and pProExHTb-E280, had an amino-terminal His6 tag and a cleavable TEV site (see Table 2.4).

2.2.11.3 Subcloning of the human *c-ets-1* gene in pQE70

Fragments 1-441 (full-length), 51-441, 73-441, 234-441, 238-441, 244-441, 280-441, 301-441 and 335-441 of human Ets-1 were amplified by standard PCR methods (2.2.8) using the cDNA of human *c-ets-1* gene (2.1.4) as template and the oligonucleotides SphI-OE1, SphI-OE51, SphI-OE73, SphI-OE234, SphI-OE238, SphI-OE244, SphI-OE280, SphI-OE301, SphI-OE335(A2) and SphI-OE335(E2) and BglII-OE441 or HindIII-OE441 as PCR primers (2.1.5), respectively. The sense primers introduce a *SphI* site and the antisense primers introduce either a *BglII* or a *HindIII* site. Each fragment was digested with *SphI*/*BglII* or *SphI*/*HindIII* (2.1.2) under standard conditions (2.2.8), purified with the QIAquick PCR Purification kit (2.2.4), and ligated (2.2.8) with linearised pET15b (2.1.4). The ligation mixtures were precipitated in n-butanol (2.2.5), resuspended in 10 µl of *dd*H_2O, transformed (2.2.2) into *E coli* XL1-Blue and plated on agar plates supplemented with 50 µg/ml ampicillin. Transformants were assayed for correct incorporation of the insert by PCR (2.2.9) and their sequence verified by sequencing (2.2.12) with the pQE-specific sequencing primer (2.1.5).

The resulting *SphI*/*BlgII* clones were termed pQE-HT-EFL, pQE-HT-E51, pQE-HT-E73, pQE-HT-E234, pQE-HT-E238, pQE-HT-E244, pQE-HT-E280, pQE-HT-E301, pQE-HT-E335(A2) and pQE-HT-E335(E2), and bore a non-cleavable His6 tag at the carboxyl terminus; the *SphI*/*HindIII* clones were termed pQE-EFL, pQE-E51, pQE-E73, pQE-E234, pQE-E238, pQE-E244, pQE-E280, pQE-E301, pQE-E335(A2) and pQE-E335(E2), and comprised no tags.

2.2.11.4 Subcloning of the human *c-ets-1* gene in pBAT4

Fragments 1-441 (full-length), 51-441, 73-441, 234-441, 238-441, 244-441, 280-441, 301-441 and 335-441 of human Ets-1 were amplified by standard PCR methods (2.2.9) using the cDNA of human *c-ets-1* gene (2.1.4) as template and the oligonucleotides NcoI-OE1, NcoI-OE51, NcoI-OE73, NcoI-OE234, NcoI-OE238, NcoI-OE244, NcoI-OE280, NcoI-OE301, NcoI-OE335(A2) and NcoI-OE335(E2) and HindIII-OE441 as PCR primers (2.1.5), respectively. The sense primers introduce an *NcoI* site and the antisense primer introduces a *HindIII* site. Each fragment was digested with *NcoI*/*HindIII* (2.1.2) under standard conditions (2.2.8), purified with the QIAquick PCR Purification kit (2.2.4), and ligated (2.2.8) with linearised pBAT4 (2.1.4). The ligation mixtures were precipitated in n-butanol (2.2.5), resuspended in 10 µl of *dd*H_2O, transformed (2.2.2) into *E coli* XL1-Blue and plated on agar plates supplemented with 50 µg/ml ampicillin.

Transformants were assayed for correct incorporation of the insert by PCR (2.2.9) and their sequence verified by sequencing (2.2.12) with the T7 promoter sequencing primer (2.1.5).

The resulting clones were termed pBAT4-EFL (full-length), pBAT4-E51 (51-441), pBAT4-E73 (73-441), pBAT4-E234 (234-441), pBAT4-E238 (238-441), pBAT4-E244 (244-441), pBAT4-E280 (280-441), pBAT4-E301 (301-441), pBAT4-E335(A2) (335-441, with A2) and pBAT4-E335(E2) (335-441, with E2), and comprised no tags (see Table 2.4).

Subcloning of the human *usf1* gene in pETM11, pETM30 and pProExHTb

Fragments 194-260 and 194-310 of human USF1 were amplified by standard PCR methods (2.2.9) using the cDNA of human *usf1* gene (2.1.4) as template and the oligonucleotides NcoI-OU194 and KpnI-OU260 or KpnI-OU310 as PCR primers (2.1.5), respectively. The sense primer introduces an *NcoI* site and the antisense primers introduce a *KpnI* site. Each fragment was digested with *NcoI*/*KpnI* (2.1.2) under standard conditions (2.2.8), purified with the QIAquick PCR Purification kit (2.2.4), and ligated (2.2.8) with either of the linearised vectors pETM11, pETM30 or pProExHTb (2.1.4). The ligation mixtures were precipitated in n-butanol (2.2.5), resuspended in 10 µl of *dd*H_2O, transformed (2.2.2) into *E coli* XL1-Blue and plated on agar plates supplemented with 25 µg/ml (for pETM11 and pETM30) or 50 µg/ml ampicillin (for pProExHTb). Transformants were assayed for correct incorporation of the insert by PCR (2.2.9) and their sequence verified by sequencing (2.2.13).

The resulting clones were termed pETM11-SU (194-260) and pETM11-LU (194-310), and had an amino-terminal HT and a cleavable TEV site; pETM30-SU (194-260) and pETM30-LU (194-310), with a GST and a His6 tag amino-terminal tags and a TEV recognition site; and pProExHTb-SU (194-260) and pProExHTb (194-310), with an amino-terminal HT and a TEV cleavage site (see Table 2.4).

2.2.12 Colony screening

2.2.12.1 PCR screening

Colony screening was performed by PCR (2.2.9) either from a liquid culture of the colony to assay or directly from a colony on selective agar (2.2.1). In both cases, the cells (5-10 ml of liquid culture, or a loopful of cells from a plate) were mixed with 1x PCR buffer (2.1.9.1) and incubated for 5-10 min at 95 ºC to disrupt the cells and liberate the DNA. Then, the other components of the PCR

reaction were added and the reaction continued. Taq polymerase (2.1.2) was used for colony screening. The PCR products were analysed on a 2% (w/v) agarose gel (2.2.6) with ethidium bromide or SYBR Green staining (2.1.9.1).

	Construct								
Vector	**E335A/E**	**E301**	**E280**	**E244**	**E238**	**E234**	**E73**	**E51**	**FL**
pET15b	Y	-	-	Y	Y	Y	-	-	-
PETM11	Y	Y	Y	Y	Y	Y	Y	Y	Y
PETM30	Y	Y	Y	Y	Y	Y	Y	Y	Y
pQE70	Y	Y	Y	Y	Y	Y	Y	Y	Y
pQE70-HT	Y	Y	Y	Y	Y	Y	Y	Y	Y
PBAT4	Y	Y	Y	Y	Y	Y	Y	Y	Y
PACYC9d	-	-	Y	Y	-	-	-	-	-
pProExHTb	Y	Y	Y	-	-	-	-	-	-

	SU	**LU**
pET15b	Y	Y
PETM11	Y	Y
PETM30	Y	Y
pProExHTb	Y	Y

Table 2.4 Constructs subcloned and assayed for protein expression. Variety of expression constructs subcloned during this study. The upper panel concerns with Ets-1 clones, and the bottom panel with the USF1 clones (left) and Kreisler (right). The top row lists the different Ets-1 fragments, E335A/E (Ets-1 335-441, with an A or an E at position 2); E301 (Ets-1 301-441); E280 (Ets-1 280-441); E244 (Ets-1 244-441); E238 (Ets-1 238-441); E234 (Ets-1 234-441); E73 (Ets-1 73-441); E51 (Ets-1 51-441); FL (Ets-1 1-441, full-length); the row on top of the bottom panel lists the USF1 clones: SU ('short USF1,' USF1 194-260) and LU ('long USF1,' 194-310). A 'Y' labels all cloned constructs; otherwise, '-' labels not cloned constructs.

2.2.12.2 Digestion

An alternative method for colony screening is the restriction digestion of miniprepared plasmid DNA (2.2.8).

Plasmid DNA was purified with the QIAquick DNA Miniprep kit (2.2.4), restricted with suitable enzymes (2.1.2) and run on a 1% (w/v) agarose gel (2.2.6). Digestion products were stained with ethidium bromide or SYBR Green staining (2.1.9.1). This method was used only with inserts whose size was larger than 500 bp, and the endonucleases (2.1.2) were the same as for cloning the fragment.

2.2.12.3 Quick lysis

The 'quick lysis' method takes advantage of the different electrophoretic mobility of plasmids with and without an insert; if the size of the insert is larger than 500 bp, the plasmid that contains it is

visibly shifted up with respect to the control plasmid.

For this method, 5-10 ml LB cultures were grown overnight (2.2.1), centrifuged the next morning in an Eppendorf tabletop centrifuge at 5 krpm for 10 min, resuspended in quick lysis solution (2.1.9.1) and incubated at 37 ºC for 15 min for complete cell lysis. Then, the cell lysate is centrifuged at 14 krpm for 30 min and the supernatant is loaded on a 1% (w/v) agarose gel for analysis (2.2.6). Because genomic DNA is also released upon cell lysis, the supernatants are highly viscous and care needs to be exercised for correct sample application.

2.2.13 DNA sequencing

All the sequencing was carried out by SeqLab. Samples were prepared as recommended by the company; in detail, 200-500 ng of pure DNA was re-purified with the QIAquick Plasmid Miniprep kit (2.2.4), ethanol-precipitated (2.2.5) to remove salts, and lyophilised in a SpeedVac (2.2.14). The primers used for sequencing are listed in section 2.1.5.

2.2.14 Lyophilisation of DNA samples

DNA samples for sequencing (2.2.13) were lyophilised (freeze-dried) in a SpeedVac machine with the temperature control off. DNA samples were ethanol-precipitated (2.2.5) in colourless 1-ml Eppendorf tubes and placed on the SpeedVac holders with the lid open, then vacuum was applied with the rotor at full speed during 2-4 h. Lyophilised samples were sealed with Parafilm and stored at -20 ºC.

2.3 Protein Methods

2.3.1 Storage of proteins

2.3.1.1 Flash-freezing

Protein samples were aliquoted in 50 µl fractions and immersed in liquid nitrogen. Depending on the particular sample, up to 50% (v/v) glycerol was added to the protein sample as cryoprotectant. All samples were stored at -80 ºC.

2.3.1.2 Lyophilisation

Elution fractions from RPC chromatography (2.3.10) were frozen in liquid nitrogen in 50-ml Falcon tubes and lyophilised in a lyophiliser machine. The dry-frozen pellet was sealed with Parafilm and

stored at -80 ºC.

2.3.2 Determination of protein concentration

Protein concentration was measured by the Bradford method (Bradford, 1976) with the Bradford solution provided by BioRad. A suitable volume of a protein sample (from 1-800 µl depending on purpose) was mixed with *dd*H_20 to a final volume of 800 µl, and then added to 200 µl of Bradford solution. The concentration was determined by measuring the absorbance at 595 nm with a Pharmacia spectrophotometer by interpolation into a BSA standard curve. Absorbance in the ultraviolet regime was not used for protein determination since protein-DNA complexes absorb in the 260-280 nm region, masking the signal from the protein.

2.3.3 Protein concentration by ultrafiltration

Protein samples were concentrated in different devises according to their volume. Samples larger than 60 ml in volume were concentrated either in an Ultrafree PF-60 (Millipore) or in a Device 8200 (Millipore), with membranes of 5-10 kDa mwco. Samples below 60 ml and up to 12 ml were concentrated in 12 ml Centriprep (Millipore) centrifugal filter unit, 3 or 10 kDa mwco. Samples between 12 and 2 ml were concentrated in 2 ml Centricon units (Millipore), 3 or 10 kDa mwco, or in 4 ml Ultrafree units with 5 kDa mwco. Usually, protein concentration was performed at 4 ºC.

2.3.4 Protein precipitation

Before running a protein gel (2.3.5) with samples containing GuHCl, this needs to be removed since it will form a precipitate when treated with SDS. This can be achieved by trichloroacetic acid (TCA) precipitation.

Protein samples in 6 M GuHCl were diluted with ddH20 up to 100 µl. Then, 1 vol of 20% (v/v) TCA was added to the protein sample, and these were incubated 30 min on ice. Next, they were centrifuged in an Eppendorf tabletop centrifuge at 4 ºC for 15 min, the supernatant was carefully removed and 300 µl of ice-cold acetone was added, followed by another centrifugation at 4 ºC for 5 min. Finally, the supernatants were removed and the pellets dried. Samples could then be resuspended in SDS PAGE sample buffer (2.1.9) for protein electrophoresis (2.3.5).

2.3.5 Protein gel electrophoresis

2.3.5.1 Novex system

Almost all protein electrophoresis was conducted on a Novex system (Novex, Invitrogen). It consists

of an electrophoresis chamber with a pair of electrodes and an additional cassette for electroblotting (2.3.6) and a programmable control unit integrated with the power supply.

Concentrated buffer stocks and most types of precast gels (eg NuPAGE 10% and 4-12% Bis-Tris gels for denaturing electrophoresis and 6-10% TBE gels for native electrophoresis) were provided by the manufacturer. Denaturing gels were run at 200 V and 125 mA during 35 min, and native gels at 100 V and 6-12 mA for 90 min.

2.3.5.2 Phast system

The Phast system (Pharmacia) utilises smaller gels and has better resolution than other systems, and is therefore ideal to visualise much diluted or very small samples.

Native PhastGel gradient gels were run according to the specification of the manufacturer. Protein bands were developed with Coomassie blue staining and DNA bands with ethidium bromide or SYBR Green (2.1.9).

2.3.6 Immunodetection by Western blot

Western blot or immuno blot is the transfer of biomolecules to the surface of a thin membrane matrix, where they stay immobilised and can then be developed.

Western blots were performed with the supplies and reagents from Novex in the manner specified by the manufacturer. In brief, either a denaturing or native protein gel is electroblotted onto a PVDF (polyvinylidene difluoride) membrane via the XCell II blot module, using 0.5x NuPAGE transfer buffer (2.1.9) and, after completion of the transfer, the membrane is developed. For development, alkaline phosphatase-linked secondary antibodies were used, which catalyse the formation of a purple chromophore when supplied with NBT (nitroblue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3-indolylphosphate p-toluidine salt).

2.3.7 Electrophoretic mobility shift assay

2.3.7.1 Novex 8-10% TBE gels

The commonest electrophoretic mobility shift assay (EMSA) or, simply, band shift assay, was performed by running 10-20 μg of protein-DNA complex on a 8 or 10% TBE gel (2.3.5).

ddH_2O was added to each of the samples to total a volume of 20 μl, and then 4 μl of 6x TBE

loading buffer (2.1.9). Samples were kept on ice at all times, and the electrophoresis chamber was surrounded with ice to avoid excessive heating. Bands were visualised with Coomassie and ethidium bromide or SYBR Green (2.1.9).

2.3.7.2 ECL kit

The ECL kit (Pharmacia) offered a more sensitive mechanism to detect DNA-protein complexes, which involves the use of oligonucleotides with one or more biotinylated bases, which are recognised by streptavidin-bound antibodies, and this in turn by a secondary antibody covalently bonded to horse radish peroxidase (HRP) (2.3.6).

Bands were developed by chemiluminescence, as outlined by the manufacturer. Briefly, 1 ml of Solution 2 and 1 ml of Solution 1 were spread over the membrane; after 1 min, the membrane was drained off and developed on a ChemiDoc system (BioRad).

2.3.8 Solubilisation of protein inclusion bodies

Inclusion bodies are intracellular aggregates that accumulate overexpressed proteins in an insoluble form. These proteins are generally misfolded, thus upon cell lysis they need to be solubilised and refolded to acquire full functionality.

Inclusion bodies were solubilised with the denaturing buffers UB, GU or NL (2.1.9), adjusting the final concentration of protein to 5-10 mg/ml. The most complete solubilisation was achieved with GuHCl and N-lauryl sarcosine (2.1.9), whereas even 8 M urea was rather ineffective to solubilise thoroughly the inclusion bodies of the insoluble Ets-1 constructs. To determine the best solubilisation conditions, systematic experiments were carried out whereby an aliquot of inclusion body pellet was vortexed in 100 µl of increasing concentrations of chaotrope (2-6 M GuHCl, 5-8 M urea or 1-2% [v/v] N-lauryl sarcosine) and then centrifuged in an Eppendorf tabletop centrifuge; the more complete the solubilisation the higher the effectiveness of the chaotrope.

2.3.9 Toxicity test

To assess whether a new construct is cytotoxic to E coli a plasmid stability test, or toxicity test, is performed (based on 'pET System Manual,' 8th ed, 1999, Novagen). The toxicity test consists in growing the cells transformed with the new construct on two LB plates (2.1.8) with 1 mM IPTG (2.1.1) and in the presence or absence of antibiotics, and on two control LB plates without IPTG and in the presence or absence of antibiotics. As the induction of protein expression detains cell

division, cells actively expressing the construct will form colonies on the two control LB plates without IPTG, while < 2-0.1% of colonies will develop on the LB plates with IPTG.

Cells transformed with a new construct were grown in 100 ml LB medium with appropriate antibiotics at 37 ºC until an OD_{600} of 0.6-0.7; at this point, an aliquot of the culture was diluted 10^4-10^6 and plated on four LB plates as explained above. The plates were inspected the next morning and the approximate number of colonies on each of the plates was counted. In one case where the number of colonies on the LB plate with antibiotics and 1 mM IPTG was moderately high (5% of the corresponding plate without IPTG), DNA was miniprepared (2.2.4) and examined by gel electrophoresis on a 1% (w/v) agarose gel (2.2.6) and by sequencing (2.2.13).

2.3.10 Protein production

2.3.10.1 Expression tests

The following strategy (or, where indicated, a modification thereof) was applied with every new expression construct: First, a 10-15 ml LB culture with appropriate antibiotics was grown at 37 ºC overnight (2.2.1); next morning, cells were centrifuged in a tabletop centrifuge at 5 krpm for 10 min, at 4 ºC, washed with 1x PBS (2.1.9) and resuspended in new LB medium with freshly prepared antibiotics; then, 1 ml of the suspension served as inoculum for a 25-ml culture, which grew at 37 ºC at 250-280 rpm until OD_{600} was 0.6-0.7; an aliquot was kept on ice as non-induced control, and the rest was induced by adding 0.2-1 mM IPTG (2.1.1); after 1-2 h, cells were harvested by centrifugation at 5 krpm for 10 min (Sorvall GSA), and cell pellets kept on ice before processing. Depending on the expression test, some modifications were introduced to the basic protocol (on a case-by-case basis); some of them were: Cells grown in different media, as Terrific broth (2.1.8), or in buffered LB medium (with 50 mM MOPS or Tris); temperature was adjusted to 16-23 ºC to enhance protein solubility; larger ratio of exposed surface to volume to increase aeration; acceleration of the shaker angular speed to increase aeration.

2.3.10.2 Small-scale production

For purposes of analysis (eg assessment of solubility properties, initial purification schemes, molecular weight estimation), small-scale expression cultures were grown as follows: 50-100 ml LB cultures with appropriate antibiotics were inoculated with 1-3 ml of an overnight preculture (previously centrifuged and resuspended in fresh medium) in a 500 ml flask, and cells grew for 3-4 h at 37 ºC with moderate to vigorous agitation (225-275 rpm), until OD_{600} was 0.6-0.8. At this point, 0.2-1 mM IPTG (2.1.1) was added to initiate expression. Cells were harvested before 5-6 h,

centrifuged at 5 krpm for 15 min (Sorvall GSA), washed with 1x PBS (2.1.9) and either processed immediately or frozen in liquid nitrogen and stored away at -80 ºC. The yield of these small-scale production experiments depends on the particular combination of expression vector and insert.

2.3.10.3 Large-scale production

For preparative scale protein production, 2-8 litres of LB medium with appropriate antibiotics were inoculated with 100-200 ml of an overnight preculture (previously centrifuged and resuspended in fresh medium), in 5 litre flasks with no more than 1.5 litres per flask. Cells grew at 37 ºC with vigorous stirring (275-300 krpm) until OD_{600} was 0.6-0.8. At this point, 0.2-1 mM IPTG (2.1.1) was added to initiate expression. Cells were harvested after 5-6 h, centrifuged at 5 krpm for 15 min (Sorvall GSA), washed in 1x PBS (2.1.9) and either processed immediately or frozen in liquid nitrogen and stored away at -80 ºC. The yield of these large-scale productions depends on the particular combination of expression vector and insert.

2.3.10.4 Cell lysis

Cell pellets (either recently centrifuged or thawed) were lysed by sonication. Cells were first thoroughly resuspended in 20 ml lysis buffer per litre of culture, and T4 lysozyme (2.1.2) and 0.2% (w/v) CHAPS were added to ease cell breakage. After incubation on ice for 30 min, cells were sonicated for 2-5 min at 25% maximum power for as many times as necessary to disrupt the cells completely (typically, 3-4 times), which was monitored by measuring the protein concentration of the released supernatant. Once lysed, the cell suspension was centrifuged at 18 krpm for 1 h at 4 ºC, and the supernatant and the insoluble fraction saved for further analyses.

2.3.10.5 Cell-free expression

In two particular cases, expression tests were carried out in a cell-free system. A cell-free system, or cell-free translation system, is an aqueous solution that contains the transcription and translation machinery (ribosomes and accessory factors) and energy sources (ATP, GTP) necessary to translate an mRNA into a protein (Zubay, 1973, with modifications). As there are no cells involved, the cell-free systems avoid culture growth and manipulation and are faster in producing yield; nonetheless, at present they produce too small yields when compared with cell-based systems.

Expression tests with pBAT4-E244 and coexpression of pBAT4-E244 and pET15b-LU was performed in a Rapid Translation System 500 (RTS500, Roche) at the default temperature of 30 ºC

during 3 h. The reaction device was rinsed with 70% (v/v) ethanol to disinfect it, and filled with the reaction mixture (50 μl, prepared as recommended by the manufacturer). At all times during the test, the reaction was stirred to ensure correct aeration and exchange of nutrients and waste products. Results were analysed by protein gel electrophoresis (2.3.5).

2.3.11 Protein purification

2.3.11.1 Solvent exchange

Solvents were fully exchanged by dialysis or desalting chromatography. Solvent exchange by dialysis involved the use of a SpectraPor 3-10 kDa mwco dialysis membrane or a Slide-A-Lyzer dialysis cassette (Pierce) with similar cutoff and capacities 0.5-3 ml and 3-15 ml. Sample volumes were in all cases restricted to less than 20 ml and the dialysis buffer (usually, larger than 100x the sample volume) was replaced 2-3 times to ensure full exchange.

Another method for solvent exchange is to use a desalting column, such as the PD10 columns (Pharmacia). PD10 columns are made up of Sephadex G-25, which has a molecular exclusion limit of 5 kDa, and, therefore, retain solvent molecules but not proteins. For desalting chromatography, protein samples were 2.5 ml in volume or less; after equilibration of the PD10 column with 20 vol of the appropriate buffer, the protein sample is applied and, once completely absorbed, enough buffer is added to make up 2.5 ml of sample volume. The protein eluted between 2-5.5 ml.

2.3.11.2 Chromatography

Purification of proteins and protein complexes was performed by fast performance liquid chromatography (FPLC) and high performance/pressure liquid chromatography (HPLC), in any of these machines: Äkta purifier, Äkta primer, Äkta FPLC or a Biosys machine. In all cases, the buffers were carefully degassed by stirring under vacuum or by sonication and purged with molecular nitrogen.

2.3.11.2.1 Unspecific DNA affinity

The unspecific DNA affinity column was prepared from sonicated salmon sperm, attached to the surface of the matrix by chemical bonding. Samples had to be diluted or buffer exchanged to reduce ionic strength. The affinity resin was purchased as a powder (Pharmacia), which was conditioned by first soaking it in 20% (v/v) ethanol, and then washing it repeatedly with *dd*H_2O after casting the resin onto a plastic column. Before loading the sample, the DNA column was equilibrated with 20

vol DNA buffer A (2.1.9.2) and wash with 5 vol DNA buffer B (2.1.9.2). Next, column was re-equilibrated in buffer A. The sample was loaded three times to ensure maximum binding, and the flow-through kept for analysis. Unbound material was washed off with 10 vol of buffer A. Elution was conducted in a single step, with 4 vol of buffer B. Samples and supplies were kept at 4 ºC at all times.

2.3.11.2.2 Heparin sepharose

Heparin is a highly sulphated glycosaminoglycan that mimics the binding properties of DNA. Thus, it is used to purify DNA-binding proteins. At the same time, and given the high density of charges on heparin surface, separation can also be achieved as an ion exchanger. The two modes of binding, as DNA-mimicking surface or as an ion exchanger, explain their efficiency. It also explains why ionic strength has to be kept low and pH within neutrality.

Samples were diluted or buffer exchanged before purification with heparin sepharose. The resin was either prepared from powder (Pharmacia) or as a pre-packed HiTrap heparin column (Pharmacia). In either case, the column was equilibrated with 10-20 vol of Heparin buffer A (2.1.9.2), wash with Heparin buffer B (2.1.9.2), and re-equilibrated back to buffer A. Samples were loaded three times to ensure maximum binding, and flow-through kept for analysis. Elution was conducted in a single step, with 4 vol of buffer B, in the gravity flow mode; as a HiTrap column, elution was conducted by a 0-60% gradient on buffer B. Samples and supplies were kept at 4 ºC at all times.

2.3.11.2.3 Ni-NTA affinity chromatography

Immobilised metal affinity chromatography (IMAC) makes use of the binding properties of metals towards proteins for purification purposes; nickel-nitriloacetic (Ni-NTA) resin (Qiagen) contains chelated nickel, which is able to specifically bind to stretches of polyhistidine in proteins. Many expression systems include His6 tags either at the N- or the C-terminus or on both. 1 ml of resin needs to be used per every 20 mg of tagged protein, since the maximum capacity of 50 mg protein/ml resin is highly dependent on cases.

Once the resin was cast on a BioRad disposable plastic column with frit, it was rinsed with *dd*H_20 to remove the preservative 20% (v/v) ethanol and equilibrated with Ni-NTA loading buffer (with 5-10 mM imidazole) (2.1.9.2). The samples should not contain DTT or other reducing agents in high concentration, or be too dense with cell debris or DNA, because those would strip the nickel

off the column. Once equilibrate, the sample was loaded three times, and the colour of the resin was inspected since its turning white indicates protein binding, while no change of colour usually correlates with no binding. After loading, the column was washed with loading buffer for 10-20 vol, then with HS buffer (2.1.9.2) for 5 vol and finally with the loading buffer for 5 vol. Elution (2.1.9) was achieved by 4 vol of elution buffer (2.1.9.2). Elution samples were collected drop-wise for further processing.

2.3.11.2.4 GST affinity chromatography

The glutathione-S-transferase affinity (GST) resin (Pharmacia) is an agarose support derivatised with glutathione. The GSTTrap columns (Pharmacia) provide the same support for HPLC purposes.

Samples had to be diluted or buffer exchanged before being loaded onto a GST affinity column, since high ionic strengths would abrogate binding. The 5-ml GSTTrap (Pharmacia) column format was the preferred, since the kinetics of binding of GST to the resin is rather slow and tight control of loading conditions was necessary. Reducing conditions were ensured with 10 mM DTT in all steps. The GSTTrap column was preconditioned with 10 vol of *dd*H_2O (to remove the preservative) and 10 vol of GST loading buffer (2.1.9.2). The sample was loaded onto the column at a flow rate of 0.1-0.2 ml/min, and passed twice or three times over the column. The column was washed by flowing 5 vol of GST loading buffer and 5 vol of GST wash buffer (2.1.9.2). Elution was achieved with 4 vol of GST elution buffer (2.1.9.2). Eluate fractions were pooled together and kept at 4 °C for further processing.

2.3.11.2.5 Anion exchange and cation exchange

Ion exchange chromatography is based on the electrostatics properties of proteins, which can bind to the charged groups exposed in the surface of the exchanger. To ensure electrostatic binding, the total ionic strength needs to be reduced.

Generally, 100 mM NaCl or KCl was include in the anion exchange buffers. The columns for anion exchange chromatography (AIEX) and cation exchange chromatography (CIEX) were MonoQ HR 10/10 and MonoS HR 10/10 (Pharmacia). After sample loading and column equilibration with AIEX/CIEX buffer A (2.1.9.2), elution was achieved by a linear gradient from 0-100% AIEX/CIEX buffer B (2.1.9.2). Elution fractions containing the protein of interest were pooled together.

2.3.11.2.6 Gel filtration

Gel filtration (or size exclusion) (GF) is based on the sieving properties of porous, inert matrices, when flowing a macromolecule solution through them. The matrices are characterised by an exclusion limit that sets the molecular weight of the largest protein that is still delayed in the column.

GF was conducted in both analytical and preparative scale. Analytical runs were done on a Superdex 200 HR 10/30 and a Superdex 75 HR 10/30 (Pharmacia), and preparative runs on a Superdex 200 HR 26/60 (Pharmacia). The two HR 10/30 columns were run at a flow-rate of 0.25 ml/min, and the preparative column at 2-3 ml/min.

2.3.11.2.7 Reverse phase

Reverse phase chromatography (RPC) separates molecules according to their hydrophobicity via their ability to establish strong interactions with a hydrophobic matrix when applied in a polar solvent. These molecules are eluted by a gradient of an apolar compound, usually ACN (2.1.9.2).

This chromatography was typically utilised as a polishing step for proteins purified from inclusion bodies, as a necessary step to lyophilise the pure protein for future use. All RPC runs were performed with a Resource RPC column. Samples were loaded in 6 M GuHCl and washed with RPC buffer A (0% ACN) (2.1.9) to remove GuHCl (2.1.9.2) and protein contaminants. Elution was carried out with a 0-60% linear gradient of ACN.

2.3.12 Refolding strategies

A major aim of a refolding protocol is to favour the refolding reaction over aggregation. In its simplest form, this involves keeping protein concentration at or below 0.1 mg/ml.

2.3.12.1 Solubilisation

Protein inclusion bodies were obtained by cell lysis and clarification, followed by centrifugation at 21 krpm for 1 h in a Sorvall SS34 rotor, at 4 ºC. Then, the pellets were solubilised in UB or GU buffer (2.1.9.2) to make a 10 mg/ml of fully unfolded protein. This solution was refolded according to one of the following protocols.

2.3.12.2 Refolding by dialysis

The most widely used technique for refolding is dialysis.

For this, refolding buffers were prepared (based on TN and HN, 2.1.9.2). The first protocol to be tried was one-step dialysis from 6 M GuHCl or 8 M urea down to 0 M chaotrope; second, multiple-step dialysis, eg from 6 M GuHCl down to 3 M, down to 1 M, down to 0 M chaotrope, in successive exchanges of dialysis buffer.

2.3.12.3 Refolding by dilution

This refolding strategy is similar to the refolding by dialysis, the difference being that here the elimination of the chaotrope is almost instantaneous upon 20-fold (or even larger) dilution. The dilution buffer may be TN or HN (2.1.9.2) or a variant thereof.

Denatured protein samples at 5-10 mg/ml was dropped in a volume 20x larger of refolding buffer in 50-100 ml aliquots, at 4 ºC and with soft shaking. When the concentration of GuHCl reached 0.6 M or the protein concentration became close to 0.1 mg/ml, addition of further denatured protein solution was stopped. Once addition was finished, the refolding reaction had to be dialysed or buffer exchanged to remove the chaotrope and stabilise the refolded protein.

2.3.12.4 Refolding by gel filtration

The principle of refolding by gel filtration is that the buffer in which the protein was initially dissolved is fully removed immediately after sample application (Ferre-D'Amare et al, 1995).

This strategy was applied in analytical scale on a Superdex 200 HR 10/30, loading 100 ml of denatured protein in 6 M GuHCl. The eluate could be concentrated for further purification.

2.3.12.5 On-column refolding

This method makes use of the observation that proteins may refold efficiently by passing them over an affinity column.

This refolding protocol was used with several His6-tagged Ets-1 constructs, which were loaded onto a Ni-NTA column in 6 M GuHCl. The chaotrope was removed by subsequent addition of refolding buffer.

2.3.12.6 Additives for refolding

Some of the refolding protocols described above were carried out with modified TN and HN buffers (2.1.9.2) containing additives shown to aid proper refolding. In the refolding procedures based on dialysis (2.3.12.3) and dilution (2.3.12.3), the refolding buffers TN and HN (2.1.9.2) were used as such or supplemented with: 0.1-1 M NaCl, 5-20% (v/v) glycerol, 5-20% (v/v) sucrose, 1-5 mM L-arginine, 50 mM L-glycine, cations (Ca^{2+}, Mg^{2+}, Zn^{2+}, MES, HEPES), anions (PO_4^{2-}, SO_4^{2-}, Tris), redox compounds (GSH:GSSG in a 10:1 molar ratio, at 5 mM), detergents (N-lauryl sarcosine), 0.1 M buffer salt (Tris, acetate, citrate, MES, HEPES) at pH ranging from 4.5 to 9.

2.3.12.7 Refolding in presence of DNA

Ligands can be used to stabilise proteins or to help them refold. Refolding of Ets-1 and USF1 made use of their DNA binding elements in a 1:1-1.5 molar ratio of protein to DNA. Either of the refolding buffers TN or HN was used. Denatured USF1 (either SU or LU) solution in 6 M GuHCl was added to 20 ml of refolding buffer in the presence of excess DNA, in small aliquots (100-200 μl), until a final concentration of 20-30 μM in complex. Excess protein usually precipitated within hours and upon concentration. Therefore, filtration of the sample was necessary to remove aggregated material (2.3.3). The remaining GuHCl was removed by dialysis (2.3.11) or via a desalting column (2.3.11). To refold Ets-1, a 10 mg/ml solution of the denatured protein in 6 M GuHCl was added in 20-100 μl aliquots to a pre-formed USF1/DNA complex. The refolding reaction kept for 1 h on ice, and then it was concentrated in a 12 ml Centriprep unit with 3kDa mwco until the volume reduced to 1-2 ml, poured in a 4 ml Ultrafree device with 5 kDa mwco and was further concentrated to about 10-12.5 mg/ml. At this concentration, the sample started to precipitate.

2.3.13 Tag removal

Tags are useful aids for expression and/or purification, but they are generally best removed for crystallisation. Except the pQE-HT constructs, all other tagged constructs provided a proteolytic site for their removal; pET15b introduces a thrombin site and pETM11, pETM30 and pProExHTb a TEV site. Proteolytic cleavage of the tags was performed in all cases after the initial purification step (Ni-NTA or GST affinity chromatography, 2.3.11.2) upon overnight dialysis, at 4 ºC. The ratio of protease to substrate was adjusted to 1:50-100 (w/w).

2.3.14 Filtration of protein samples

To remove precipitated protein during purification steps, either centrifugation or filtration were

used. Protein solutions were filtered with 50 ml Steriflip units (Millipore), which consist of a 50 ml Falcon tube attached to a 0.22 μm filter. Protein concentration was measured before and after every filtration to estimate the loss of material.

2.3.15 Limited proteolysis

Limited proteolysis involves the incomplete digestion of a protein by a protease, which can be examined as a function of time or of protease concentration. The partial digestion of a protein defines those fragments resistant to proteolysis and, therefore, which are expected to have a compact structure. In contrast, flexible regions are prone to be digested. The identity of the fragments can be analysed by Edman degradation. Information from these experiments can aid the design of constructs for structural biology.

Time courses of trypsin digestion were performed on E280ΔS/LUΔS/DNA, LUΔS/DNA and E280ΔS. Trypsin (La Roche, lyophilised) was resuspended in trypsin buffer (2.1.9.2) to a concentration equivalent to 10 U/μl. The reactions were performed at 15 ºC in 20 ml samples with protein or protein complexes at a concentration of 0.5 mg/ml (10 μg of protein per sample) by addition of 1 μl of trypsin. Reactions were stopped with 4 ml of 6x electrophoresis sample buffer (2.1.9.2) and boiling at 70 ºC for 15 min. Protein gel electrophoresis of the digestions permitted the analysis of the results (2.1.9.2).

2.3.16 Dynamic Light Scattering

Dynamic light scattering (DLS) is concerned with the counting statistics of single photons. A correlation function of the intensity scattered by a protein solution over time with itself analyses the seemingly random 'noise' of the signal. This analysis renders the diffusion coefficient as the fundamental parameter describing the mobility of the particles in solution. Then, by assuming a geometrical model for the shape of the particle (eg a hard sphere or an elongated ellipsoid) and the viscosity of the solvent, the distribution of apparent radii of gyration can be calculated. Indeed, the polydispersity, ie the quotient of the radius of gyration over its root mean square deviation expressed as percentage units, is a strong indicator of the presence of multiple particles in solution.

DLS experiments were conducted in a DynaPro machine (Protein Solutions) thermostated at 20 ºC, and data were processed with the software provided by the manufacturer. Samples were centrifuged in an Eppendorf tabletop centrifuge at maximum speed (~14 krpm), which can remove

dust without removing high molecular weight aggregates. This is important to obtain an unbiased analysis of the polydispersity of the sample. Protein samples were concentrated to 1 mg/ml and measurements done with 20 µl of protein solution in a 0.1 mm quartz cuvette, by duplicate and with greater than 25 data points per scan. All experiments were performed at 20 ºC, except when explicitly stated otherwise.

2.3.17 Crystallisation

The crystallisation of proteins is a highly cooperative, self-assembly process by virtue of which a supersaturated solution of a protein gives rise to crystal nuclei that can further grow by addition of new molecules to their surface. The sitting drop method was used the most for convenience, because it is fast to set and seal. The sitting drop method is a crystallisation method based on vapour diffusion. Equilibration proceeds by evaporation of volatile compounds (water and organic molecules). If diffusion occurs from the droplet towards the reservoir, the droplet reduces its volume and, therefore, all the components will be concentrated.

Crystallisation solutions and supplies are described in the Material section (2.1.10). A concentrated solution of a protein or protein-complex (10-12.5 mg/ml for all Ets-1/USF1 ternary complexes and 1.5-2 mg/ml for E280S and E301S) was centrifuged in an Eppendorf tabletop centrifuge at 13 krpm for 5 min at 4 ºC to eliminate precipitates or dust. Drops were commonly set in sitting drop Linbro plates (2.1.10), but also in hanging drop plates (2.1.10). For each crystallisation plate (2.1.10), 1 µl of protein solution was mixed with 1 µl of reservoir per well. The plate was sealed with sealing tape (2.1.10) (or each well sealed with a siliconised cover slide, if in hanging drop plates) and stored at 20 ºC.

Each day during the first week, the plates were examined; then, once or twice every week during the first and second month; henceforth, once every two weeks. Initial screens included the following commercial screens: Cryo, Lite, Low Ionic Strength, PEG/Ion (Haptom Reseach) (2.1.10) and Cryo I and II (Emerald Biosciences, 2.1.10), which provide suitable conditions for the crystallisation of protein-DNA complexes; if sufficient sample was available, the crystal screens I and II were used (Hampton Research, 2.1.10). Additionally, a systematic grid screening was performed with selected complexes to fine tune initial crystallisation conditions.

2.4 References

Bradford, MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-54.

Brinboim, HG, et al (1979) *Nucleic Acids Res* **7**: 1512-1513.

Brunner, M., et al (1987) Promoter recognition and promoter strength in the Escherichia coli system. *EMBO J* **6**(10): 3139-44.

Bullock (1987) *BioTechniques* **5**: 376-399.

Carstens, et al (1999) *Strategies* **12**: 49-51.

Deng, WP, et al (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal Biochem* **200**(1): 81-8.

Dower, WJ, et al (1988) High efficiency transformation of E coli by high voltage electroporation. *Nucleic Acids Res* **16**(13): 6127-45.

Hanahan, D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**(4): 557-80.

Morrison, DA (1977) Transformation in Escherichia coli: cryogenic preservation of competent cells. *J Bacteriol* **132**(1): 349-51.

Mullis, KB, et al (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**: 335-50.

Sambrook, J, et al (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press.

Sambrook, J, et al (2001) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press.

Sarkar, G, et al (1990) The 'megaprimer' method of site-directed mutagenesis. *Biotechniques* **8**(4): 404-7.

Short, JM, et al (1988) Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res* **16**(15): 7583-600.

Sieweke, MH, et al (1996) MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* **85**(1): 49-60.

Sieweke, MH, et al (1998) Cooperative interaction of ets-1 with USF-1 required for HIV-1 enhancer activity in T cells. *EMBO J* **17**(6): 1728-39.

Singer, VL, et al (1999) Comparison of SYBR Green I nucleic acid gel stain mutagenicity and ethidium bromide mutagenicity in the Salmonella/mammalian microsome reverse mutation assay (Ames test). *Mutat Res* **439**(1): 37-47.

Studier, FW, et al (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**(1): 113-30.

Studier, FW, et al (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**: 60-89.

Wood, WB (1966) Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol* **16**(1): 118-33.

Yanisch-Perron, C, et al (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**(1): 103-19.

Zubay, G (1973) In vitro synthesis of protein in microbial systems. *Annu Rev Genet* **7**: 267-87.