Solution Structures of Proteins and Complexes Involved in the Regulation of Eukaryotic Gene Expression

Thesis Submitted for the Degree of

Doctor of Philosophy

At the University of Leicester

by

Lorna Catherine Waters, BSc (Essex)

Department of Biochemistry

University of Leicester

January 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

Firstly, I would like to offer my immense gratitude to my supervisor Dr. Mark Carr for providing me with the opportunity to conduct this research, and for all his guidance and patience. I also extend my thanks to Dr. David Heery and Dr. Karl Heinz Klempnauer for their expertise and input. In addition, I would like to thank my committee members Dr. Christine Wells and Dr. Peter Moody for their advice. I would also like to offer my thanks to the Biotechnology and Biological Sciences Research Council for funding this studentship.

I would like to offer my sincere thanks to all those who have directly contributed to the work described in this thesis. In particular, I would like to acknowledge; Dr. Vaclav Veverka, Dr. Phil Renshaw, Dr. Ronnie Dosanjh, and Dr. Fred Muskett. Without their instruction and assistance I would not have been able to complete this research. I would also like to offer my thanks to Dr. Baigong Yue and Thore Schmedt, who are responsible for the expression and purification of some of the proteins reported here.

My thanks also go to the members of the Henry Wellcome Laboratories for Structural Biology at the University of Leicester, both past and present. Most especially I would like to thank, Kirsty, Ojay, Dariush, Tina, Ian, and Cat, who have been both friends and helpful colleagues. I also express my thanks to Dr Tom Frenkiel and Dr. Geoff Kelly (MRC Biomedical NMR Centre, NIMR) for their assistance acquiring NMR spectra, and to Dr Sharad Mistry and Dr Andrew Bottrill (PNACL, University of Leicester) for acquiring the HPLC and Mass Spectrometry data reported in this thesis.

My sincerest thanks go to my parents for the love, support and patience that they have shown me throughout my life. Without their encouragement I would not be here now.

Finally, to all my friends and family who I have not been able to mention above, thank you for your support, understanding, friendship and love.

Lorna Catherine Waters, Leicester, January 2007.

Lorna Catherine Waters

Title: Solution Structures of Proteins and Complexes Involved in the Regulation of Eukaryotic Gene Expression

ABSTRACT

The regulation of eukaryotic gene expression is dependent on the formation of specific protein-protein and protein-nucleic acid complexes. The crucial roles played by these protein complexes highlights the importance of the work reported here, which focuses on the structures and interactions of two specific regulatory complexes. The initial work focused on the interaction between the general coactivator CBP (SID domain) and members of the p160 family of coactivators (AD1 domain), which is a key step in the activation of transcription by nuclear receptors. The solution structure of the CBP SID / SRC1 AD1 complex described in this thesis shows that the two helical domains are intimately associated, with the first helix in SRC1 AD1 and the first three helices in CBP SID forming a four helix bundle, which is capped by the fourth helix of the AD1 domain. Comparisons with the structure of the related CBP SID / ACTR AD1 complex showed that while the CBP SID domain adopts a similar fold in complex with different p160 proteins, the topologies of the AD1 domains are strikingly different, a feature that is likely to contribute to functional specificity of these complexes. The second part of the work described here focused on the interaction between the MA-3 domains of the tumour suppressor Pdcd4 and the translation factor eIF4A, which has been shown to inhibit cap-dependent translation. The C-terminal MA-3 domain (Pdcd4 MA-3_C) was shown to consist of three atypical HEAT repeats capped by a final helix. This domain was found to interact with the N-terminal domain of eIF4A through a conserved surface region. The comparison of NMR spectra obtained from Pdcd4 MA-3_C and the tandem MA-3 region strongly suggests that the tandem MA-3 region is composed of two equivalent domains connected by a semi-flexible linker. The high resolution structural information obtained provides important insights into the interactions and functional specificity of the protein complexes studied.

CONTENTS

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Carr, M. and Heery, D. M. (2006). Structural Diversity in p160/CREB-binding Protein Coactivator Complexes. *J. Biol. Chem.*, **281:** 14787-14795.

Waters, L. C., Bohm, M., Veverka, V., Muskett, F. W., Frenkiel, T. A., Kelly, G. P., Prescott, A., Dosanjh, N. S., Klempnauer, K. H. and Carr, M. D. (2006). NMR Assignment and Secondary Structure Determination of the C-terminal MA-3 Domain of the Tumour Suppressor Protein Pdcd4. *J. Biomol. NMR*, **36:** S5:18.

Waters, L. C., Veverka, V., Bohm, M., Schmedt, T., Choong, P. T., Muskett, F. W., Klempnauer, K. H. and Carr, M. D. (2007). Structure of the C-terminal MA-3 Domain of the Tumour Suppressor Protein Pdcd4 and Characterisation of its Interaction with eIF4A. *Oncogene*, (Published Online)

ABBREVIATIONS

5' UTR	5' untranslated region of mRNA
α1 - 7	α helices 1 - 7
Aα1 / 2 / 3	ACTR helix 1 / 2 / 3
ACTR	Activator for thyroid hormone and retinoid receptors
AD1 / 2	Activation domain 1 / 2
AF-1 / 2	Activation function-1 / 2
AP-1	Activator protein-1
Cα1 / 2 / 3 / 3'	CBP helix 1 / 2 / 3 / 3'
CANDID	Combined automated NOE assignment and structure
CARM1	Coactivator-associated arginine methyltransferase 1
CBP	CREB binding protein
CBP80	Cap-binding protein 80
CD	Circular dichroism
CDK	Cyclin dependent kinase
c-Jun	Cellular Jun
CREB	cAMP response element binding protein
CTD	C-terminal domain
CYANA	Combined assignment and dynamics algorithm for NMR applications
DAP5/NAT1/p97	Death associated protein 5/novel APOBEC-1 target no. 1/p97
DBD	DNA binding domain
DEAD	Amino acid sequence motif present in RNA helicases
DTT	1,4-dithiothreitol
E1A	Early region 1A
eIF	Eukaryotic initiation factor

eIF4Gc	C-terminal region of eIF4G
$eIF4G_M$	Middle region of eIF4G
ERα	Estrogen receptor α
F1 / 2 / 3	Frequency domain 1 / 2 / 3
FPLC	Fast protein liquid chromatography
GST	Glutathione-S-transferase
GTF	General transcription factor
H2A / 2B / 3 / 4	Histone 2A / 2B / 3 / 4
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEAT	Huntingtin, elongation factor 3, A subunit of protein phosphatase 2A, target of rapamycin kinase
HMQC	Heteronuclear multiple quantum coherence
HMT	Histone methyltransferases
HSQC	Heteronuclear single quantum coherence
IAD	IRF association domain
IBiD	IRF3 interaction domain
IPTG	Isopropyl-1-thio-ß-D-galactopyranoside
IRF-3	Interferon regulatory factor-3
JNK	Jun N-terminal kinase
KID	Kinase inducible domain
KIX	KID interaction domain
LB	Luria Bertani
LBD	Ligand binding domain
LXXLL	Amino acid motif of sequence LXXLL (where X is any residue)
$m^7G(5')ppp(5')N$	5' 7-methyl-5'guanosine triphosphate-5'-N ($m^7G(5')ppp(5')N$)

MA-3	Domain present in the mouse apoptosis 3/PDCD4 gene
$MA\text{-}3_C$	C-terminal MA-3 domain of Pdcd4
$MA\text{-}3_M$	N-terminal MA-3 domain of Pdcd4
MAPK	Mitogen-activated protein kinase
mIF4G	Middle of eIF4G domain
Mnk1	MAP kinase activated protein kinase 1
MOZ	Monocytic leukaemia zinc finger protein
NH	Backbone amide group
NID	Nuclear receptor interaction domain
Ni-NTA	Nickel-nitrilotriacetic acid
NTD	N-terminal domain
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NuRD	Nucleosome remodelling and histone deacetylation complex
p160	Protein of molecular weight 160 Da
p300	Protein of molecular weight 300 kDA
PABP	Poly(A)-binding protein
Pdcd4	Programmed cell death protein 4
Pdcd4 (157-449)	Pdcd4 dual MA-3 domain protein (residues 157-449)
PIC	Pre-initiation complex
Pol II	RNA polymerase II
PRMT1	Protein arginine methyltransferase1
r.m.s.d.	Root mean squared deviation
Sα1 / 2' / 2 / 3	SRC1 helix 1 / 2' / 2 / 3
SID	SRC1 interaction domain
SRC1	Steroid receptor coactivator 1

SWI/SNF	switch/sucrose non-fermentation
TAFs	TBP-association factors
TALOS	Torsion angle likelihood obtained from shift and sequence similarity
TATA box	Promoter DNA sequence motif
TAZ	Transcriptional adaptor putative zinc finger
TBP	TATA-binding protein
TFIIA - H	RNA polymerase II transcription factors A - H
TIF2	Transcription intermediary factor 2
TOCSY	Total correlation spectroscopy
TPA	12-*O*-tetradecanoylphorbol-13-acetate
TRAP	Thyroid receptor-associated protein complex
TRIS	Tris-(hydroxymethyl)-methylamine
W2	Domain containing two invariant tryptophans
WATERGATE	Water suppression by gradient-tailored excitation
ZZ	Zinc finger motif
Φ	Phi, the amide nitrogen alpha carbon dihedral angle
Ψ	Psi, the alpha carbon carboxyl carbon dihedral angle

CHAPTER 1

GENERAL INTRODUCTION

1.1 Regulation of Eukaryotic Gene Expression

The regulation of gene expression in multi-cellular organisms is critical for many essential cellular events, including: cell growth, proliferation, differentiation, apoptosis, and homeostasis. The breakdown or misregulation of expression is associated with many major diseases including: cancer, polio, Huntington's disease and HIV. Regulation at all levels along the gene expression pathway involves the formation of specific protein-protein and protein-nucleic acid complexes. The essential roles played by these complexes emphasises the importance of the work described in this thesis, which focuses on the structures and interactions of two protein-protein complexes involved in the regulation of gene expression: the complex formed between the interaction domains of the transcriptional coactivators CBP and SRC1 and the complex formed between the translational regulators Pdcd4 and eIF4A.

As indicated above the expression of genes can be regulated at a number of levels:

- The structure of chromatin affects the ability of the transcription initiation machinery to access specific genes.
- The initiation of transcription is regulated at many levels by external stimuli, transcription factors and coregulators.
- The mRNA transcripts are processed and modified by a number of enzymes, which are involved in the addition of the 5' cap and 3' polyadenylate tail, and splicing.
- mRNA transcripts are selectively transported from the nucleus to the cytoplasm.
- The stability of certain mRNA transcripts is dependent on the presence of sequences that signal for rapid degradation in the cytoplasm.

- Translation initiation is regulated by eukaryotic initiation factors, signal transduction pathways and translational regulators.

Whilst all of these processes are important, only chromatin structure, and the regulation of transcription initiation and translation initiation relate to the focus of this thesis and hence are discussed in more detail. A general introduction to these processes is presented below, with more specific introductions relating to the structural studies of CBP and Pdcd4 contained within the relevant experimental chapters.

1.2 Histone Modifications and Chromatin Remodelling

The condensation of DNA in chromatin suppresses gene activity partly through the inaccessibility of promoter sequences to the transcription apparatus (Cheung *et al.*, 2000). Chromatin is composed of repeating nucleosome units, which consist of approximately 146 base pairs of DNA wound around an octamer of eight histones (two molecules of each of the core histones: H2A, H2B, H3 and H4) (Luger *et al.*, 1997). The structure of a nucleosome unit is shown in figure 1.1. The additional binding of histone H1 to the outside of the octamer and the linker DNA helps to form more compact higher order structures. The four core histones consist of globular domains with unstructured amino and carboxy terminal tails that are subject to various reversible posttranslational modifications, including lysine acetylation, arginine and lysine methylation, serine and threonine phosphorylation, lysine sumoylation and lysine ubiquitination (reviewed in Davie & Murphy, 1990; Kimura *et al.*, 2005; Nowak & Corces, 2004; Sterner & Berger, 2000; Zhang & Reinberg, 2001). These modifications may serve as epigenetic markers, which are able to influence the patterns of gene expression and have been implicated in both the recruitment of regulatory proteins (discussed below) and in transcriptional silencing (reviewed in Mellor, 2006; Nightingale *et al.*, 2006; Rosenfeld *et al.*, 2006).

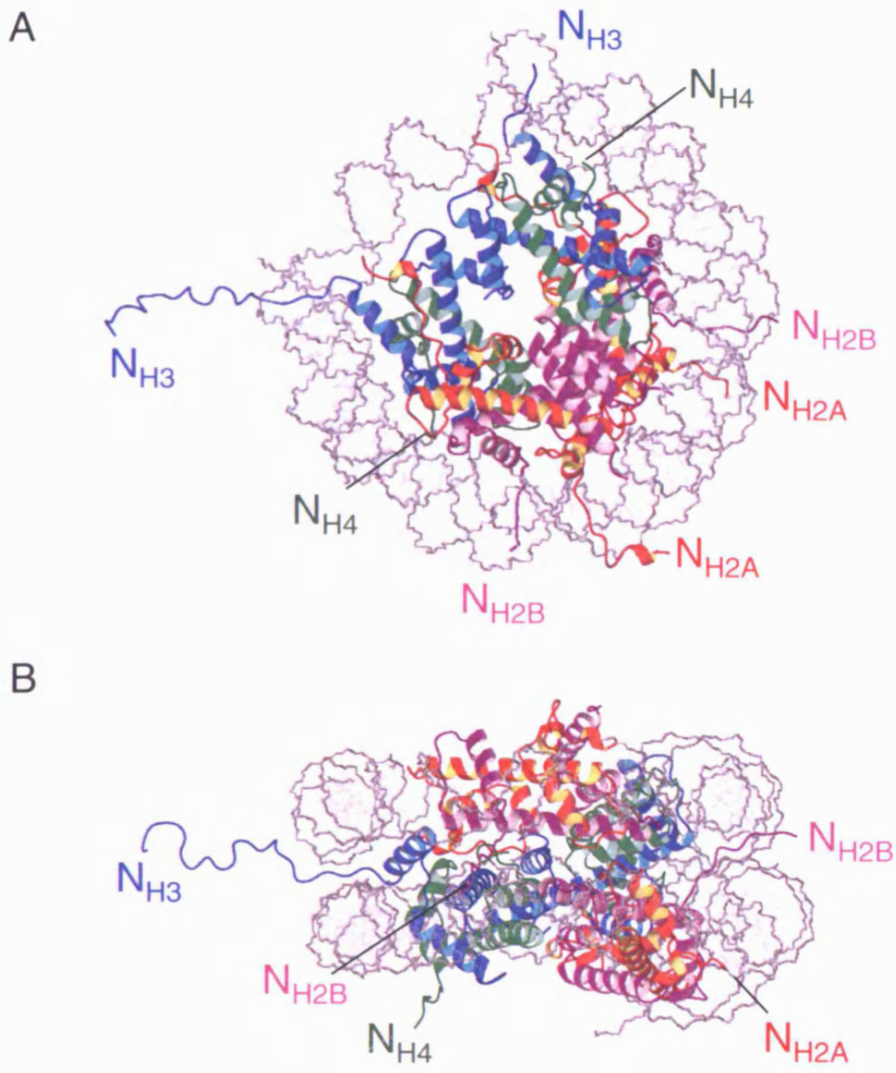


Figure 1.1 Structure of the nucleosome core particle (Luger *et al.*, 1997) (PDB code 1AOI). The nucleosome is composed of 146 base pairs of DNA (shown in grey) wound around a histone octamer, which contains two molecules of histones H2A, H2B, H3 and H4. The ribbon representation of the backbone topology of histones H2A are shown in red, H2B in purple, H3 in blue and H4 in green. Where visible, the unstructured N-terminal tails are labelled. It should be noted that a number (between 3 and 37) of the N-terminal residues are absent from the histone structures. The nucleosome unit shown in panel B has been rotated by approximately -90° about the horizontal axis as compared to panel A.

Research into histone modifications has, until recently, focused primarily on histone acetylation, a process which is generally associated with transcriptional activation (Hebbes *et al.*, 1988). The acetylation of lysine residues neutralises the ε-amino groups positive charge, which is believed to weaken the internucleosomal interactions and destabilise the chromatin structure (Hong *et al.*, 1993). Histone acetyltransferases (HATs), such as CBP (discussed in section 2.1.3), play an important role in the control of transcription and hence cellular control. Their misregulation is involved in the development of some human tumours (reviewed in Timmermann *et al.*, 2001) and they are also targeted by many viral proteins (reviewed in Caron *et al.*, 2003). However, the precise roles of histone acetylation are not fully understood. This is highlighted by specific examples where histone acetylation is associated with transcriptional repression and deacetylation with transcriptional activation. For example, the acetylation of K12 of histone 4 (H4) is increased in heterochromatic regions (Turner *et al.*, 1992), and the HDAC Hos2 has been shown to function as a transcriptional activator (Kurdistani & Grunstein, 2003).

Histone methylation has been shown to both inhibit and activate transcription (reviewed in Nightingale *et al.*, 2006; Santos-Rosa & Caldas, 2005). For example, the arginine methyltransferases CARM1 (coactivator-associated arginine methyltransferase) and PRMT1 (protein arginine methyltransferase1) have been shown to enhance transcription (Koh *et al.*, 2001), whilst the methylation of K27 of H3 is associated with transcriptional repression (Cao *et al.*, 2002; Kuzmichev *et al.*, 2002). Histone phosphorylation also plays contradictory roles in transcriptional regulation. For example, Phosphorylation of S10 of H3 results in either the condensation of chromatin, or transcriptional activation, depending on the context (reviewed in Nowak & Corces, 2004). The ubiquitination or sumoylation of lysine residues are generally associated with transcriptional activation and inhibition respectively. For example, ubiquitination of H2B in yeast is associated with transcriptional

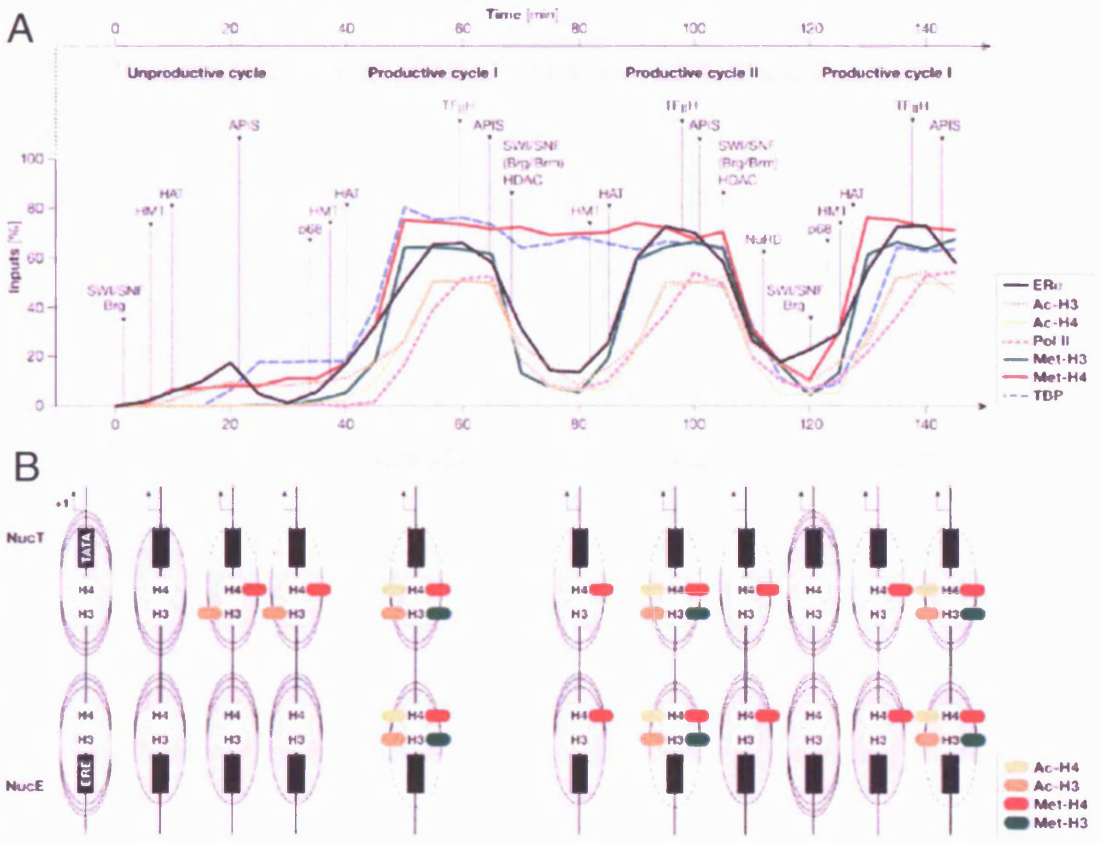
activation. Sumoylation is associated with the recruitment the transcriptional corepressors HDAC1 and HP1 (reviewed in Santos-Rosa & Caldas, 2005).

Specific patterns of histone modifications have been observed in nucleosomes. For example, the phosphorylation of S10 of H3 has been shown to promote the acetylation of K14 of H3, whilst the ubiquitination of K123 of yeast H2B is required for the subsequent methylation of K4 and K79 of H3 (reviewed in Nightingale *et al.*, 2006; Santos-Rosa & Caldas, 2005). The existence of these histone modification patterns, in conjunction with the contradictory roles of specific modifications and the residue specificity of histone modifying enzymes, is consistent with a hypothesis known as the 'histone code' (reviewed in Kimura *et al.*, 2005; Sterner & Berger, 2000; Strahl & Allis, 2000). The code holds that residue specific posttranslational modifications of histones influences subsequent cofactor binding and modifications to either the same or different histone molecules (Fischle *et al.*, 2003; Jenuwein & Allis, 2001; Strahl & Allis, 2000; Turner, 2000).

The modified histones serve as recognition sites for the recruitment of effectors that stimulate further modifications or nucleosome remodelling, in which the nucleosome slides along the DNA exposing or occluding local DNA areas to interactions (Agalioti *et al.*, 2002; Kimura *et al.*, 2005; Mellor, 2006). These chromatin interacting proteins interact through common domains such as chromodomains, which bind methylated histones (Bannister *et al.*, 2001) and bromodomains, which bind to acetylated histones (Dhalluin *et al.*, 1999). Bromodomains are present in many histone modifying proteins and chromatin-remodelling complexes, for example CBP/p300 and BRG1 (brahma-related gene 1) (Margueron *et al.*, 2005; Ragvin *et al.*, 2004).

Studies of human estrogen receptor α (hERα) induced transcription initiation at the pS2 promoter of MCF-7 cells showed that cycles of histone modifications were observed

during the cycles of transcription initiation (Mellor, 2006; Metivier *et al.*, 2003; Reid *et al.*, 2003). These modifications are shown in figure 1.2.



Metivier *et al.*, 2006

Figure 1.2 Histone modifications observed at the pS2 promoter during cycles of transcription activation (Metivier *et al.*, 2006). Panel A shows the recruitment of cofactors to the pS2 promoter of MCF-7 cells observed during estrogen receptor α (ERα) mediated transcription initiation, as determined by chromatin immunoprecipitation assay. Steroid deprived cells were treated with α-amanitin to prevent basal transcription. The pS2 promoter was subsequently stimulated with 10 nM estradiol. The y axis represents the amount of immunoprecipitated pS2 promoter expressed as a percentage of the input. The details of the transcriptional cycles are discussed in section 2.1.1. The periodic association of HATs, HDACs, HMTs and SWI/SNF (Brg/Brm), as well as other important proteins/complexes are indicated.

Also shown are the levels of acetylated H3 K14 and H4 K16, and of methylated H3 R17 and H4 R3. Panel B shows the corresponding nucleosome dynamics. The location of the modified histones in nucleosome E (NucE) and nucleosome T (NucT), which are associated with the estrogen response element (ERE) and TATA box respectively, are shown. Increased accessibility of either the TATA box or the ERE is shown by dashed lines. The schemes are based on Metivier *et al* (2003), Reid *et al* (2003) and Metivier *et al* (2006). Abbreviations used are: Acetylated histone 3 K14 (Ac-H3), acetylated histone 4 K16 (Ac-H4), AAA ATPase proteins independent of 20S (APIS), estrogen response element (ERE), histone methyltransferase (HMT), methylated histone 3 R17 (Met-H3), methylated histone 4 R3 (Met-H4), nucleosome remodelling and deacetylating complex (NuRD), switch/sucrose non-fermentation (SWI/SNF).

1.3 Overview of the Regulation of Transcription Initiation

Initiation of transcription is one of the most complex control points of gene expression. It requires the coordinated functions of a vast number of proteins including the basal transcription machinery; sequence specific DNA binding transcription factors, chromatin remodelling enzymes and cofactors that act as scaffolds and bridges.

In order to initiate the transcription of protein encoding genes, RNA polymerase II (Pol II) requires the assistance of several general transcription factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, which together form the pre-initiation complex (PIC) (reviewed in Thomas & Chiang, 2006). With the exception of TFIIB, all the GTFs are in fact multiprotein complexes, composed of between two and fifteen different proteins. Similarly, Pol II is also a multiprotein complex composed of 12 different protein subunits. The assembly of the PIC begins with the recruitment of TFIID, which is composed of the TATA-binding protein (TBP) and a number of TBP-association factors (TAFs: TAF1 -

TAF15), to the TATA box, initiator element (Inr) and/or the downstream promoter element (DPE) of the core promoter (reviewed in Lee & Young, 2000; Lemon & Tjian, 2000; Thomas & Chiang, 2006). The other GTFs and Pol II are subsequently recruited through one of two possible pathways: the sequential assembly pathway or the Pol II holoenzyme pathway (reviewed in Lemon & Tjian, 2000; Thomas & Chiang, 2006).

In the sequential assembly pathway, the GTFs and Pol II are sequentially recruited to promoter bound TFIID. The initial step in this pathway involves the recruitment of TFIIA followed by TFIIB. In the case of TATA box containing promoters, these two GTFs help to stabilize the TBP-TATA complex, as shown in figure 1.3 panels A and B. TFIIB also plays an important role in the subsequent recruitment of the Pol II - TFIIF complex. In addition to interacting with TFIIB, TFIIF plays several other important roles in PIC assembly, including, stabilizing the Pol II–TFIIB-TBP-promoter complex and recruiting TFIIE. TFIIE in turn recruits the final GTF, TFIIH, to the PIC. TFIIH possesses DNA-dependent ATPase activity, ATP-dependent helicase activity and kinase activity, which are all required for transcription initiation. Cyclin-dependent kinase 7 (CDK7), a component of the TFIIH protein complex, has been shown to phosphorylate the C-terminal domain of Pol II, an event that is believed to be an important step in the release of Pol II from the PIC (reviewed in Boeger *et al.*, 2005; Bushnell *et al.*, 2004; Lemon & Tjian, 2000; Nikolov & Burley, 1997; Thomas & Chiang, 2006). The proposed structural model of the PIC is shown in figure 1.3 panel C.

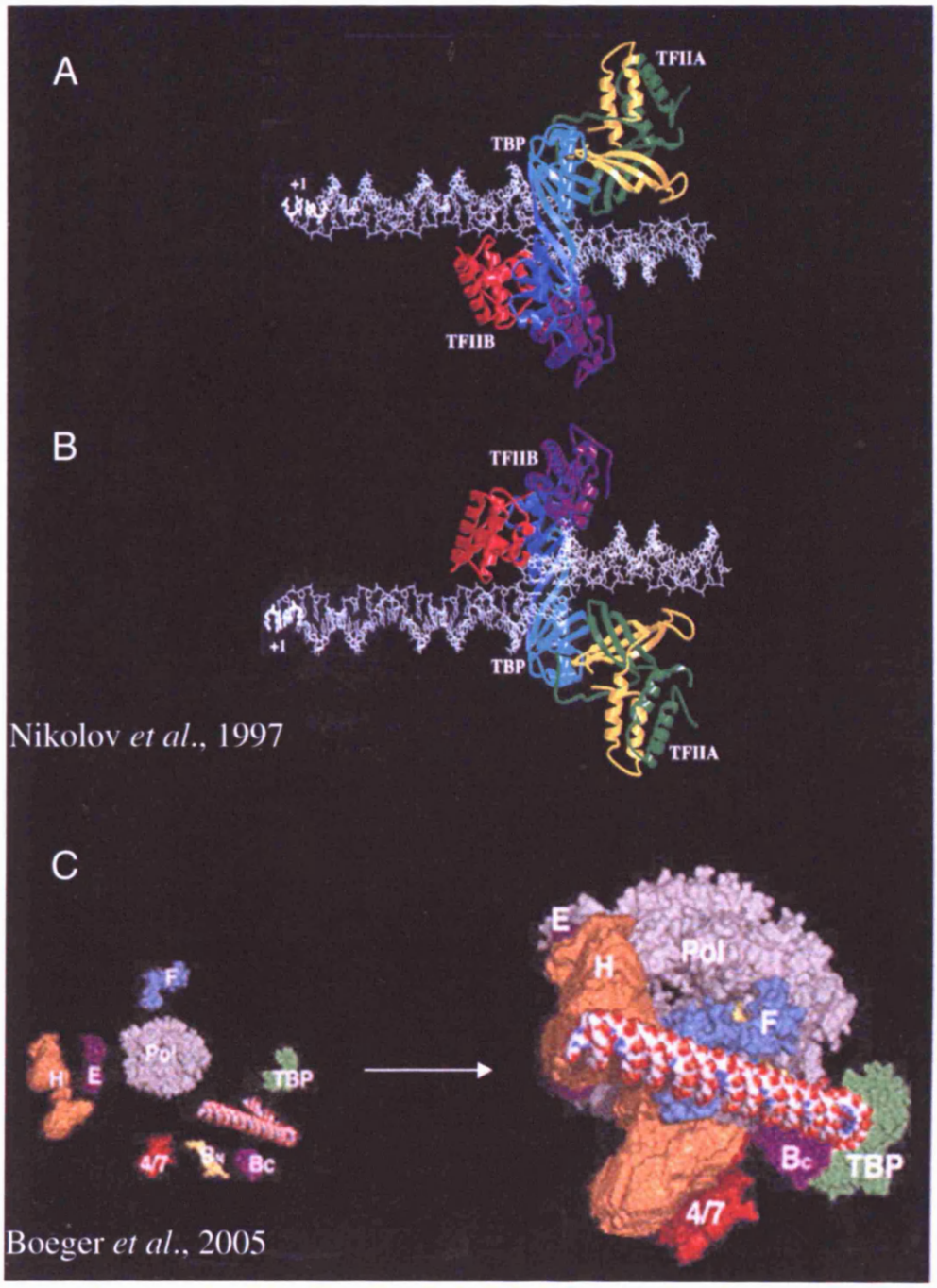


Figure 1.3 Model of the promoter bound general transcription factor and RNA polymerase II complexes. Panels A and B show the model of the TFIIA-TFIIB-TBP-DNA complex, based on the structures of the C-terminal (c) TFIIB-TBP-TATA (Nikolov *et al.*, 1995) and the TFIIA-TBP-TATA complexes (Geiger *et al.*, 1996; Tan *et al.*, 1996), proposed by (Nikolov & Burley, 1997). The transcription start site is labelled as +1. The TBP N-terminus is shown in light blue, the TBP second repeat in

second repeat in dark blue, TFIIA small subunit in green, TFIIA large subunit in yellow, cTFIIB first repeat in red, cTFIIB second repeat in magenta and DNA in grey. Panel C shows a model of the PIC, which was proposed on the basis of crystallography (TBP, TFIIB, Pol II, Rpb4, Rpb7) and electron microscope (TFIIE, TFIIF and TFIIH) structures (Boeger *et al.*, 2005; Bushnell *et al.*, 2004). The individual structures are shown on the left side of the panel and the assembled complex on the right. TFIIB is shown in yellow (N-terminal) and purple (C-terminal), TBP in green, TFIIE in purple, TFIIF in blue, TFIIH in orange, Pol II in grey, polymerase subunits Rbp4 and Rbp7 in red and the DNA in red, blue and white.

The Pol II holoenzyme pathway was proposed as a result of the discovery that Pol II could be recruited as part of a preassembled complex, containing a subset of GTFs and cofactors. For example, Wu *et al* reported the isolation of a complex composed of Pol II, TFIIB, TFIIE, TFIIF, TFIIH, GCN5, SWI/SNF and Mediator complex (Wu & Chiang, 1998). Many different complexes have been observed, depending on the experimental methods and materials used, however, none of these complexes contain TFIID. The holoenzyme pathway is therefore believed to consist of two steps, the initial recruitment of TFIID to the promoter, followed by the recruitment of the Pol II complex (reviewed in Lemon & Tjian, 2000; Thomas & Chiang, 2006).

The formation of the promoter bound PIC is sufficient for a basal level of transcription. However, the coordinated assembly of additional transcription factors and coactivator complexes are required for higher levels of regulated transcriptional activity. Transcription factors are recruited to DNA promoter or enhancer sequences, which are found up to thousands of base pairs upstream or downstream of the core promoter. Transcription factors are composed of distinct functional domains; a DNA binding domain and, in the

majority of cases, one or more activation domains (Ptashne & Gann, 1997; Triezenberg, 1995). The DNA binding domains generally conform to one of a number of structural motifs including the basic helix-loop-helix, the Cys_4 type and Cys_2-His_2 zinc fingers, the basic leucine zipper and the helix-turn-helix. The majority of the activation domains that have been studied, have been shown to be regulated by either ligand binding or posttranslational modifications. They function by recruiting coregulators, which are required to regulate transcription at nearby promoters (Pugh & Tjian, 1990). A schematic representation of the transcription factor/coactivator complexes formed at the promoter and enhancer sequences is shown in figure 1.4.

Coactivators enhance gene transcription by several methods, including the posttranslational modification of transcriptional regulators and histones, by remodelling chromatin structure, by acting as a scaffold for the formation of multiprotein complexes, or as a bridge between the enhancer bound transcriptional activators and the PIC (reviewed in Perissi & Rosenfeld, 2005; Spiegelman & Heinrich, 2004). The role of some cofactors in transcriptional activation has been shown to be context dependent (Chiba *et al.*, 1994; Jepsen *et al.*, 2000; Metivier *et al.*, 2003; Underhill *et al.*, 2000). For example, the ATP-dependent remodelling complex SWI/SNF has been shown to play a role in both transcriptional activation and repression (Metivier *et al.*, 2003; Underhill *et al.*, 2000). Transcriptional repressors also play a key role in the regulation of gene expression. They function by competing with activators for DNA binding, sequestering activators, interacting with the core transcription machinery and modifying histones and transcriptional regulators (reviewed in Jepsen & Rosenfeld, 2002; Perissi & Rosenfeld, 2005; Rosenfeld *et al.*, 2006).

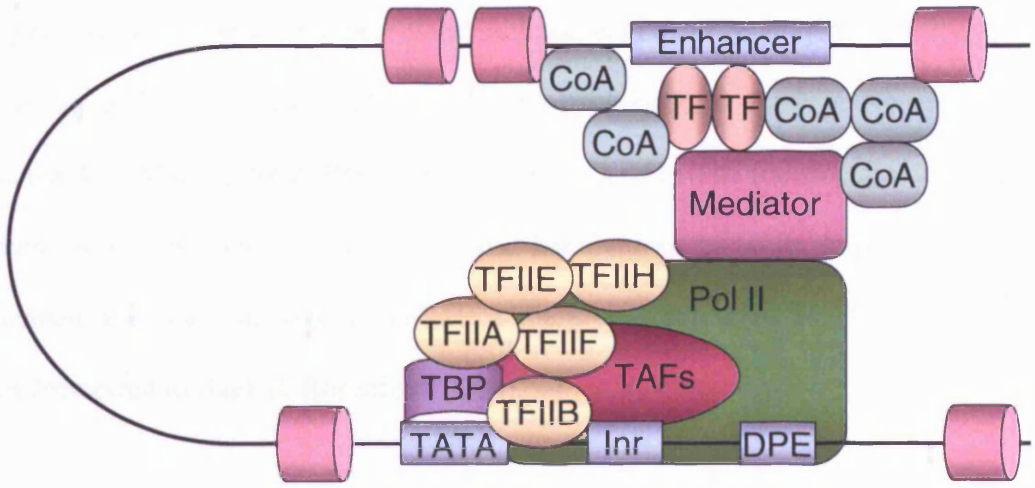


Figure 1.4 A schematic representation of the protein-protein complexes that are required for eukaryotic transcription initiation. (Adapted from Taatjes *et al.*, 2004). The pre-initiation complex (PIC) is primarily composed of RNA polymerase II (PolII), general transcription factors (TFIIA/B/E/F/H) and TBP-associated proteins (TAFs), which assemble on promoter elements, including the TATA box, the initiator (INR) and downstream promoter element (DPE). The Mediator complex serves to bridge the PIC and enhancer bound transcription factors (TF) and associated coactivators (CoA) complexes. Nucleosome units are represented by pink cylinders.

The coordinated control of the cellular levels and activities of transcription factors and coactivators are crucial for the normal development of organisms, as a result their misregulation is associated with developmental disorders and cancer (reviewed in Lonard & O'Malley, 2005; Spiegelman & Heinrich, 2004). One of the primary regulatory mechanisms is through the binding of ligands and coregulatory complexes (reviewed in Lonard & O'Malley, 2005; Perissi & Rosenfeld, 2005). A number of signalling pathways, as well as other coregulators, are able to regulate transcription initiation by posttranslationally modifying transcription factors and coregulators. These modifications,

which include ubiquitination, acetylation, sumoylation, methylation, and phosphorylation, have been shown to influence processes such as, degradation, enzyme activity, protein-protein interactions and subcellular localisation (reviewed in Baek & Rosenfeld, 2004; McKenna & O'Malley, 2002; Perissi & Rosenfeld, 2005). The translocation of regulators between the nucleus and cytoplasm has also been shown to be an important regulatory mechanism and may function to influence the concentration of regulators in different tissues (reviewed in Baek & Rosenfeld, 2004).

1.4 Regulation of Translation Initiation

1.4.1 Overview of Cap-Dependent Translation Initiation

Translation initiation is a highly regulated multi-step process that is catalysed by a number of eukaryotic initiation factors (eIFs) (reviewed in Gingras *et al.*, 1999; Hershey, 2000a; Marintchev & Wagner, 2004; Preiss & Hentze, 2003). One of the key regulatory points of translation initiation is the recruitment of the 43S preinitation complex, by the eIF4F complex, to the 5' mRNA cap. As a result the different components of the eIF4F complex, especially eIF4E, are targeted by a number of translational regulators and signal transduction pathways. The structure, function and regulation of the eIF4F complex are discussed in more detail in section 1.4.2.

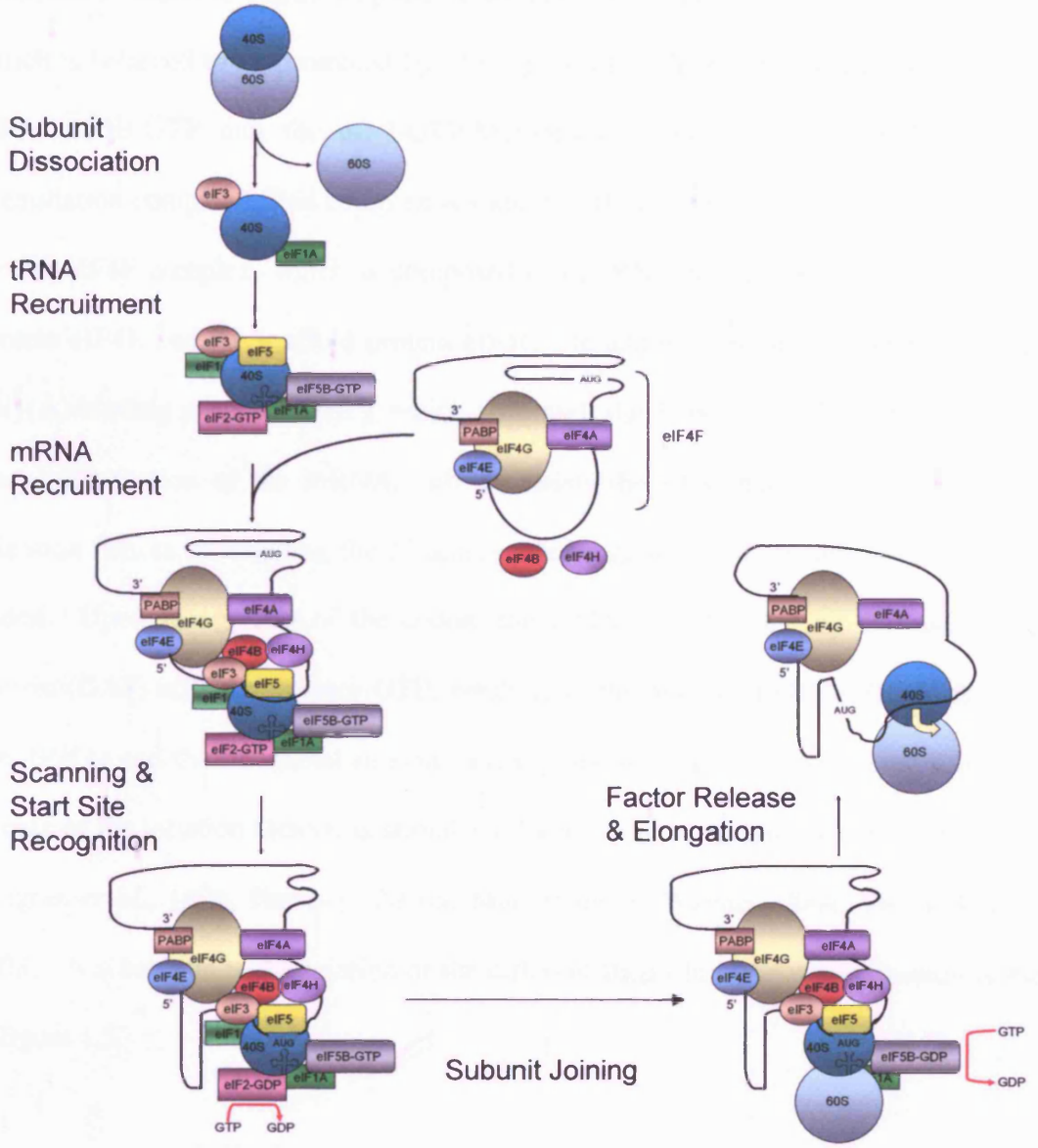


Figure 1.5 Schematic representation of the translation initation pathway. (Adapted from Hershey, 2000a; Marintchev & Wagner, 2004; Preiss & Hentze, 2003). Initiation begins with the dissociation of the ribosomal subunits, followed by the formation of the 43S complex. The 43S preinitiation complex is recruited to the 5' mRNA by the eIF4F complex, which also assists the ribosomal subunit in scanning. Upon start site recognition the ribosomal subunits rejoin and the initiation factors are released in GTP dependent processes, marking the end of initiation.

Translation initiation begins with the dissociation of the 60S and 40S ribosomal subunits, which is believed to be promoted by eIF3 and eIF1A. The 40S subunit then binds eIF1, eIF5, eIF5B-GTP and the eIF2-GTP-Met-tRNAi ternary complex to form the 43S preinitiation complex. This complex is subsequently recruited to the 5' end of the mRNA by the eIF4F complex, which is composed of the RNA helicase eIF4A, the cap-binding protein eIF4E and the scaffold protein eIF4G. In addition, eIF4G also interacts with the poly(A)-binding protein (PABP), which also binds the 3' poly(A) mRNA tail, resulting in the circularisation of the mRNA. eIF4F assists the ribosomal subunit and associated initiation factors in scanning the 5' untranslated region (UTR), in search of an initiation codon. Upon recognition of the codon, the GTPase eIF2, aided by GTPase-activating protein (GAP) eIF5, hydrolyses GTP, resulting in the release of eIF2-GDP from both the Met-tRNAi and the ribosomal subunit. Finally, the joining of the ribosomal subunits and release of the intiation factors, is stimulated by the GTPase activity of eIF5B (reviewed in Gingras *et al.*, 1999; Hershey, 2000a; Marintchev & Wagner, 2004; Preiss & Hentze, 2003). A schematic representation of the different stages in translation initiation is shown in figure 1.5.

In addition to the cap-dependent translation, which is described above, a second initiation mechanism, known as internal ribosome entry (IRES) exists for some mRNAs, allowing them to bind directly to the initiation codon (Belsham & Sonenberg, 1996; Jackson, 2000). The work described in this thesis does not relate to IRES-mediated translation and hence it will not be discussed.

1.4.2 Structure and Function of the eIF4F Complex

The translation initiation factor eIF4G plays a central role in cap-dependent translation as it both serves as a scaffold for the assembly of the cap binding complex (eIF4F) and recruits

the 43S ribosomal subunit to the 5' cap via an interaction with eIF3 (reviewed in Hershey, 2000a; Marintchev & Wagner, 2004). The N-terminal region of eIF4G, which is believed to be unstructured in the free protein, contains binding sites for the poly(A)-binding protein (PABP) and the cap binding protein eIF4E (Groft & Burley, 2002; Gross *et al.*, 2003; Imataka *et al.*, 1998; Mader *et al.*, 1995; Marcotrigiano *et al.*, 1999; Tarun & Sachs, 1996). Through its interactions with both eIF4E and PABP, eIF4G bridges the mRNA 5' and 3' ends (Tarun & Sachs, 1996). This circularisation of the mRNA serves to protect the 43S subunit from nucleases and translational repressors, is necessary to transfer signals from the 3' to 5' ends and may also function to channel the dissociated 40S subunit back to the cap (reviewed in Marintchev & Wagner, 2004; Preiss & Hentze, 2003). This interaction between eIF4G, eIF4E and PABP also increases the affinity of eIF4E for the 5' mRNA cap (Borman *et al.*, 2000; Haghighat & Sonenberg, 1997; Wei *et al.*, 1998).

The middle region of eIF4G contains an interaction domain often referred to as an mIF4G (middle of eIF4G) domain, which binds to eIF4A, eIF3 and mRNA (Imataka & Sonenberg, 1997; Lamphear *et al.*, 1995; Marcotrigiano *et al.*, 2001; Morino *et al.*, 2000). The C-terminal region of eIF4G from higher eukaryotes contains two further interaction domains. Although this C-terminal region is not considered essential for translation initation, it is required for robust translation (Morino *et al.*, 2000). The first of the C-terminal region domains, is an extended MA-3 domain, which interacts with eIF4A (Bellsolell *et al.*, 2006; Imataka & Sonenberg, 1997; Morino *et al.*, 2000). The adjacent W2 domain, which is homologous to the W2 HEAT domains of eIF2Bε and eIF5 (Bellsolell *et al.*, 2006; Boesen *et al.*, 2004), interacts with the MAP kinase activated protein kinase, Mnk1 (Morino *et al.*, 2000; Pyronnet *et al.*, 1999; Waskiewicz *et al.*, 1999). Mnk1 has been shown to phosphorylate eIF4E, regulating the rate of translation initiation (reviewed in Scheper & Proud, 2002). All three of these domains are composed of between 3 and 5 atypical HEAT

repeats. The arrangement of these atypical HEAT domains of eIF4G shows homology to that of the nuclear cap-binding protein 80 (CBP80), suggesting that the domains of eIF4G might form a similar compact structure (Marintchev & Wagner, 2004; Mazza *et al.*, 2001). This would allow for extensive cross talk between the different interaction domains (reviewed in Marintchev & Wagner, 2004; Prevot *et al.*, 2003). The domain organisation and structure of the interaction domains of human eIF4G are shown in figure 1.6.

Two eIF4G genes/proteins have been characterised in human, eIF4GI and eIF4GII, which share 46% sequence identity (reviewed in Hernandez & Vazquez-Pianzola, 2005; Prevot *et al.*, 2003). Biochemical and functional analysis of the human eIF4G proteins have revealed only very minor differences, for example, eIF4GI is selectively recruited to capped mRNA at the onset of cell differentiation (Caron *et al.*, 2004). In this thesis, given the high sequence homology and the functional similarities between the two human eIF4G proteins, they are generally referred to collectively as eIF4G. Likewise, unless otherwise stated, the residue numbers refer to the human eIF4GI isoform.

eIF4A is a member of the DEAD box RNA helicase protein family (Linder *et al.*, 1989). Its helicase activity is stimulated by the binding of eIF4A to eIF4G or when bound to eIF4B or eIF4H (Abramson *et al.*, 1987; Richter-Cook *et al.*, 1998; Rogers *et al.*, 2001; Rozen *et al.*, 1990). eIF4A catalyses the unwinding of mRNA secondary structure at the 5' UTR, allowing the recruitment of the 40S ribosomal subunit to the 5' cap and the subsequent scanning (reviewed in Hershey, 2000a; Rogers *et al.*, 2002).

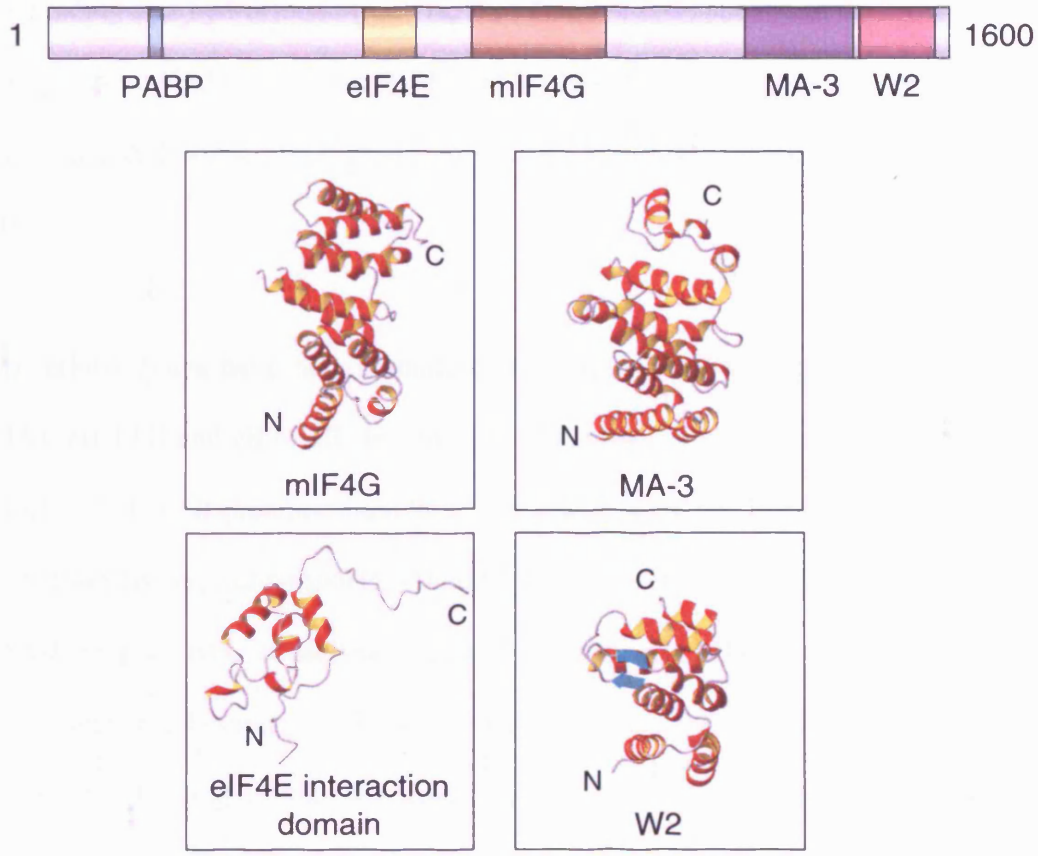


Figure 1.6 Schematic diagram of the functional regions of the translation initiation factor eIF4GI. The PABP interaction domain (residues 174-199), eIF4E interaction domain (557-646), mIF4G (752-993), extended MA-3 domain (1234-1427) and W2 domain (1438-1566) are all indicated. The ribbon representation of the backbone topology of the eIF4E interaction domain (Gross *et al.*, 2003) (PDB code 1RF8), mIF4G (Marcotrigiano *et al.*, 2001) (1HU3), extended MA-3 domain and W2 domain (Bellsolell *et al.*, 2006) (1UG3) are also shown. The N and C-terminals are labelled. The structure of the PABP interaction domain, which consists of a short helical region has also been solved (Groft & Burley, 2002) (1LJ2).

The DEAD box proteins share nine highly conserved sequence motifs (between 3 and 9 residues in length), which are required for RNA binding, ATPase activity and helicase

acitivity (Pause & Sonenberg, 1992). Motifs I, II and probably VI and Q are required for ATP binding and hydrolysis, motifs Ia, Ib, IV and V are believed to be involved in RNA binding and unwinding, and motifs III and VI link ATP hydrolysis with helicase activity (Gorbalenya & Koonin, 1993; Pause *et al.*, 1993; Pause & Sonenberg, 1992; Tanner *et al.*, 2003).

Three eIF4A genes have been identified in mammals: eIF4AI (generally referred to as eIF4A), eIF4AII and eIF4AIII (reviewed in Hernandez & Vazquez-Pianzola, 2005). The eIF4AI and eIF4AII proteins share 91% amino acid sequence identity and are believed to be functionally indistinguishable (Nielsen & Trachsel, 1988; Yoderhill *et al.*, 1993). eIF4AIII shares ~65% conserved sequence identity with eIF4AI. However, despite the level of sequence homology, eIF4III is unable to bind the C-terminal region of eIF4G, or facilitate the binding of the 40S ribosomal subunit and associated factors to mRNA. Instead it has been suggested that eIF4AIII may play an inhibitory role in translation (Li *et al.*, 1999). eIF4AIII has also been shown to form part of the exon junction complex, which plays a role in the posttranscriptional regulation of mRNA (reviewed in Bono *et al.*, 2006; Hernandez & Vazquez-Pianzola, 2005).

Crystal structures of yeast eIF4A and human eIF4AIII have shown that the protein consists of two separate compact domains, an N- and C- terminal domain, as shown in figure 1.7 (Andersen *et al.*, 2006; Bono *et al.*, 2006; Caruthers *et al.*, 2000). Distinct open (inactive) and closed (active) interdomain conformations of eIF4A have been observed in these structures (figure 1.7). The closed conformation is formed in the presence of ATP and RNA, whose binding sites are located at the interface. The binding of eIF4Gm to the C-terminal domain (CTD) of eIF4A is believed to stabilise this closed active form of eIF4A (Oberer *et al.*, 2005). The additional binding of the extended MA-3 domain of eIF4G

plays a modulatory role on translation (Ali & Jackson, 2001; Morino *et al.*, 2000). The interaction between eIF4A and eIF4G is not constant during translation initiation and it is believed that eIF4A cycles in and out of the eIF4F complex during RNA unwinding (Pause *et al.*, 1994).

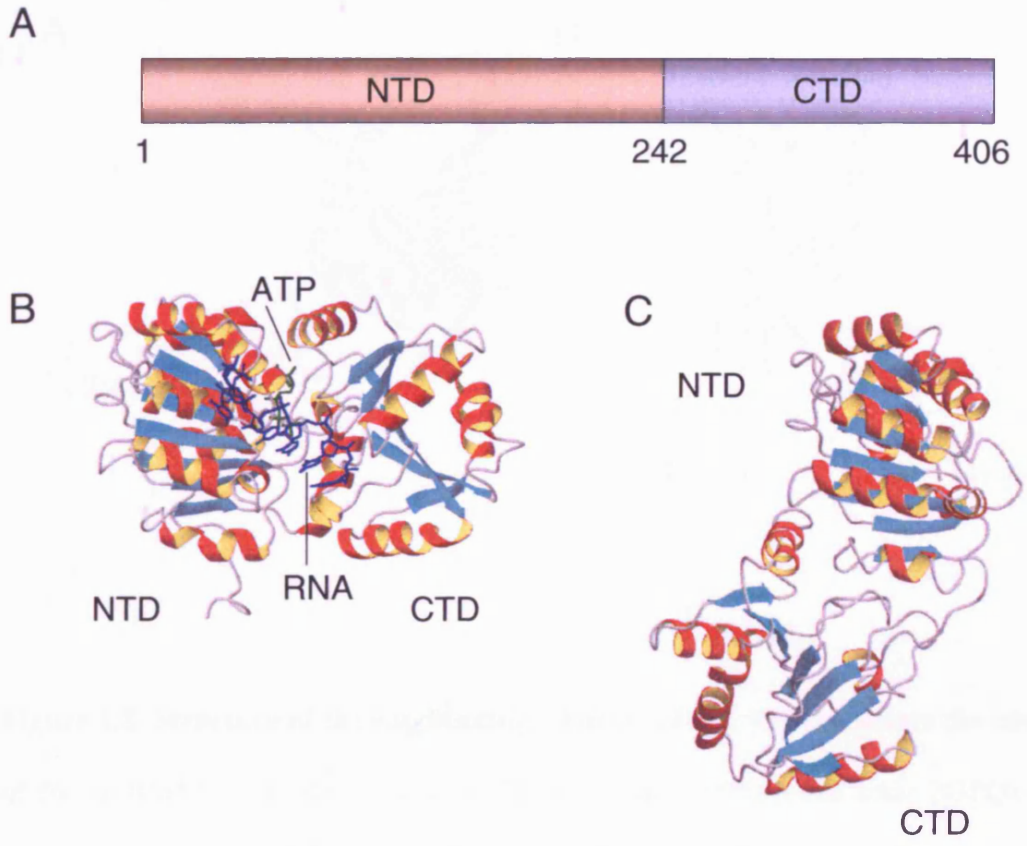


Figure 1.7 Structure of the translation initiation factor eIF4A. Panel A shows a schematic representation of the domain organisation of mouse eIF4AI. Panels B and C show ribbon representations of the backbone structure of human eIF4AIII (Andersen *et al.*, 2006) (PDB codes 2HYI and 2HXY respectively). Panels B shows the ATP and RNA bound closed interdomain orientation. Panel C shows the open conformation, with the N-terminal domain shown in the same orientation as in panel B. The ATP, RNA and the N- and C-terminal domains (NTD and CTD) are highlighted.

The orientation of eIF4A with respect to the ribosomal subunit is unknown. However, it seems likely that eIF4A would be located on the 3' side of the subunit, allowing it to stabilise the single stranded conformation of mRNA (reviewed in Marintchev & Wagner, 2004).

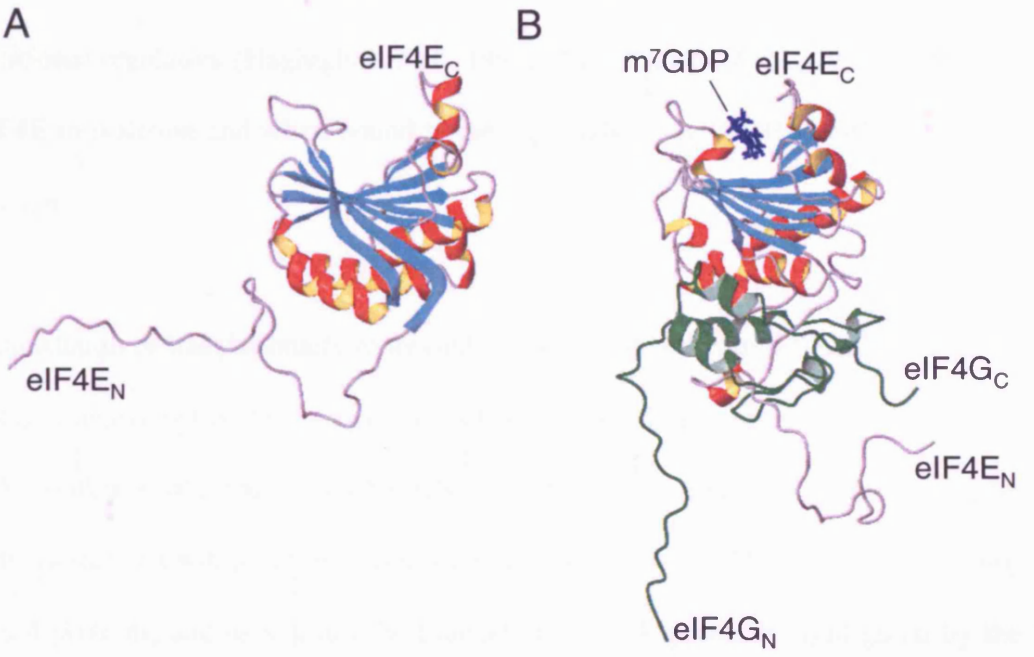


Figure 1.8 Structure of the cap binding protein eIF4E. Panel A shows the structure of the isolated human eIF4E protein (Volpon *et al.*, 2006) (PDB code 2GPQ). The structure of yeast eIF4E in complex with eIF4G (residues 393-490) and the cap analogue m^{7}GDP has also been solved and is shown in panel B (Gross *et al.*, 2003) (1RF8). eIF4G binding induces conformational changes in eIF4E, which stabilise cap binding. Similarly cap binding is believed to stabilise eIF4G binding. The N- and C-termini of the proteins and m^{7}GDP are labelled. eIF4E is shown in red and cyan, eIF4G in green and m^{7}GDP in dark blue.

As previously mentioned, eIF4E recruits the eIF4F complex and indirectly the 43S subunit, to the 5' end of mRNA (reviewed in Hershey, 2000a; Hershey, 2000b; Marintchev &

Wagner, 2004). eIF4E has been shown to interact with a 100 residue sequence from the N-terminal region of eIF4G (figure 1.6) (Gross *et al.*, 2003). This interaction induces conformational changes in the N-terminal region and the cap binding surface of eIF4E, which stabilises the interaction between eIF4E and the 5' 7-methyl-5'guanosine triphosphate-5'-N (m^7G(5')ppp(5')N) cap structure of mRNA (Gross *et al.*, 2003; Volpon *et al.*, 2006). eIF4E can also interact with the 4E-binding proteins (4E-BPs), which act as translational regulators (Haghighat *et al.*, 1995; Marcotrigiano *et al.*, 1999). The structure of eIF4E in isolation and when bound to the cap analogue m7GDP and eIF4G is shown in figure 1.8.

The translation of translationally repressed mRNAs with structured 5' untranslated regions (UTRs) is dependent on the components of the cap-binding complex (eIF4F), particularly eIF4A (Svitkin *et al.*, 2001). Such mRNAs include those coding for transcription factors, growth factors, growth promotion genes and protooncogenes. The need to tightly regulate the eIF4 proteins, and as a result the translation of such genes, is highlighted by the fact that elevated levels of eIF4A, eIF4E and eIF4G have been observed in human tumours and tumour cell lines (Bauer *et al.*, 2001; De Benedetti & Harris, 1999; Eberle *et al.*, 1997). Translational regulators have been shown to interact with all three components of the eIF4F complex and PABP (reviewed in Hershey, 2000b; Marintchev & Wagner, 2004). They function by either posttranslationally modifying the eIF4F proteins, affecting their binding to partner proteins, or by directly competing with the initiation factors for binding. For example, the 4E-BPs, Pdcd4 and DAP5/NAT1/p97, which all share homology with regions of eIF4G, have been shown to compete with eIF4G for binding to either eIF4E or eIF4A (Haghighat *et al.*, 1995; Imataka & Sonenberg, 1997; Marcotrigiano *et al.*, 1999; Yang *et al.*, 2003a). Some regulators have been shown to indirectly modulate translation initiation by affecting the binding of specific translational regulators (reviewed in

Marintchev & Wagner, 2004).

1.5 Aims of the Project

The regulation of eukaryotic gene expression involves the formation of a vast number of protein-protein complexes. The aim of this project was to obtain high resolution structural data for two particular protein-protein complexes that are involved in the regulation of gene expression. This data may provide important insights into how the proteins/complexes function, as well as revealing the specificity of their interactions.

(1) The structure of the complex formed between the interaction domains of the transcriptional coactivators CBP and SRC1.

The coactivator function of the scaffold protein CBP is partly dependent on its ability to recruit coregulator complexes to DNA bound transcription factors, including ligand bound nuclear receptors. In addition it has been shown to acetylate both transcriptional regulators and histones, and serves as a bridge between transcription factor/coregulator complexes and the basal transcription machinery (reviewed Chan & La Thangue, 2001). CBP is indirectly recruited to nuclear receptors through its interaction with the AD1 domain of the p160 proteins (Kamei *et al.*, 1996; Sheppard *et al.*, 2001). The interaction between the SID domain of CBP and the p160 AD1 domains is essential for ligand-dependent transcription mediated by the steroid hormone receptors (Kim *et al.*, 2001; Sheppard *et al.*, 2001). The original aim of the work described in the first experimental chapter of this thesis was to determine the solution structure of the complex formed by the CBP SID domain and the AD1 domain of the p160 protein SRC1. As mentioned above the interaction between CBP SID and the p160 AD1 domains is vital for the recruitment and assembly of multiprotein complexes that are required for nuclear receptor mediated

transcription initiation. The structure of the CBP SID / SRC1 AD1 complex would therefore be likely to provide important insights into the functions and interactions of CBP. Abnormalities in the expression or misregulation of both CBP and the p160s have been strongly linked to human disease including cancer (reviewed in Goodman & Smolik, 2000; Kalkhoven, 2004; Xu & Li, 2003). Therefore detailed structural knowledge of the CBP SID / p160 AD1 complex could potentially form the basis for the development of therapeutic agents that could regulate nuclear receptor mediated transcription. During the initial stages of this project the structure of the CBP SID domain in complex with the ACTR AD1 domain was published (Demarest *et al.*, 2001). The structure determined for CBP SID / SRC1 AD1 was compared with the published CBP SID / ACTR AD1 complex, with the aim of understanding the molecular basis for the functional specificity of different CBP / p160 complexes.

In addition to binding p160s, the SID domain also facilitates interactions with a number of other unrelated nuclear proteins including the transcription factors Ets1, Ets-2, p53 and IRF-3, and viral activators such as E1A, KSHV IRF-1 and Tax (Kamei *et al.*, 1996; Lin *et* al., 2001; Livengood *et al.*, 2002; Scoggin *et* al., 2001; Sheppard *et al.*, 2001). Competition between proteins for binding to the CBP SID domain has been shown to contribute to the negative cross-talk between different signalling pathways (Matsuda *et al.*, 2001). The CBP SID / p160 AD1 complexes were also compared to the recently published structure of the CBP SID / IRF-3 IAD complex with the aim of understanding the ability of the CBP SID domain to interact with a number of diverse transcriptional proteins and any potential functional differences between the different CBP SID complexes.

(2) The structure of the C-terminal MA-3 domain of Pdcd4 and the characterisation of its interaction with the translation initiation factor eIF4A.

A number of studies have highlighted the potential roles of the tumour suppressor protein Pdcd4 in normal cell physiology and cancer prevention (Afonja *et al.*, 2004; Chen *et al.*, 2003; Jansen *et al.*, 2004; Jansen *et al.*, 2005). The cellular functions of Pdcd4 have just started to emerge, with recent work indicating that Pdcd4 plays essential roles in the regulation of both transcription and translation (Bitomsky *et al.*, 2004; Yang *et al.*, 2004; Yang *et al.*, 2003a). The regulation of translation involves specific interactions with the eukaryotic initiation factors eIF4A and eIF4G, which are mediated via the two tandem MA-3 domains (Yang *et al.*, 2004; Yang *et al.*, 2003a). MA-3 domains are present in over 200 eukaryotic proteins, including eIF4G. At the commencement of this project no structure or detailed interaction data existed for any region of Pdcd4 or MA-3 domain. The initial aim of this work was to solve the solution structure of the C-terminal MA-3 domain of Pdcd4 (Pdcd4 MA-3_C), and subsequently use chemical shift mapping experiments to characterize the interaction between Pdcd4 MA-3_C and eIF4A. This will provide important insights into the general structure of MA-3 domains, and will for the first time define the regions of MA-3 domains involved in binding eIF4A, and potentially other partner proteins. The structural and interaction data will significantly enhance our understanding of the interaction between Pddc4 and eIF4A and the role of Pdcd4 in the regulation of cap dependent translation. These results should also lead to a clearer picture of the functions of Pdcd4 that are associated with its role as a tumour suppressor and could potentially pave the way for the development of novel therapeutics for cancer treatment.

CHAPTER 2

SOLUTION STRUCTURE OF THE COMPLEX FORMED BETWEEN THE CBP SID / SRC1 AD1 DOMAINS

2.1 Introduction

2.1.1 Nuclear Receptor-Mediated Transcriptional Activation

Nuclear receptor proteins activate the transcription of specific genes in response to hormonal or metabolic stimuli. They function by recruiting coactivator complexes, which act as scaffolds or bridges, or by posttranslationally modifying chromatin or regulatory proteins (reviewed in Perissi & Rosenfeld, 2005; Rosenfeld *et al.*, 2006). The nuclear receptor superfamily includes ligand-inducible transcription factors, such as the steroid, retinoid and thyroid hormone receptors, as well as orphan receptors for which no ligand has yet been identified (Giguere, 1999; Laudet, 1997). Nuclear receptor proteins have a conserved domain structure (reviewed in Aranda & Pascual, 2001; Jordan, 2003), which includes a highly conserved zinc finger DNA binding domain (DBD) (region C) (Luisi *et al.*, 1991; Schwabe *et al.*, 1993), a hinge region (region D) and a ligand binding domain (LBD) (region E). In addition to binding ligand, the LBD also mediates receptor dimerisation. The conserved regions of the human estrogen receptor α (hERα) are illustrated in figure 2.1.

Nuclear receptors also contain two transactivation domains, AF-1 and AF-2, which can function either independently or synergistically to facilitate interactions with transcriptional coactivators (Lees *et al.*, 1989; Tora *et al.*, 1989). The ligand independent AF-1 (activation function-1) is located in the A/B region and is regulated by phosphorylation, which is mediated by several signalling pathways, including the mitogen-activated protein kinase (MAPK) signalling cascade, and cyclin dependent kinases

(Bunone *et al.*, 1996; Kato *et al.*, 1995; Legoff *et al.*, 1994; RochetteEgly *et al.*, 1997; Rogatsky *et al.*, 1999). Phosphorylation of AF-1 is believed to regulate nuclear receptor activity by enhancing the ligand independent recruitment of coactivators such as the p160 protein steroid receptor coactivator 1 (SRC1) (Dutertre & Smith, 2003).

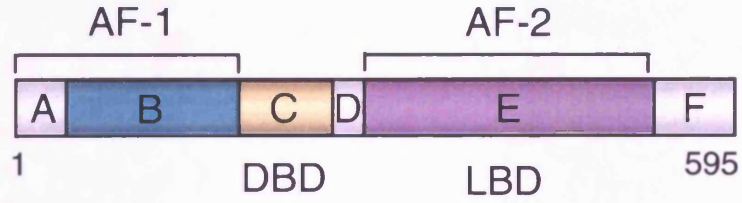


Figure 2.1 Schematic representation of the functional regions within the human estrogen receptor α. Nuclear Receptors contain six conserved regions: A (residues 1-38 of hERα), B (39-179), C (180-262), D (263-301), E (302-552) and F (553-595). The DNA binding domain is located within region C and the ligand binding domain within region E. The activation functions AF-1 and AF-2 are contained within the variable A/B region and ligand binding domain (region E) respectively.

AF-2 (activation function-2) is located in the LBD and is regulated by ligand binding (Hollenberg & Evans, 1988; Lees *et al.*, 1989; Webster *et al.*, 1988). Agonist binding activates transcription either by stabilising, or causing conformational changes to the LBD, (Bourguet *et al.*, 1995; Nolte *et al.*, 1998; Renaud *et al.*, 1995) (reviewed in Nagy & Schwabe, 2004) allowing coactivators, such as the p160 proteins, to bind through the common sequence motif LXXLL (where L is a leucine residue and X is any other residue) (Heery *et al.*, 1997). This motif adopts an α helical structure on binding to the LBD of nuclear receptors (Nolte *et al.*, 1998; Shiau *et al.*, 1998). Antagonist binding repositions

helix 12, such that it occludes the coactivator recognition site, thereby preventing coactivator binding (Brzozowski *et al.*, 1997; Shiau *et al.*, 1998). A comparison of the structures of the agonist and antagonist bound LBD of the hERα is shown in figure 2.2.

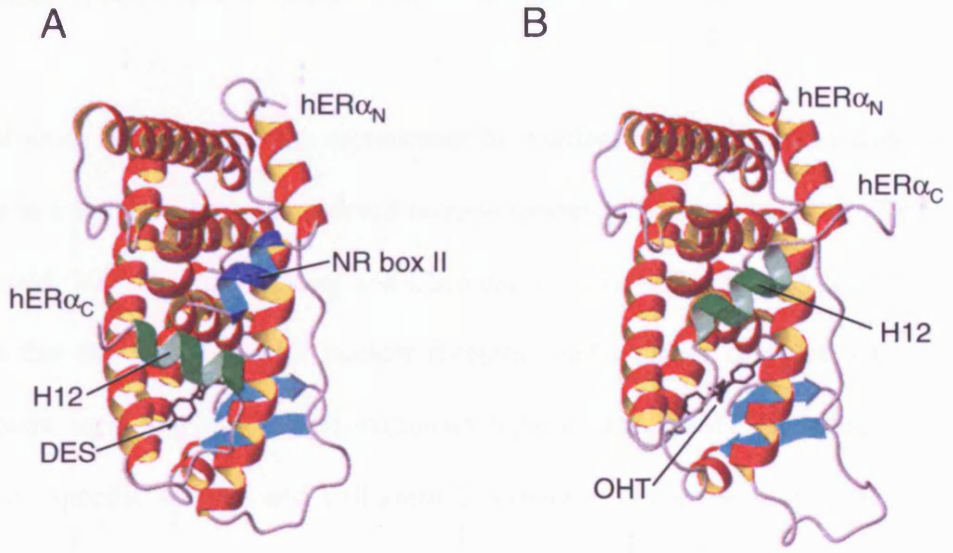


Figure 2.2 Structure of the ligand binding domain of human estrogen receptor α. Ribbon representations of the backbone topology of the hERα (human estrogen receptor α) LBD bound to the agonist diethylstilbestrol (DES) and the antagonist 4-hydroxytamoxifen (OHT) are shown in panels A and B respectively (Shiau *et al.*, 1998). Agonist binding stabilises the active orientation of helix 12 (H12), which is shown in green, allowing the p160 coactivators to bind. The nuclear receptor binding motif II (NR box II) of the p160 TIF2 is shown in blue. In panel B H12 is repositioned by antagonist binding such that it occludes the coactivator recognition site.

In the absence of ligand, classical steroid receptors, such as the glucocorticoid receptor, reside in the cytoplasm. However, heterodimeric nuclear receptors, such as the retinoic acid receptor and the thyroid receptor, are constitutively bound to DNA and often recruit repressive complexes to the target promoter where they generate a repressive chromatin

environment (reviewed in McKenna & O'Malley, 2002). In addition to inducing the exchange of corepressors for coactivators (reviewed in Glass & Rosenfeld, 2000), ligand binding can also result in the nuclear translocation of receptors and the formation of either homo- or heterodimers, which are able to associate with the target promoter on specific sequences termed response elements (reviewed in Aranda & Pascual, 2001).

Recent work suggests that the recruitment of coactivator complexes to nuclear receptors occurs in a sequential, highly ordered process (reviewed in Metivier *et al.*, 2006; Perissi & Rosenfeld, 2005). Cross linking and Chromatin immunoprecipitation (ChIP) assays have shown that the recruitment of nuclear receptors and cofactor complexes to a number of promoters, including pS2, mouse mammary tumour virus (MMTV), cathepsin D (CATD), prostrate specific antigen and kallikrein 2 promoter, occurs in a cyclical pattern. This allows ligand levels to be continually sampled allowing a rapid response to changes in ligand concentration (Kang *et al.*, 2002b; Metivier *et al.*, 2003; Nagaich *et al.*, 2004; Shang *et al.*, 2000).

The best characterised example of cyclical recruitment is the hERα at the pS2 promoter of MCF-7 cells (Metivier *et al.*, 2003). A summary of the ChIP assay data and the resulting proposed model of transcriptional activation at the pS2 promoter are shown in figures 1.2 and 2.3 respectively. An initial cycle was also observed, which fails to initiate transcription but serves to generate a permissive chromatin environment at the core promoter through the recruitment of the ATP-dependent remodelling complex SWI/SNF. It is unknown if this cycle exists under normal physiological conditions *in vivo*, or whether it is a result of the treatment of the cells with α-amanitin, which is required to inhibit basal transcription initiation. A similar cycle was also observed in the absence of ligand. It has been suggested that this cycle functions to keep the chromatin poised, such that an

immediate response to estrogen can occur. The existence of this cycle is somewhat surprising, as in the absence of ligand, the steroid hormone receptors are believed to reside predominantly in the cytoplasm.

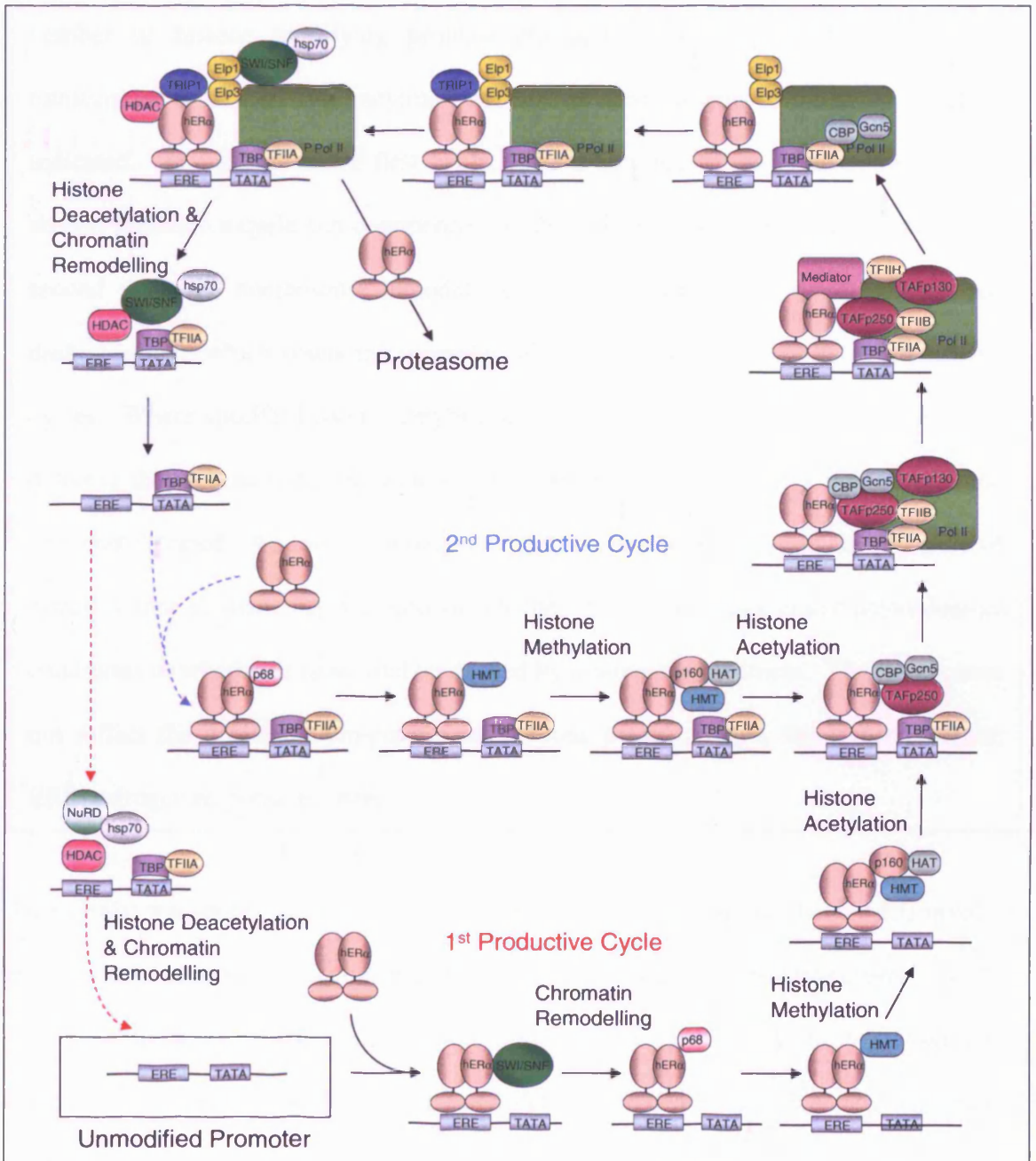


Figure 2.3 Ordered cyclic recruitment by the estrogen receptor α at the pS2 promoter. The figure shows a summary of the model of cyclic ERα and cofactor recruitment proposed by Metivier *et al,* based on kinetic ChIP experiments

(Metivier *et al.*, 2003; Metivier *et al.*, 2006; Reid *et al.*, 2003). In the presence of ligand two similar, transcrptionally productive cycles were observed. The first cycle begins with the recruitment of ligand bound ERα and the remodelling complex SWI/SNF to the unmodified promoter (shown in the box in the bottom left corner). A number of histone modifying proteins are then recruited prior to the general transcription machinery; the enzymatic acivities of these proteins on the nucleosomes is indicated. At the end of the first cycle, ERα is targeted to the proteasome and the second productive cycle can commence (via the blue dashed arrow). At the end of the second cycle the nucleosome remodelling complex NuRD is recruited (via the red dashed arrow), which resets the promoter, allowing for subsequent sets of productive cycles. Where specific histone acetyltransferases and histone methyltransferases were detected they are named, otherwise HMT or HAT is used to signify one of multiple enzymes detected. An initial unproductive cycle was also observed in the presence of estrogen (not shown). It is unknown whether this cycle exists under physiological conditions or whether it is an artefact caused by α-amanitin treatment. The figure does not reflect the actual protein-protein interactions present. Elps, Elongator proteins; ERE, estrogen response element.

Two similar transcriptionally productive cycles were found to follow. Each cycle involves the sequential recruitment of coactivator complexes that serve to remodel chromatin and/or recruit components of the basal transcription machinery. The cycles begin with the recruitment of either the RNA helicase p68 or SWI/SNF, followed by complexes containing histone methyltransferases (HMTs), members of the p160 protein family and HATs. The subsequent recruitment of components of the basal transcription machinery and the TRAP/mediator complex results in the phosphorylation of Pol II and the initiation of transcription (Metivier *et al.*, 2003). At the end of each cycle hERα becomes

ubiquitinated and targeted to the proteasome allowing further cycles to occur (Reid *et al.*, 2003). In addition, histone deacetyltransferase (HDAC)-SWI/SNF complexes remodel the local chromatin structure inducing a restrictive environment (Metivier *et al.*, 2003). The ATP-dependent remodelling complex, NuRD (nucleosome remodelling and histone deacetylation complex) is specifically recruited at the end of the second transcriptionally productive cycle to remodel the nucleosome associated with the TATA box (Metivier *et al.*, 2003).

Cyclical recruitment of transcriptional complexes has also been observed at other promoters and for other nuclear receptors although different sequences of receptor/coactivator recruitment were observed (Agalioti *et al.*, 2002; Kang *et al.*, 2002b; Shang *et al.*, 2000; Sharma & Fondell, 2002; Vaisanen *et al.*, 2005). For example, on the CATD promoter, TRAP/mediator and the p160s were shown to be recruited simultaneously by the ERα, whilst CBP was not recruited until after Pol II (Shang *et al.*, 2000). Unlike the ER and androgen receptor, the thyroid receptor is not targeted to the proteasome at the end of each cycle, but persists on thyroid hormone responsive promoters (Sharma & Fondell, 2002).

In addition to activating transcription, some nuclear receptors have been shown to inhibit transcription in a ligand-dependent manner, either by binding negative response elements or corepressors, or by antagonising the activities of transcription factors (Delage-Mourroux *et al.*, 2000; Kamei *et al.*, 1996; White *et al.*, 2004) (reviewed in Gronemeyer *et al.*, 2004; Rosenfeld *et al.*, 2006). For example, competition between glucocorticoid receptor and retinoic acid receptor for limiting amounts of CBP has been shown to inhibit Activator protein-1 (AP-1) activation (Kamei *et al.*, 1996). In addition, a selective repressor of estrogen receptor activity (REA) is recruited to hormone bound ER, preventing the ER

from binding to the p160 proteins (Delage-Mourroux *et al.*, 2000).

2.1.2 p160 Steroid Receptor Coactivator Family

2.1.2.1 Overview of the p160 Proteins

The p160 steroid receptor coactivator proteins (SRC) are single chain polypeptides (approximately 160 kDa), which function as transcriptional coactivators for a number of transcription factors, including the nuclear receptors, AP-1, MyoD and CREB (Lee *et al.*, 1998b; Onate *et al.*, 1995; Torchia *et al.*, 1997; Wu *et al.*, 2005). Research into the p160 proteins has primarily focused on their role in ligand-dependent, nuclear receptor mediated transcription initiation. The p160s enhance transcription by recruiting HATs and HMTs to promoters via interactions with nuclear receptors (reviewed in Xu & Li, 2003; Xu & O'Malley, 2002).

The p160 gene family contains three homologous members: SRC1, also known as NCoA-1 (nuclear receptor coactivator 1) (Onate *et al.*, 1995), TIF2 (transcription intermediary factor 2), also known as GRIP1 (glucocoticoid receptor-interacting protein), SRC2 or NCoA-2 (Hong *et al.*, 1996; Voegel *et al.*, 1996), and ACTR (activator for thyroid hormone and retinoid receptors), also known as p/CIP (p300/CBP interacting protein), RAC3 (receptor-associated coactivator 3), AIB1 (amplified in breast cancer1), TRAM-1 (thyroid hormone receptor activator molecule 1), SRC3 or NCoA-3 (Anzick *et al.*, 1997; Torchia *et al.*, 1997).

The amino acid sequence of the three p160 proteins is highly conserved (sequence identity 43-48%) (Xu & Li, 2003), and the proteins share a conserved domain organisation, which is shown for the p160 protein SRC1 in figure 2.4. The most conserved portion is the N-terminal region which encompasses the basic helix-loop-helix (bHLH) and Per/Ah receptor

nuclear translocator (ARNT)/Sim (PAS) domains (Xu & Li, 2003). This region is involved in protein-protein interactions. For example, the bHLH/PAS domains of TIF-2 have been shown to interact with several myogenic factors leading to the coactivation of myogenin and MEF-2C mediated transactivation (Chen *et al.*, 2000), whilst the interaction of STAT6 with the PAS-B domain of SRC1 is required for the transcriptional activation of IL-4 (Litterst & Pfitzner, 2001; Litterst & Pfitzner, 2002). In addition, the bHLH/PAS domains were shown to mediate the formation of p160 heterodimers on the ER-binding fragment-associated antigen-9 (EBAG9) promoter (Zhang *et al.*, 2004). The nuclear receptor interaction domain (NID) is located in the central region of the p160 proteins. The interaction between the NID and the AF-2 domain of ligand bound nuclear receptors is mediated by one of three LXXLL motifs (where L is leucine and X any amino acid) (Heery *et al.*, 1997; Torchia *et al.*, 1997). Each LXXLL motif forms part of an amphipathic α helix, which is able to bind to a conserved hydrophobic cleft in the surface of the LBD of ligand bound nuclear receptors, as shown in figure 2.2 (Shiau *et al.*, 1998). The glutamine (Q) -rich region of the p160s has been shown to interact with the AF-1 of the androgen and estrogen receptors, and is therefore believed to play a role in ligand independent transcriptional initiation (Bevan *et al.*, 1999; Ma *et al.*, 1999; Webb *et al.*, 1998). The p160 proteins also contain two transcriptional activation domains (AD1 and AD2). The AD1 domain interacts with acetyltransferases, including CBP and p300, and acts as a potent transcriptional activator in mammalian and yeast cells (Chen *et al.*, 1997; Kalkhoven *et al.*, 1998; Kamei *et al.*, 1996; McInerney *et al.*, 1998; Sheppard *et al.*, 2001; Sheppard *et al.*, 2003; Voegel *et al.*, 1998), while the C-terminal AD2 domain has been shown to interact with the arginine methyltransferases CARM1 (Chen *et al.*, 1999a) and PRMT1 (Koh *et al.*, 2001). These acetyltransferases and methyltransferases are involved in chromatin remodelling, as well as the posttranslational modification of other transcriptional regulators (Chen *et al.*, 1999b; Xu *et al.*, 2001). The C-terminal regions of

SRC1 and ACTR have also been reported to possess weak HAT activity, although they show no sequence similarity to other known HAT domains (Chen *et al.*, 1997; Spencer *et al.*, 1997). However, the HAT activity of SRC1 is believed to be relatively unimportant in nuclear receptor mediated transcription (Liu *et al.*, 2001), though it is required for the coactivation of the myogenic factor MyoD (Wu *et al.*, 2005).

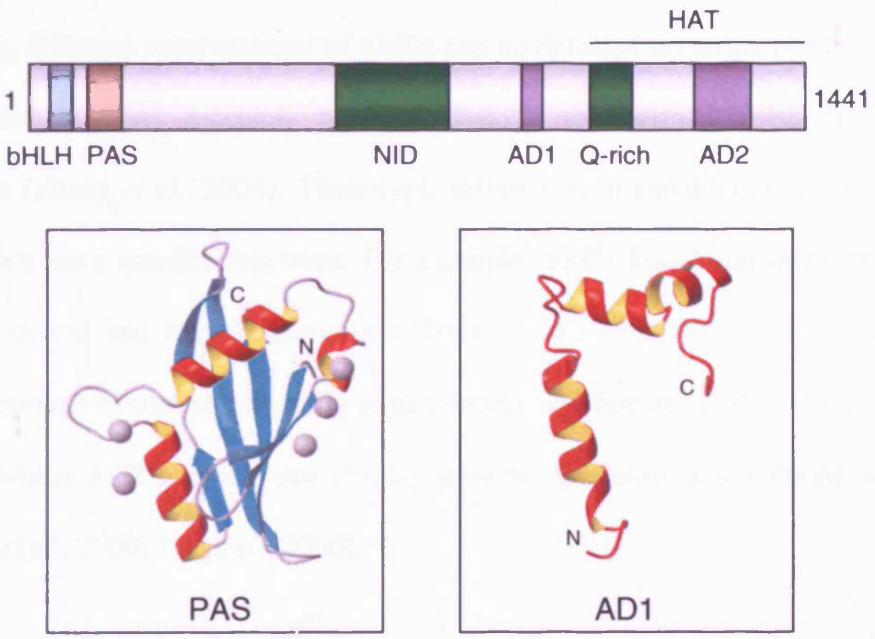


Figure 2.4 Schematic diagram of the functional regions of human SRC1. The basic helix-loop-helix (bHLH) (residues 24-81), Per/ARNT/Sim region (PAS) (109-385), nuclear receptor interaction domain (NID) (570-780), activation domain-1 (AD1) (926-970), glutamine rich region (Q-rich) (1053-1185) and activation domain-2 (AD2) (1240-1345) are indicated. The ribbon representation of the backbone topology of the PAS-B (residues 257-385) (Razeto *et al.*, 2004) (PDB code 1OJ5) and AD1 (Demarest *et al.*, 2002) (1KBH) domains are also shown. The C-terminal region which has been reported to possess weak HAT activity is also indicated.

The viability of p160 knock-out mice, together with analysis of p160 expression patterns (Gehin *et al.*, 2002; Nishihara *et al.*, 2003; Xu *et al.*, 1998), indicate that the p160s can partially substitute for each other functionally. Despite this partial redundancy between the p160s, they have also been shown to have specific functions. For example, SRC1 and ACTR have been shown to interact with thymine DNA glycosylase, whereas TIF2 does not (Lucey *et al.*, 2005). SRC1 and ACTR have also been shown to function as coactivators of MyoD-mediated transcription, whereas TIF-2 functions as a corepressor (Wu *et al.*, 2005). Interestingly, different combinations of p160s can be detected on target promoters for the androgen and estrogen receptor, possibly through the formation of distinct p160 heterodimers (Zhang *et al.*, 2004). Phenotypic differences in knock-out mice also indicate that the p160s have specific functions. For example, SRC1 knock-out mice are partially resistant to steroid and thyroid hormones (Weiss *et al.*, 1999; Xu *et al.*, 1998), TIF2 knock-out animals exhibit significantly reduced male and female fertility levels (Gehin *et al.*, 2002), whilst ACTR knock-outs display growth retardation and reduced adult body size (Wang *et al.*, 2000; Xu *et al.*, 2000).

Chromosomal rearrangements involving the p160 proteins have been implicated in the development of several specific cancers. For example, a translocation resulting in the production of a MOZ-TIF2 fusion protein is associated with acute myeloid leukaemia (Carapeti *et al.*, 1998). This fusion protein contains the C-terminal region of TIF2, including the AD1 and AD2 domains, and the acetyltransferase catalytic domain of MOZ (monocytic leukaemia zinc finger protein). A translocation involving a similar region of SRC1 and the transcription factor PAX3 has recently been identified in alveolar rhabdomyosarcoma (Wachtel *et al.*, 2004). In addition, the overexpression of ACTR has been implicated in the development of primary breast and ovarian cancers (Anzick *et al.*, 1997; Bautista *et al.*, 1998).

2.1.2.2 Steroid Receptor Coactivator 1

SRC1 was originally identified as a coactivator of ligand-bound human progesterone receptor, and was subsequently shown to be a coactivator of thyroid hormone receptor and estrogen receptor (Kamei *et al.*, 1996; Onate *et al.*, 1995; Takeshita *et al.*, 1996). The two main splice variants of human SRC1, SRC1A and SRC1E, are widely expressed in different cell lines and differ in their ability to enhance estrogen receptor mediated transcription (Hayashi *et al.*, 1997; Kalkhoven *et al.*, 1998).

SRC1 is expressed in many tissue types, including, brain, testis, lung, liver, kidney and heart (Torchia *et al.*, 1997; Xu *et al.*, 1998). It has been shown to play an important role in brain development and function (Nishihara *et al.*, 2003). For example, SRC1 is critically involved in the hormone-dependent development of normal male reproductive behavior and brain morphology (Auger *et al.*, 2000). In addition, adult SRC1 knockout mice have been shown to exhibit moderate motor learning dysfunction and delayed development of cerebellar purkinje cells (Nishihara *et al.*, 2003).

2.1.2.3 Activation Domain 1

The AD1 domain of the p160 proteins is a potent, CBP/p300-dependent, transcriptional activator (Chen *et al.*, 1997; Kalkhoven *et al.*, 1998; Sheppard *et al.*, 2001; Torchia *et al.*, 1997; Voegel *et al.*, 1998). In transiently transfected cells, only the NID and AD1 domains of SRC1 were required to enhance ligand-dependent nuclear receptor mediated transcription (Sheppard *et al.*, 2001). The AD1 domain of human SRC1 has been mapped to residues 926-970, whilst 926-960 have been shown to be sufficient for binding to the SRC1 interaction domain (SID) of CBP (Sheppard *et al.*, 2001). Similar domain boundaries have been observed for TIF2 (1041-1106) and ACTR (1039-1088) (Chen *et al.*, 1997; Voegel *et al.*, 1998). The amino acid sequence of the AD1 domains is well

conserved among the p160 proteins, with approximately 54% homology as illustrated in figure 2.5.

Figure 2.5 Multiple sequence alignment of the AD1 domains of human SRC1, TIF2 and ACTR. The alignment illustrates the high degree of sequence homology between the p160 AD1 domains. Residues with conserved sequence identity are highlighted in red, those which are conserved in two domains are highlighted in yellow. The consensus sequence is shown below. The alignment was prepared using ClustalW and ESPript.cgi (http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl)

The AD1 domain has been shown to be intrinsically disordered in isolation, but upon complex formation with the CBP SID domain folds into a highly helical structure (Demarest *et al.*, 2002; Lin *et al.*, 2001). The structure of the AD1 domain of ACTR, which was solved in complex with CBP SID (Demarest *et al.*, 2002) is shown in figure 2.4.

Phosphorylation of residues, outside of the AD1 domain (T24, S543, S857, S860 and S867 of ACTR), has been shown to enhance the interaction of ACTR with CBP (Wu *et al.*, 2004). This may be as a result of additional binding between other regions of ACTR and CBP and/or conformational changes in ACTR that affect the affinity of the AD1 domain for CBP. Only one of these residues, S543 (S517 of SRC1) has been shown to be phosphorylated in SRC1 (Rowan *et al.*, 2000), suggesting that the specific phosphorylation

of non-conserved residues could play an important role in the regulation of the individual p160 proteins.

2.1.3 CREB binding protein (CBP)

2.1.3.1 Overview of the homologous coactivators CBP and p300

The transcriptional coactivator CBP (cAMP Response Element Binding (CREB)- Binding Protein) and its homologue p300 were originally identified in protein interactions assays using CREB (Chrivia *et al.*, 1993) and the adenoviral transforming protein E1A (Stein *et al.*, 1990) respectively. Through their function as coactivators, CBP/p300 participate in the regulation of diverse cellular processes including proliferation, differentiation and apoptosis (reviewed in Chan & La Thangue, 2001; Giordano & Avantaggiati, 1999; Goodman & Smolik, 2000).

The coactivator function of CBP/p300 is dependent on its ability to interact with a variety of different proteins which are involved in transcription, including transcription factors such as CREB, C/EBP, c-Myb and c-Jun (Bannister *et al.*, 1995; Chrivia *et al.*, 1993; Dai *et al.*, 1996; Mink *et al.*, 1997) and coregulators such as pCAF and the p160s (Chen *et al.*, 1997; Torchia *et al.*, 1997; Yang *et al.*, 1996; Yao *et al.*, 1996). The interactions are mediated by a number of conserved interaction domains that are linked by highly polar flexible linkers (Dyson & Wright, 2005). A schematic of the domain organisation of CBP, together with a partial list of binding partners, is shown in figure 2.6.

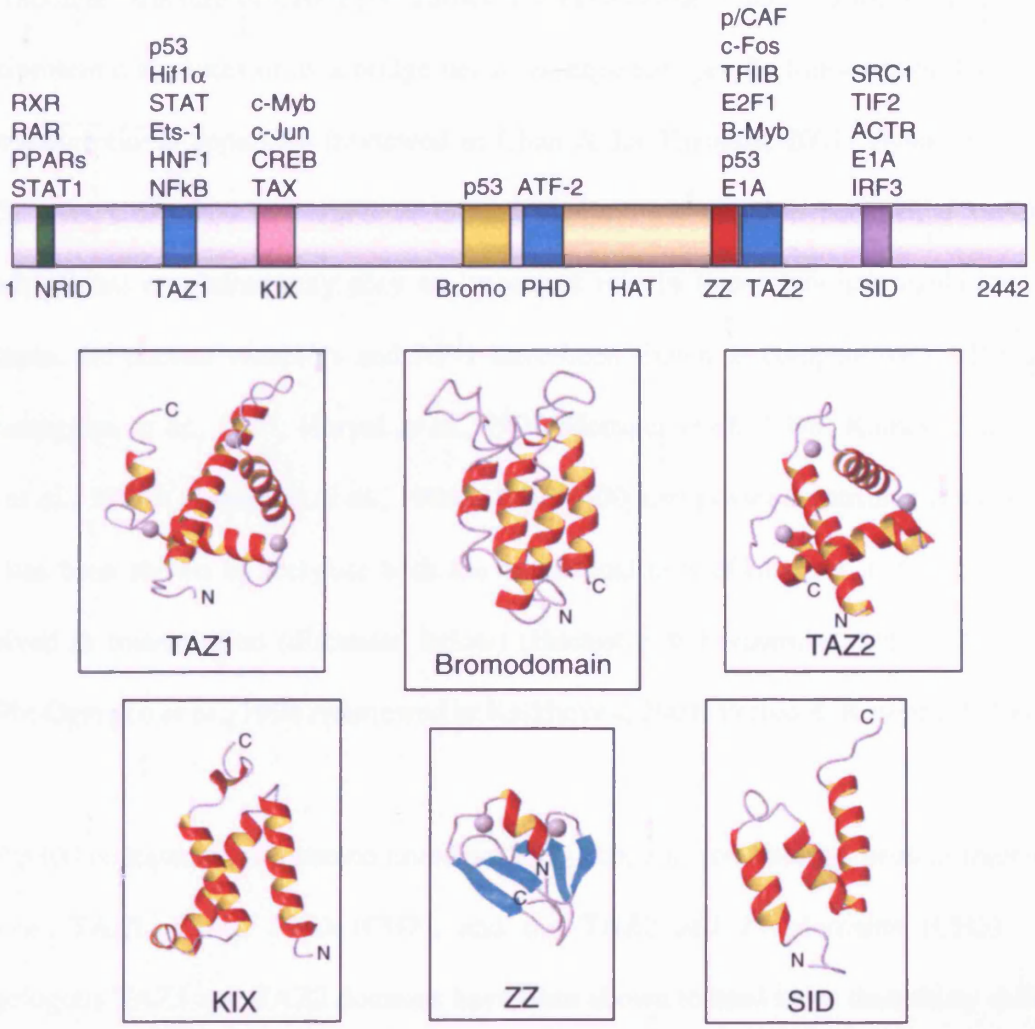


Figure 2.6 Schematic views of the functional regions of mouse CBP with a partial list of proteins that bind the conserved domains. The nuclear receptor interaction domain (NRID) (residues 60-74), TAZ1 (346-438), KIX (586-672), Bromo (1081-1196), PHD (1225-1318), HAT (1319-1694), ZZ (1695-1752), TAZ2 (1764-1855) and SID (2059-2117) domains are indicated. Ribbon representations are shown for the backbone topologies of the TAZ1 domain (De Guzman *et al.*, 2005) (PDB code 1U2N), KIX domain (Radhakrishnan *et al.*, 1997) (1KDX), Bromodomain (Mujtaba *et al.*, 2004) (1JSP), ZZ domain (Legge *et al.*, 2004) (1TOT), TAZ2 domain (De Guzman *et al.*, 2000) (1FS1) and SID domain (Demarest *et al.*, 2002) (1KDH).

The modular structure of CBP/p300 allows it to act as either a scaffold for the assembly of multiprotein complexes or as a bridge between sequence-specific transcription factors and the transcriptional apparatus (reviewed in Chan & La Thangue, 2001; Dyson & Wright, 2005). As CBP/p300 is present in limited amounts competition between a variety of transcriptional regulators may play an important role in transcriptional regulation. For example, the nuclear receptors and AP-1 have been shown to compete for CBP binding (Avantaggiati *et al.*, 1997; Horvai *et al.*, 1997; Hottiger *et al.*, 1998; Kamei *et al.*, 1996; Lee *et al.*, 1998a; Nakajima *et al.*, 1996). CBP/p300 also posseses intrinsic HAT activity and has been shown to acetylate both the N-terminal tails of histones and other proteins involved in transcription (discussed below) (Bannister & Kouzarides, 1996; Chen *et al.*, 1999b; Ogryzko *et al.*, 1996) (reviewed in Kalkhoven, 2004; Perissi & Rosenfeld, 2005).

CBP/p300 contains three cysteine-histidine (CH) rich, zinc ion binding, protein interaction regions: TAZ1 (CH1), PHD (CH2), and the TAZ2 and ZZ domains (CH3). The homologous TAZ1 and TAZ2 domains have been shown to bind more than thirty different transcription factors, and a number of viral proteins, examples of which are shown in figure 2.6. The TAZ1 domain, which interacts with HIF1α and CITED2, has been shown to play a role in regulating the hypoxic response (Arany *et al.*, 1996; Bhattacharya *et al.*, 1999; De Guzman *et al.*, 2004; Freedman *et al.*, 2002). The CH2 region contains a highly conserved plant homeodomain type zinc finger, which forms an integral part of the enzymatic HAT domain of CBP (Bordoli *et al.*, 2001; Kalkhoven *et al.*, 2002). The HAT domain of CBP/p300 has been shown to acetylate both histones and transcriptional regulators. For example, CBP has been shown to acetylate ACTR resulting in its dissociation from the nuclear receptor LBD (Chen *et al.*, 1999b). Contained within the HAT domain is a 60 residue highly polar/charged sequence which is believed to regulate its catalytic activity (Thompson *et al.*, 2004). The bromodomain mediates the interaction

of CBP/p300 with acetylated proteins, including the acetylated C-terminal region of p53, MyoD and histones (Mujtaba *et al.*, 2004; Polesskaya *et al.*, 2001; Ragvin *et al.*, 2004). The KIX domain binds the transactivation domain of a number of transcription factors, many of which are unstructured in isolation, including the phosphorylated KID domain of CREB and c-Jun, as well as the viral proteins Tat and TAX (Campbell & Lumb, 2002; Radhakrishnan *et al.*, 1997; Vendel & Lumb, 2003; Vendel *et al.*, 2003). The C-terminal SID domain (also referred to as the IRF3 interaction domain (IBiD)) mediates the interaction of CBP/p300 with a number of diverse nuclear proteins including the p160 proteins, Ets-1, Ets-2, E1A, IRF-3, TAX, p53 and CAS (Kamei *et al.*, 1996; Lin *et al.*, 2001; Livengood *et al.*, 2002; Matsuda *et al.*, 2004; Ryan *et al.*, 2006; Scoggin *et al.*, 2001; Sheppard *et al.*, 2001). CBP/p300 is generally believed to be indirectly recruited to steroid hormone receptors via its interaction with the p160 proteins (Coulthard *et al.*, 2003; Sheppard *et al.*, 2001), however, CBP/p300 is also able to interact directly with nuclear receptors via their N-terminal nuclear receptor interaction domain (NRID) (Chakravarti *et al.*, 1996; Heery *et al.*, 2001; Kamei *et al.*, 1996). As with the p160 nuclear receptor interaction domain this interaction is also mediated by an LXXLL motif (Heery *et al.*, 2001).

CBP and p300, are believed to have many overlapping roles, for example during embryonic development where CBP and p300 knockout mice show similar phenotypes (Chan & La Thangue, 2001; Goodman & Smolik, 2000; Yao *et al.*, 1998). However, they have also been shown to have unique functions, for example, p300, but not CBP, is required for the differentiation of F9 cells (Chan & La Thangue, 2001; Goodman & Smolik, 2000; Kawasaki *et al.*, 1998; Ugai *et al.*, 1999). The HAT activities of CBP and p300 has also been shown to exhibit distinct substrate specificity profiles (McManus & Hendzel, 2003). For example, CBP, but not p300, has been shown to acetylate K12 of

histone 4, whilst p300, but not CBP, was shown to acetylate K8 of histone 4. As shown in figure 2.3, CBP is specifically recruited with Gcn5 and TAF250 to the pS2 promoter by ERα during both transcriptionally productive cycles (Metivier *et al.*, 2003).

CBP/p300 is regulated by several signal transduction pathways, which compete for the limited amounts of CBP/p300 in order to regulate the transcription of particular responsive genes, for example the Ras pathway (reviewed in Chan & La Thangue, 2001; Goodman & Smolik, 2000). The activity of CBP/p300 has been shown to be positively or negatively regulated by a number of kinases including calcium calmodulin (CaM) kinase IV, protein kinase A (PKA), MAPK, CDC2 and CDK2 (Ait-Si-Ali *et al.*, 1998; Banerjee *et al.*, 1994; Chawla *et al.*, 1998; Hu *et al.*, 1999; Janknecht & Nordheim, 1996; Nakajima *et al.*, 1996; Perkins *et al.*, 1997).

Abnormalities in the expression, or mis-regulation of CBP/p300 has been strongly linked to human disease (reviewed in Goodman & Smolik, 2000; Kalkhoven, 2004). For example, haploinsufficiency of CBP is believed to result in the developmental disorder Rubinstein Tabi syndrome, which results in multiple abnormalities including growth and mental retardation, and increased tumour risk (Miller & Rubinstein, 1995; Rubinstein & Taybi, 1963). Mutations in CBP/p300 have also been observed in a number of tumours, which is consistent with reports that CBP and p300 posses tumour suppressor activity (reviewed in Iyer *et al.*, 2004; Kishimoto *et al.*, 2005). In addition, a number of viral proteins, such as the viral oncoprotein E1A, SV40 large T antigen and HTLV Tax, have been shown to target CBP/p300 in order to affect the transcription of genes involved in cellular processes including cell growth and differentiation (Bex *et al.*, 1998; Eckner *et al.*, 1994; Eckner *et al.*, 1996; Stein *et al.*, 1990; Yang *et al.*, 1996; reviewed in Goodman & Smolik, 2000; Iyer *et al.*, 2004). Several chromosomal translocations that generate fusions

of CBP, or its partner proteins, with other proteins are associated with hematologic malignancies (reviewed in Iyer *et al.*, 2004; Kalkhoven, 2004). For example, the interaction of CBP with the MOZ-TIF2 fusion protein is an important feature of MOZ-TIF2 action in acute myeloid leukaemia. This interaction, which is dependent on the AD1 domain of TIF2, has been shown to reduce the cellular levels of CBP, disrupting the activity of CBP-dependent activators, such as the nuclear receptors and p53 (Collins *et al.*, 2006; Kindle *et al.*, 2005, reviewed in Troke *et al.*, 2006).

2.1.3.2 SRC1 Interaction Domain (SID)

The interaction between the C-terminal SID domain of CBP/p300 and the p160 AD1 domain is essential for ligand dependent transcription mediated by steroid hormone receptors (Kim *et al.*, 2001; Sheppard *et al.*, 2001). The SID domain has been mapped to residues 2058-2117 (Demarest *et al.*, 2002; Sheppard *et al.*, 2001). The secondary and tertiary structure of the isolated domain has previously been studied by far UV circular dichroism and NMR (Demarest *et al.*, 2002; Heery, Unpublished data; Lin *et al.*, 2001). Although the CBP SID domain was shown to contain helical secondary structure, it is not believed to contain any stable tertiary structure (Demarest *et al.*, 2004; Demarest *et al.*, 2002). Upon binding to partner proteins, such as the p160s, CBP SID folds to form a stable, highly helical protein domain (Demarest *et al.*, 2002). The structure of the CBP SID domain (2058-2117) in complex with the ACTR AD1 domain and a shorter CBP SID construct (2067-2112) in complex with the IAD (IRF association domain) domain of IRF-3 have been solved (Demarest *et al.*, 2002; Qin *et al.*, 2005). These structures suggest that the domain can exist in different conformations depending on its binding partner. The structure of the CBP SID domain, taken from the complex with ACTR is shown in figure 2.6 (Demarest *et al.*, 2002).

As mentioned above the SID domain interacts with a number of coactivators, transcription factors and viral proteins, which contain virtually no sequence similarity. However, a semi conserved leucine rich motif, which corresponds to the amphipathic helix Aα1 of the ACTR AD1 domain, is believed to play a key role in binding CBP SID (Matsuda *et al.*, 2004). Competition between proteins for binding to the CBP SID domain contributes to the negative cross talk between different signalling pathways (Matsuda *et al.*, 2004).

The interactions between CBP SID and SRC1 AD1, E1A and Ets-2 have been investigated by mutational analysis (Matsuda *et al.*, 2004; Sheppard *et al.*, 2001). Whilst very similar results were observed for the SRC1 AD1 and E1A interactions, subtle differences were observed for the interaction between CBP SID and Ets-2 (Sheppard *et al.*, 2001).

As can be seen above structural studies have focused on the SID domain of CBP. The SID domains of CBP and p300 share 68% amino acid sequence identity and 82% sequence homology. With the exception of the substitution of Q2082 for a leucine, all the non-similar amino acid substitutions are located in the N- and C-termini of the domain, in regions that appear to be unstructured (Demarest *et al.*, 2002). Given the high homology between the two proteins, it seems likely that the SID domain of p300 would form very similar complexes to those observed for CBP.

2.1.4 Aims

The aims of the work described in this chapter were to determine the solution structure of the complex formed between the interaction domains of the transcriptional coactivators CBP and SRC1, and compare the features of the complex to that reported for the closely related CBP SID / ACTR AD1 complex. The boundaries of the interacting regions of CBP SID (2059-2117) and SRC1 AD1 (920-970) were based on the results of previous detailed

mapping experiments (Demarest *et al.*, 2002; Sheppard *et al.*, 2001). Expression and purification of the isolated SRC1 AD1 domain had previously been shown to be impossible and precluded preparation of the complex by mixing the individual purified domains. This problem was also reported in structural studies of the closely related ACTR AD1 domain, and presumably indicates that the isolated AD1 domain is unfolded and therefore very susceptible to proteolysis.

2.2 Materials and Methods

2.2.1 Protein Expression and Purification

Uniformly ^{15}N and $^{15}N/^{13}C$ labelled samples of the CBP SID domain (mouse residues 2059-2117) complexed with the SRC1 AD1 domain (human residues 920-970) were expressed and purified in collaboration with Dr. B Yue. The samples were prepared from a modified pET22b *E. coli* expression vector (a gift from Peter Wright, Scripps Institute), which allowed coexpression of the two domains. The vector introduced a C-terminal hexa-histidine tag to the SRC1 domain linked via a 5 residue region. The labelled CBP SID and SRC1 AD1 domains were coexpressed in minimal media containing 0.6 g/l >99% ^{15}N ammonium sulphate and unlabelled or 2 g/l >99% ^{13}C glucose as the sole nitrogen and carbon sources. The medium was also supplemented with 50 mg/L unlabelled methonine as the host strain used (B834) is a methionine auxotroph. Transformed cells were grown at 37°C and expression of the two proteins forming the complex was induced in mid log phase by the addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG) to 1 mM. The cells were harvested at 3 hours post induction by centrifugation (6000 g for 20 minutes), resuspended in 20 mM sodium phosphate, 100 mM sodium chloride and 5mM imidazole buffer at pH 7 and then lysed by sonication. Insoluble cell debris was removed by centrifugation at 13,000 g for 30 minutes and the resulting supernatant containing the CBP

SID / SRC1 AD1 complex was filtered through a 0.2 μm millipore filter in preparation for chromatography. The initial step in the purification of the complex involved affinity chromatography on a Ni-NTA column. After loading the lysate supernatant, and a simple wash step to remove non-bound material, the CBP SID / SRC1 AD1 complex was eluted from the column, using a two step linear gradient of imidazole (5-250mM). Fractions containing the complex were then pooled and subjected to a final polishing purification step by gel filtration on a Superdex 75 16/60 pre-packed column (Amersham Biosciences). The purified complex obtained was judged to be greater than 95% pure by SDS-PAGE (Invitrogen 4-12% Bis-Tris NuPAGE gel system).

2.2.2 Reverse Phase High Performance Liquid Chromatography (HPLC)

Reverse phase HPLC analysis of the complex on a C4 column, was used to determine the stoichiometry of the purified complex. The individual domains were resolved on the C4 column by applying a linear gradient of acetonitrile (1.6-80%) and the relative amounts of each were quantified by their absorbance at 215 nm.

2.2.3 Circular Dichroism (CD) Spectroscopy

Far UV CD spectra were obtained from 0.4 mM samples of the CBP SID / SRC1 AD1 complex in a 20 mM sodium phosphate, 100 mM sodium chloride buffer at pH 7. Samples were placed in a 0.1 mm path length cell and spectra were recorded from 180 to 250 nm at a resolution of 1 nm and scan speed of 20 nm/min, with each spectrum representing the average of 10 accumulations. Prior to analysis CD spectra were corrected for buffer absorbance and the raw data converted to molar CD per residue. The software package CDPro was used to determine the proportions of regular secondary structure in the complex indicated by the CD data (Sreerama & Woody, 2000).

2.2.4 NMR Spectroscopy

NMR spectra were acquired from 0.35 mL samples of 1.5 mM CBP SID / SRC1 AD1 complex in a 20 mM sodium phosphate, 100 mM sodium chloride, 10 μM EDTA and 0.02% (w/v) sodium azide buffer at pH 7, containing either 90% D_2O/10% H_2O or 100% D_2O as appropriate. All NMR data were acquired at 25°C on either an 800 MHz Varian Inova or a 600 MHz Bruker Avance spectrometer (table 2.1). The 2D and 3D spectra recorded to obtain sequence specific assignments for CBP SID and SRC1 AD1 in complex were: ^{1}H TOCSY (Braunschweiler & Ernst, 1983) with mixing times of 40 and 55 ms, and NOESY (Macura & Ernst, 1980) with an NOE mixing time of 100 ms; ^{15}N/^{1}H HSQC (Bodenhausen & Ruben, 1980); TOCSY-HSQC (Marion *et al.*, 1989a) with a mixing time of 50 ms and NOESY-HSQC (Marion *et al.*, 1989b) with an NOE mixing time of 100 ms; ^{13}C/^{1}H HCCH-TOCSY (Bax *et al.*, 1990) with a mixing time of 20 ms, HMQC-NOESY (Zuiderweg *et al.*, 1990) with an NOE mixing time of 100 ms; and ^{15}N/^{13}C/^{1}H HNCACB (Wittekind & Mueller, 1993) and CBCA(CO)NH (Grzesiek & Bax, 1993). Typical acquisition times in F_1 and F_2 for the 3D experiments were 11-13 ms for ^{15}N, 7.5-9.5 ms for ^{13}C and 15 ms for ^{1}H, and an acquisition time of 75 ms in F_3 (^{1}H). The majority of the 3D spectra were collected over approximately 88 hours, 2D ^{1}H experiments over 8.5-24 hours and ^{15}N/^{1}H HSQC spectra over about 30 minutes. Typical acquisition times in 2D experiments were either 70 ms (^{15}N) or 35 ms (^{1}H) in F_1 and 250 ms in F_2 (^{1}H). The WATERGATE method was used to suppress the water signal when required (Piotto *et al.*, 1992). The 3D NMR data were processed using NMRPipe (Delaglio *et al.*, 1995) with linear prediction used to extend the effective acquisition times by up to 1.5 - 2 fold in F_1 and F_2. The spectra were analysed using the XEASY package (Bartels *et al.*, 1995).

Table 2.1 NMR experimental details

Experiment	Solvent	Spectrometer
$^{15}N/^{13}C/^{1}H$ HNCACB	10% D_2O/90% H_2O	600 MHz Bruker Avance
$^{15}N/^{13}C/^{1}H$ CBCA(CO)NH	10% D_2O/90% H_2O	600 MHz Bruker Avance
$^{15}N/^{1}H$ TOCSY-HSQC	10% D_2O/90% H_2O	600 MHz Bruker Avance
$^{15}N/^{1}H$ NOESY-HSQC	10% D_2O/90% H_2O	800 MHz Varian Inova
$^{13}C/^{1}H$ HCCH-TOCSY	100% D_2O	600 MHz Bruker Avance
$^{13}C/^{1}H$ NOESY-HSQC	100% D_2O	800 MHz Varian Inova
^{1}H-^{1}H TOCSY	100% D_2O	600 MHz Bruker Avance
^{1}H-^{1}H NOESY	100% D_2O	600 MHz Bruker Avance

2.2.5 Sequence Specific Assignments

Sequence-specific backbone resonance assignments (N, NH, Cα and Cβ) were obtained for the CBP SID / SRC1 AD1 complex from the identification of intra and inter-residue connectivities in HNCACB, CBCA(CO)NH and $^{15}N/^{1}H$ NOESY-HSQC spectra. Initially, intra and inter-residue amide nitrogen/proton to Cα and Cβ peaks were identified in the triple resonance spectra and used to search for signals from neighbouring residues in the sequence. The identification of signals from adjacent residues was confirmed wherever possible by the observation of NH to NH NOEs in the NOESY-HSQC spectra.

Assignments were then extended to the side chain signals using correlations observed primarily in $^{15}N/^{1}H$ TOCSY-HSQC and $^{13}C/^{1}H$ HCCH-TOCSY, with additional supporting evidence provided in some cases by $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ HMQC-NOESY spectra. $^{1}H/^{1}H$ TOCSY and NOESY spectra were used to assign aromatic side chain protons.

2.2.6 Secondary Structure Determination

The Chemical Shift Index (CSI) method was used to predict the secondary structure of the CBP SID / SRC1 AD1 domains on the basis of their Cα, Cβ, and Hα chemical shifts (Wishart & Sykes, 1994; Wishart *et al.*, 1992). In addition, the pattern of sequential and medium range NOEs involving backbone amide signals, observed in the $^{15}N/^{1}H$ NOESY-HSQC spectrum, was used to confirm the locations of helical regions (Wuthrich, 1986).

2.2.7 Structural Calculations

The family of converged CBP SID / SRC1 AD1 structures was determined in a two stage process using the program CYANA (Guntert *et al.*, 1997). Initially, the Combined Automated NOE assignment and structure determination protocol (CANDID) (Herrmann *et al.*, 2002) was used to automatically assign the NOE cross peaks identified in three dimensional ^{15}N- and ^{13}C-edited NOESY spectra and to produce preliminary structures of the protein. After which several cycles of simulated annealing combined with redundant dihedral angle constraints (REDAC) were used to produce the final converged CBP SID / SRC1 AD1 structures (Guntert & Wuthrich, 1991).

The input for the CANDID stage consisted of essentially complete ^{15}N, ^{13}C and ^{1}H resonance assignments for the non-exchangeable groups in the CBP SID / SRC1 AD1 complex, two manually picked three-dimensional NOE peak lists corresponding to all NOEs involving amide protons (1179) and all NOEs between aliphatic protons (2371), and one manually picked two-dimensional NOE peak list corresponding to all NOEs involving aromatic sidechain protons (127). In addition, the CANDID stage included 80 torsion angle constraints for CBP SID and 64 torsion angle constraints for SRC1 AD1 as determined by the protein backbone dihedral angle prediction programme TALOS (Cornilescu *et al.*, 1999). Hydrogen bond constraints (between i and i+4) were also added

for regions indicated to contain regular helical secondary structure by both the NOE and chemical shift data (Guntert *et al.*, 1997; Wishart & Sykes, 1994; Wishart *et al.*, 1992; Wuthrich, 1986), which corresponds to residues A2067-L2075, Q2082-K2092 and A2099-A2107 of CBP SID and E929-S941 of SRC1 AD1. The peak lists were prepared using XEASY and the intensity of peaks were obtained using the 'interactive integration' option within the programme (Bartels *et al.*, 1995). CANDID calculations were carried out using the default parameter settings in CYANA 1.0.6 with chemical shift tolerances set to 0.02 ppm (direct and indirect ^{1}H) and 0.3 ppm (^{15}N and ^{13}C).

The final converged CBP SID / SRC1 AD1 structures were produced from 100 random starting coordinates using a standard torsion angle-based simulated annealing protocol combined with 5 cycles of redundant dihedral angle constraints (REDAC) (Carr *et al.*, 2003; Muskett *et al.*, 1998). The calculations were based upon 1855 non-redundant, NOE-derived upper distance limits, assigned to unique pairs of protons using CANDID, 104 hydrogen bond constraints (4 constraints for each of the 26 hydrogen bonds) for residues suggested to adopt regular helical structure by both the NOE and chemical shift data, and 142 Φ and Ψ torsion angle constraints derived from TALOS (Guntert *et al.*, 1997; Wishart & Sykes, 1994; Wishart *et al.*, 1992). Typical Ψ and Φ angle constraints, within helical regions, were between $\pm$ 10° and $\pm$ 30°. A complete list of the dihedral angle constraints used in the structural calculations is shown in appendix A2 sections 1 and 2.

Analysis of the family of structures obtained was carried out using the programs CYANA and MOLMOL (Guntert *et al.*, 1997; Koradi *et al.*, 1996).

2.3 Results

2.3.1 Expression and Purification of the CBP SID / SRC1 AD1 Complex.

A pET22b dual expression vector, containing sequences encoding the CBP SID (2059-2117) and SRC1 AD1 (920-970) domains, was used to coexpress the polypeptides in *E. coli* (in collaboration with Dr. B Yue). The complex was purified in two stages. The presence of a polyhistidine tag at the C-terminus of SRC1 AD1 allowed the complex to be purified by anion exchange chromatography, as shown in figure 2.7. The pooled His-tagged CBP SID / SRC1 AD1 complex fractions were subjected to a final polishing purification step by gel filtration, as shown in figure 2.8. Typical yields were about 10 mg/l.

The gel filtration elution profile clearly shows that the CBP SID and SRC1 AD1 polypeptides associate with high affinity, with over 95% of the two domains forming the complex. Based on these results, the upper limit of the dissociation constant (K_d) for the CBP SID / SRC1 AD1 complex can be calculated as ~ 1 X 10^{-7} M. This result is comparable to the K_d of 3.4 X 10^{-8} M (34 nM) that was calculated by isothermal titration calorimetry for the CBP SID / ACTR AD1 complex (Demarest *et al.*, 2002).

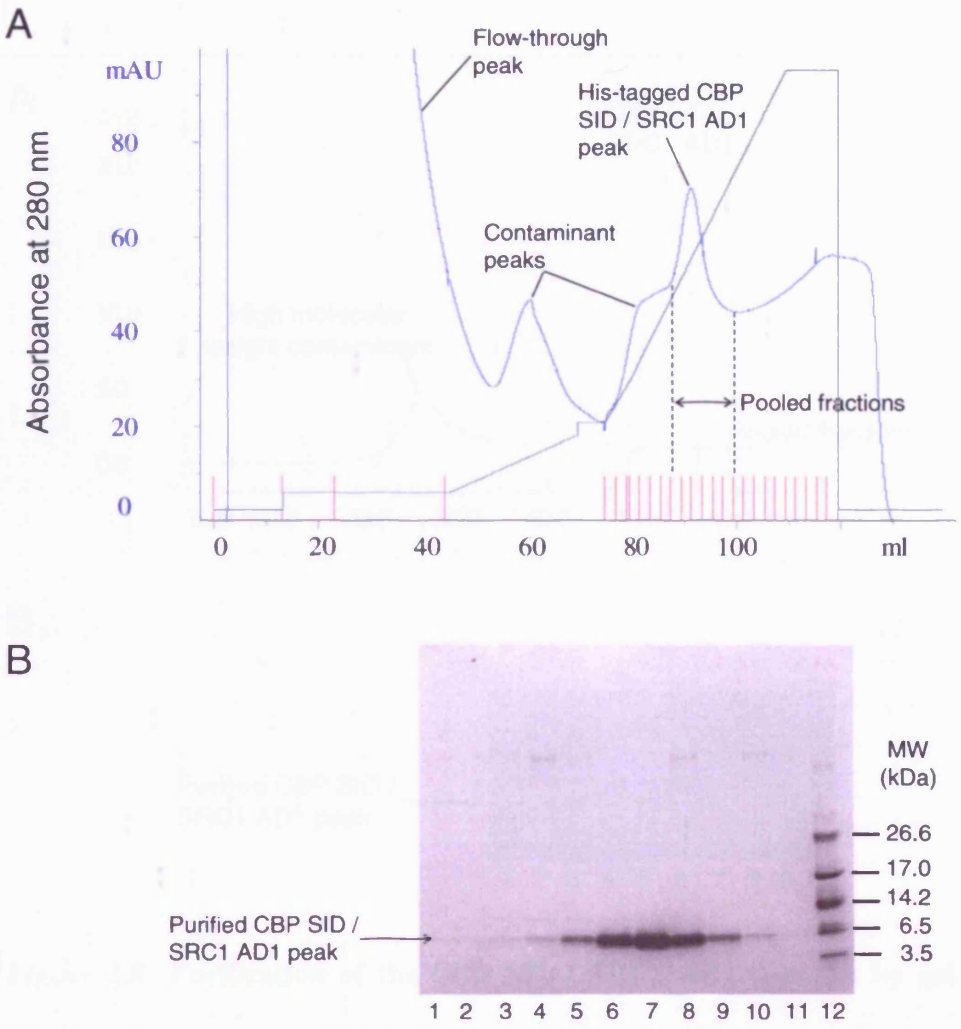


Figure 2.7 Purification of the His-tagged CBP SID / SRC1 AD1 complex. Panel A shows a typical FPLC profile attained for the Ni-NTA column based purification of the His-tagged CBP SID / SRC1 AD1 complex. The complex was eluted over a two step linear gradient from 5-250 mM imidazole (indicated in green). The flow-through, contaminant and CBP SID / SRC1 AD1 elution peaks are labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying FPLC elution fractions. Lanes 1-11 show samples removed from the second contaminant (1-4) and CBP SID / SRC1 AD1 (5-11) elution peaks. The CBP SID (6692.7 Da) and His-tagged SRC1 AD1 (7432.3 Da) protein bands are not resolved from each other. Fractions from lanes 5-10 were pooled for further purification. Lane 12 contains ultra low weight molecular markers (Sigma).

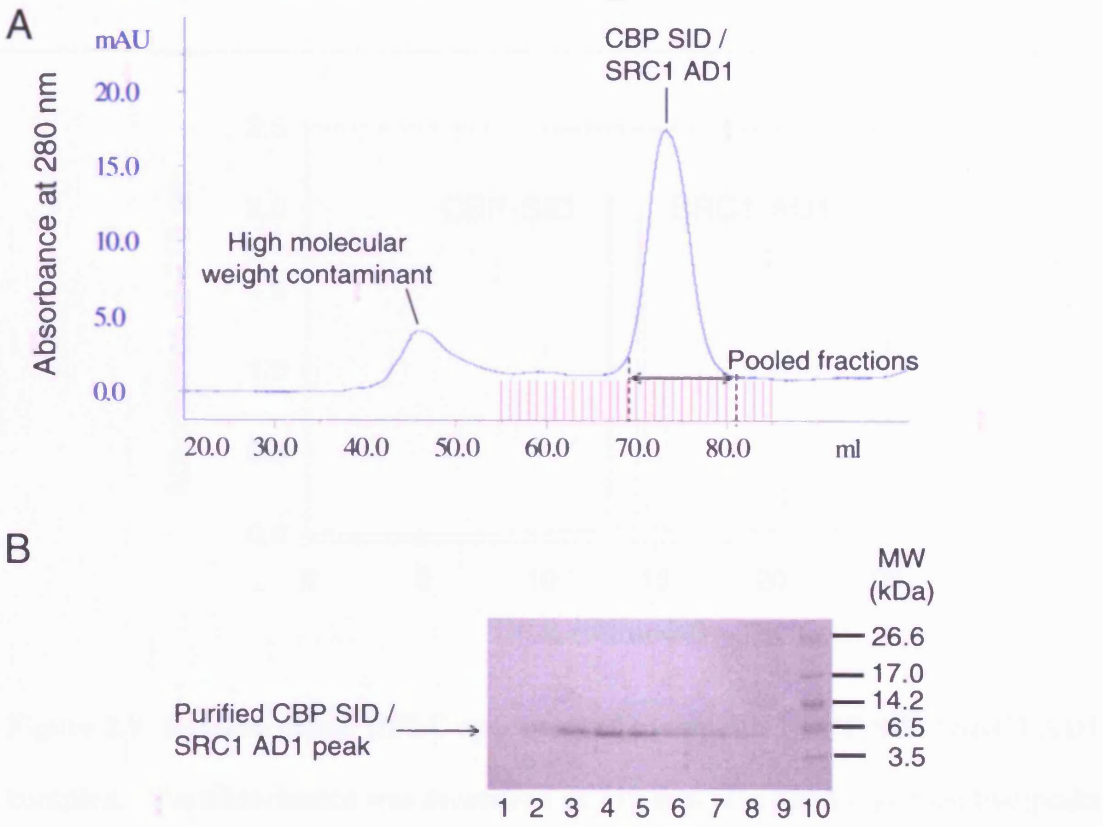


Figure 2.8 Purification of the CBP SID / SRC1 AD1 complex by gel filtration. Panel A shows a typical FPLC elution profile attained by gel filtration on a superdex 75 16/60 column. The contaminant and CBP SID / SRC1 AD1 elution peaks are labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying FPLC elution fractions. Lanes 1-9 show samples removed from the CBP SID / SRC1 AD1 elution peak. The CBP SID (6692.7 Da) and His-tagged SRC1 AD1 (7432.3 Da) protein bands are not resolved from each other. Fractions from lanes 2-8 were pooled. Lane 12 contains ultra low weight molecular markers (Sigma).

2.3.2 Reverse Phase HPLC

Under denaturing conditions the CBP SID and SRC1 AD1 domains were readily separated by reverse phase HPLC and quantification of the peptide bond absorbance from both

fractions confirmed the formation of a 1:1 complex, as shown in figure 2.9.

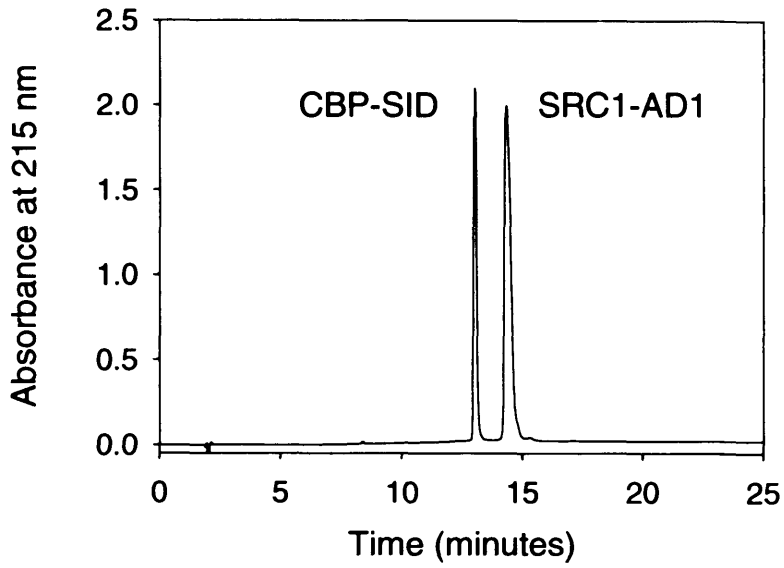


Figure 2.9 Reverse phase HPLC spectrum of the purified CBP SID / SRC1 AD1 complex. The absorbance was measured at 215 nm. The areas under the two peaks were used to confirm the expected ratio of CBP SID to SRC1 AD1 in the purified complex. The identities of the two major fractions were confirmed by mass spectrometry to correspond to the intact CBP-SID and SRC1-AD1 domain.

2.3.3 Circular Dichroism Spectroscopy

The far UV CD spectra acquired for the CBP SID / SRC1 AD1 complex are typical of a structured helical protein, with characteristic negative elipticity peaks at approximately 209 and 221 nm and a large positive peak at 195 nm (Sreerama & Woody, 1994), as shown in figure 2.10. Analysis of the spectra using the CD Pro software package, suggested that the complex contained around 54% (±7.3%) helical, 20% (±6.1%) turn and 25% (±5.7%) random coil secondary structure (± standard deviation) (Sreerama & Woody, 2000).

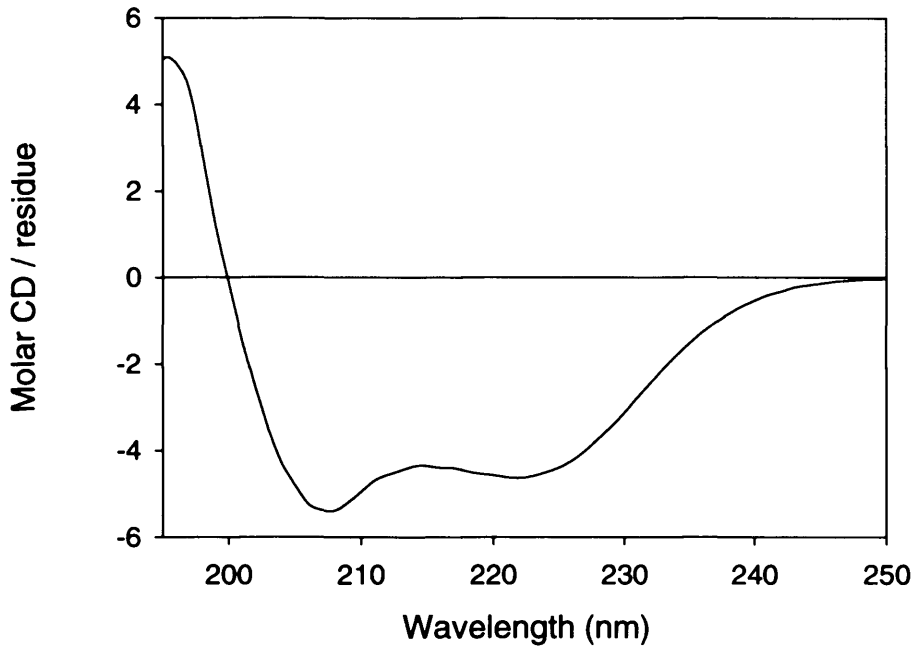


Figure 2.10 Far UV circular dichroism spectrum for the CBP SID / SRC1 AD1 complex. Analysis of the spectrum revealed that the complex is predominantly helical (54%).

2.3.4 Sequence Specific Assignments

Very comprehensive sequence-specific resonance assignments were obtained for the CBP SID / SRC1 AD1 complex despite the relatively poor dispersion observed in spectra, which is illustrated by the HSQC spectrum shown in figure 2.11. For example, backbone amide assignments were obtained for all non-proline residues in the complex except: N2060, R2061, and Q2117 in CBP SID and N927 in SRC1 AD1 (96%) and for all Cα and Cβ signals apart from the two unlabelled methionines and two residues in CBP SID (P2059 and Q2117) (96%). A representative set of ^{15}N strips from the HNCACB spectrum of the CBP SID / SRC1 AD1 complex used to assign the backbone atoms (NH, N, Cα and Cβ) are shown in figure 2.12. The corresponding set of representative ^{15}N strips from the $^{15}N/^{1}H$ NOESY-HSQC spectrum, used to confirm the backbone assignments, are shown in figure 2.13.

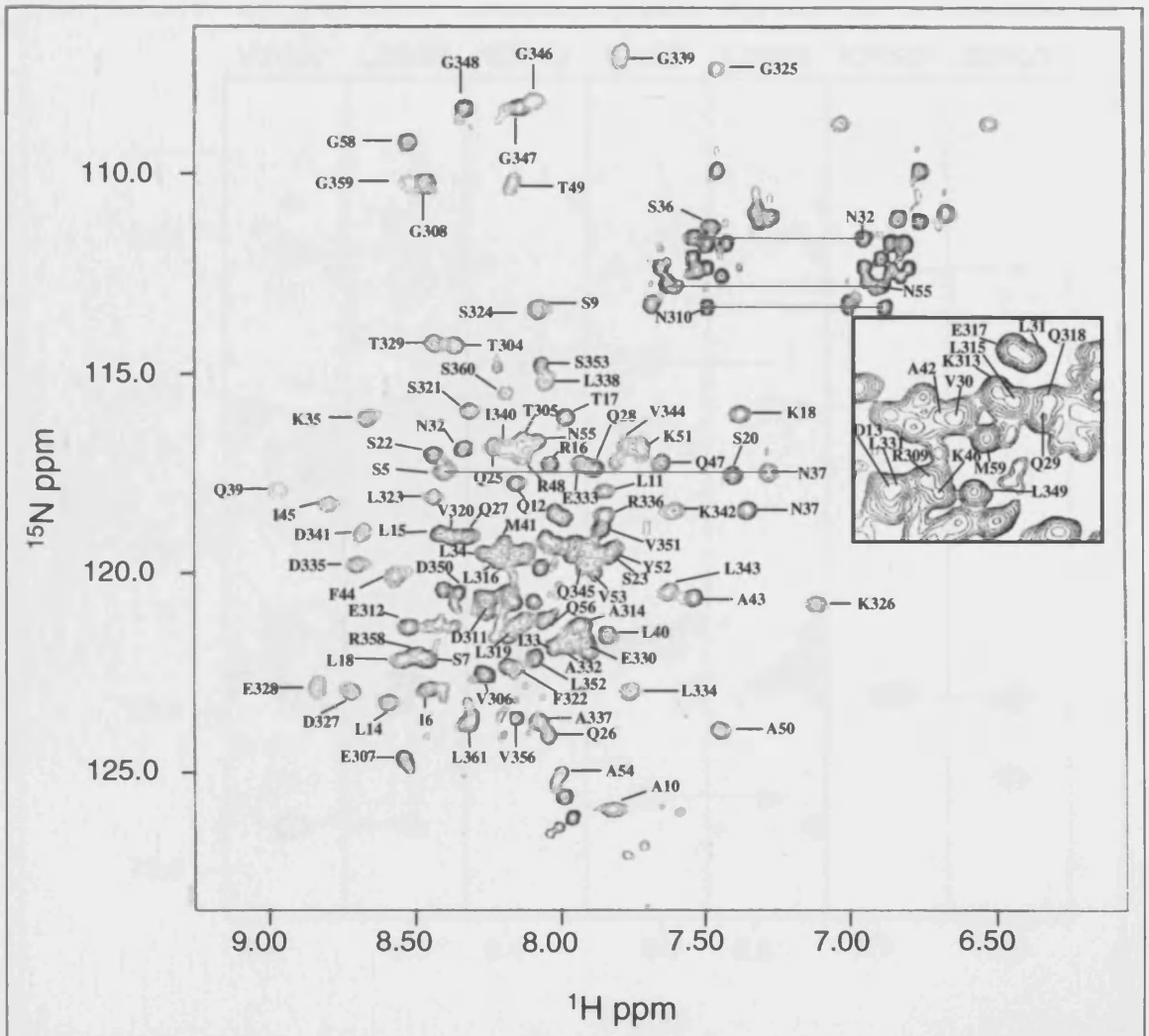


Figure 2.11 $^{15}N/^{1}H$ HSQC spectrum of the CBP SID / SRC1 AD1 complex. The assignments of the signals from backbone amide groups in both domains are indicated by residue type and number, with the overlapped region between 7.85 and 8.35 ppm in ^{1}H and 118.5 and 121.5 ppm in ^{15}N shown in the expanded region to the right of the complete spectrum. For clarity residues are numbered according to their position in the domain. Thus for CBP residues P2059-Q2117 are numbered P2-Q60 and for SRC1 residues P920-S970 are numbered P303-S353, with the C terminal linker to the His-tag numbered 354-361. The region between 6.5 and 7.7 ppm in ^{1}H and 108.5 and 113.5 ppm in ^{15}N contains, with the exception of S36, signals from the side chain amide groups of asparagine and glutamine residues. Assignments obtained for the sidechain NH_2 groups of asparagine residues are also indicated on the spectrum.

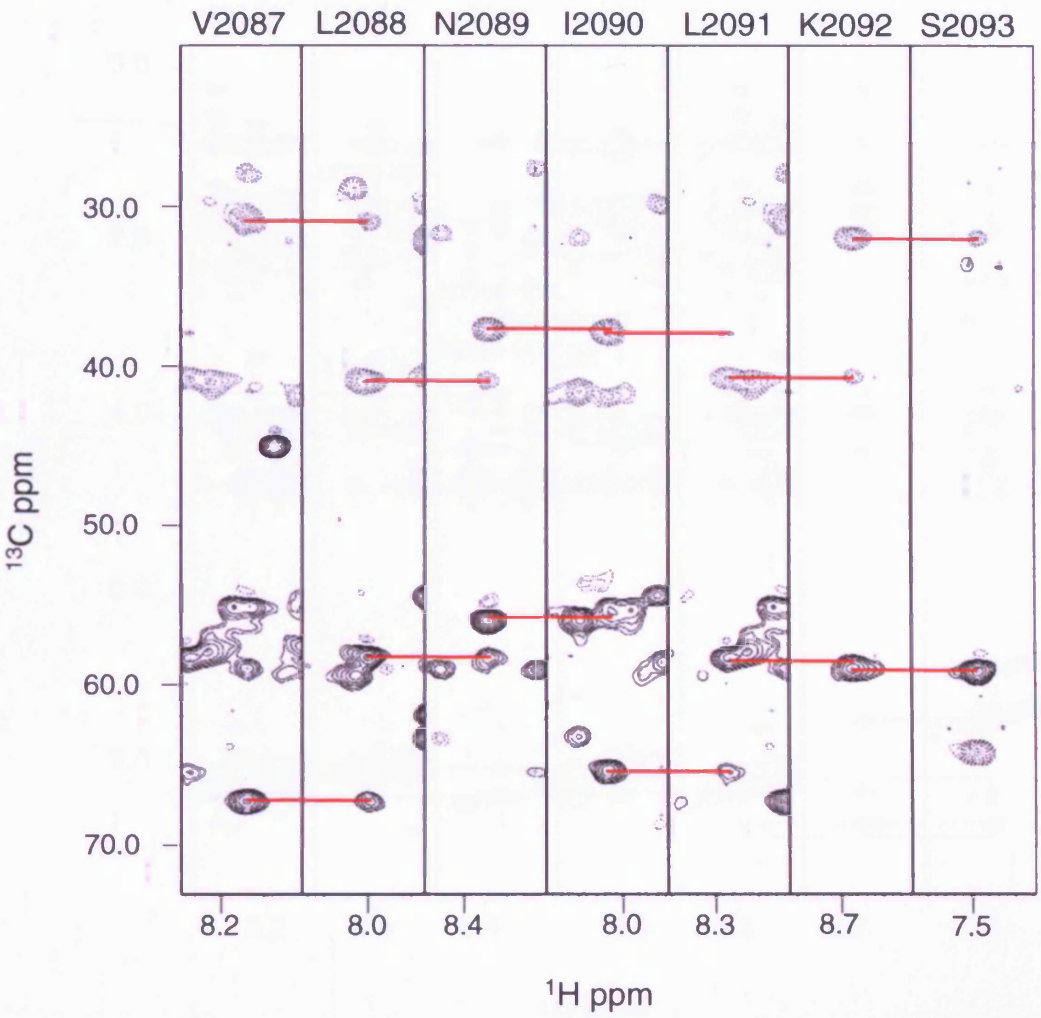


Figure 2.12 ^{15}N strips from the HNCACB spectrum of the CBP SID / SRC1 AD1 complex illustrating the intra- and inter-residue cross peaks linking residues V2087 to S2093. Cα signals are represented by solid contours and Cβ by broken contours. Intra-residue connectivities between amide nitrogen/proton and Cα and Cβ show stronger signals than those observed for sequential connectivities. Red lines indicate sequential Cα and Cβ connections. Ambiguities caused by the overlap of intra- and inter-residue cross peaks, or the absence of peaks were resolved using the CBCA(CO)NH spectrum.

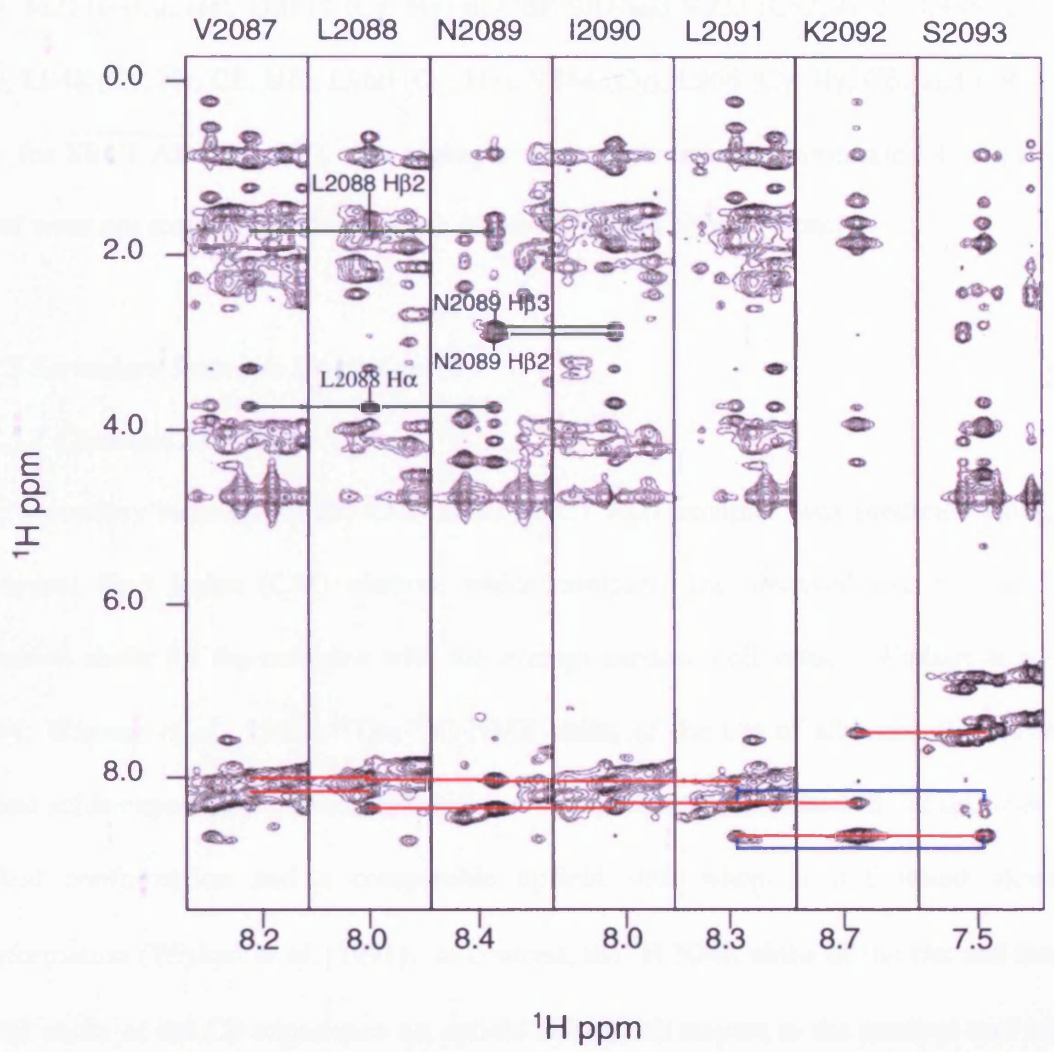


Figure 2.13 ^{15}N strips from the NOESY-HSQC spectrum of the CBP SID / SRC1 AD1 complex illustrating typical NOE connectivities linking residues V2087 to S2093. Red lines indicate sequential NH-NH NOEs, blue lines indicate i to i+2 NH-NH NOEs and green lines indicate examples of sequential aliphatic to NH NOEs.

In the case of Hα and Hβ signals, assignments were obtained for all residues except: P2059, L2071 (only Hβ absent), M2098 and Q2117 in CBP SID; and G942, L948 and D952 in SRC1 AD1 (94%). For the remaining non-exchangeable aliphatic side chain signals (^{13}C and ^{1}H) complete assignments were obtained apart from: P2059 (Cγ, Hγ, Cδ, Hδ), L2068 (Cγ ,Hγ), P2081 (Hδ3), Q2083 (Cγ, Hγ), L2088 (Cγ, Hγ), M2098 (Cγ, Hγ, Cε,

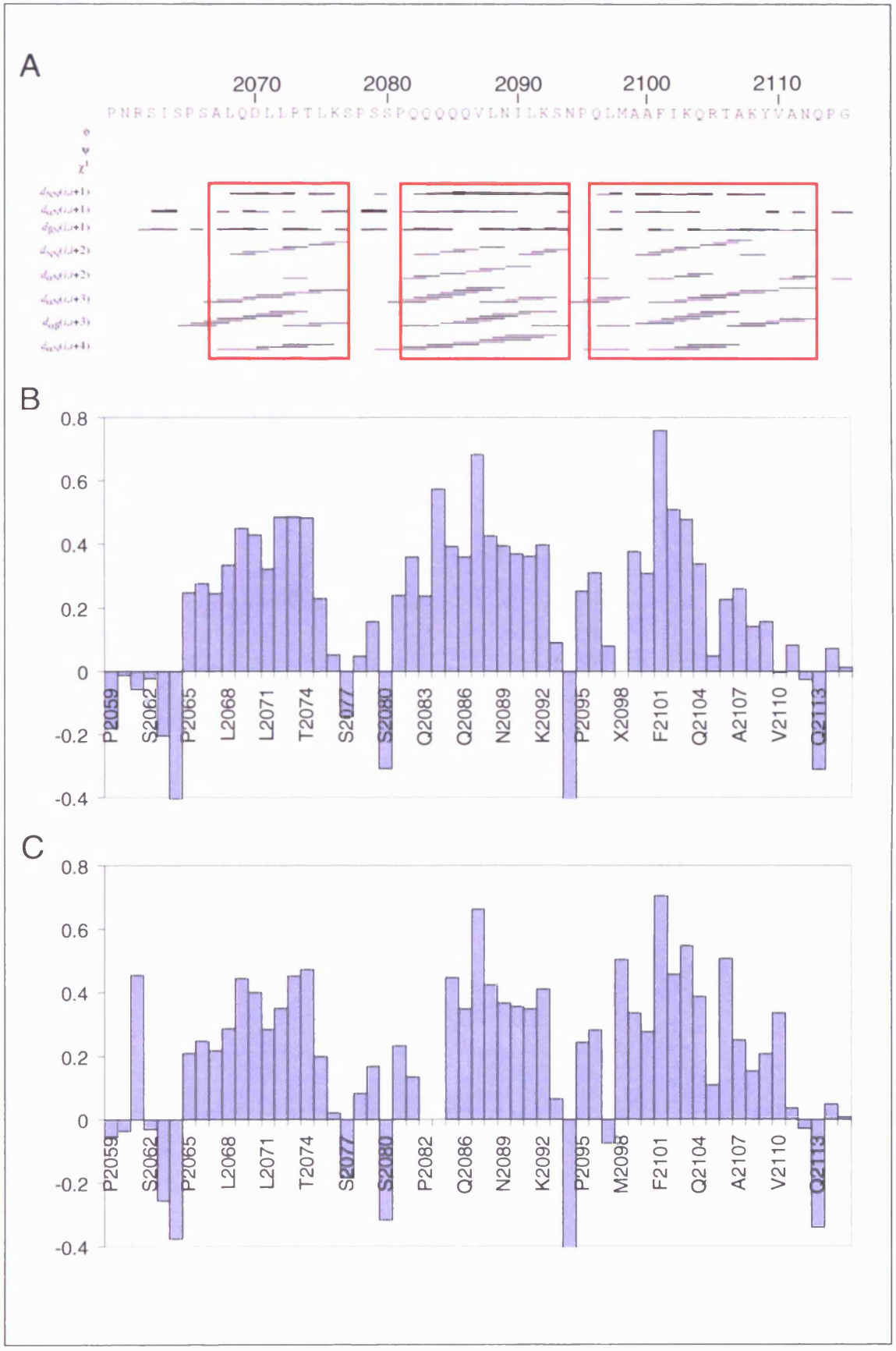
Hε), M2116 (Cε, Hε), Q2117 (Cγ, Hγ) in CBP SID and V923 (Cγ2, Hγ2), L936 (Cγ, Hγ, Cδ), L948 (Cγ, Hγ, Cδ, Hδ), L960 (Cγ, Hγ), V344 (Cγ), L966 (Cγ, Hγ, Cδ) and L969 (Cγ, Hγ) for SRC1 AD1 (91.5%). No attempts were made to assign aromatic ^{13}C signals, as these were not required for the analysis of the ^{13}C edited NOESY spectra.

2.3.5 Secondary Structure Prediction

2.3.5.1 Chemical Shift Index

The secondary structure of the CBP SID / SRC1 AD1 complex was predicted using the Chemical Shift Index (CSI) method, which compares the observed Cα, Cβ, and Hα chemical shifts for the complex with the average random coil values (Wishart & Sykes, 1994; Wishart *et al.*, 1992). The ^{13}C NMR shifts of the Cα of all naturally occurring amino acids experience a downfield shift (with respect to the random coil value) when in a helical conformation and a comparable upfield shift when in a β-strand extended conformation (Wishart *et al.*, 1991). In contrast, the ^{1}H NMR shifts of the Hα and the ^{13}C NMR shifts of the Cβ experience an upfield shift (with respect to the random coil value) when in a helical conformation and a comparable downfield shift when in a β-strand extended conformation (Spera & Bax, 1991; Wishart *et al.*, 1992). However, due to the overlap observed for the Cβ shifts their chemical shifts can only be used identify stretches of β-sheet (Wishart & Sykes, 1994).

The consensus CSI results suggestted that the CBP SID domain contained three helices located between residues S2066-L2075, Q2082-K2092 and A2099-A2107, and that the SRC1 AD1 domain contained two helices between residues E929–S941 and A949–A954. The CSI data is summarised in figure 2.14.


A
2070
2080
2090
2100
2110
PNRSISPSALQDLLRTLKSPSSPQQQQQVLNILKSNPQLMAAFIKQRTAKYVANQPG
B
0.8
0.6
0.4
0.2
0
-0.2
-0.4
P2059
S2062
P2065
L2068
L2071
T2074
S2077
S2080
Q2083
Q2086
N2089
K2092
P2095
X2098
F2101
Q2104
A2107
V2110
Q2113
C
0.8
0.6
0.4
0.2
0
-0.2
-0.4
P2059
S2062
P2065
L2068
L2071
T2074
S2077
S2080
P2082
Q2086
N2089
K2092
P2095
M2098
F2101
Q2104
A2107
V2110
Q2113

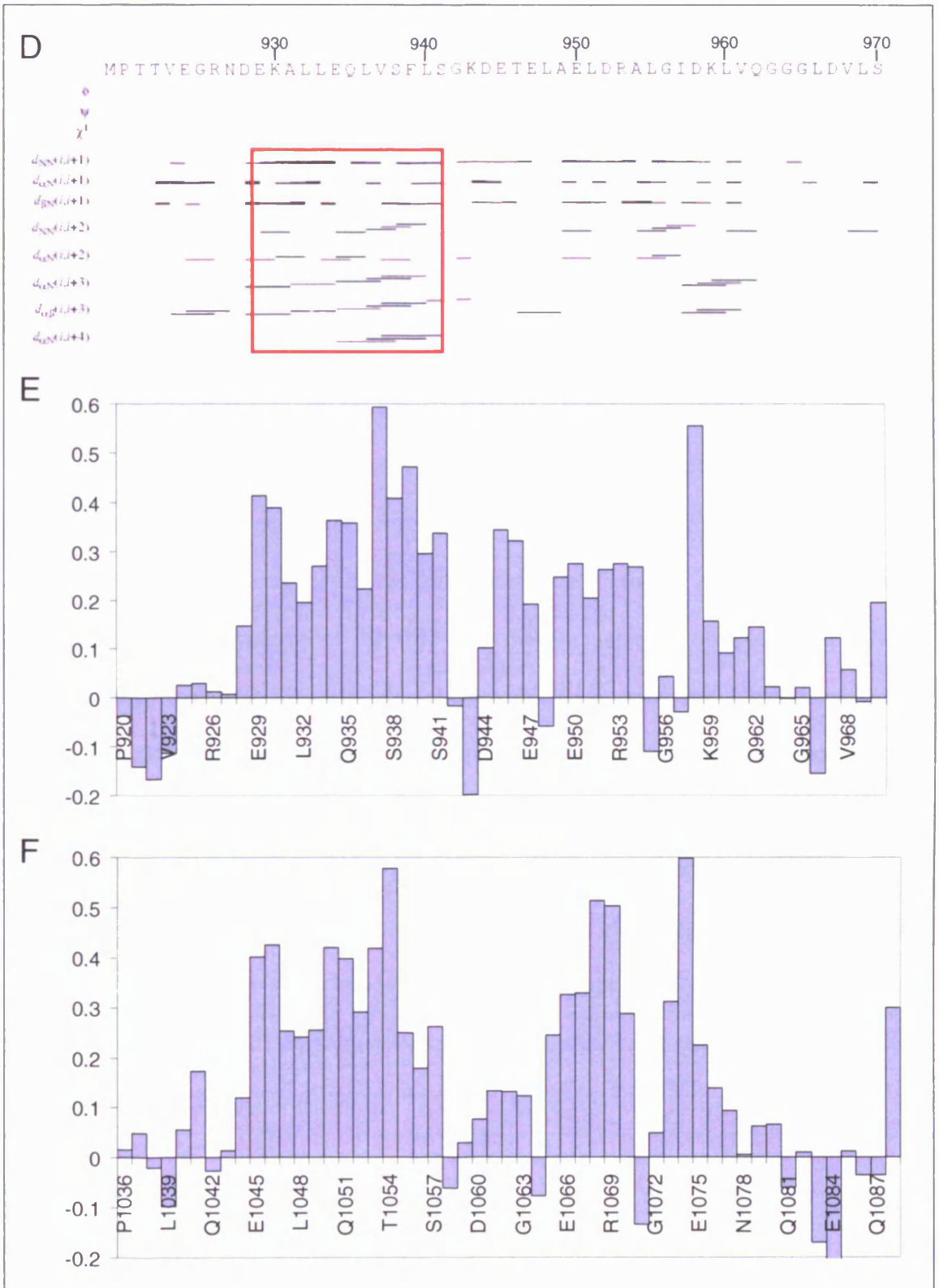


Figure 2.14 Predicted secondary structure of the CBP SID / SRC1 AD1 complex. Panels A and D show the pattern of NH-NH and Hα-NH NOEs observed for the CBP SID domain and SRC1 AD1 domain respectively. Regions that contain

characteristic helical pattern of NOEs are contained within the red boxes. Panels B and E show the average deviation of the observed Cα and Hα shifts from the average random coil values. Panel B shows the deviations observed for the CBP SID domain and panel E those seen for the SRC1 AD1 domain. To compensate for the larger secondary shifts observed for the Cα signals, and for the expected opposite shifts observed for Cα and Hα signals in helical regions, the average deviation was calculated as: $(((\delta C\alpha_{observed}-\delta C\alpha_{random\ coil}) \times 0.16) - (\delta H\alpha_{observed} - \delta H\alpha_{random\ coil}))/2$. Helical regions are characterised on the histogram by positive secondary shifts. For comparison the deviations observed for the CBP SID / ACTR AD1 complex are also shown, with the deviations for CBP SID shown in panel C and for ACTR AD1 in panel F. The secondary structure of the CBP SID domain, when in complex with SCR1 AD1, is predicted by both the NOE pattern and the chemical shift deviations to be composed of three helical regions. Similar results were obtained for the chemical shift deviations of the CBP SID domain when in complex with ACTR AD1. The secondary structure of the SRC1 AD1 domain was shown by both methods to contain one N-terminal helix. However, the chemical shift deviations also suggested that the SRC1 AD1 contains a further three helical regions. Similar results were observed for the homologous ACTR AD1 domain, however, there are minor differences in the positions of the predicted helices.

2.3.5.2 NOE Pattern

Regions of proteins which contain regular secondary structure display characteristic patterns of inter-residue NOEs. For example α-helical structures are characterised by strong sequential and weaker i to (i+2) NH-NH NOEs, as well as by i to (i+3) and i to (i+4) ^1H to ^1H NOEs (Wuthrich, 1986).

The pattern of NOEs that were calculated by CYANA are summarised in figure 2.14 panels A and D. Analysis of the pattern indicates that the CBP SID domain contains 3 helices between residues, A2067-S2077, Q2082-N2094 and Q2096-N2112. These results are similar to those obtained by the CSI method, the main difference being that the third helix is extended by 2-3 residues at both the N- and C-termini. In contrast to the CSI secondary structure prediction for the SRC1 AD1 domain, which suggested that the domain contained two helical regions, the NOE pattern suggests that the SRC1 AD1 domain contains only one well defined helix between residues E929 and S941. The NOE pattern for the remaining region of the SRC1 AD1 domain is inconclusive.

2.3.5 Structural Calculations

The CANDID protocol proved effective at determining unique assignments for the NOEs identified in the three dimensional ^{15}N- and ^{13}C-edited NOESY and the aromatic to aliphatic region of the two dimensional NOESY. At the end of the final cycle, unique assignments were obtained for 89.8% (1059/1179) of the NOE peaks picked in the ^{15}N/^{1}H NOESY-HSQC spectra, 89.3% (2117/2371) in the ^{13}C/^{1}H HMQC-NOESY spectra and 92.1% (117/127) in the NOESY spectrum. The uniquely assigned NOE peaks produced 1759 non-redundant ^{1}H to ^{1}H upper distance limits, which were used as the principle constraints in the final round of structural calculations.

The final family of CBP SID / SRC1 AD1 complex structures was determined using a total of 2005 NMR-derived structural constraints (an average of 18.2 per residue), including 1759 NOE-based upper distance limits (328 intra residue, 514 sequential (i, i+1), 643 medium range (i, $i \leq 4$) and 274 long range (i, $i \geq 5$)), 142 backbone torsion angle constraints (71 Φ and 71 Ψ) and 104 hydrogen bond constraints in regions of regular helical structure.

Following the final round of CYANA calculations, 37 satisfactorily converged structures were obtained from 100 random starting structures. The converged structures contain no distance or van der Waals violation greater than 0.5 Å and no dihedral angle violations greater than 5°, with an average value for the CYANA target function of 6.17 ± 0.95 $Å^2$. The sums of the violations for the upper distance limits, lower distance limits, van der Waals contacts and torsion angle constraints were 20.7 ± 1.84 Å, 1.5 ± 0.24 Å, 14.8 ± 1.55 Å and 33.8 ± 6.53° respectively. Similarly, maximum violations for the converged structures were 0.40 ± 0.05 Å, 0.24 ± 0.08 Å, 0.28 ± 0.04 Å and 3.44 ± 0.64° respectively. The family of converged CBP SID / SRC1 AD1 complex structures, together with the NMR constraints, have been deposited in the Protein Data Bank (accession code 2C52). The NMR constraints and structural statistics for the CBP SID / SRC1 AD1 complex are summarised in table 2.2.

The solution structure of the CBP SID / SRC1 AD1 complex is determined to high precision, which is clearly evident from the superposition of the protein backbone shown for the family of converged structures in figure 2.15 (best fit for residues 2063-2113 (CBP-SID) and 928-963 (SRC1-AD1)) and is reflected in low root mean standard deviation (r.m.s.d.) values to the mean structure for both the backbone and all heavy atoms of 0.49 ± 0.10 Å and 0.95 ± 0.09 Å respectively.

Table 2.2 NMR constraints and structural statistics for the CBP SID/SRC1 AD1 complex

a) No. of Constraints used in Final Structural Calculation		
Intraresidue NOEs	328	
Sequential NOEs (*i*, *i*+1)	514	
Medium-range NOEs (*i*, *i*≤4)	643	
Intramolecular long-range NOEs (*i*, *i* ≥5)	166	
Intermolecular long-range NOEs (*i*, *i* ≥5)	108	
Torsion angles	142	(71Φ and 71Ψ)
Hydrogen bonding	104	
b) Maximum and Total Constraint Violations in 37 Converged CBP SID/SRC1 AD1 Structures		
Upper distance limits (Å)	0.40 ± 0.05	20.7 ± 1.84
Lower distance limits (Å)	0.24 ± 0.08	1.5 ± 0.24
van der Waals contacts (Å)	0.28 ± 0.04	14.8 ± 1.55
Torsion angle ranges (°)	3.44 ± 0.64	33.8 ± 6.53
Average CYANA target function (Å^2)	6.17 ± 0.95	
c) Structural Statistics for the Family of Converged CBP SID/SRC1 AD1 Structures		
Residues within the		
favourably allowed		71
additionally allowed		20
generously allowed		6
disallowed		2
regions of the Ramachandran plot (%)		
Backbone atom r.m.s.d. for structured region (residues 2063-2113 of CBP SID and 928-963 of SRC1 AD1)		0.49 ± 0.10 Å
Heavy atom r.m.s.d. for structured region (residues 2063-2113 of CBP SID and 928-963 of SRC1 AD1)		0.95 ± 0.09 Å

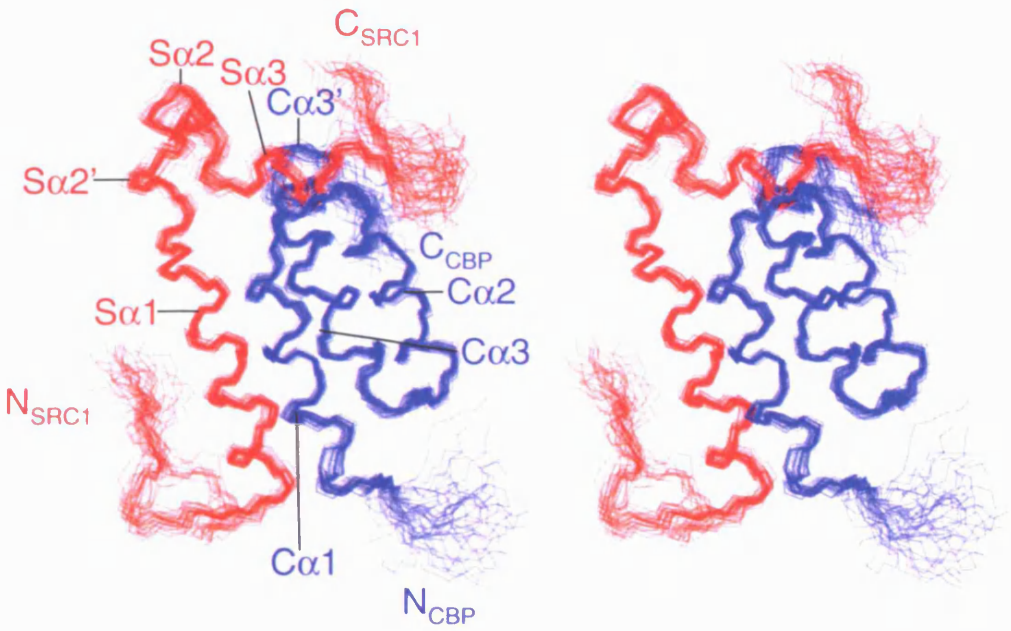


Figure 2.15 Stereo view of the solution structure of the CBP SID / SRC1 AD1 complex. The CBP SID domain is shown in blue and the SRC1 AD1 domain is shown in red. Shown is the best fit superposition of the family of 37 converged structures obtained. The four helices of CBP SID, Cα1, Cα2, Cα3, and Cα3' and the four helices of SRC1 AD1, Sα1, Sα2', Sα2 and Sα3 are labeled, as is the N and C termini of both domains.

2.1 Discussion

2.4.1 Structural Features of the CBP SID / SRC1 AD1 Complex

The backbone topology of the CBP SID / SRC1 AD1 complex is illustrated by the ribbon diagrams shown in figure 2.16. The complex is primarily composed of eight helices linked by turns and loops, including four α-helices in the CBP SID domain (Cα1 S2066-K2076, Cα2 Q2082-S2093, Cα3 P2095-T2106 and Cα3' Y2109-N2112) and three in the SRC1 AD1 domain (Sα1 E929-S941, Sα2' E945-L948 and Sα3 I957-Q962). SRC1 AD1 also

contains a short 3_{10}-helix (Sα2 E950-L955). These helices constitute just over 50% of the polypeptide in the complex, which fits with the CD data prediction of 54% helical secondary structure.

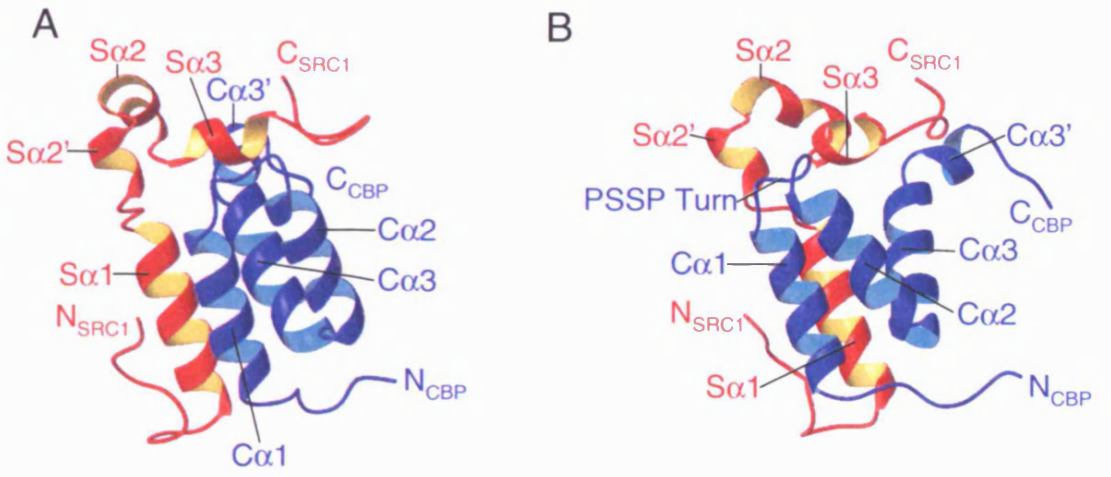


Figure 2.16 Backbone topology of the CBP SID / SRC1 AD1 complex. The CBP SID domain is shown in blue and the SRC1 AD1 in red. Panel A shows a ribbon representation of the backbone topology of the two domains in the complex in the same orientation as in figure 2.15. In panel B the ribbon representation of the complex is rotated by -50° about the vertical axis compared to the view shown in panel A. The four helices of CBP SID, Cα1, Cα2, Cα3, and Cα3' and the four helices of SRC1 AD1, Sα1, Sα2', Sα2 and Sα3 are labelled, as is the PSSP turn and the N and C termini of both domains.

The two domains are intimately associated in the complex, with Cα1, Cα2, Cα3 and Sα1 forming a four-helix bundle (figure 2.16). The middle helices in the SRC1 AD1 domain (Sα2' and Sα2) pack together on one side of the complex to stabilise the corner region of the 'L' shaped domain and are not involved in the interface with CBP SID. In contrast, the

fourth helix of SRC1 AD1 (Sα3) serves to cap the four-helix bundle region and forms extensive contacts with the third helix of CBP SID (Cα3) and with the well defined five residue turn that links the N-terminal two helices of CBP SID (Cα1 and Cα2). As a result of its amino acid sequence (SPSSP), this turn is referred to as the PSSP turn.

The N-terminal helix of CBP SID (Cα1) contains a leucine rich motif and lies almost parallel to Cα2. The N-terminus of Cα2 contains a five residue polyglutamine stretch which proceeds the hydrophobic residue rich C-terminal half of the helix. The third helix of CBP SID (Cα3) is inclined away from Cα1 and Cα2 exposing a hydrophobic groove between Cα1 and Cα3, which accommodates the first helix of SRC1 AD1 (Sα1). The C-terminal helix of CBP SID (Cα3') is separated from Cα3 by a short region of irregular structure, with the two helices resembling a single bent helix. The stabilising interactions between the CBP SID helices Cα1 and Cα2, and between Cα2 and Cα3 appear to involve no ionic or hydrogen bonds, but rely solely on favourable van der Waals contacts, primarily involving residues; A2067, L2071 and T2074 from Cα1; Q2083, V2087, I2090 and L2091 of Cα2; and L2097, F2101 and T2106 from Cα3. Residue N2094 also makes a number of van der Waals interactions with the adjacent two helices Cα2 and Cα3. Likewise the SRC1 AD1 domain appears to be primarily stabilised by van der Waals interactions, which involve residues A931 and Q935 from Sα1; K943 from the unstructured region between Sα1 and Sα2'; T946 from Sα2'; A949 which links Sα2' to Sα2 and G956 which links Sα2 to Sα3. This extensive network of stabilising van der Waals interactions is summarised in figure 2.17. Two potential salt bridges, formed between E945 and K959 and between D952 and K959, have also been identified. The mutation of one of these residues, D952 to alanine has been shown to abrogate the binding of CBP SID to SRC1 AD1 (Waters *et al.*, 2006).

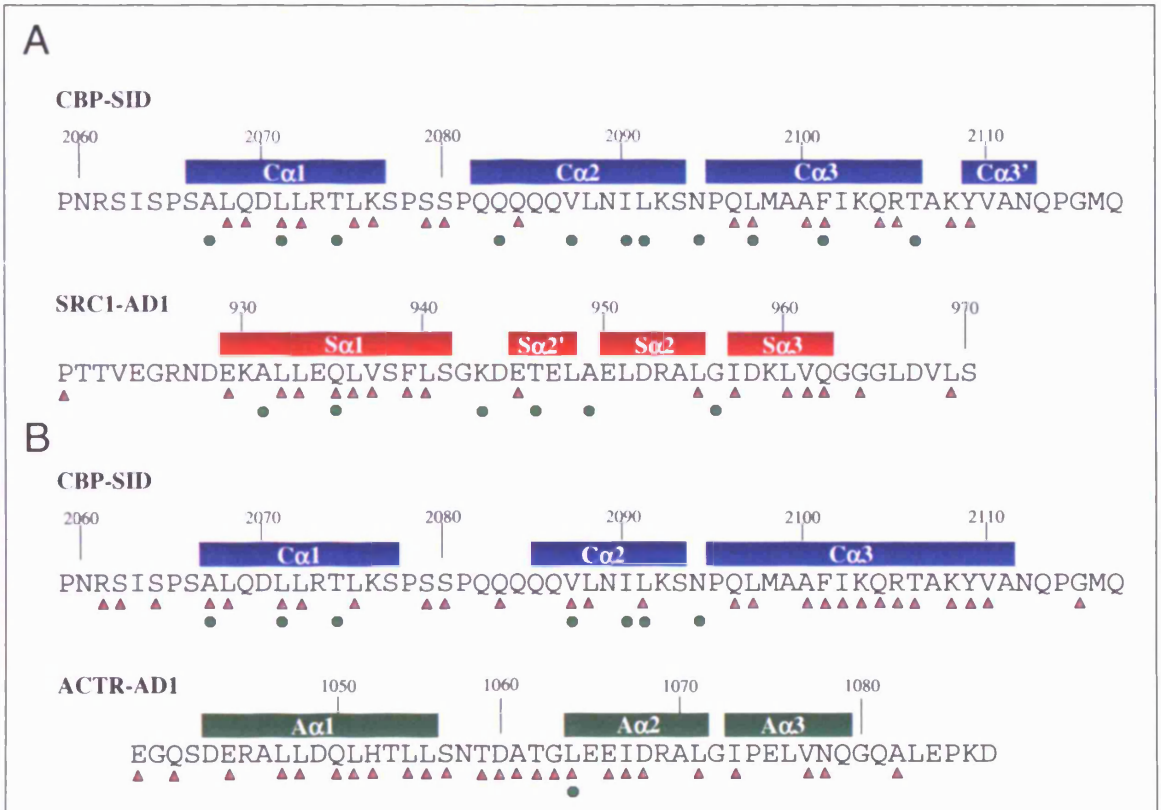


Figure 2.17 A summary of the intra- and inter-domain interactions observed in the CBP SID / SRC1 AD1 complex and the CBP SID / ACTR AD1 complex. Panel A shows the amino acid sequences of the CBP SID and SRC1 AD1 domains, the position of the helices are indicated above the sequences. Residues involved in inter-domain van der Waals contacts responsible for stabilising the complex are indicated by purple triangles, and residues involved in stabilizing intra-domain van der Waals contacts are indicated by green circles. For comparison the amino acid sequences and intra- and inter-domain contacts observed in the CBP SID / ACTR AD1 complex are also shown in panel B.

The interface between the two domains is fairly substantial, corresponding to a solvent inaccessible surface of 1019 $Å^2$ on CBP SID and 1088 $Å^2$ on SRC1 AD1, which is consistent with the tight binding observed. The binding of the N-terminal helix of SRC1

AD1 in the hydrophobic groove situated between Cα1 and Cα3 is also stabilised by van der Waals interactions. The interdomain stabilising van der Waals interactions are summarised in figure 2.17. The favourable contacts between the N-terminal helices of CBP SID and SRC1 AD1 mainly involve interactions between buried side chains found in the leucine rich motifs of Cα1 (LXXLLXXL corresponding to L2068, L2071, L2072, L2075) and Sα1 (LLXXLXXFL corresponding to L932, L936). These leucine rich motifs are similar to the LXXLL motifs known to mediate protein-protein interactions between many transcription factors and coregulators (Heery *et al.*, 1997). Additional favourable van der Waal interactions are found between residues Q935 and F939 from Sα1 and the leucine residues of Cα1. A potential salt bridge between K2076 of Cα1 and D944 which is located between Sα1 and Sα2' may also stabilise this region of the complex. The mutation of residue D944 to alanine has been shown to prevent the binding of the AD1 domain to CBP SID, suggesting that this potential salt bridge may play a role in stabilising the complex (Waters *et al.*, 2006).

Stabilising van der Waals contacts are also found between SRC1 AD1 Sα1 and the N-terminal end of CBP SID Cα3, involving the leucine rich motif (L933, L936, L940) and residues E929 and V937 of Sα1, and residues Q2096, L2097, A2100, F2101 and Q2104 of Cα3. Point mutations in Cα3 (mutants F2101P and K2103P), which are likely to disrupt the helical structure of Cα3, have been shown to result in the complete loss of binding between CBP SID and SRC1 AD1 (Matsuda *et al.*, 2004; Sheppard *et al.*, 2001). In addition the mutation of F2101, which is located in the core of CBP SID / SRC1 AD1 complex, to serine also prevented the binding of CBP SID to SRC1 AD1 (Sheppard *et al.*, 2001).

The final helix of SRC1 AD1 (Sα3) fits into a second hydrophobic groove located between the well defined PSSP turn linking Cα1 and Cα2, and the C-terminus of Cα3 (figure 2.16). This extension of the interface between SRC1 AD1 and CBP SID has been shown to play an important role in stabilising the binding of CBP SID to SRC1 AD1. The mutation of key Sα3 residues (particularly D958A and L960A) were shown to abrogate the binding of CBP SID to SRC1 AD1 (Waters *et al.*, 2006). This interaction is again stabilised purely by van der Waals contacts, which involve residues L955, I957, L960, V961, Q962 and G964 of SRC1 and S2079, S2080, Q2084, R2105, K2108 and Y2109 of CBP.

The final five C-terminal residues of SRC1 AD1 loop back towards Sα3 and Cα3', as a result van der Waals contacts are observed between L969 and Y2109 of Cα3'. In addition, a potential interdomain salt bridge was identified between R2105 of Cα3 and D967. However, this loop is primarily stabilised by van der Waals interactions involving residues within the C-terminal linker region introduced by the expression construct used and Q962 of Sα3. In the absence of these interactions the contacts between L969 and Y2109 may not be present and this C-terminal region of SRC1 AD1, including Sα3, is likely to be solvent exposed and available to make interactions with additional proteins, or domains of CBP or SRC1.

Some features of the NMR data obtained for the CBP SID / SRC1 AD1 complex clearly show the presence of a second minor conformational state, which is in exchange with the major species whose structure is reported here. This is most clearly indicated by the presence of cross peaks between signals from backbone amide groups in $^{15}N/^{1}H$ TOCSY-HSQC spectra (figure 2.18), which can only arise through chemical exchange processes (Feeney *et al.*, 1991). In the majority of cases no NOEs arising from the minor conformation were observed in the $^{15}N/^{1}H$ NOESY-HSQC spectrum. However, in a few

cases some weak amide to aliphatic NOEs were observed. The intensity of these peaks suggest that less than 10% of the complex is in the minor conformation. The amide nitrogen/proton chemical shifts for the major and minor conformations were well resolved, and only the NOEs arising from the major conformation were included in the peak lists used for structural calculations.

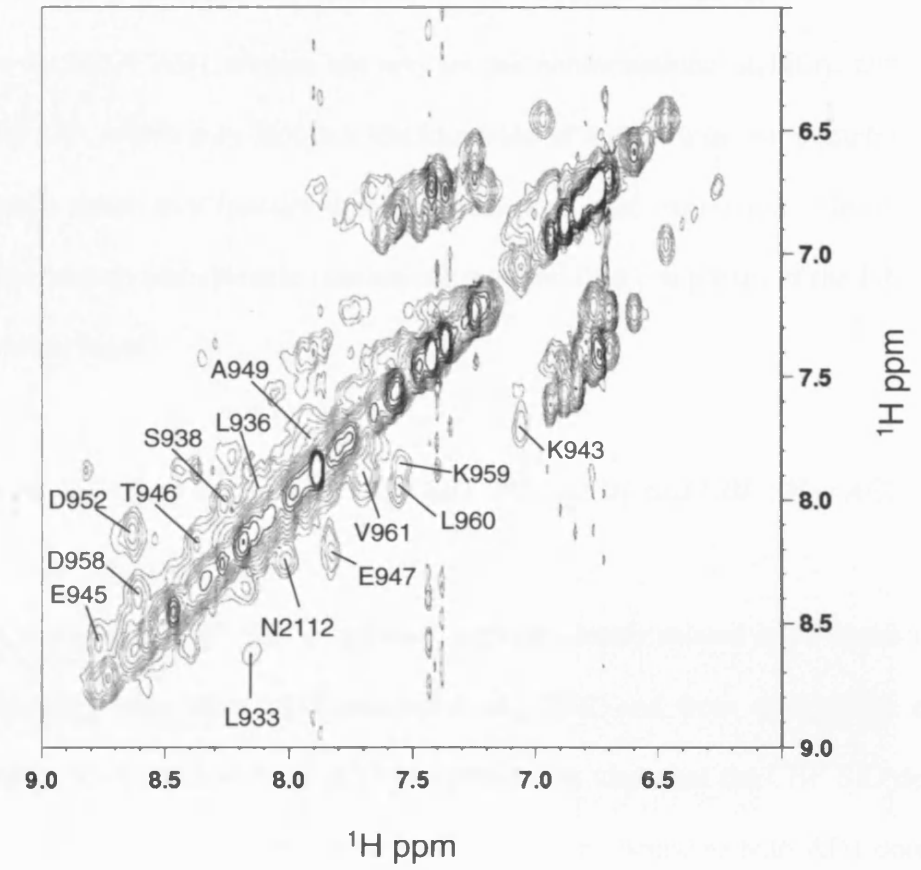


Figure 2.18 Two-dimensional ^{1}H/^{1}H projection of the ^{15}N/^{1}H TOCSY-HSQC spectrum of the CBP SID / SRC1 AD1 complex. The backbone amide region of the ^{15}N/^{1}H TOCSY-HSQC spectrum clearly shows the presence of a number of cross peaks between signals from backbone amide groups. These peaks arise through chemical exchange processes, which predominantly involve the C-terminal two-thirds of the SRC1 AD1 domain. The majority of the cross peaks present in the region between 6.5 and 7.5 ppm are between signals from sidechain amide groups. The identity of a number of the signals is indicated on the spectrum.

The conformational exchange also results in the broadening of a significant number of backbone amide signals, as evident in the $^{15}N/^{1}H$ HSQC spectrum shown in figure 2.11.

Interestingly, only a couple of the TOCSY exchange peaks involve residues in CBP SID, whereas almost all of the residues between L936-V961 of SRC1 AD1 are affected, which corresponds to the C-terminal six residues of Sα1 through to the end of Sα3. This suggests that the SRC1 AD1 domain has very limited conformational stability, even when bound to CBP SID, which may facilitate the formation of a more transient complex that is inherently better suited to a functional role in control of gene expression. However, the possibility that other protein/protein interactions may stabilise complexes of the full length proteins is not excluded.

2.4.2 Structural Homology between the CBP SID / SRC1 AD1 and CBP SID / ACTR AD1 Complexes

The solution structure of CBP SID in complex with the closely related AD1 domain from ACTR has recently been reported (Demarest *et al.*, 2002) and from comparison of the structures shown for both CBP SID / AD1 complexes it is clear that the CBP SID domain adopts a very similar secondary and tertiary structure, when bound to both AD1 domains, as shown in figure 2.19 panel A. Comparisons of the backbone atom coordinates yield an r.m.s.d. of 1.79 Å for the superposition of residues I2063, P2065-A2067, D2070-S2079 and Q2085-K2108. One of the major differences between the CBP SID domain structure in complex with SRC1 AD1 and ACTR AD1 is the conformation of a region formed from five glutamine residues (2082-2086) (figures 2.19A and 2.20). In the CBP SID / SRC1 AD1 complex these residues form the N-terminus of an extended second helix, whereas in the ACTR AD1 complex, the N-terminal three glutamine residues form part of an ill defined, proposed flexible loop (figure 2.19A and 2.20). This difference probably reflects

the lack of many NMR signal assignments for residues in this region of the ACTR complex, resulting in few NMR constraints on the conformation, rather than a true difference in the structure of this region.

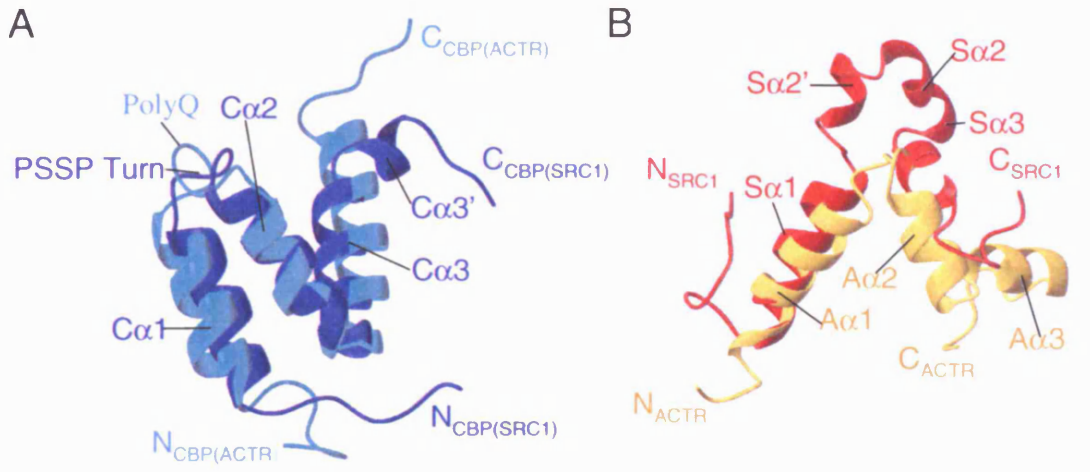


Figure 2.19 Comparison of the backbone topologies of the CBP SID and p160 AD1 domains in the CBP SID / p160 AD1 complexes. Panel A shows a comparison of the backbone topology of the CBP SID domain taken from the SRC1 (blue) and ACTR (cyan) AD1 complexes. The CBP SID domain was overlaid on residues I2063, P2065-A2067, D2070-S2079 and Q2085-K2108. Panel B shows a comparison of the backbone topologies of the AD1 domains of SRC1 (red) and ACTR (yellow). The AD1 domains were overlaid on residues I2063, P2065-A2067, D2070-S2079, Q2085-K2108 of CBP SID and residues D928-S941 and D1044-S1057 of SRC1 and ACTR respectively. The position of the helices, PSSP turn, PolyQ loop and the N- and C-termini are also labelled.

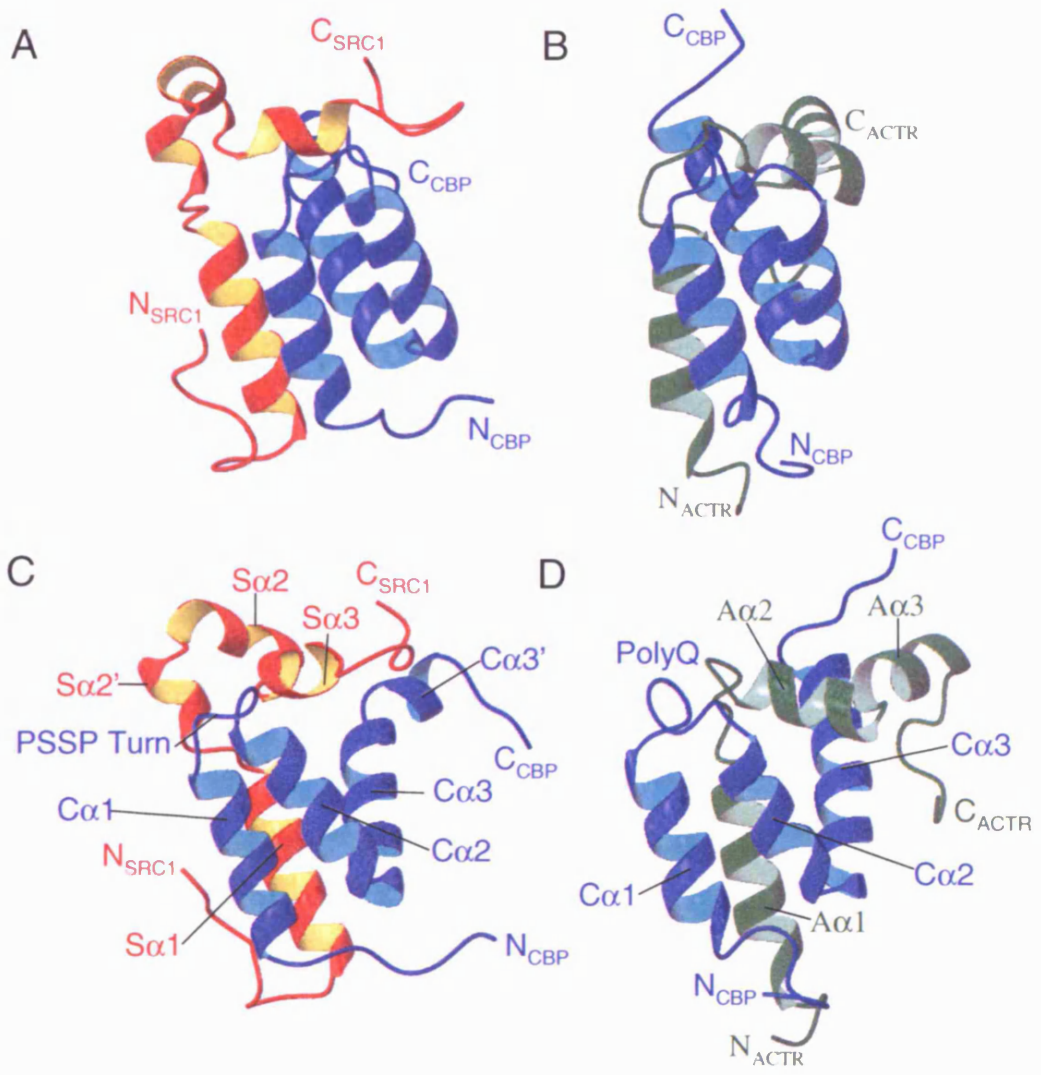


Figure 2.20 Comparison of the solution structures of the closely related CBP SID / SRC1 AD1 and CBP SID / ACTR AD1 complexes. Panels A & B show equivalent views of the backbone topology of the CBP SID / SRC1 AD1 and CBP SID / ACTR AD1 complexes respectively, which were obtained by superimposing the CBP SID domain from both complexes (residues I2063, P2065-A2067, D2070-S2079, Q2085-K2108). The CBP SID domain is shown in blue, SRC1 AD1 is shown in red and ACTR AD1 in green. The alternative views of the complexes shown in panels C and D were obtained by rotating the structures by -50° about the vertical axis. The position of the helices, PSSP turn, PolyQ loop and the N- and C-termini are labelled.

In contrast to the CBP SID domain, the structural similarity between SRC1 AD1 and ACTR AD1 when bound to CBP SID is confined to just the N-terminal helix, which forms the fourth helix of the four-helix bundle (figures 2.19B and 2.20).

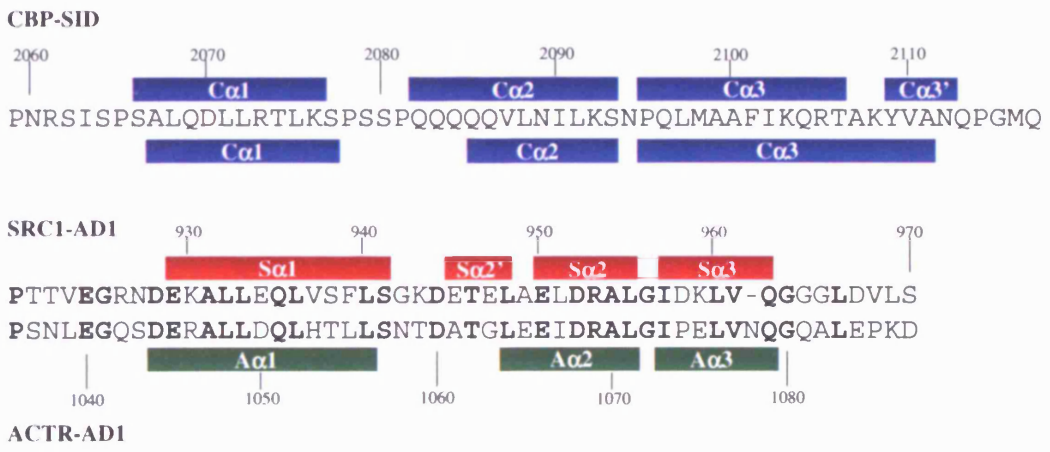


Figure 2.21 Comparison of the secondary structures of the CBP SID / SRC1 AD1 and CBP SID / ACTR AD1 complexes. The amino acid sequence of the CBP SID domain is shown at the top of the figure. The positions of the four α-helices of CBP SID present in the CBP SID / SRC1 AD1 complex are indicated above the sequence, whilst the three α-helices identified in the CBP SID / ACTR AD1 complex are indicated below the sequence. The lower panel contains an optimized sequence alignment of the AD1 domains from SRC1 and ACTR, with positions occupied by identical residues highlighted in bold, and clearly illustrates the extensive sequence homology. The positions of the four SRC1 AD1 helices present in the CBP SID / SRC1 AD1 complex are indicated in red, whilst the three α-helices of ACTR AD1 identified in CBP SID / ACTR AD1 complex are indicated in green.

The remaining residues of both AD1 domains adopt strikingly distinct structures, to the extent that SRC1 AD1 contains three further helical regions (residues Sα2' E945-L948 Sα2 E950-L955 and Sα3 I957-Q962), whilst ACTR AD1 contains only two (residues Aα2 L1064-L1071 and Aα3 I1073-Q1079). The secondary structure of the two complexes are summarised in figure 2.21. This difference is somewhat surprising since SRC1 and ACTR share 53% sequence identity and 67% homology over the AD1 domain (figure 2.5).

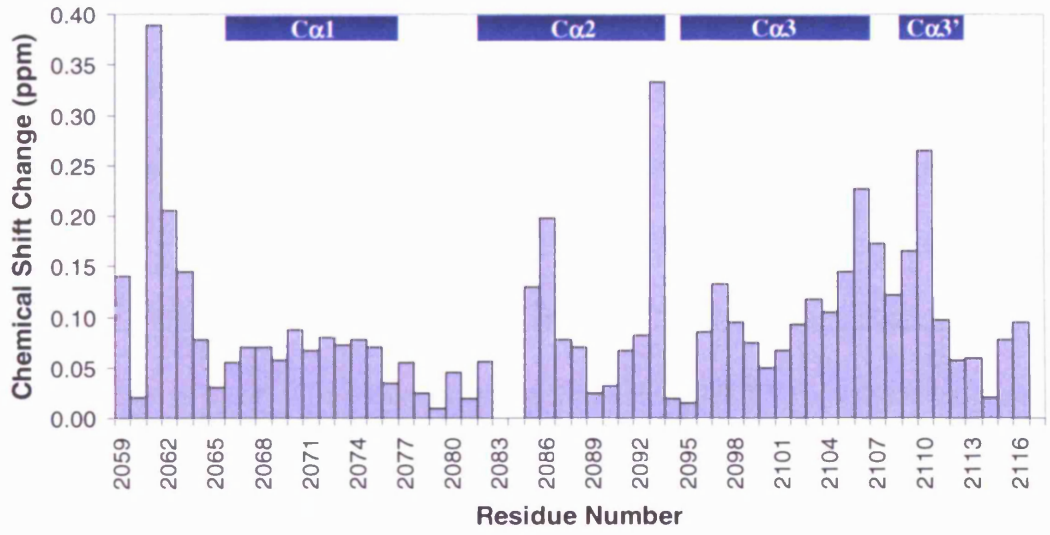


Figure 2.22 Overall changes in the chemical shifts of CBP SID backbone signals when bound to either SRC1 or ACTR. The differences in the chemical shifts of backbone amide proton and nitrogen, Hα and Cα were calculated. To compensate for the increased chemical shift range of nitrogen and carbon compared to proton, the overall change was calculated by: $(\Delta^{1}H\alpha + \Delta^{1}HN + (\Delta^{15}N \times 0.2) + (\Delta^{13}C\alpha \times 0.1)) / 4$.

It is important to note that the striking structural differences are reflected in significant chemical shift differences for NMR signals for CBP SID residues that make distinct contacts in the two complexes; in particular, the C-terminus of Cα3 and Cα3', and the

polyQ region, as shown in figure 2.22. Significant chemical shift differences were also noted between the N-terminal five residues of CBP SID. In the CBP SID / SRC1 AD1 complex these residues are solvent exposed, whilst in the ACTR AD1 complex they form part of an extended N-terminal interdomain interface. It is unknown whether these structural differences in the N-terminal regions would be present in the full length proteins.

In the CBP SID / ACTR AD1 complex, Aα2 and Aα3 essentially wrap around Cα3 making extensive contacts, as shown in figures 2.17 and 2.20. The sequence of ACTR AD1 Aα2 corresponds to that of the SRC1 AD1 helix Sα2 (figure 2.21). However, unlike in the SRC1 AD1 domain, where Sα2 packs against the previous SRC1 AD1 helix (Sα2'), Aα2 fits into the groove between the PSSP turn and Cα3. In the CBP SID / SRC1 AD1 structure this groove is instead filled by Sα3, whose sequence corresponds to that of Aα3 (figure 2.21). This C-terminal helix of ACTR AD1 (Aα3) packs against the adjacent hydrophobic face of Cα3. The sequence of the Sα3 helix (IDKLV) is somewhat divergent from that of ACTR and TIF2 (IPELV). Interestingly one of the divergent residues of SRC1 AD1, K959, could potentially form intradomain salt bridges with residues D952 and E945, which may help to stabilise the corner region of SRC1 AD1. In contrast, neither of the divergent residues of ACTR are involved in the formation of inter- or intradomain interactions (Waters *et al.*, 2006).

Another notable difference in this region of the complex is the length of Cα3, which is 17 residues long (P2095-A2111) in CBP SID / ACTR AD1, whilst in the CBP SID / SRC1 AD1 complex this region forms two shorter helices (Cα3: P2095-T2106 & Cα3': Y2109-N2112) which are 12 and 4 residues long respectively. The second of these helices, Cα3' is inclined away from the rest of the four-helix bundle to a greater extent than the C-terminus of Cα3 in the ACTR AD1 complex, which allows Sα3 to fit into the hydrophobic

groove.

In the CBP SID / ACTR AD1 complex a salt bridge, between D1068 in Aα2 and R2105 in Cα3 and two coordinating hydrogen bonds (between residues D1060 and R2105, and D1068 and Y2109), were identified at the molecular interface of the CBP SID / ACTR AD1 complex. These interactions are believed to contribute to the specificity of binding (Demarest *et al.*, 2004; Demarest *et al.*, 2002). Despite the fact the equivalent residues are conserved in the CBP SID / SRC1 AD1 complex the residues are spatially distant and no such interactions are present. The position of residues R2105, Y2109, D944 and D952 (equivalent to D1060 and D1068 respectively) in the CBP SID / SRC1 AD1 complex are highlighted in figure 2.23. Three other potential hydrogen bonds were identified in the CBP SID / ACTR AD1 complex, however, these interactions are also absent from the CBP SID / SRC1 AD1 complex.

As would be expected from the tertiary structures of the two complexes, CBP SID has a significantly larger interface with ACTR AD1 than SRC1 AD1 (corresponding to solvent inaccessible surfaces of 1681 $Å^2$ and 1019 $Å^2$ respectively). Likewise the SRC1 AD1 domain interface with CBP SID is significantly smaller than that of ACTR AD1 (corresponding to solvent inaccessible surfaces of 1088 $Å^2$ and 1839 $Å^2$ respectively).

To summarise, the backbone fold for CBP SID and the first helix of the SRC1 or ACTR AD1 domain is very similar in the closely related complexes, which is reflected in a backbone atom r.m.s.d. of 2.05 Å for the superposition of residues I2063, P2065-A2067, D2070-S2079, Q2085-K2108 of CBP-SID and residues D928-S941 and D1044-S1057 of SRC1 AD1 and ACTR AD1 respectively. The subsequent regions of the AD1 domains adopt very distinct folds, though both domains contain a conserved C-terminal helix that

interacts with CBP SID and appears to be essential for stabilistation of the complex.

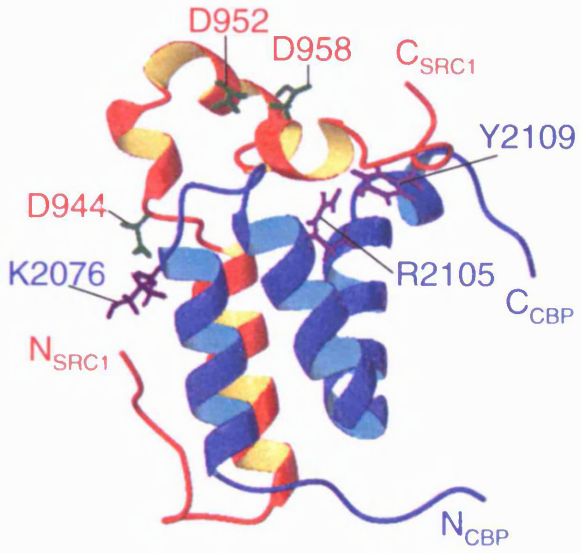


Figure 2.23 Relative location of key residues from the CBP SID / ACTR AD1 complex in the CBP SID / SRC1 AD1 complex. In the CBP SID / ACTR AD1 complex a salt bridge was formed between residues R2105 and D1068 (equivalent to D952 of SRC1), whilst hydrogen bonds were detected between R2105 and D1060 (equivalent to D944) and between Y2109 and D1068. The position of residues R2105, Y2109, D952 and D958 in the CBP SID / SRC1 AD1 complex are highlighted on the complex. It is evident from the location of these residues, that these interactions are absent in the CBP SID / SRC1 AD1 complex. A potential interdomain salt bridge was identified between K2076 and D944. The positions of these residues are also shown.

2.4.3 Comparison of the Secondary and Tertiary Structure of the isolated and AD1 bound CBP SID Domain

The NMR structure of a shorter CBP SID construct (mouse residues 2067-2112) has been solved in isolation (Lin *et al.*, 2001). The domain was shown to consist of three α helices, Cα1 (residues A2067-K2076), Cα2 (V2087-K2092) and Cα3 (Q2096-R2105), which fold

around a hydrophobic core so that Cα1 and Cα3 lie almost parallel to each other. The Cα2 and Cα3 helices of the isolated SID domain are shorter than those observed in the CBP SID / p160 AD1 complexes. The region between Cα1 and Cα2, which is referred to as L1, contains the PSSP and polyQ sequences. As is evident in figure 2.24, the isolated SID domain has a very different topology from that observed in the CBP SID / p160 AD1 complexes.

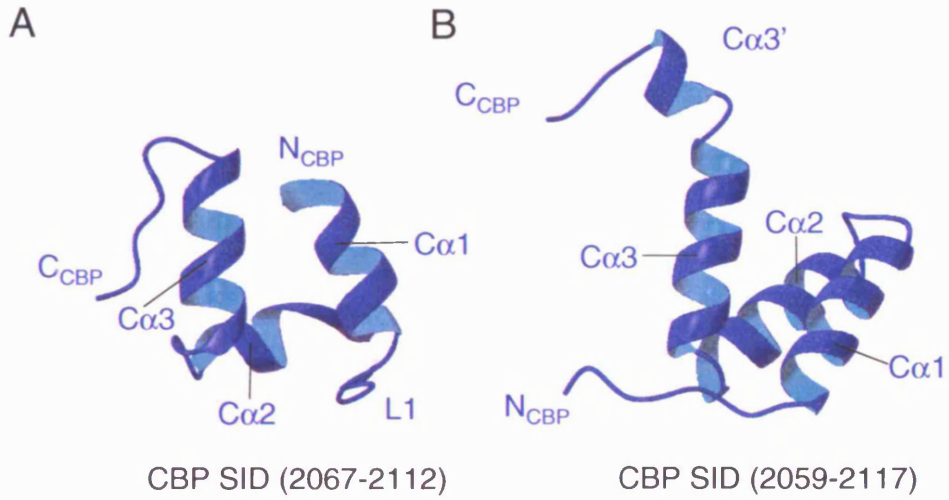


Figure 2.24 Comparison of the isolated and AD1 bound CBP SID domain. The structures of the CBP SID domain obtained in isolation (Lin *et al.*, 2001) and in complex with SRC1 AD1 are shown in panels A and B respectively. The very different conformation of the CBP SID domain is clearly evident from the comparison of the backbone topologies. The backbone atoms of the CBP SID domains were superimposed over residues L2088-T2106. Part of the L1 region is absent from the structure, as no assignments were obtained for four of the five glutamine residues in this region. The N- and C-termini and positions of the helices are labelled, as is the L1 region of the isolated SID domain.

Interestingly denaturation and differential scanning calorimetry studies have shown that the isolated CBP SID domain (residues 2059-2117) is not cooperatively folded (Demarest *et al.*, 2004; Demarest *et al.*, 2002). This suggests that although the isolated domain contains a significant amount of helical secondary structure it has no stable tertiary structure. In addition, Demarest *et al* reported that they were only able to obtain a stable fold of the isolated CBP SID domain at very high salt concentrations (750mM NaCl or 500mM Na_2SO_4) but the domain was not able to bind ACTR (Demarest *et al.*, 2002). It has therefore been suggested that the isolated SID domain exists as a molten globule and that the reported structure may represent one of many possible conformations.

2.4.4 Structural Flexibility of the CBP SID Domain Permits Complex Formation with Multiple Partners

A number of transcription factors, coactivators and viral proteins, which contain virtually no sequence homology, have been shown to bind the CBP SID domain. It has been proposed that the interaction of these diverse proteins with the CBP SID domain is mediated via a semi conserved leucine rich motif (LLXXLXXLL), which is contained within Sα1 of SRC1 AD1 and Aα1 of ACTR AD1 (Matsuda *et al.*, 2004). In the case of SRC1 AD1, the penultimate leucine in the LLXXLXXLL motif is substituted for phenylalanine. As in the case of SRC1 and ACTR, this sequence is believed to form part of an amphipathic helix, which, upon interaction with CBP SID, forms the fourth helix of the four-helix bundle. Mutation of the conserved leucine residues of SRC1 Sα1 (residues L932/L933/L935) to alanine has been shown to prevent CBP SID binding (Waters *et al.*, 2006). Similarly, the deletion of the conserved motif of Ets2, E1A or CAS has previously been shown to abrogate binding (Matsuda *et al.*, 2004; Ryan *et al.*, 2006).

The crystal structure of the CBP SID domain (mouse residues 2066-2112) in complex with

the IRF association domain (IAD) of the transcription factor IRF-3 has recently been reported (Qin *et al.*, 2005). The secondary structure of the CBP SID domain was composed of three α-helices, Cα1 (residues S2066-T2074), Cα2 (S2080-S2093) and Cα3 (N2094-R2105). The helical boundaries are similar to those observed for the N-terminal three helices of the CBP SID / SRC1 AD1 complex, however the C-terminal helix (Cα3') seen in the CBP SID / SRC1 AD1 complex (corresponding to the C-terminal six residues of Cα3 in the CBP SID / ACTR AD1 complex) is absent. In common with the CBP SID / SRC1 AD1 complex, the polyQ region is contained within Cα2. Despite the similar secondary structure, the tertiary structure of the CBP SID domain in complex with the IAD domain is distinct from that seen for the CBP SID / AD1 complexes, as shown in figure 2.25 (Demarest *et al.*, 2002; Waters *et al.*, 2006). Instead of forming three quarters of a four-helix bundle, the three α-helices and the loop connecting Cα1 and Cα2 fold to form a rectangular framework, which interacts with IRF-3 via a flat surface.

The interaction between the IAD and SID domains is mediated by hydrophobic contacts involving the three helices of CBP SID and two amphipathic α-helices, H3 and H4, of the IAD domain. Notably neither of these IAD helices contain the conserved leucine rich motif that is proposed to mediate binding between the CBP SID domain and partner proteins (Matsuda *et al.*, 2004). The stabilising van der Waals interactions primarily involve residues: T2074 of Cα1, L2075 of the loop between Cα1 and Cα2, Q2082, Q2083, Q2086, V2087 and I2090 of Cα2, N2094, Q2096, L2097, A2100, F2101 and Q2104 of Cα3, and I326, I330, T333, C371, A374, L375, M378, V381, G382 and G383 of the IAD domain. The complex is also stabilised by two interdomain hydrogen bonds between residues E203 and S2080 and between S221 and E2086.

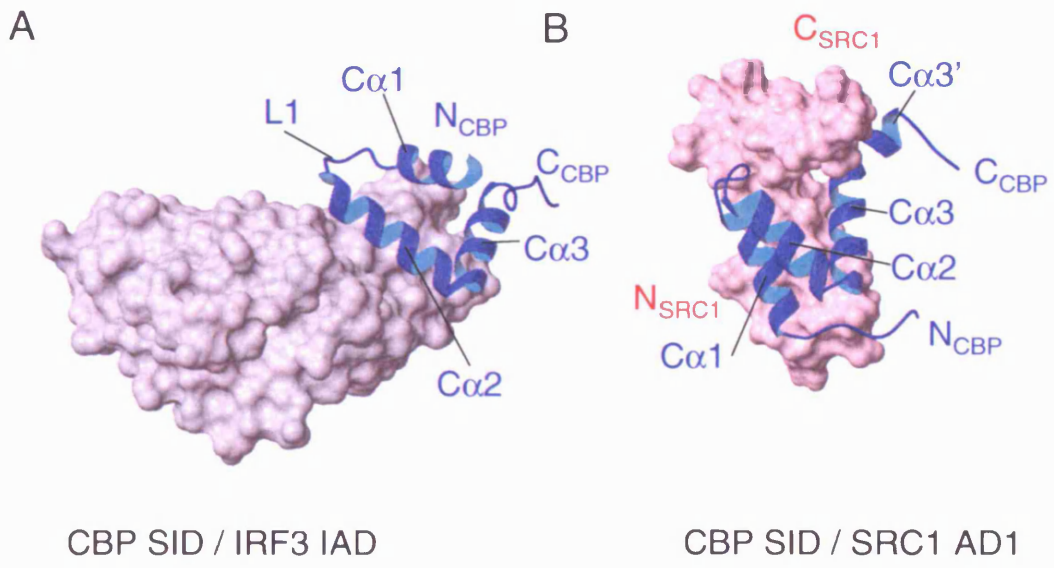


Figure 2.25 Comparison of the CBP SID / SRC1 AD1 and CBP SID / IRF-3 IAD structures. The equivalent views of the backbone topology of the CBP SID domain in complex with the IRF-3 IAD domain and the SRC1 AD1 domain are shown in panels A and B respectively. The backbone atoms of the CBP SID domains were superimposed over residues K2076-K2103 of helices Cα2 and Cα3. The surface representation of the IAD domain is shown in grey and the AD1 domain in pink. The CBP SID domain forms two distinct conformations when in complex with the different domains. The position of the helices, and the N- and C-termini of the CBP SID domains are labelled.

The ability of the CBP SID domain to fold into distinct conformations could potentially allow the domain to interact with distinct groups of transcriptional regulators that possess little if any sequence or structural similarity. It has also been suggested that the marked difference in the positioning of the N- and C-termini of the CBP SID domain could impact on other regions of the CBP protein, potentially affecting the binding of additional transcriptional regulators or the acetyltransferase activity of CBP (Qin *et al.*, 2005).

2.4.5 Conclusion

Structural studies of the CBP SID domain have shown a high degree of structural plasticity in the functional complexes formed by this region of CBP. For example, the domain can adopt at least two distinct folds when in complex with different partner proteins, allowing it to interact with a number of diverse classes of transcriptional regulators, and potentially influencing its coactivator activities. Comparison of the NMR structures obtained for the closely related CBP SID / p160 AD1 complexes revealed that whilst the CBP SID domain adopts a similar fold, the topologies of the p160 AD1 domains are distinct. These structural differences may reflect functional differences between the p160 proteins. For example, the different solvent accessible surfaces could potentially create different interaction surfaces, allowing for the formation of p160 specific higher order complexes.

CHAPTER 3

SOLUTION STRUCTURE OF THE C-TERMINAL MA-3 DOMAIN OF PDCD4 AND THE CHARACTERISATION OF ITS INTERACTION WITH THE TRANSLATION INITIATION FACTOR eIF4A

3.1 Introduction

3.1.1 Programmed Cell Death 4

Pdcd4 (Programmed Cell Death Protein 4) is a highly conserved, eukaryotic protein, which has recently been shown to play critical roles in the regulation of both transcription and translation (Bitomsky *et al.*, 2004; Yang *et al.*, 2004; Yang *et al.*, 2003a). The protein was initially discovered in a screen for genes activated during apoptosis (Shibahara *et al.*, 1995) and then subsequently identified as a tumour suppressor in studies of a mouse keratinocyte model of tumour promotion, in which high levels of Pdcd4 were found to render cells resistant to transformation by the tumour promoter TPA (Cmarik *et al.*, 1999; Yang *et al.*, 2001). Pdcd4 has been shown to inhibit the activation of Activator protein-1 (AP-1)-responsive promoters by complexes containing c-Jun, providing a possible explanation for its ability to suppress TPA-induced transformation (Yang *et al.*, 2001; Yang *et al.*, 2003b).

Recently, loss of Pdcd4 expression has been strongly implicated in the development and progression of both human lung cancer and aggressive malignant breast cancer (Afonja *et al.*, 2004; Chen *et al.*, 2003). The expression of Pdcd4 has been shown to be reduced in many renal-, lung- and glia- derived tumours (Jansen *et al.*, 2004). In addition, Pdcd4 has been shown to suppress tumour development in a mouse model of skin carcinogenesis (Jansen *et al.*, 2005). This suggests that Pdcd4 might be a potential molecular target for cancer prevention. Although the exact mechanism of Pdcd4s anti-tumour activity is unknown, it has been suggested that it may involve processes such as the induction of apoptosis and the indirect suppression of the expression of carbonic anhydrase and cyclin-

dependent kinase 1/cdc2 (Afonja *et al.*, 2004; Goke *et al.*, 2004; Jin *et al.*, 2006; Lankat-Buttgereit *et al.*, 2004; Zhang *et al.*, 2006).

Human (also referred to as H731-L) and mouse Pdcd4 are 469 residue proteins, which contain at least three domains, an N-terminal RNA-binding region (residues 1-157) and two MA-3 domains (residues 157-287 and 319-449) (Bohm *et al.*, 2003; Ponting, 2000). Pdcd4 also contains two potential nuclear localisation signals, which are found close to the N- and C-termini, and two nuclear export signals, which are found within the middle MA-3 domain (MA-3_M) (Bohm *et al.*, 2003). The domain structure of mouse Pdcd4 is shown in figure 3.1.

Figure 3.1 Schematic diagram of the functional regions of mouse Pdcd4. The RNA-binding region (residues 1-157), middle MA-3 domain (MA-3_M) (157-287) and C-terminal MA-3 domain (MA-3_C) (319-449) are indicated.

Characterisation of the cellular functions of Pdcd4 is the focus of many ongoing investigations, however, recent work indicates essential roles in the control of both transcription and translation, mediated via specific protein-protein and protein-RNA interactions (discussed in sections 3.1.2 and 3.1.3) (Bitomsky *et al.*, 2004; Yang *et al.*, 2004; Yang *et al.*, 2003a). The best characterised functional interaction is between the MA-3 domains of Pdcd4 and the eukaryotic initiation factors eIF4A and eIF4G, with

complex formation resulting in the inhibition of cap-dependent translation (Goke *et al.*, 2002; Kang *et al.*, 2002b; Yang *et al.*, 2004; Yang *et al.*, 2003a).

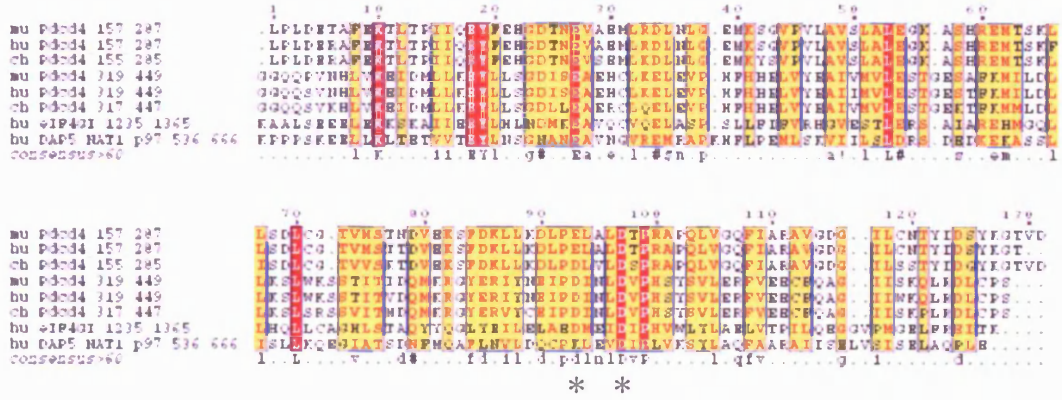


Figure 3.2 Amino acid sequence conservation between the MA-3 domains of Pdcd4 and eIF4G proteins. The multiple sequence alignment of the MA-3 domains of Pdcd4 (mouse, human and chicken), human eIF4G and its homologue DAP-5/NAT1/p97 illustrates the high degree of sequence homology between the MA-3 domains. Residues with conserved sequence identity are highlighted in red, those which are conserved sequence similarity in five or more domains are highlighted in yellow. The positions of two functionally important, semi-conserved aspartate residues (D414 and D418 of Pdcd4 MA-3_C) are indicated (*). The consensus sequence is shown below. Amino acids with conserved sequence identity are shown in uppercase; those with conserved sequence similarity in over 60% of the sequences are shown in lowercase. Similar residues were grouped as follows: AVILM, FYW, KRH, DE, STNQ, PG and C. The symbol '!' is used to denote either I or V, '$' denotes L or M, and '#' denotes anyone of NDQE. The alignment was prepared using ClustalW and ESPript.cgi (http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl)

To date, MA-3 domains have been found in over 200 eukaryotic proteins, including eIF4G,

and appear to be involved in mediating specific protein-protein interactions (http://www.sanger.ac.uk/Software/Pfam/)(Finn *et al.*, 2006).

As Pdcd4 has been shown to regulate both nuclear and cytoplasmic processes, a number of studies have focused on determining its subcellular localisation, but a somewhat confusing picture has emerged (Goke *et al.*, 2004; Yang *et al.*, 2003a; Yoshinaga *et al.*, 1999; Zhang *et al.*, 2006). Two recent studies have shown that Pdcd4 is able to shuttle between the nucleus and the cytoplasm but under normal cell growth conditions it appears to be predominantly localised to the nucleus (Bohm *et al.*, 2003; Palamarchuk *et al.*, 2005). It was further shown that the phosphorylation of residue S457 of Pdcd4 by the serine/threonine kinase Akt promotes the nuclear translocation of Pdcd4 (Palamarchuk *et al.*, 2005). In addition, the phosphorylation of either residue S67 or S457 by Akt has been shown to significantly reduce the ability of Pdcd4 to inhibit AP-1 dependent activation of gene expression (Palamarchuk *et al.*, 2005).

Recent results suggested that S67 of Pdcd4 is also phosphorylated by the mTOR-ribosomal protein S6 kinase 1 (S6K1) in response to mitogens. This promotes the phosphorylation of S71 and S76, by an unidentified kinase, and the subsequent degradation of Pdcd4 by the ubiquitin ligase $SCF^{\beta TRCP}$. It has been suggested that this degradation of Pdcd4 allows for the efficient translation of proteins required for cell growth and proliferation (Dorrello *et al.*, 2006).

3.1.2 Inhibition of AP-1 Dependent Transactivation by Pdcd4

As previously mentioned, Pdcd4 has been shown to inhibit the transcription factor AP-1, which is known to regulate a wide range of cellular processes including; proliferation, differentiation, invasion, metastasis, hypoxia, angiogenesis and apoptosis (reviewed in

Eferl & Wagner, 2003; Shaulian & Karin, 2002). The AP-1 complex is formed from different dimeric combinations of proteins from the Jun (c-Jun, JunB, JunD) and Fos (c-Fos, FosB, Fra-1, Fra-2) subfamilies, which form either Jun-Jun homodimers or Jun-Fos heterodimers. Work by Bitomsky *et al.* suggested that Pdcd4 inhibits AP-1 dependent transcription by interacting with both c-Jun and Jun N-terminal kinase (JNK), thereby preventing JNK-dependent phosphorylation of c-Jun and the recruitment of the coactivator p300 by AP-1 complexes (Bitomsky *et al.*, 2004). Phosphorylation of c-Jun residues S63 and S73 is believed to enhance transcription by inducing the exchange of AP-1 bound co-repressor complexes for coactivators such as CBP and p300 (Bannister *et al.*, 1995; Derijard *et al.*, 1994; Hibi *et al.*, 1993; Lee *et al.*, 1996; Ogawa *et al.*, 2004; Smeal *et al.*, 1991).

A second mechanism for the inhibition of AP-1 dependent transcription by Pdcd4 has been proposed, in which Pdcd4 indirectly inhibits AP-1 by targeting the expression of a kinase found upstream of JNK. Recent microarray data showed that Pdcd4 inhibits the transcription of mitogen activated protein kinase kinase kinase kinase 1 (MAP4K1), resulting in the inhibition of JNK dependent phosphorylation of c-Jun. It was suggested that Pdcd4 works indirectly by inhibiting the translation of a transcription factor/coactivator that is required for the expression of MAP4K1 (Yang *et al.*, 2006).

Interestingly, it has been shown that both nuclear and cytoplasmic Pdcd4 are capable of inhibiting AP-1 dependent transactivation (Palamarchuk *et al.*, 2005). It would seem likely that nuclear Pdcd4 acts by binding to c-Jun and JNK, thus inhibiting c-Jun phosphorylation. However, cytoplasmic Pdcd4 could function by either interacting with cytoplasmic JNK, again directly inhibiting c-Jun phosphorylation, or through inhibiting the translation of proteins required for AP-1 dependent transactivation.

3.1.3 Pdcd4 dependent Translational Regulation

The eIF4F cap-binding complex (eIF4A, eIF4E and eIF4G) plays a key role in cap-dependent translation, as it recruits the small ribosomal subunit to the 5' mRNA cap and is subsequently involved in scanning for the initiation codon. The translation of mRNAs with structured 5'UTRs is dependent on the helicase activity of eIF4A (Svitkin *et al.*, 2001), which is stimulated by binding to eIF4F (Abramson *et al.*, 1987; Richter-Cook *et al.*, 1998; Rogers *et al.*, 2001). The recruitment of eIF4A to eIF4F is mediated by two domains contained in the middle and C-terminal regions of eIF4G (discussed in section 1.4.2) (Imataka & Sonenberg, 1997; Morino *et al.*, 2000).

Pdcd4 has been shown to suppress translation by competitively inhibiting the interaction between eIF4A and the C-terminal regulatory region of eIF4G (eIF4Gc: residues 1040-1560), which is likely to result in the decreased recruitment of eIF4A to the eIF4F complex (Laronde-Leblanc *et al.*, 2006; Yang *et al.*, 2004; Yang *et al.*, 2003a). The interaction of eIF4Gc with eIF4A is mediated by an MA-3 domain, homologous to those found in Pdcd4 (Bellsolell *et al.*, 2006; Imataka & Sonenberg, 1997; Morino *et al.*, 2000). Recent data has shown that the C-terminal MA-3 domain of Pdcd4 (Pdcd4 MA-3_C) effectively competes with eIF4Gc for binding to eIF4A. In addition, it was shown that the isolated Pdcd4 MA-3_C domain is sufficient to inhibit the translation of 5' UTR structured mRNA (Laronde-Leblanc *et al.*, 2006).

Mutational studies have shown that the interactions between eIF4A and the MA-3 domains of either Pdcd4 or eIF4Gc involve two conserved acidic residues (Pdcd4 MA-3_M: E249 & D253, Pdcd4 MA-3_C: D414 & D418, eIF4Gc: D1329 & D1333), which are highlighted in figure 3.2 (Yang *et al.*, 2004). The substitution of one of these conserved residues in either Pdcd4 MA-3 domain, or removal of one of the MA-3 domains, has been shown to

dramatically reduce binding to eIF4A. Similarly, substitution of conserved residues in either domain of eIF4A results in reduced affinity for Pdcd4 (Zakowicz *et al.*, 2005). These results clearly suggest that the two MA-3 domains of Pdcd4 and the two globular domains of eIF4A are involved in forming a tight complex.

Pdcd4 also appears to bind tightly to the middle region of eIF4G (eIF4Gm; residues 672-1065), but to a site distinct from that involved in eIF4A binding (Yang *et al.*, 2003a). This interaction is also believed to contribute to the inhibitory effect of Pdcd4 on translation, possibly by locking eIF4A into the eIF4F complex and preventing the required eIF4A-eIF4F association/dissociation cycles (Pause *et al.*, 1994).

The ability of Pdcd4 to suppress the helicase activity of eIF4A means that the translation of mRNAs with structured 5'UTRs are highly susceptible to translational inhibition by Pdcd4 (Yang *et al.*, 2004; Yang *et al.*, 2003a). Although no specific translational targets of Pdcd4 have yet been identified, a recent study has suggested that Pdcd4 may inhibit the translation of a transcription factor/coactivator that is required for the transcription of MAP4K1 (Yang *et al.*, 2006).

3.1.4 Aims

The main aim of the work described in this chapter was to study the interaction of the C-terminal MA-3 domain of Pdcd4 and eIF4A. At the commencement of this project no structure existed for any MA-3 domain or region of Pdcd4. Therefore the initial aim was to determine the solution structure of the Pdcd4 MA-3_C and subsequently use NMR mapping experiments to characterise its interaction with eIF4A.

3.2 Materials and Methods

3.2.1 Protein Expression and Purification

3.2.1.1 Expression and Purification of Pdcd4 MA-3$_C$

Unlabelled and uniformly ^{15}N and $^{15}N/^{13}C$ labelled samples of Pdcd4 MA-3$_C$ (mouse residues 319-449) were prepared from a pGex6P2 based *E. coli* expression vector, which introduced an N-terminal glutathione-S-transferase (GST) tag linked via a 5 residue sequence. The unlabelled samples were grown in LB medium and the labelled samples were grown in minimal media containing 0.6 g/l >99% ^{15}N ammonium sulphate and/or 2 g/l >99% ^{13}C glucose as the sole nitrogen and carbon sources. A ^{13}C labelled protein sample which was used for the $^{13}C/^{1}H$ NOESY-HSQC experiment was prepared as described above, but with the addition of non-isotopically labelled aromatic amino acids to the minimal media (His, Phe, Trp and Tyr at 50 mg/l). Transformed cells were grown at 37°C and the expression of protein was induced in mid log phase by addition of IPTG to 0.45 mM. The cells were induced for four hours at 30°C and then harvested by centrifugation at 6600g for 20 minutes. The cell pellet was resuspended and lysed in 50 mM tris-(hydroxymethyl)-methylamine (Tris), 0.5 mM EDTA and 0.1% Triton X100 (v/v) buffer at pH 8, to which was added 100 µg/mL lysozyme and 0.1 mM PMSF (phenylmethylsulfonyl fluoride) to inhibit protease activity. In addition, 5 mM magnesium chloride and 0.065 mg/L DNase I were added in order to break down the DNA. Insoluble debris was removed from the cell lysate by centrifugation at 13,000g for 20 minutes. The resulting supernatant, which contained the GST fusion protein, was dialysed into 150 mM sodium chloride, 50 mM Tris, 2 mM 1,4-dithiothreitol (DTT) and 1 mM EDTA column running buffer at pH 7.4 in preparation for purification by affinity chromatography. The dialysate was mixed with 10 ml of S-linked glutathione agarose beads and incubated with constant agitation for 1 hour at 4°C to promote binding. After incubation the agarose beads were washed twice with the column running buffer to remove non-bound material

and then packed into a column. Using an AKTA FPLC the column was washed with a further 3-4 column volumes of the column running buffer. The GST-tagged Pdcd4 MA-3_C protein was then eluted by the application of 10 mM reduced glutathione in 150 mM sodium chloride, 50 mM Tris, 2 mM DTT and 1 mM EDTA buffer at pH 7.4. FPLC fractions containing the fusion protein were pooled and dialysed into PreScission Protease cleavage buffer; 150 mM sodium chloride, 50 mM Tris, 1 mM DTT and 1 mM EDTA, at pH 7. The GST tag was then removed by cleavage with PreScission Protease (150 U/L) (Pharmacia) and the two products separated by gel filtration on a superdex 75 16/60 column (Amersham Biosciences). The Pdcd4 MA-3_C fractions were eluted in 100 mM sodium chloride, 25 mM sodium phosphate, 0.5 mM DTT and 10 μM EDTA buffer at pH 6.5. The purified protein was judged to be greater than 95% pure by SDS-PAGE (Invitrogen 4-12% Bis-Tris NuPAGE gel system). The identity of the purified protein was confirmed by mass spectroscopy.

3.2.1.2 Expression and Purification of the Pdcd4 Dual MA-3 Domain Protein.

A protein encompassing the two MA-3 domains of Pdcd4 (residues 157-449) was prepared by a visiting Ph.D student (Thore Schmedt). The protein was prepared from a pGex6P2 based *E. coli* expression vector. Uniformly ^{15}N and $^{15}N/^{13}C$ labelled samples were prepared essentially as described for Pdcd4 MA-3_C, however, the transformed cells were induced for 20 hours at 15 °C. In addition, the cleaved Pdcd4 (157-449) and GST were dialysed in to 300 mM sodium chloride, 50 mM Tris, 2 mM DTT and 1mM EDTA buffer at pH 7.4. The GST tag was removed from the sample by re-incubating the two proteins with the glutathione agarose beads. The purified Pdcd4 (157-449) was then dialysed into the NMR buffer.

3.2.1.3 Expression and Purification of eIF4A Proteins

The N-terminal histidine-tagged full length mouse eIF4A and the isolated N-terminal domain of eIF4A (residues 1-242) were prepared from pET28a-based *E. coli* expression vectors. Transformed cells were grown at 37°C in LB medium and expression of the protein was induced in mid log phase by addition of IPTG to 0.45 mM. The cells were harvested at 3 hours post induction by centrifugation at 6600g for 20 minutes. The cell pellet was resuspended in 50 mM Tris, 0.5 mM EDTA and 0.1% Triton X100 (v/v) buffer at pH 8, to which was added 100 μg/ml lysozyme and 0.1 mM PMSF. The cells were then lysed by sonication. Insoluble debris was removed from the cell lysate by centrifugation at 13,000g for 30 minutes. The resulting supernatant, which contained the His-tagged eIF4A proteins, was dialysed into 100 mM sodium chloride, 20 mM Tris and 15 mM imidazole column running buffer at pH 8 and filtered through a 0.2 μm millipore filter in preparation for affinity chromatography. The dialysate was loaded onto a 10 mL Ni-NTA column and then washed with 2 column volumes of column running buffer. His-tagged eIF4A proteins were eluted from the column using a linear imidazole gradient (15-300 mM). FPLC fractions containing the eIF4A proteins were pooled and dialysed into 125 mM sodium chloride, 25 mM sodium phosphate, 0.5 mM DTT and 10 μM EDTA, at pH 7 and then subjected to a polishing purification step by gel filtration chromatography on a Superdex 75 16/60 pre-packed column (Amersham Biosciences). The purified protein was judged to be greater than 95% pure by SDS-PAGE (Invitrogen 4-12% Bis-Tris NuPAGE gel system).

The isolated C-terminal domain of eIF4A (residues 237-406) was prepared by a visiting Ph.D student (Thore Schmedt) essentially as described above, but using a higher IPTG concentration of 0.9 mM. The identities of the purified eIF4A proteins were confirmed by mass spectroscopy.

3.2.2 Circular Dichroism (CD) Spectroscopy

Far UV CD was employed to monitor possible changes in the secondary structure of the full length or isolated domains of eIF4A upon increasing sample temperature. Spectra were obtained from 15 μM samples of full length eIF4A in 100 mM sodium chloride, 25 mM sodium phosphate, 0.5 mM DTT and 10 μM EDTA buffer at pH 7 and the isolated N- and C-terminal domains of eIF4A in a 125 mM sodium chloride, 25 mM sodium phosphate, 5 mM DTT and 10 μM EDTA buffer at pH 7. CD spectra were recorded at 5 °C intervals over a temperature range of 5 to 85 °C. Samples were placed in a 1 mm path length cell and spectra were recorded from 190 to 250 nm at a resolution of 1 nm and scan speed of 20 nm/min, with each spectrum representing the average of 10 accumulations. The CD spectra were corrected for buffer absorbance and the raw data converted to molar CD per residue. In addition, changes in the intensity of the maximum negative peak at 221 nm were recorded as a function of increasing temperature, with the temperature rising incrementally at 1 °C/min.

3.2.3 NMR Spectroscopy

3.2.3.1 Experiments Acquired to Determine the Structure of Pdcd4 MA-3_C

NMR spectra were acquired from 0.35 ml samples of 0.6 – 1.0 mM Pdcd4 MA-3_C in a 25 mM sodium phosphate, 100 mM sodium chloride, 0.5 mM DTT, 10 μM EDTA and 0.02% (w/v) sodium azide buffer at pH 6.5, containing either 10% D_2O/90% H_2O or 100% D_2O as appropriate. All NMR data were acquired at 25°C on either 600 MHz Bruker Avance/DRX systems, or an 800 MHz Varian Inova spectrometer (table 3.1). The 2D and 3D spectra recorded to obtain sequence specific assignments and calculate the structure of Pdcd4 MA-3_C were: ^{1}H-^{1}H TOCSY (Braunschweiler & Ernst, 1983) with mixing times of 40 and 60 ms, ^{1}H-^{1}H NOESY (Macura & Ernst, 1980) with an NOE mixing time of 100 ms; ^{15}N/^{1}H HSQC (Bodenhausen & Ruben, 1980); TOCSY-HSQC (Marion *et al.*, 1989a)

with a mixing time of 55 ms; NOESY-HSQC (Marion *et al.*, 1989b) with an NOE mixing time of 100 ms; $^{13}C/^{1}H$ HSQC; HCCH-TOCSY (Bax *et al.*, 1990) with a mixing time of 13 ms, NOESY-HSQC (Zuiderweg *et al.*, 1990) with an NOE mixing time of 100 ms and $^{15}N/^{13}C/^{1}H$ HNCACB (Wittekind & Mueller, 1993) and CBCA(CO)NH (Grzesiek & Bax, 1993). Typical acquisition times in F_1 and F_2 for the 3D experiments were 9.2-24 ms for ^{15}N, 6.6-10.0 ms for ^{13}C and 15-18 ms for ^{1}H, and with an acquisition time of 80 ms in F_3 (^{1}H). The majority of the 3D spectra were collected over approximately 72 hours, 2D ^{1}H experiments over 24 hours and $^{15}N/^{1}H$ and $^{13}C/^{1}H$ HSQC spectra over about 1 hour and 20 minutes respectively. Typical acquisition times in 2D experiments were either 60 ms (^{15}N), 10 ms (^{13}C) or 35-42 ms (^{1}H) in F_1 and 85-125 ms in F_2 (^{1}H). The WATERGATE method was used to suppress the water signal when required (Piotto *et al.*, 1992).

Table 3.1 NMR experimental details

Experiment	Solvent	Spectrometer
$^{15}N/^{13}C/^{1}H$ HNCACB	10% D_2O/90% H_2O	600 MHz Bruker Avance
$^{15}N/^{13}C/^{1}H$ CBCA(CO)NH	10% D_2O/90% H_2O	600 MHz Bruker Avance
$^{15}N/^{1}H$ TOCSY-HSQC	10% D_2O/90% H_2O	600 MHz Bruker Avance
$^{15}N/^{1}H$ NOESY-HSQC	10% D_2O/90% H_2O	800 MHz Varian Inova
$^{13}C/^{1}H$ HCCH-TOCSY	100% D_2O	600 MHz Bruker Avance
$^{13}C/^{1}H$ NOESY-HSQC	100% D_2O	600 MHz Bruker DRX (with cryoprobe)
^{1}H-^{1}H TOCSY	100% D_2O	600 MHz Bruker DRX (with cryoprobe)
^{1}H-^{1}H NOESY	100% D_2O	600 MHz Bruker DRX (with cryoprobe)

The 3D NMR data were processed using either NMRPipe (Delaglio *et al.*, 1995) or Topspin (Bruker Biospin Ltd) with linear prediction used to extend the effective

acquisition times by up to 2 fold in F_1 and F_2. The spectra were analysed using the Sparky package (Goddard *et al.*, unpublished).

3.2.3.2 Chemical Shift Mapping Experiments

$^{15}N/^{1}H$ and $^{13}C/^{1}H$ HSQC spectra (Bodenhausen & Ruben, 1980) of Pdcd4 MA-3_C were acquired in the presence and absence of eIF4A to identify any changes in the positions of signals induced by eIF4A binding. In these experiments, unlabelled full length eIF4A or the N- or C-terminal domains alone, were added to 100μM ^{15}N or ^{13}C-labelled Pdcd4 MA-3_C in a 25 mM sodium phosphate, 125 mM sodium chloride, 5 mM DTT, 10 μM EDTA and 0.02% (w/v) sodium azide buffer at pH 7, containing either 10% D_2O/90% H_2O ($^{15}N/^{1}H$ HSQC) or 100% D_2O ($^{13}C/^{1}H$ HSQC). The full length eIF4A was titrated to a concentration of 150 μM (limiting solubility of eIF4A) and the isolated N-and C-terminal domains to between 400μM and 1mM. $^{15}N/^{1}H$ HSQC spectra were collected as described previously, but with an acquisition time of 75 ms in F_2 (^{1}H) for a period of 6.5 hours. $^{13}C/^{1}H$ HSQC spectra of 100μM ^{13}C-labeled Pdcd4 MA-3_C (with unlabeled aromatics) in the presence and absence of unlabeled 400 μM eIF4A N-terminal domain were collected as described previously, but with an acquisition time of 75 ms in F_2 (^{1}H) for a period of 12 hours. All spectra were acquired on a 600 MHz Bruker DRX spectrometer with a cryoprobe. The spectra were processed and analysed as described in section 3.2.3.1.

3.2.3.3 Spectra Acquired for the Comparison of the Pdcd4 MA-3_C and Pdcd4 Dual MA-3 Domain (157-449) Proteins

$^{15}N/^{1}H$ HSQC (Bodenhausen & Ruben, 1980) of 200 μM Pdcd4 (residues 157-449) were collected as described in section 3.2.3.1. $^{15}N/^{13}C/^{1}H$ HNCO (Grzesiek & Bax, 1992) spectra of 300 μM Pdcd4 (residues 157-449) and 200 μM Pdcd4 MA-3_C were acquired to identify shifts in the position of Pdcd4 MA-3_C signals in the dual MA-3 domain (157-449)

protein. All spectra were acquired on a 600 MHz Bruker Avance spectrometer. Typical acquisition times for the HNCO were 75 ms in F_3 (^{1}H), 14.5 ms in F_2 (^{15}N), and 8.6 ms in F_1 (^{13}C), with the spectra collected over approximately 40 hours. The 2D and 3D NMR data were processed and analysed as described in section 3.2.3.1.

3.2.4 Sequence Specific Assignments

Sequence specific backbone and sidechain assignments were obtained for the Pdcd4 MA-3_C domain, essentially as previously described in section 2.2.5.

3.2.5 Structural Calculations

The family of converged Pdcd4-MA3_C structures was determined in a two stage process using the program CYANA (Guntert *et al.*, 1997), as previously described in section 2.2.7. The input for the CANDID stage consisted of essentially complete ^{15}N, ^{13}C and ^{1}H resonance assignments for the non-exchangeable groups in Pdcd4 MA-3_C (Waters *et al.*, 2006), two manually picked three-dimensional NOE peak lists corresponding to all NOEs involving amide protons (1346) and all NOEs between aliphatic protons (2497), and one manually picked two-dimensional NOE peak list corresponding to all NOEs involving aromatic sidechain protons (411). In addition, the CANDID stage included 174 torsion angle constraints for Pdcd4 MA-3_C as determined by the protein backbone dihedral angle prediction programme TALOS (Cornilescu *et al.*, 1999). Typical Ψ and Φ angle constraints, within helical regions, were between ± 15° and ± 20°. A complete list of the dihedral angle constraints used in the structural calculations is shown in appendix A2.3. CANDID calculations were carried out using the default parameter settings in CYANA 2.1 with chemical shift tolerances set to 0.03 ppm (direct and indirect ^{1}H) and 0.3 ppm (^{15}N and ^{13}C).

The final converged Pdcd4 MA-3_C structures were produced from 100 random starting coordinates using a standard torsion angle-based simulated annealing protocol combined with 5 cycles of redundant dihedral angle constraints (REDAC) (Carr *et al.*, 2003; Muskett *et al.*, 1998). The calculations were based upon 2545 non-redundant, NOE-derived upper distance limits, assigned to unique pairs of protons using CANDID, 87 Φ and Ψ torsion angle constraints derived from TALOS. In addition, hydrogen bond constraints, involving 17 residues with slowly exchanging backbone amide signals (residues 337, 338, 366-370, 372, 381, 384-388, 429-431), and where the hydrogen bond acceptor was unambiguous in preliminary structures, were added to the final round of calculations. Analysis of the final family of structures obtained was carried out using the programs CYANA and MOLMOL (Guntert *et al.*, 1997; Koradi *et al.*, 1996).

3.2.6 Chemical Shift Mapping of the eIF4A Binding Site on Pdcd4 MA-3_C

The minimal shift approach was used to estimate the changes in the positions of Pdcd4 MA-3_C NMR signals resulting from eIF4A binding (Muskett *et al.*, 1998; Williamson *et al.*, 1997). The minimum change in the position for all backbone amide peaks between the free and eIF4A-bound Pdcd4 MA-3_C was obtained by using Microsoft Excel to calculate the combined chemical shift difference in ^{15}N and ^{1}H for each assigned peak in the $^{15}N/^{1}H$ HSQC spectrum of the free protein compared to all peaks observed in the HSQC spectra of the complexes formed with eIF4A. The combined amide proton and nitrogen chemical shift difference (Δδ) was defined according to the calculation $\Delta\delta = \sqrt{(\Delta\delta_{HN})^2 + (\Delta\delta_N \alpha_N)^2}$ where $\Delta\delta_{HN}$ and $\Delta\delta_N$ correspond to the differences in ^{1}H and ^{15}N chemical shifts between pairs of compared HSQC peaks and α_N is a scaling factor of 0.2 required to account for differences in the range of amide proton and nitrogen chemical shifts. For each individual HSQC peak, the minimal shift induced by ligand binding was taken as the lowest possible combined shift value (Δδ). The minimum change in the position of peaks arising from the

well resolved Cα/Hα and methyl signals in $^{13}C/^{1}H$ HSQC spectra was similarly determined, but using a scaling factor of 0.1 for the Cα shifts and 0.12 for the methyl carbon shifts. The minimal shift approach was also used to determine any significant changes in the positions of Pdcd4 MA-3_C signals in HNCO spectra of the dual MA-3 domain protein (157-449), with a scaling factor of 0.35 used for carbonyl signals.

3.3 Results

3.3.1 Protein Expression and Purification

3.3.1.1 Expression and Purification of the Pdcd4 MA-3_C Domain.

A pGEX expression vector, containing the sequence encoding the Pdcd4 MA-3_C domain (319-449) was used to express the polypeptide in *E. coli*. The domain was purified in two stages. The GST-Pdcd4 fusion protein was initially purified by affinity chromatography using S-linked glutathione agarose resin, as shown in figure 3.3. The GST tag was then removed by cleavage with PreScission Protease (figure 3.4), and the two products were separated by gel filtration, as shown in figure 3.5. A small amount of the cleaved Pdcd4 MA-3_C and GST appeared to form a disulphide linked complex, however, this was also separated from the cleaved Pdcd4 MA-3_C and GST by gel filtration. Typical yields were about 4 mg/l. Electrospray mass spectroscopy was used to determine the mass of the purified ^{15}N-labelled protein (15578.1 Da). This corresponds closely to that of the expected full length Pdcd4 MA-3_C domain minus the C-terminal serine (15575.7 Da).

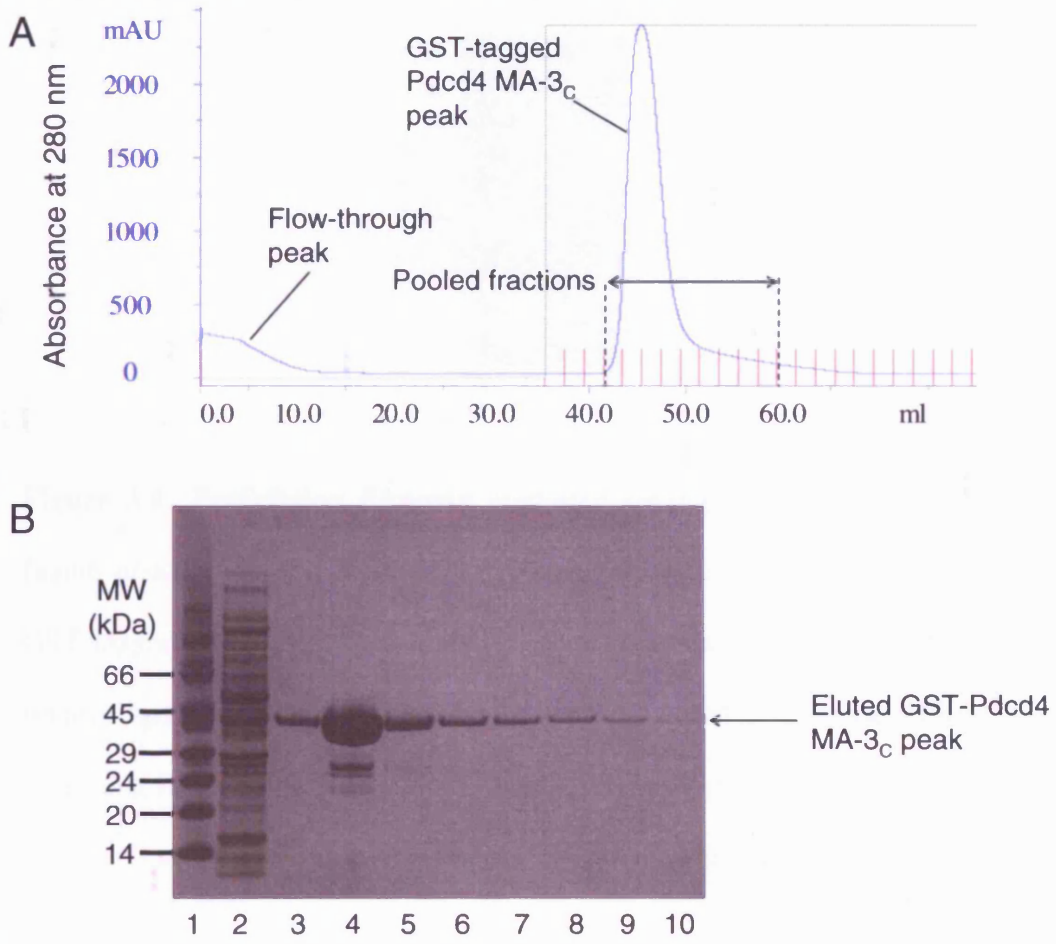


Figure 3.3 Purification of the GST-tagged Pdcd4 MA-3$_C$ domain. Panel A shows a typical FPLC profile attained for the glutathione agarose column based purification of the GST-tagged Pdcd4 MA-3$_C$ domain. The Pdcd4 MA-3$_C$ domain was eluted by the application of elution buffer containing10 mM reduced glutathione, as indicated on the profile in green. The flow-through and Pdcd4 MA-3$_C$ elution peaks are labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying FPLC elution fractions. Lane 2 shows a sample removed from the initial flow-through, lanes 3-10 show samples removed from the Pdcd4 MA-3$_C$ elution peak. Samples were taken from every other fraction. Fractions from lanes 3-7 were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).

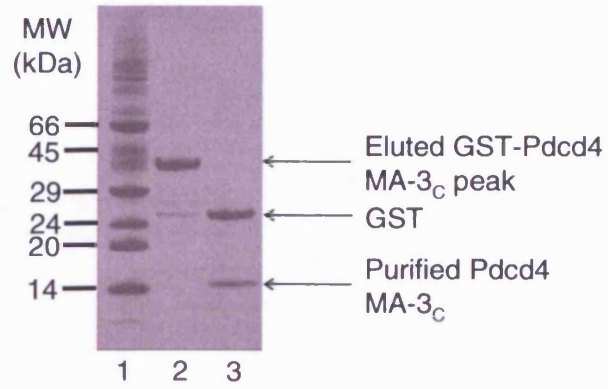


Figure 3.4 PreScission Protease mediated cleavage of the GST-Pdcd4 MA-3_C fusion protein. Lane 2 of the SDS PAGE gel shows a sample taken from the purified GST tagged Pdcd4 MA-3_C sample. Lane 3 shows a sample taken after the fusion protein had been incubated overnight with PreScission Protease (150 U/l). No remaining fusion protein is visible in lane 3, confiming that the all the protein had been cleaved. Lane 1 contains high molecular weight markers (Sigma).

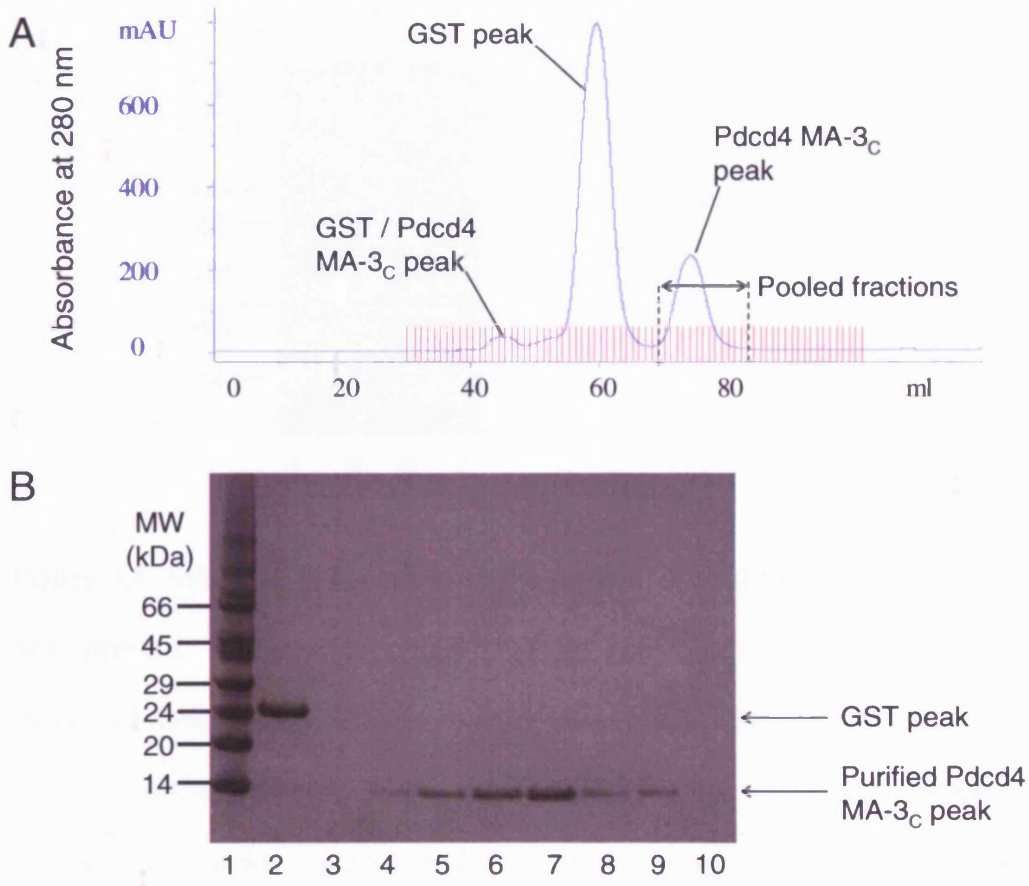


Figure 3.5 Purification of the Pdcd4 MA-3$_C$ domain by gel filtration. Panel A shows a typical FPLC elution profile obtained by gel filtration on a Superdex 75 16/60 column. The disulphide linked GST / Pdcd4 MA-3$_C$ complex, GST and Pdcd4 MA-3$_C$ elution peaks are labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying FPLC elution fractions. Lane 2 shows a sample taken from the GST elution peak and lanes 3-10 show samples removed from the Pdcd4 MA-3$_C$ elution peak. Fractions from lanes 3-10 were pooled. Lane 1 contains high molecular weight markers (Sigma).

3.3.1.2 Expression and Purification of the Pdcd4 Dual MA-3 Domain Protein (157-449).

The Pdcd4 dual MA-3 domain protein (157-449) was purified as described in section 3.2.1.2. A gel showing the purified protein is shown in figure 3.6.

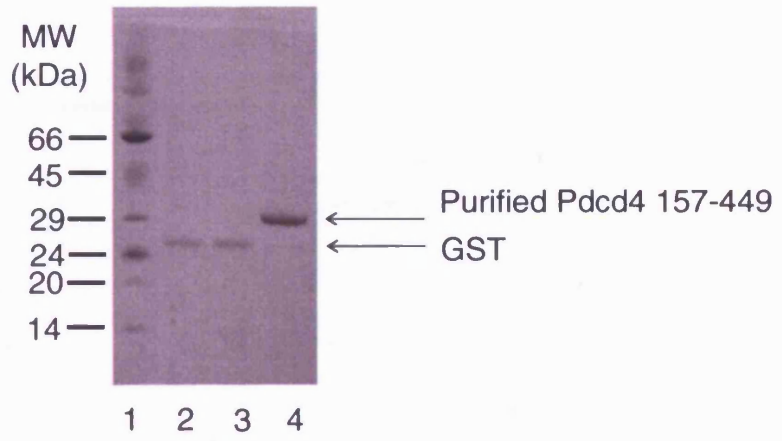


Figure 3.6 SDS PAGE gel showing the purified Pdcd4 dual MA-3 domain (157-449) protein. The cleaved Pdcd4 (157-449) protein and GST were separated by affinity chromatography. Lane 4 shows a sample of purified Pdcd4 (157-449) protein after incubation with glutathione agarose beads, lanes 2 and 3 show samples of the cleaved GST that was eluted from the glutathione agarose beads by the application of 10 mM reduced glutathione elution buffer. No remaining fusion protein is visible in lanes 2-4 confirming that all the protein had been cleaved. Lane 1 contains high weight molecular weight markers (Sigma).

3.3.1.3 Expression and Purification of the eIF4A Proteins

A pET28a expression vector, containing the sequence encoding either the full length eIF4A protein or just the N-terminal domain, was used to express the polypeptides in *E. coli*. The proteins were purified in two stages. The N-terminal hexa-histidine tag allowed the proteins to be purified by affinity chromatography, as shown in figure 3.7. The pooled His-tagged eIF4A protein fractions were subjected to a final polishing purification step by gel filtration, as shown in figure 3.8. Typical yields for full length eIF4A were 5 mg/mL, and for the N-terminal domain were 40 mg/L. SDS PAGE gels showing the purified N-terminal domain and full length eIF4A are shown in figures 3.8 and 3.9 respectively. A gel

showing the isolated C-terminal domain after purification is also shown in figure 3.9.

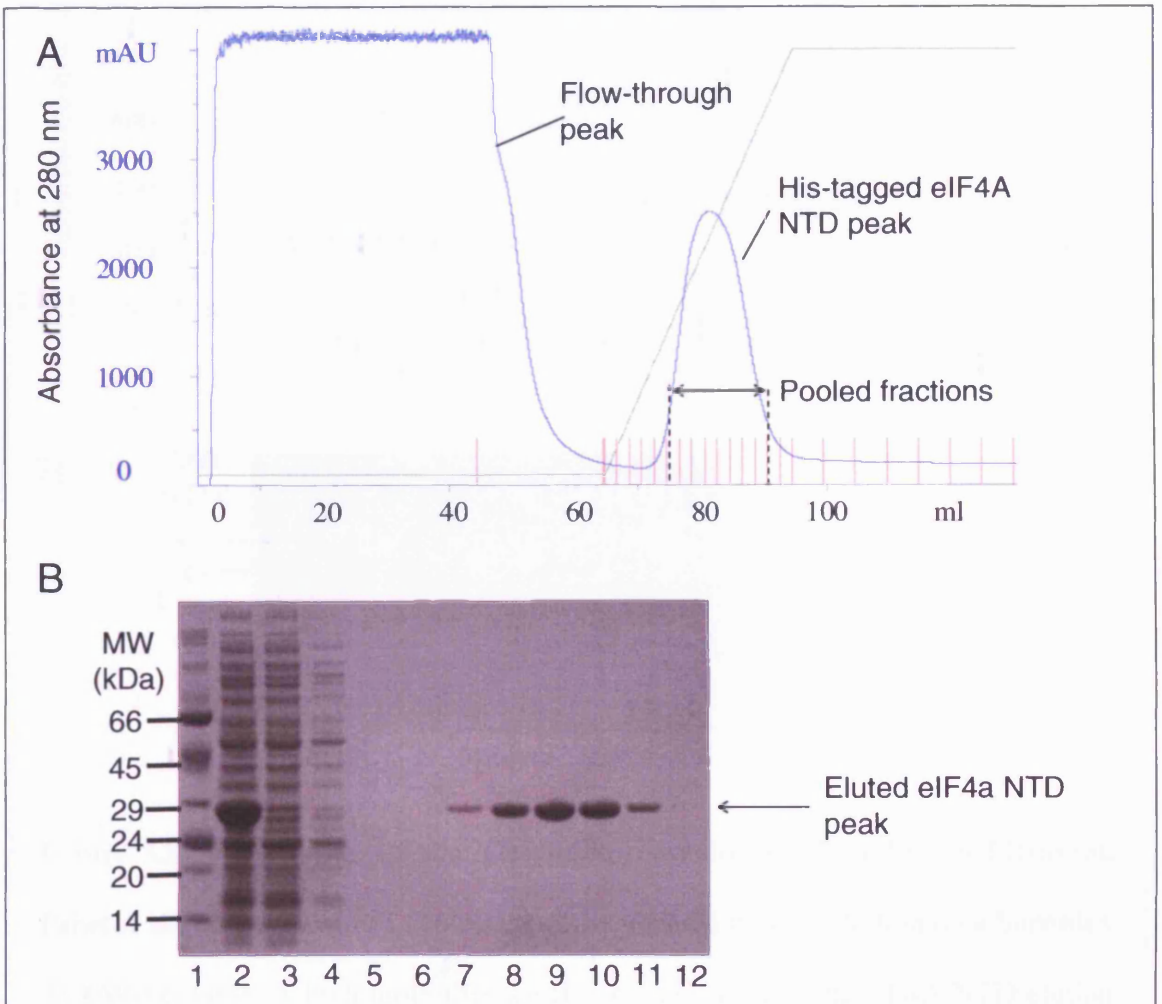


Figure 3.7 Purification of the His-tagged N-terminal domain of eIF4A. Panel A shows a typical FPLC profile attained for the Ni-NTA column based purification of the His-tagged eIF4A N-terminal domain (NTD). The eIF4A NTD was eluted over a linear gradient from 15 mM to 300 mM imidazole (indicated in green). The flow-through and eIF4A NTD elution peaks are labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2-4 shows samples removed from the loaded cell lysate, column flow-through and wash respectively, lanes 5-12 show samples removed from the eIF4A NTD elution peak. Fractions from lanes 7-11 were pooled for further purification. Lane 1 contains high molecular weight markers (Sigma).

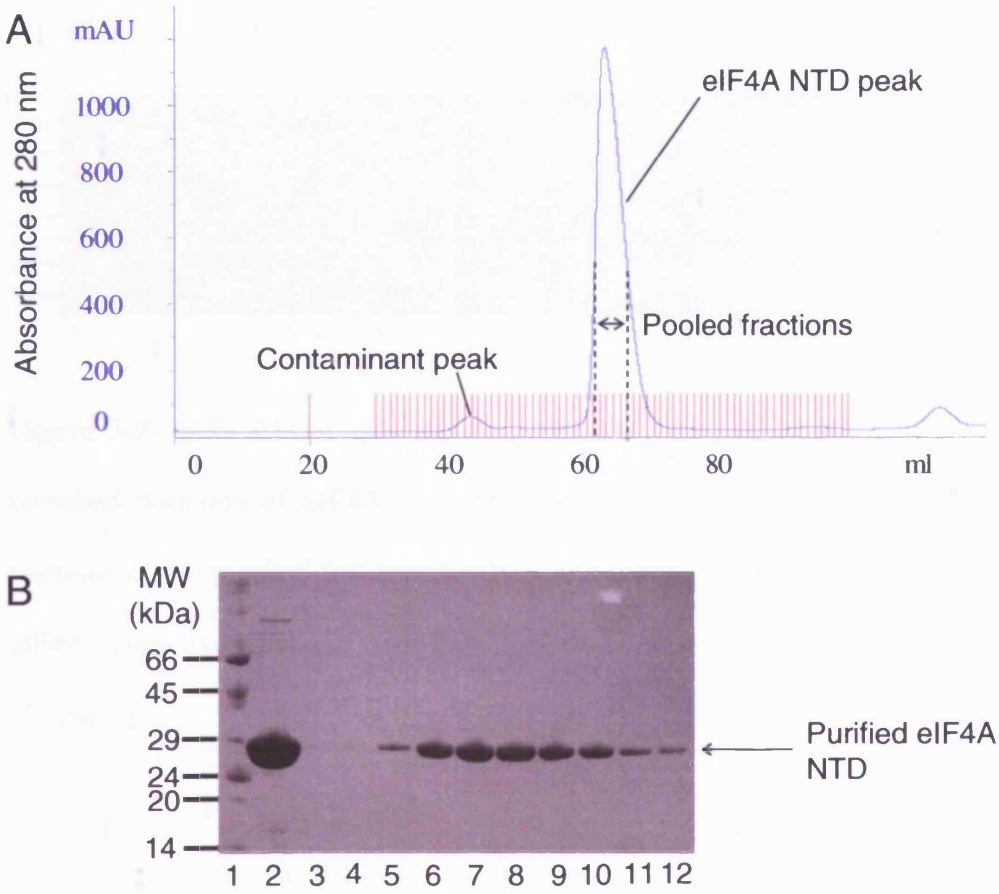


Figure 3.8 Purification of the N-terminal domain of eIF4A by gel filtration. Panel A shows a typical FPLC elution profile attained by gel filtration on a Superdex 75 16/60 column. A high molecular weight contaminant and the eIF4A NTD elution peaks are labelled. The pooled fractions are also highlighted. Panel B shows an SDS PAGE gel displaying the FPLC elution fractions. Lanes 2 and 3 show samples taken from the load and contaminant elution peak respectively, and lanes 4-12 show samples removed from the eIF4A NTD elution peak. Fractions from lanes 6-9 were pooled for structural studies. Lane 1 contains high molecular weight markers (Sigma).

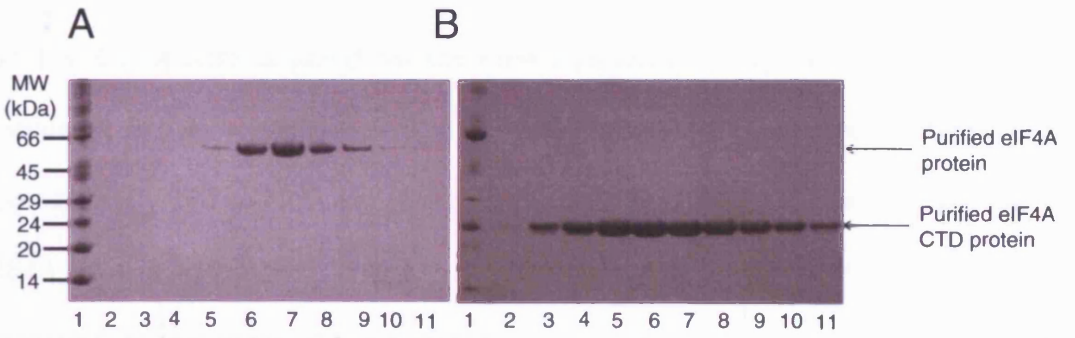


Figure 3.9 SDS PAGE gels showing the purified full length and isolated C-terminal domains of eIF4A. Panels A and B show gels displaying gel filtration fractions of the purified full length eIF4A and isolated C-terminal domain (CTD) of eIF4A respectively. Lane 1 of each gel contains high molecular weight markers (Sigma).

The identity of the purified eIF4A proteins were confirmed by electrospray mass spectroscopy, as shown in table 3.2. The mass of the purified full length eIF4A protein was approximately 36 Da larger than expected, suggesting that the protein may have been acetylated. The masses of the purified N- and C-terminal domains of eIF4A were within the 0.025% error range of the expected masses.

Table 3.2 Mass spectroscopy results obtained for the eIF4A proteins

	Expected Mass (Da)	Mass of Purified Protein (Da)
Full-length eIF4A	48317.2	48353.6
eIF4A N-terminal domain	29119.3	29125.2
eIF4A C-terminal domain	22108.0	22111.9

3.3.2 Circular Dichroism Spectroscopy

Far UV CD spectra acquired for the eIF4A proteins are typical of structured proteins containing helical secondary structure, with characteristic negative elipticity peaks at approximately 210 and 221 nm (Sreerama & Woody, 1994), as shown for the full length eIF4A protein in figure 3.10 panel B. The series of far UV CD spectra recorded at increasing temperatures showed that the full length and isolated N- and C-terminal domains of eIF4A are stable to thermal denaturation to about 30, 40 and 45°C respectively. The change in the intensity of the negative peak at 221 nm was recorded as a function of increasing temperature, as shown figure 3.10 panels A, C and D. The cooperative sigmoidal unfolding curves obtained for all three eIF4A proteins show that they contain stable tertiary structure and almost certainly form fully folded proteins.

3.3.3 Sequence Specific Assignments

Pdcd4 MA-3_C gives rise to well-resolved spectra, as illustrated by the $^{15}N/^{1}H$ HSQC spectrum shown in Figure 3.11, which allowed essentially complete backbone resonance assignments to be made. Backbone amide assignments were obtained for all residues in Pdcd4 except: H326, K329, E330, V356, H360, I397, D418, V419 and S422 (94%) and for all Cα and Cβ signals. A representative set of ^{15}N strips from the HNCACB spectrum of Pdcd4 MA-3_C used to assign the backbone atoms (NH, N, Cα and Cβ) are shown in figure 3.12. The corresponding set of representative ^{15}N strips from the $^{15}N/^{1}H$ NOESY-HSQC spectrum, used to confirm the backbone assignments, are shown in figure 3.13. Assignments were also obtained for all Hα and Hβ signals. For the remaining non-exchangeable side chain signals (^{13}C and ^{1}H) complete assignments were obtained apart from: K329 (Cε), H358 (Hε1, Hδ2), H361 (Hε1, Hδ2), L372 (Cγ, Hγ), I383 (Cγ1), Q400 (Cγ, Hγ2, Hγ3), I412 (Hγ12, Hγ13) I438 (Cγ1) and F434 (Hζ) (98%). The comprehensive ^{15}N, ^{13}C and ^{1}H resonance assignments obtained for Pdcd4 have been deposited at the

BioMagResBank database. (accession number 6900).

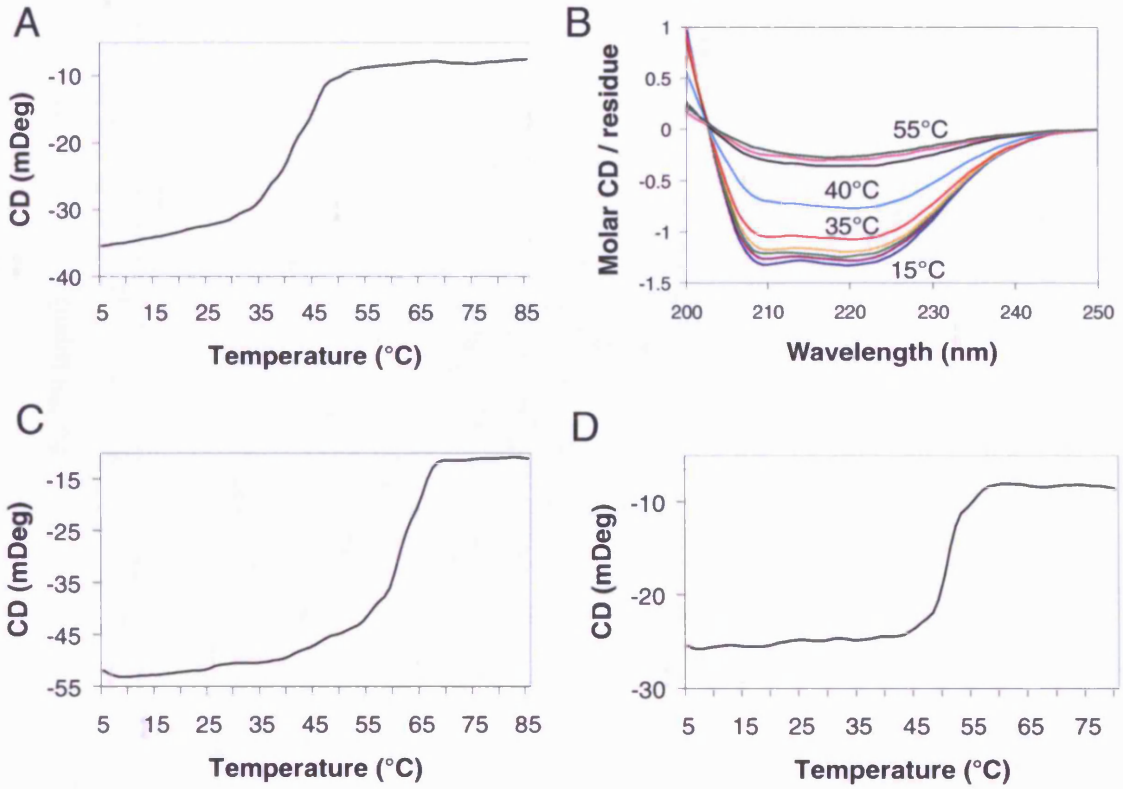


Figure 3.10 Far UV circular dichroism analysis of the temperature stability of eIF4A proteins. Panels A, C and D show the effects of increasing temperature on the helical secondary structure of full length eIF4A (panel A), and the isolated N-terminal (C) and C-terminal (D) domains, reflected as the change in CD intensity at 221 nm. Panel B shows overlays of far UV spectra acquired for the full length eIF4A protein over a range of temperatures. The spectra are the result of 10 accumulations. The spectra acquired at 15 (blue), 20 (pink), 25 (green), 30 (yellow), 35 (red), 40 (cyan), 45 (black), 50 (magenta), and 55°C (dark green) are shown. Examination of the low temperature spectra (15-30°C) shows that the eIF4A proteins contain predominantly helical secondary structure. It is evident from the overlays of the CD spectra and the sigmoidal denaturation curves that the eIF4A proteins unfold in a cooperative manner, and that they are resistant to thermal denaturation to between 30 and 45 °C.

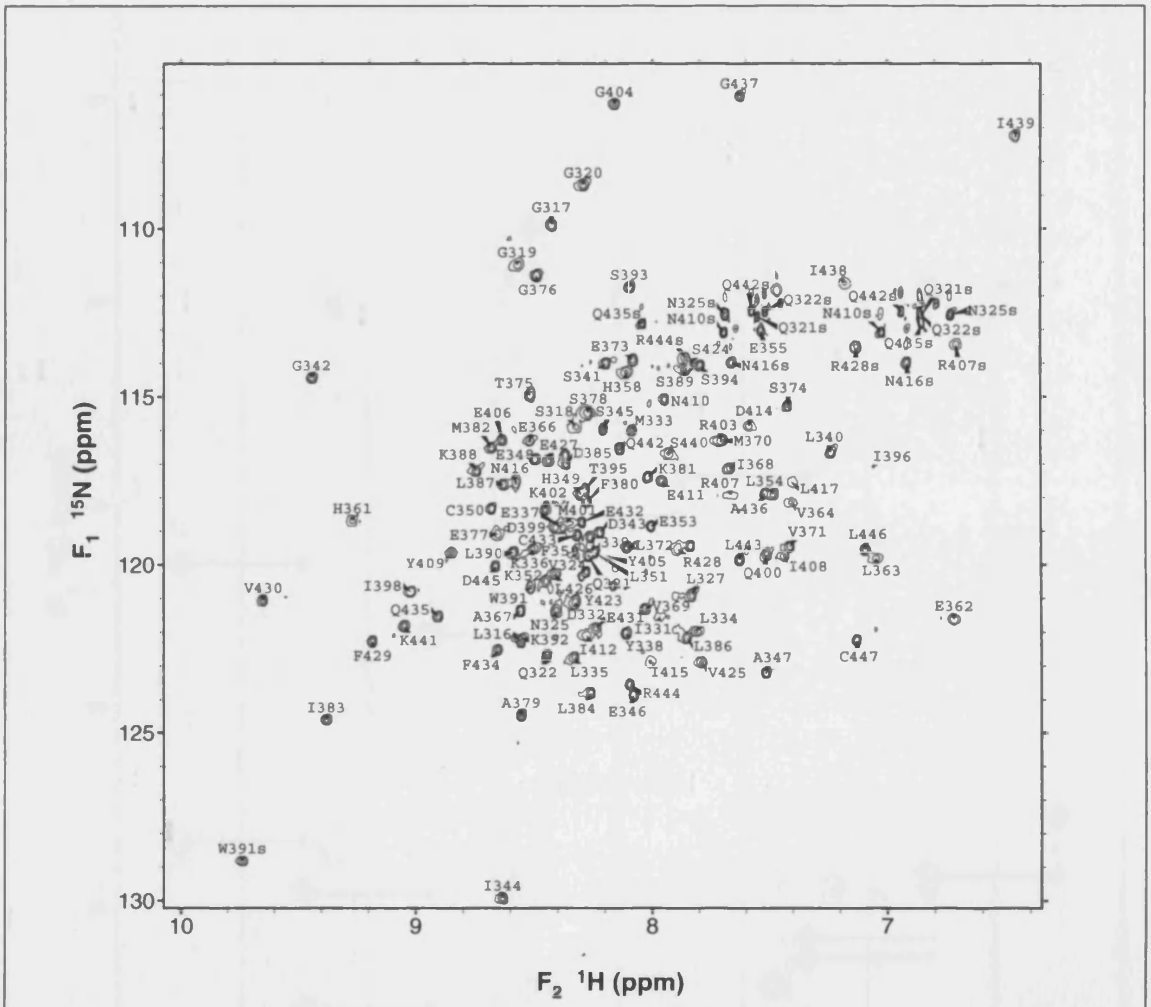


Figure 3.11 $^{15}N/^{1}H$ HSQC spectrum of the C-terminal MA-3 domain of Pdcd4. The assignments of the signals from backbone amide groups are indicated by residue type and number. In addition, sidechain amide group assignments are also labeled with residue type and number followed by 's'.

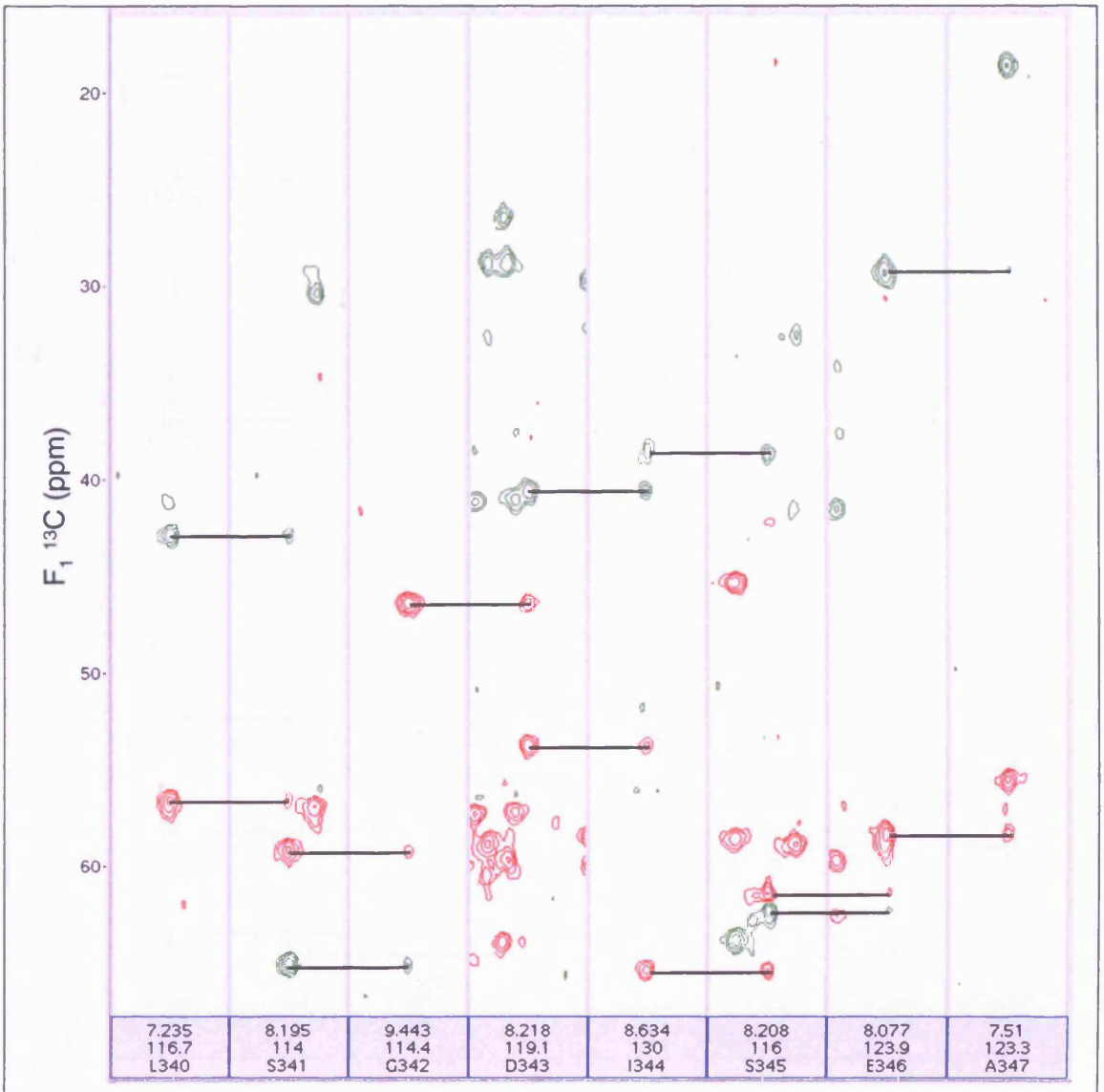


Figure 3.12 ^{15}N strips from the HNCACB spectrum of Pdcd4 MA-3$_C$ illustrating the intra- and inter-residue cross peaks linking residues L340 to A347. Cα signals are shown in red and Cβ in green. Intra-residue connectivities between amide nitrogen/proton and Cα and Cβ show stronger signals than those observed for sequential connectivities. Black lines indicate sequential Cα and Cβ connections. The F3 proton chemical shift, F2 nitrogen chemical shift and residue name and number are shown beneath the strips.

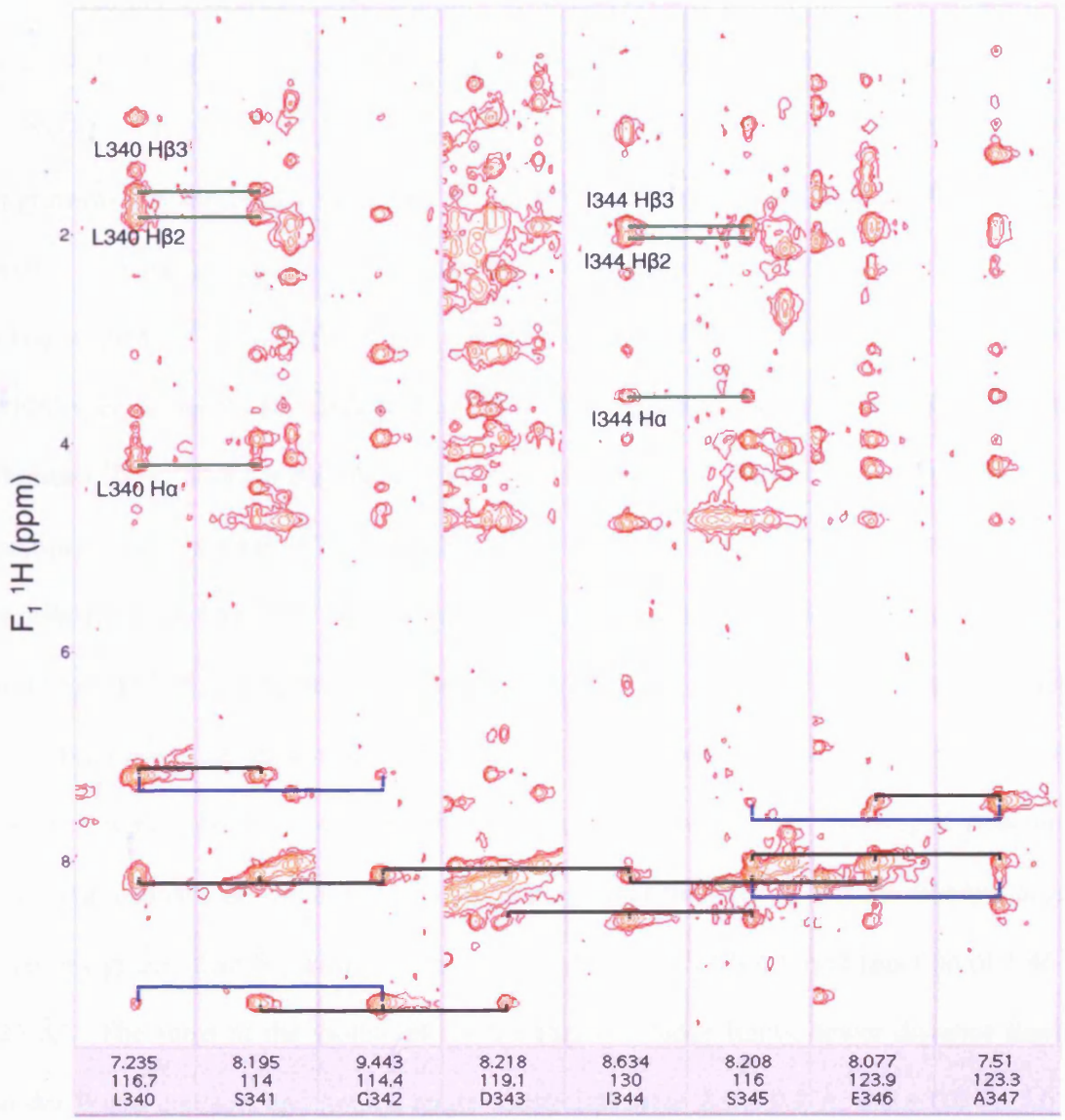


Figure 3.13 ^{15}N strips from the NOESY-HSQC spectrum of Pdcd4 MA-3$_C$ illustrating typical NOE connectivities linking residues L340 to A347. Black lines indicate sequential NH-NH NOEs, blue lines indicate i to i+2 NH-NH NOEs and green lines indicate examples of sequential aliphatic to NH NOEs. The F3 proton chemical shift, F2 nitrogen chemical shift and residue name and number are shown beneath the strips.

3.3.4 Structural Calculations for Pdcd4 MA-3$_C$

The combined automated NOE assignment and structure determination protocol (CANDID) (Herrmann *et al.*, 2002) proved very effective at determining unique assignments for the NOEs identified in the two and three dimensional NOESY -based spectra. Unique assignments were obtained for 95.3% (1283/1346) of the NOE peaks picked in the $^{15}N/^{1}H$ NOESY-HSQC spectra, 94.2% (2353/2497) in the $^{13}C/^{1}H$ NOESY-HMQC spectra and 88.3% (363/411) in the NOESY spectrum, which produced 2545 non-redundant ^{1}H to ^{1}H upper distance limits. The final family of Pdcd4 MA-3$_C$ structures was determined using a total of 2787 NMR-derived structural constraints (an average of 21.3 per residue), including 2545 NOE-based upper distance limits, 174 backbone torsion angle constraints and 68 hydrogen bond constraints in regions of regular helical structure (table 3.3). Following the final round of CYANA calculations, 49 satisfactorily converged structures were obtained from 100 random starting structures. The converged structures contain no distance or van der Waals violation greater than 0.5 Å and no dihedral angle violations greater than 5°, with an average value for the CYANA target function of 1.46 ± 0.25 Å^2. The sums of the violations for the upper distance limits, lower distance limits, van der Waals contacts and torsion angle constraints were 2.9 ± 0.7 Å, 0.0 ± 0.0 Å, 5.6 ± 0.8 Å and 16.2 ± 3.4 ° respectively. Similarly, maximum violations for the converged structures were 0.32 ± 0.02 Å, 0.00 ± 0.01 Å, 0.27 ± 0.02 Å and 3.03 ± 0.67° respectively. The NMR constraints and structural statistics for Pdcd4 MA-3$_C$ are summarised in table 3.3. The family of converged Pdcd4-MA3$_C$ domain structures, together with the NMR constraints, have been deposited in the Protein Data Bank (accession code 2HM8).

Table 3.3 NMR Constraints and Structural Statistics for Pdcd4 MA-3_C

NMR Constraints and Structural Statistics for Pdcd4 MA-3_C		
a) No. of Constraints used in Final Structural Calculation		
Intraresidue NOEs	614	
Sequential NOEs (i, i+1)	562	
Medium-range NOEs (i, $i \leq 4$)	599	
Long-range NOEs (i, $i \geq 5$)	771	
Torsion angles	174	(87Φ and 87Ψ)
Hydrogen bonding	68	
b) Maximum and Total Constraint Violations in 49 Converged Pdcd4 MA-3_C Structures		
Upper distance limits (Å)	0.32 ± 0.02	2.9 ± 0.7
Lower distance limits (Å)	0.00 ± 0.01	0.0 ± 0.0
van der Waals contacts (Å)	0.27 ± 0.02	5.6 ± 0.8
Torsion angle ranges (°)	3.03 ± 0.67	16.2 ± 3.4
Average CYANA target function (Å^2)	1.46 ± 0.25	
c) Structural Statistics for the Family of Converged Pdcd4 MA-3_C Structures		
Residues within the		
favourably allowed		84.6
additionally allowed		13.7
generously allowed		1.4
disallowed		0.3
regions of the Ramachandran plot (%)		
Backbone atom r.m.s.d. for structured region (residues 327-356, 359-412,421-448 of Pdcd4 MA-3_C)		0.49 ± 0.07 Å
Heavy atom r.m.s.d. for structured region (residues 327-356, 359-412,421-448 of Pdcd4 MA-3_C)		0.92 ± 0.07 Å

The solution structure of Pdcd4 MA-3_C is determined to high precision, which is clearly evident from the superposition of the protein backbone shown for the family of converged structures in figure 3.14 panel A (best fit for residues 327-356, 359-412 and 421-448) and is reflected in low root mean squared deviation (r.m.s.d.) values to the mean structure for both the backbone and all heavy atoms of 0.49 ± 0.07 Å and 0.92 ± 0.07 Å respectively.

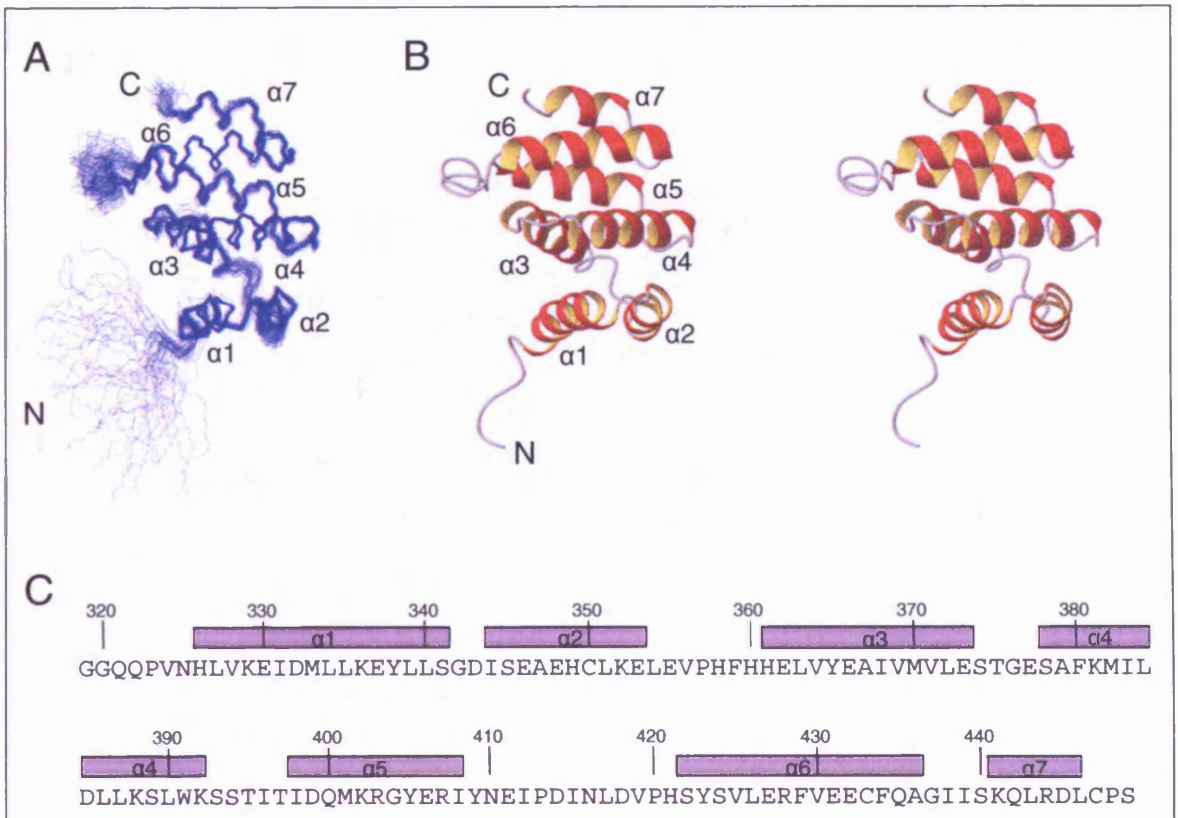


Figure 3.14 Solution structure of the Pdcd4 MA-3$_C$ domain. Panel A shows a best fit superposition of the protein backbone for the family of 49 converged structures obtained. Panel B contains a stereo view of the ribbon representation of the backbone topology of Pdcd4 MA-3$_C$ based on the structure closest to the mean and is shown in the same orientation as in panel A. The seven helical regions of Pdcd4 MA-3$_C$, α1, α2, α3, α4, α5, α6 and α7 are labelled, as are the N and C termini of the domain. For clarity the locations of the seven helical regions are also indicated above the amino acid sequence of Pdcd4 MA-3$_C$ shown in panel C.

3.3.5 Mapping of the eIF4A Binding Site

The positions of signals from backbone amide groups in proteins are highly sensitive to changes in their local environment and have been used to localize the eIF4A binding site on Pdcd4 MA-3$_C$. Typical $^{15}N/^{1}H$ HSQC spectra obtained from samples of ^{15}N-labelled Pdcd4 MA-3$_C$ (100 μM) in the presence and absence of an excess of full length eIF4A (150μM), are shown in figure 3.15 panel A.

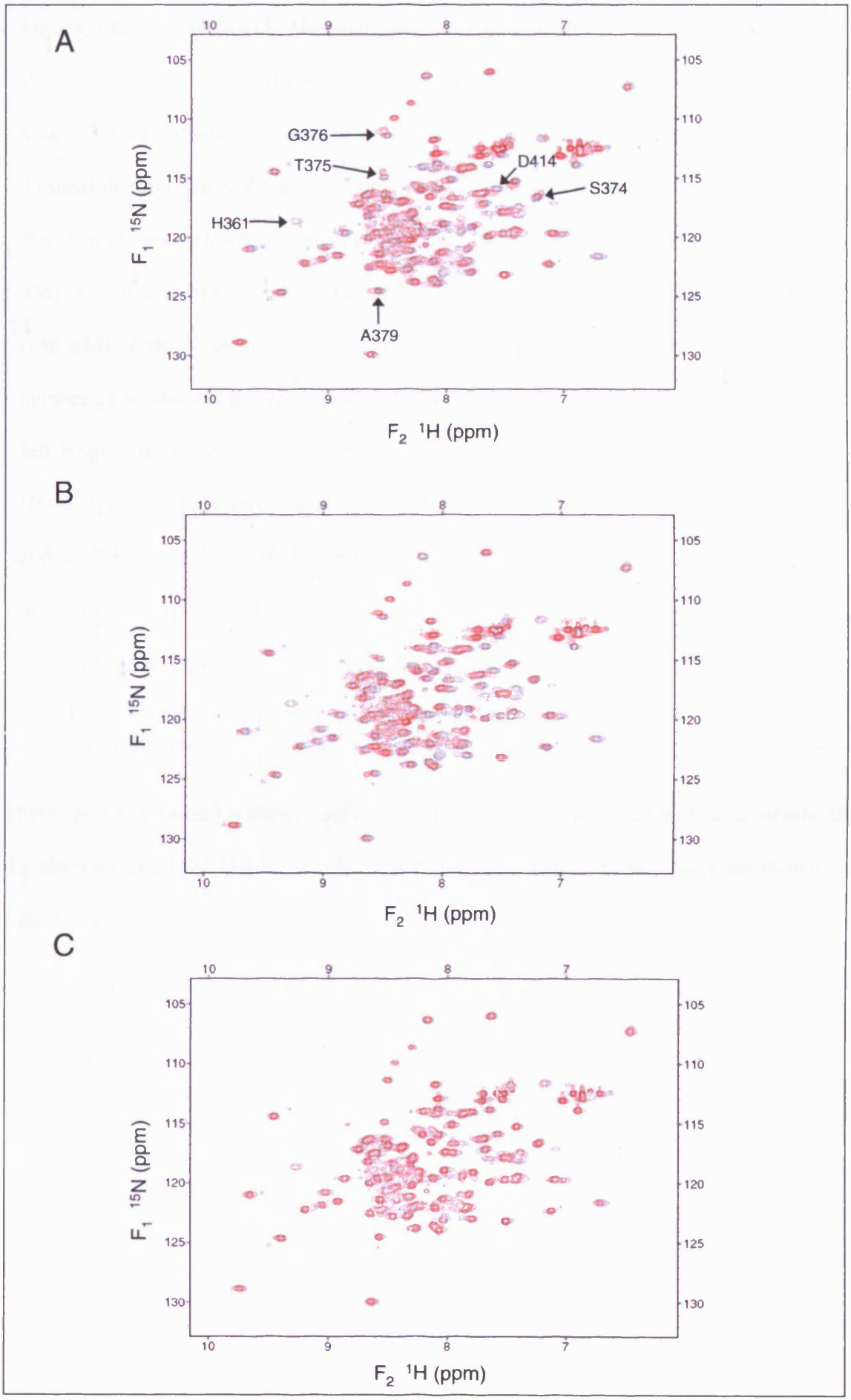

A
B
C
G376
T375
D414
S374
H361
A379
F_1 ^{15}N (ppm)
F_2 ^{1}H (ppm)
10
9
8
7
105
110
115
120
125
130

Figure 3.15 Pdcd4 MA-3$_C$ chemical shift changes induced by eIF4A binding. The above overlays of $^{15}N/^{1}H$ HSQC spectra of uniformly ^{15}N labelled Pdcd4 MA-3$_C$ were acquired in the presence and absence of full length unlabelled eIF4A (panel A) or the N-terminal domain (panel B) and clearly show that a number of peaks have shifted in the presence of either eIF4A protein. The spectrum of the free Pdcd4 MA-3$_C$ domain (100 μM) is shown in black, and for Pdcd4 MA-3$_C$ (100 μM) bound to full length eIF4A (150 μM) or the N-terminal domain of eIF4A (500 μM) in red. Peaks arising from a number of residues that were significantly shifted on formation of the Pdcd4 MA-3$_C$–full length eIF4A complex have been highlighted. Panel C shows an overlay of the HSQC spectra for the free Pdcd4 MA-3$_C$ domain (100 μM), which is shown in black, and for Pdcd4 MA-3$_C$ (100 μM) in the presence of an excess of the C-terminal domain of eIF4A (400 μM), which is shown in red. No observable chemical shift changes were observed in the presence of the C-terminal domain of eIF4A. The results indicate that Pdcd4 MA-3$_C$ interacts solely with the N-terminal domain of eIF4A.

These spectra allowed backbone amide minimal shift values to be determined for nearly all the residues of bound Pdcd4 MA-3$_C$, which are summarised in the histogram shown in figure 3.16 panel A.

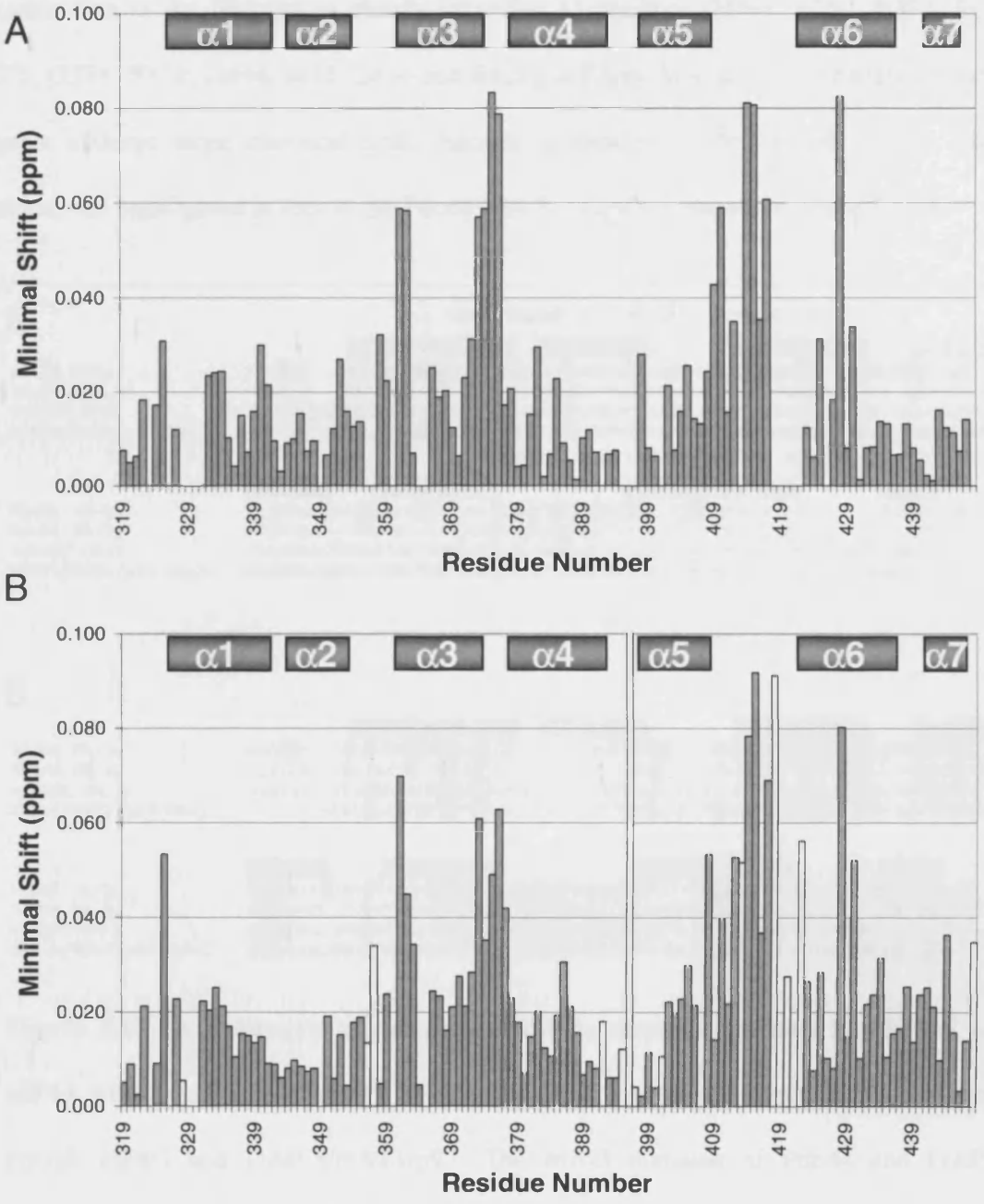


Figure 3.16 Mapping of the eIF4A binding site on Pdcd4 MA-3$_C$. Panels A and B show histograms of the minimal chemical shift changes observed for backbone amide groups of Pdcd4 MA-3$_C$ on binding of full length eIF4A (A) and the isolated N-terminal domain (B). The positions of the helices in Pdcd4 MA-3$_C$ are highlighted above the histograms. In panel B minimal shift data for Cα/Hα, is shown for a few residues where no information was obtained for the backbone amide groups (white bars). The bar for residue I396 has been truncated (minimal shift 0.113 ppm).

Examination of the histograms clearly identifies 11 residues (H361, E362, E373, S374, T375, G376, N410, D414, I415, L416 and R428) in Pdcd4 MA-3_C whose backbone amide signals undergo large chemical shift changes on binding of full length eIF4A. These residues are highlighted in red on the Pdcd4 MA-3_C sequence shown in figure 3.17A.

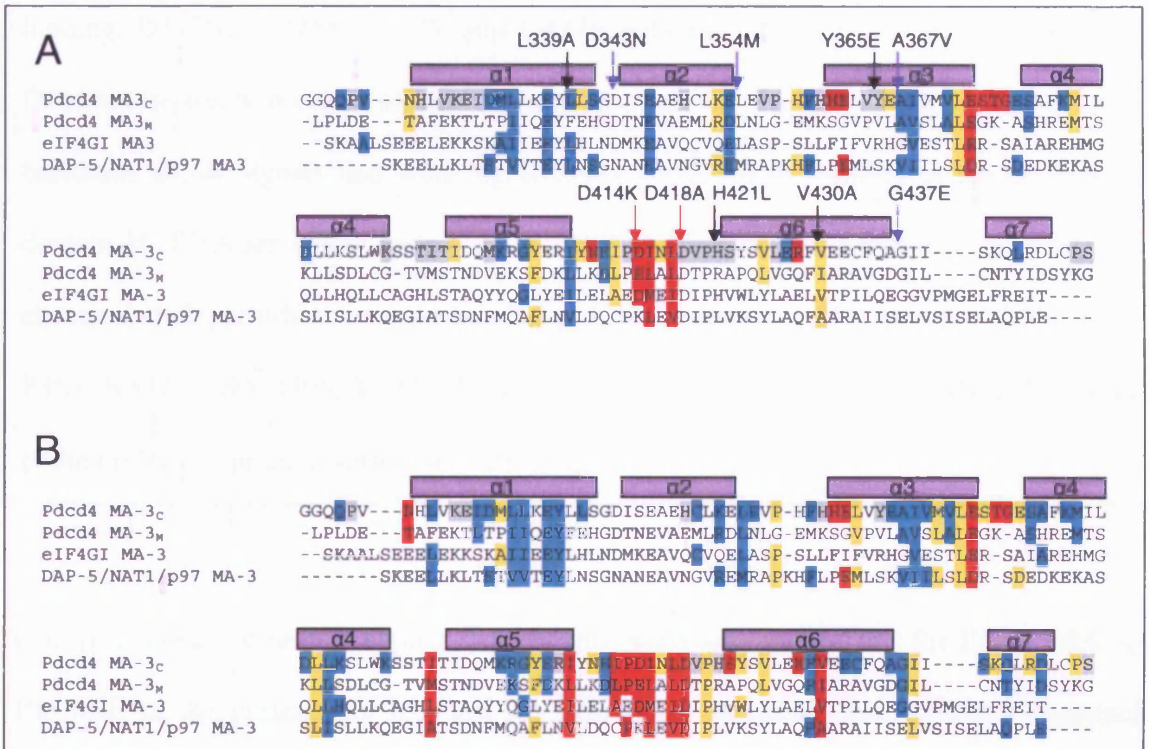


Figure 3.17 A summary of the Pdcd4 MA-3_C domain residues implicated in eIF4A binding. Panel A shows a sequence alignment of the MA-3 domains from Pdcd4, eIF4G and DAP-5/NAT1/p97. The MA-3 domains of Pdcd4 and DAP-5/NAT1/p97 are aligned on the basis of amino acid sequence similarity, whilst eIF4G MA-3 is aligned with Pdcd4 MA-3_C according to the structural based sequence alignment produced by DALI (Holm & Sander, 1993). The positions of the helices in Pdcd4 MA-3_C are shown above the alignment. Residues in Pdcd4 MA-3_C are colour coded to indicate the minimal shifts in backbone amide signals induced by eIF4A binding, with changes of between 0.015 and 0.025 ppm shown in blue, between 0.025 and 0.045 ppm in yellow and by over 0.045 ppm in red. Equivalent conserved

residues are also highlighted in the other MA-3 domains. Residues for which no chemical shift data was obtained are shown in grey. The position of a number of site-directed mutants of Pdcd4 characterised by Yang and co-workers are also indicated (Yang *et al.*, 2004). L339A, Y365E, H421L and V430A were found to have no effect on eIF4A binding. D343N, L354M, A367V and G437E reduced binding by 40-60%. D414K and D418A effectively blocked eIF4A binding. In panel B, residues of Pdcd4 MA-3_C with backbone amide signals that were significantly perturbed by binding of the N-terminal domain of eIF4A are colour coded as described for Panel A. For a few residues where no chemical shift perturbation could be obtained for backbone amide signals (V328, V356, P357, K392, T395, I396, T397, M401, P413, D418, P420, S422 and P448), the colour coding refers to minimal shifts for Cα/Hα signals.

Analogous chemical shift mapping experiments were also carried out for Pdcd4 MA-3_C (100 μM) in the presence of a 5 and 10 fold molar excess of the isolated N-terminal domain of eIF4A, as shown by the HSQC spectra shown in figure 3.15 panel B. Similar changes in the HSQC spectra of Pdcd4 MA-3_C were observed at both N-terminal domain eIF4A concentrations used (data not shown) and are summarised for the 5 fold molar excess by the histogram shown in figure 3.16 panel B. Analysis of the data clearly reveals that the same groups of residues were affected by binding of the isolated N-terminal domain of eIF4A as observed with the full length protein, with the most substantial shifts (>0.45 ppm) seen for residues N325, H361, E362, E373, T375, G376, I408, I412, D414, I415, L417, R428 and V430. These residues are highlighted in red on the Pdcd4 MA-3_C sequence shown in figure 3.17 panel B. The side chain amide group of N416 was also significantly shifted by both full length eIF4A and the isolated N-terminal domain. In contrast, no significant shifts were observed in $^{15}N/^{1}H$ HSQC spectra of Pdcd4 MA-3_C

(100 μM) acquired in the presence of a five fold excess of the isolated C-terminal domain of eIF4A, as shown in the overlay of HSQC spectra shown in figure 3.15 panel C.

We were unable to use $^{15}N/^{1}H$ HSQC spectra to monitor the effects of eIF4A binding to a number of non-proline residues in Pdcd4 MA-3_C (H326, V328, K329, E330, H349, V356, H360, Y365, K392, T395, I396, T397, M401, D418, V419, H421 and S422) since it was not possible to identify the corresponding backbone amide peak in either the free or bound Pdcd4 MA-3_C spectra. To try to establish whether any of these residues were located on the interface with eIF4A, $^{13}C/^{1}H$ HSQC spectra were acquired for Pdcd4 MA-3_C (100 μM) in the presence and absence of the isolated N-terminal domain of eIF4A (400 μM), as shown in figure 3.18. Analysis of the Cα-Hα region of the spectra provided reliable minimal shift information for 13 additional residues of Pdcd4 MA-3_C (V328, V356, P557, K392, T395, I396, T397, M401, P413, D418, P420, S422 and P448). Substantial shifts were observed for four of these residues: I396, P413, D418 and S422, as shown in figures 3.16B, 3.17B and 3.18B. In addition, several signals from methyl groups were perturbed significantly by eIF4A binding, in particular, those from I331, I408 and I412 (figure 3.18A), which also showed significant changes in their backbone amide signals.

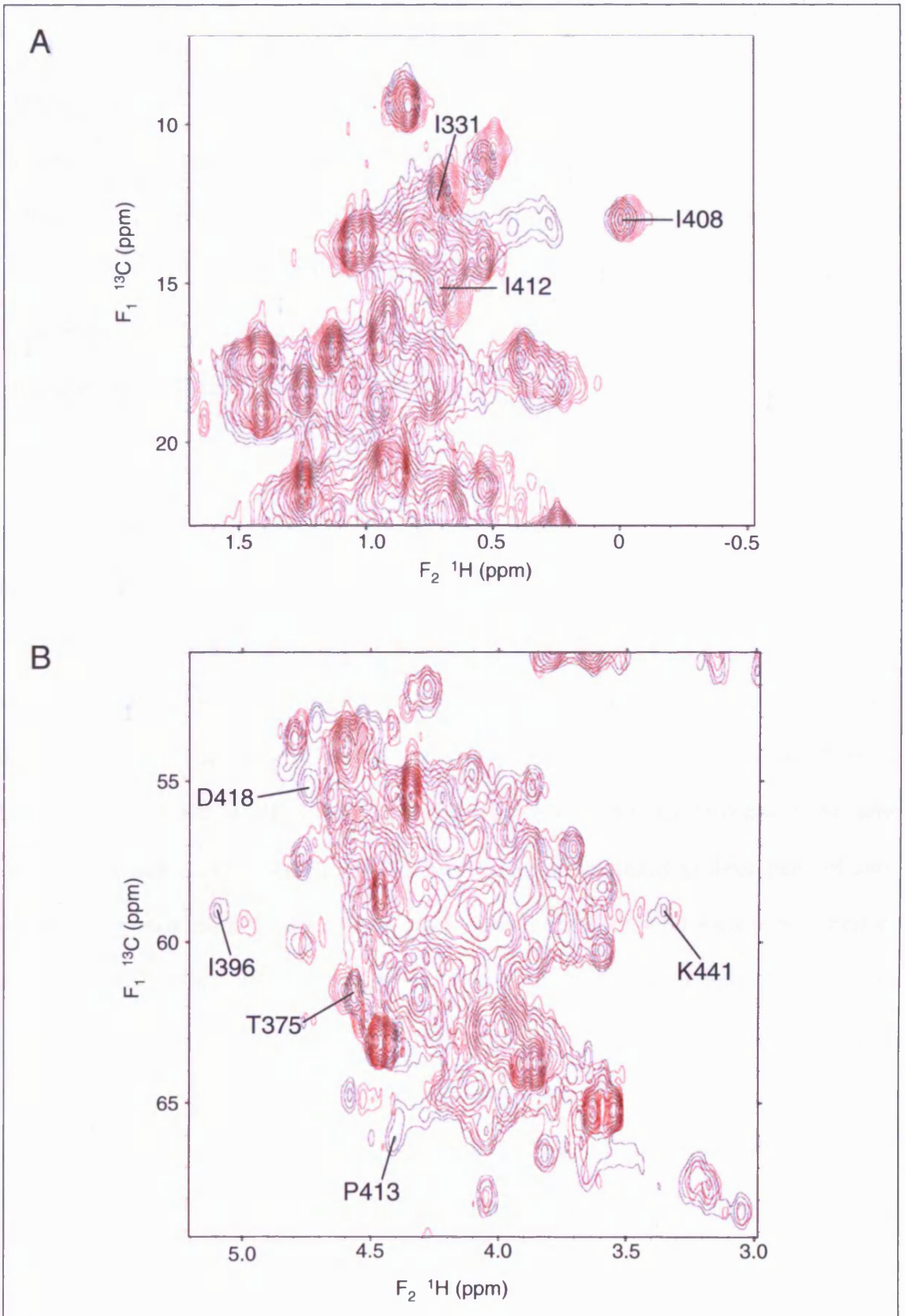


Figure 3.18 Comparison of the $^{13}C/^{1}H$ HSQC spectra of uniformly ^{13}C labelled Pdcd4 MA-3$_C$ acquired in the presence and absence of the isolated N-terminal domain of eIF4A. The above overlay of the methyl (panel A) and Cα/Hα (panel b)

Cα/Hα (panel B) regions of the $^{13}C/^{1}H$ HSQC spectra of uniformly ^{13}C labelled Pdcd4 MA-3_C acquired in the presence and absence of the isolated N-terminal domain of unlabelled eIF4A clearly shows that a number of peaks have shifted in the presence of eIF4A. The spectrum of the free Pdcd4 MA-3_C domain (100 μM) is shown in black, and for Pdcd4 MA-3_C (100 μM) bound to the N-terminal domain of eIF4A (400 μM) in red. Peaks arising from a number of residues that were significantly shifted on formation of the Pdcd4 MA-3_C - eIF4A complex have been highlighted.

3.4 Discussion

3.4.1 Structural Features of the Pdcd4 MA-3_C Domain

The backbone topology of the Pdcd4 MA-3_C domain is illustrated by the ribbon diagram shown in figure 3.14 panel B. In agreement with the secondary structure predictions, the domain is primarily composed of seven α-helices (α1 H326-S341, α2 I344-E353, α3 H361-E373, α4 S378-K392, α5 I398-I408, α6 S422-A436 and α7 K441-L446) linked by turns and loops (figure 3.14C). The six N-terminal helices are arranged as three pairs of anti-parallel helix-turn-helix hairpins, which stack upon each other, with each pair of helices offset by approximately 45° from the previous layer. The C-terminal helix, α7, is located centrally above helices α5 and α6. One notable feature of the protein backbone is the loop between α5 and α6. In contrast to the short well defined turns between α1 and α2 (2 residues) and between α3 and α4 (4 residues), the long loop between α5 and α6 (13 residues) appears to be relatively mobile and exist in multiple conformations. In part, the poor definition of this loop reflects the absence of backbone amide resonance assignments for several residues in this region (D418, V419, P420 and S422), resulting in fewer NMR structural constraints, however, the broad lines observed for many signals are strongly indicative of the presence of conformational heterogeneity (reviewed in Cavanagh *et al.*,

1995). Chemical shift and NOE data indicate that part of the loop (Y409-L417) has a propensity to adopt a helical conformation. This region of Pdcd4 MA-3_C forms part of the interaction site with eIF4A (discussed in section 3.4.5) and binding of eIF4A may stabilise the helical conformation, either extending the C-terminal end of α5 by up to three turns, or forming a separate shorter helix. The absence of Hα to Hβ (*i*, *i*+3) NOEs in the latter half of the α5 to α6 loop (residues D418 to H421), for which no backbone amide assignments are available, is consistent with the lack of helical secondary structure in this region of the loop. The amino acid sequence in this region is fairly well conserved across the different MA-3 domains (figures 3.2 and 3.17), with two aspartate residues (D414 and D418) being conserved in over 65% of MA-3 domains, and may reflect some functional importance of this region.

The stabilising interactions between the helical hairpins and between their constituent helices rely almost entirely on favourable van der Waals contacts. A hydrogen bond between the carbonyl of L390 and the amide proton of I396, which are located in the C-terminal end of α4 and the following linker to α5, helps to stabilise this region of the domain. A second hydrogen bond was also identified between the sidechain hydroxyl group of Y409 and the carbonyl of L443, which contributes to a stable interaction between α5 and α7.

3.4.2 Comparison of the NMR and Crystallography Structures of Pdcd4 MA-3_C

A high resolution crystallography structure of Pdcd4 MA-3_C (320-469) has just been reported (Laronde-Leblanc *et al.*, 2006) (PDB code 2NSZ). Very similar secondary and tertiary structures were observed for Pdcd4 MA-3_C in the solution and crystallography structures, as shown by the ribbon representations in figure 3.19. The similarity is also highlighted by comparisons of the backbone atom coordinates, which yield an r.m.s.d. of

0.98 Å for the superposition of residues V328-E353, E355-S374, G376-I412, H421-L446.

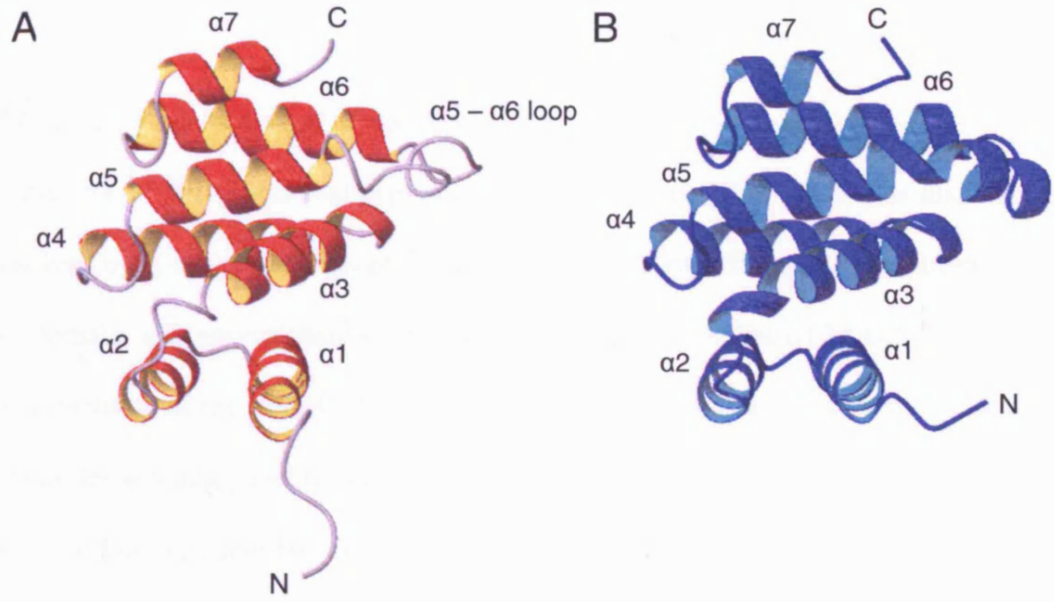


Figure 3.19 Comparison of the backbone topologies of the Pdcd4 MA-3$_C$ domain observed in the solution and crystal structures. Panels A & B show equivalent views of the backbone topology of the Pdcd4 MA-3$_C$ domain, which were obtained by superimposing the backbone atoms of residues V328-E353, E355-S374, G376-I412, H421-L446. The solution structure is shown in red (panel A) and the crystallography structure in blue (panel B). The helical regions, α5 – α6 loop and the N- and C-termini are labelled.

Two minor differences were noted in the secondary structure of the two structures. In the crystal structure one turn of 3_{10} helix is found at the N-terminus of α3 (residues P357-F359), whereas in the solution structure these residues do not form part of a regular helix. Secondly, the C-terminal end of α5 is extended by approximately three turns in the crystal structure (residues Y409-D418, figure 3.19). As mentioned in section 3.4.1, this region of the solution structure forms an ill defined loop, part of which may adopt a helical conformation on eIF4A binding. In agreement with the NMR data, analysis of the temperature factors in this region of the crystal structure suggests that it may be somewhat

mobile or disordered. For example, residues I417-H421 have an average B-factor value of 25.0 $Å^2$ (for Cα atoms), whilst the regions of well ordered secondary structure have an average B-factor value of 15.2 ± 3.8 $Å^2$.

3.4.3 MA-3 Domain Structural Homology

The structure of the C-terminal region of eIF4G (residues 1234-1566) has also recently been solved by X-ray crystallography (Bellsolell *et al.*, 2006). This region contains a single MA-3 domain with approximately 36% sequence homology to Pdcd4 MA-3_C. Comparison of the structures of the Pdcd4 MA-3_C and eIF4G MA-3 domains clearly shows that they adopt similar secondary and tertiary structures, as highlighted by the ribbon representations shown in figure 3.20 and by a backbone atom rmsd of 2.7 Å for the superposition of the two domains (residues Q321-V324, N325-S374, G376-I439, S440-P448 of Pdcd4 MA-3_C and residues S1234-A1237, E1241-R1290, S1291-V1354, L1359-L1367 of eIF4G MA-3). One major difference between the two MA-3 domain structures is the presence of an additional short helix (α5b: residues A1327-H1330 of eIF4G MA-3, equivalent to I412-I415 of Pdcd4 MA-3_C) between helices α5 and α6 of eIF4G MA-3 (figure 3.20B). As previously discussed, the NMR data obtained for the loop between α5 and α6 of Pdcd4 MA-3_C suggests that this region has a propensity to adopt a helical conformation. The very recently reported crystal structure of Pdcd4 MA-3_C also contains an additional helical section in this region (Laronde-Leblanc *et al.*, 2006). As with Pdcd4 MA-3_C, the crystallographic data suggests that this region of eIF4G MA-3 may be somewhat mobile or disordered (Bellsolell *et al.*, 2006; Laronde-Leblanc *et al.*, 2006). The other notable difference between the Pdcd4 and eIF4G MA-3 domains is the orientation of the seventh helix. The MA-3 domain of eIF4G forms part of a larger eIF4A binding region, which is comprised of five anti-parallel helical hairpins or atypical HEAT repeats, as shown in figure 1.6 (Bellsolell *et al.*, 2006). Consequently, in contrast to the final C-terminal helix

of Pdcd4 MA-3_C, which caps the previous layer of helices, α7 of eIF4G MA-3 forms part of a fourth helical hairpin, which is offset from the previous layer. Full length Pdcd4 contains an additional twenty residues beyond the C-terminus of Pdcd4 MA-3_C, however, this region is rich in glycines and polar residues, and is not predicted to adopt a helical conformation (Cuff *et al.*, 1998). In addition, no conformational data could be obtained for this region of Pdcd4 MA-3_C by crystallography, as the region is not visible in the electron density maps used to determine the crystal structure, which suggests that it is highly flexible or disordered (Laronde-Leblanc *et al.*, 2006).

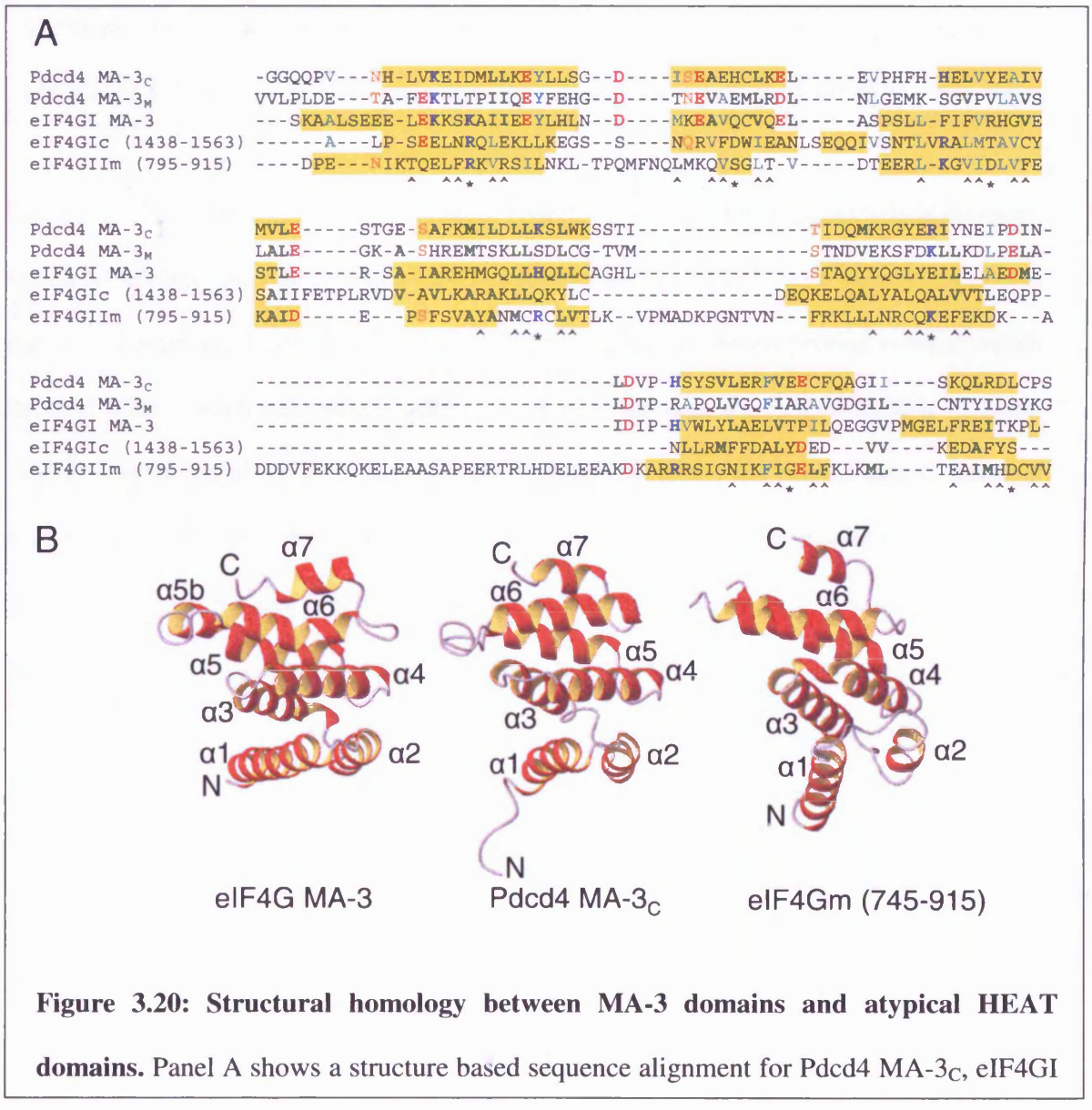


Figure 3.20: Structural homology between MA-3 domains and atypical HEAT domains. Panel A shows a structure based sequence alignment for Pdcd4 MA-3_C, eIF4GI

MA-3 (residues 1234-1367), eIF4GI W2 HEAT domain (residues 1438-1563) and $eIF4GII_M$ HEAT domain (residues 745-915), which was obtained using DALI (Holm & Sander, 1993). Pdcd4 $MA\text{-}3_M$ was aligned to Pdcd4 $MA\text{-}3_C$ on the basis of sequence conservation. Regions of helical secondary structure have been highlighted in yellow. Residues with conserved sequence similarity in three or more of the domains shown are highlighted as follows: hydrophobic (AVILM) in green, aromatic (FYW) in cyan, positively charged (KRH) in blue, negatively charged (DE) in red and polar (STNQ) in orange. The position of the conserved hydrophobic (^) and polar (*) residues found in the atypical HEAT repeats are highlighted below the alignment. Panel B shows equivalent views of ribbon representations of the backbone structure of the homologous regions of eIF4GI MA-3, Pdcd4 $MA\text{-}3_C$ and $eIF4GII_M$. Residues 851-880 of $eIF4GII_M$ are absent

Canonical HEAT repeats consist of tandem repeats of pairs of anti-parallel helical hairpins, which form right-handed, solenoid-type structures, with extended quasi-cylindrical shapes (Groves & Barford, 1999). The repeats contain a highly conserved proline residue within the first helix of each pair, which results in a marked bend in the helix (Andrade & Bork, 1995). Despite the absence of the conserved proline residues, the middle and C-terminal regions of eIF4G, including the MA-3 domain, are considered to be atypical members of the HEAT repeat protein superfamily (Bellsolell *et al.*, 2006; Marcotrigiano *et al.*, 2001). HEAT repeats are generally involved in protein-protein interactions and several atypical HEAT repeats have been shown to play an essential role in the assembly of the protein synthesis machinery (Bellsolell *et al.*, 2006; Marcotrigiano *et al.*, 2001; Mazza *et al.*, 2001; Yamamoto *et al.*, 2005). The amino acid sequences of the helices in Pdcd4 $MA\text{-}3_C$ fit the proposed consensus sequence for atypical HEAT repeats (Bellsolell *et al.*, 2006). The consensus sequence features three clusters of conserved hydrophobic residues in each helix (figure 3.20A), which allow favourable contacts within the HEAT repeat and with adjacent

HEAT repeats. There is also a conserved solvent exposed polar residue, which lies between the second and third hydrophobic clusters.

Comparison of the backbone topology of Pdcd4 MA-3_C with other known folds in the Protein Data Bank using Dali identified almost 600 structural neighbours (Holm & Sander, 1993). Four of the five closest structural homologues were of the type found in eIF4G, which are often referred to as mIF4G domains, and have been found in a number of eukaryotic initiation factors and the nuclear cap binding protein CBP80 (Ponting, 2000). This confirms previous suggestions that despite little sequence homology, MA-3 domains might be distantly related examples of mIF4G domains (Ponting, 2000).

Among the mIF4G domains of known structure, the closest structural homologue to Pdcd4 MA-3_C is the central region of eIF4G (eIF4GII$_M$, residues 745-915 (Marcotrigiano *et al.*, 2001), which is reflected in a backbone RMSD of 3.65 Å (residues 322-326, 327-341, 343, 344-352, 354-374, 378-394, 397-413, 417-420, 421-449 of Pdcd4 MA-3_C and residues 745-749, 751-765, 768, 773-781, 782-802, 804-820, 834-849, 882-885, 886-915 of eIF4GII$_M$) and is clearly apparent from the structural views shown in figure 3.20 panel B. eIF4G$_M$ is involved in binding to eIF3, RNA and eIF4A (Imataka & Sonenberg, 1997; Lamphear *et al.*, 1995; Marcotrigiano *et al.*, 2001; Morino *et al.*, 2000). The interaction between eIF4G$_M$ and the C-terminal domain of eIF4A is believed to stabilise the closed, catalytically active orientation of the two domains of eIF4A (Oberer *et al.*, 2005). The most striking difference between Pdcd4 MA-3_C and eIF4G$_M$ (745-915) is the loop between α5 and α6, which is approximately thirty residues longer than the corresponding loop in Pdcd4 MA-3_C. There is no conformational data available for this extended loop in eIF4G$_M$, as the region is not visible in the electron density maps used to determine the crystal structure (Marcotrigiano *et al.*, 2001). This region may serve a similar function to

the flexible, ill defined loop of Pdcd4 MA-3_C, which is involved in binding partner proteins (see section 3.4.5). Pdcd4 MA-3_C also shows structural homology to the C-terminal W2 domain of $eIF4G_C$, which is involved in binding to Mnk1 (Bellsolell *et al.*, 2006). This is reflected in a backbone RMSD of 3.5 Å (residues 324-354, 355-360, 361-373, 374-392, 397-416, 421-434, 438-446 of Pdcd4 MA-3_C and residues 1438-1468, 1474-1479, 1481-1493, 1499-1517, 1521-1540, 1541-1554, 1555-1563 of eIF4G). A structure based alignment of the sequences from Pdcd4 MA-3_C and the atypical HEAT domains of eIF4G are shown in figure 3.20 panel A.

3.4.4 Relative Orientation of the MA-3 Domains in Pdcd4

The two MA-3 domains of Pdcd4 share 18% sequence identity and 22% sequence similarity, which suggests they will adopt similar overall structures. In addition, secondary structure predictions suggest that Pdcd4 MA-3_M will contain virtually identical secondary structure to Pdcd4 MA-3_C. However, the predicted helices α1, α3 and α6 of Pdcd4 MA-3_M all contain a proline residue, which is not conserved in the Pdcd4 MA-3_C and eIF4G MA-3 domains (figure 3.20). These additional proline residues are likely to disrupt or cause a bend in the helices, as seen in canonical HEAT repeats. As a result some minor differences in the secondary and tertiary structures and/or the stability of Pdcd4 MA-3_M may be observed as compared to Pdcd4 MA-3_C and eIF4G MA-3. The two Pdcd4 MA-3 domains are linked by a stretch of 33 residues, which is predicted by JPRED (Cuff *et al.*, 1998) to contain an additional 16 residue long helix (data not shown). This helix, together with the expected seventh C-terminal helix of Pdcd4 MA-3_M may form a fourth HEAT repeat and thereby a somewhat extended domain. The remaining fifteen residues in the linker region are predicted to form an extended or irregular structure and the presence of six glycine residues implies significant flexibility. This region may therefore allow significant flexibility in the positioning and orientation of the two Pdcd4 MA-3 domains,

perhaps facilitating interactions with both domains of eIF4A, or other partner proteins.

Comparison of the linewidths for backbone amide signals in $^{15}N/^{1}H$ HSQC spectra obtained from the isolated Pdcd4 MA-3_C domain (18.2 ± 1.4 Hz) and the Pdcd4 dual MA-3 domain protein (157-449) (25.4 ± 3.8 Hz), suggest that there is a significant degree of independent motion of the two domains consistent with a flexible linker. Analysis of HNCO spectra for isolated Pdcd4 MA-3_C and Pdcd4 (157-449) identified small but significant shifts in signals from Pdcd4 MA-3_C residues located on the exposed surface of α1 and α2, which suggests some transient interaction between the two MA-3 domains that may be stabilised on binding of eIF4A.

Previous studies have shown that Pdcd4 contains two leucine rich nuclear export signals (NES), which are located within the middle MA-3 domain (Bohm *et al.*, 2003). Given the significant level of sequence conservation between the two Pdcd4 MA-3 domains, it seems very likely that the regions of the middle MA-3 domain corresponding to the NES motifs will adopt the structure as seen in Pdcd4 MA-3_C. NES1 (residues 243-253) is likely to be located in an unstructured loop between Pdcd4 MA-3_M α5 and α6, which suggests it may be functionally active. Whilst NES2 (residues 184-194) will probably form the C-terminal end of α2 and part of the linker to α3. This suggests that the conserved leucine residues, within NES2 will be buried and therefore not available as a functional signal.

3.4.5 Characterisation of eIF4A Binding to Pdcd4

The chemical shift changes induced by eIF4A binding have clearly highlighted a number of residues on Pdcd4 MA-3_C that are located at the interface with eIF4A, most notably E373, S374, T375 and G376, which are positioned at the C-terminal end of α3 and the following turn, and N410, I412, P413, D414, I415, L417 and D418, which are found at the

C-terminal end of α5 and the following ill defined loop (figures 3.16 and 3.17). These residues are arranged in a broad strip on the surface of Pdcd4, which runs over a quarter of the way around the molecule, as illustrated in figure 3.21 panels A and B. A few of the residues with shifted signals (Y405, I408 and S422) are hidden behind the α5-α6 loop and so observed changes here may reflect localised conformational changes in Pdcd4 MA-3_C induced by eIF4A binding. As discussed previously, the N-terminal half of the ill defined loop between α5 and α6 may adopt a helical conformation upon eIF4A binding. This conformational change could expose the hidden residues and form a groove in the surface of Pdcd4 MA-3_C between residues E373-G376 and D414-H421. A small group of residues at the N-terminal end of α3 were also affected by eIF4A binding (figures 3.16, 3.17 and 3.21). These residues are somewhat separated from the main interaction site and it seems probable that the changes arise from a slight change in the orientation of α3 induced by the interaction of eIF4A with residues at the C-terminus of α3.

Very similar patterns of chemical shift changes were observed for Pdcd4 MA-3_C in the presence of both full-length eIF4A and the N-terminal domain alone, which clearly demonstrates that only the N-terminal domain is required for binding to Pdcd4 MA-3_C. In addition, no changes were observed for Pdcd4 MA-3_C in the presence of the isolated C-terminal domain. Previous studies have shown that both the N- and C-terminal domains of eIF4A, and the two MA-3 domains of Pdcd4 are required for the formation of a very high affinity complex (Yang *et al.*, 2004; Zakowicz *et al.*, 2005). As Pdcd4 MA-3_C clearly interacts with the N-terminal domain of eIF4A, this strongly suggests that the C-terminal domain will bind to Pdcd4 MA-3_M, giving a head to tail arrangement to the two proteins in the complex.

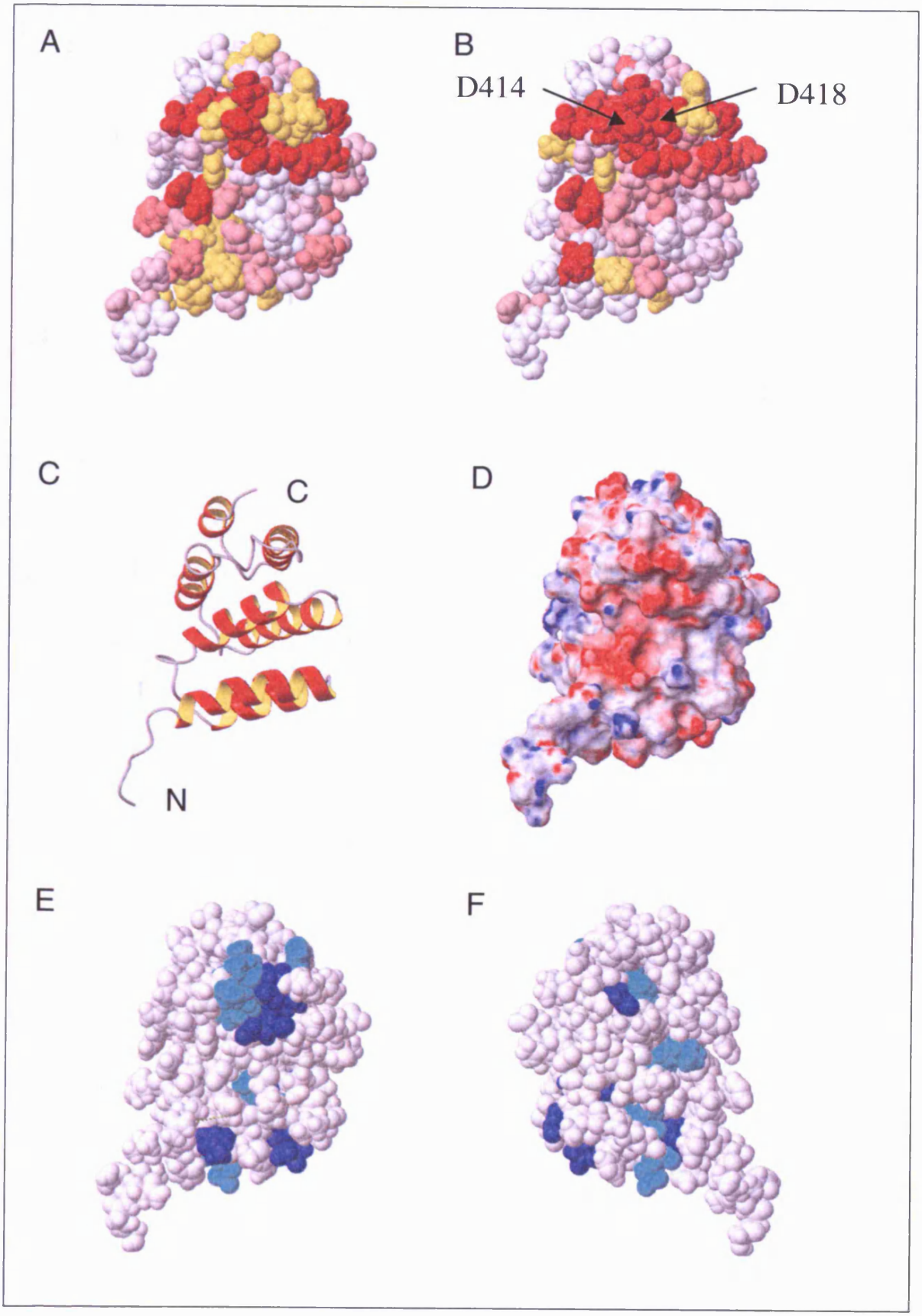

A
B
D414
D418
C
C
N
D
E
F

Figure 3.21 Location, features and properties of the eIF4A-binding site on Pdcd4 MA-3_C. Panel A shows a space filled view of Pdcd4 MA-3_C in which residues are coloured according to the perturbation of their backbone amide signals induced by eIF4A binding. Residues that showed a minimal shift change of less than 0.010 ppm are shown in white, over 0.045 ppm in red, and between 0.010 and 0.045 ppm are coloured according to the magnitude of the shift on a linear gradient between white and red. Residues for which no minimal shift data was obtained are shown in yellow. In panel B the backbone amide chemical shift changes induced by binding of the isolated N-terminal domain are highlighted using the same colour scheme as described above. Where possible, residues for which no backbone amide minimal shift data was available were coloured according to changes of their Cα/Hα signals (V328, V356, P357, K392, T395, I396, T397, M401, P413, D418, P420, S422 and P448) with the same colour scheme used. The position of D414 and D418, which have previously been reported to be involved in the interaction between Pdcd4 MA-3_C and eIF4A, are also highlighted. Panel D shows a surface view of Pdcd4 MA-3_C, which is coloured according to electrostatic potential, with areas of significant negative charge shown in red, significant positive charge in blue and neutral in white. The electrostatic potential was calculated using MOLMOL (Koradi *et al.*, 1996), with the threshold for depicting significant areas of charge chosen to obtain a neutral representation for the aromatic ring of F434. Calculations were done with the assumption that all histidine residues were uncharged. Panels E and F show space filled views of Pdcd4 MA-3_C in which residues are highlighted on the basis of sequence conservation, with equivalent residues identical in the MA-3 domains of Pdcd4 and eIF4G shown in blue and those conserved to a close homologue shown in cyan. Panel F is rotated by 180° about the Y axis from the view shown in panel E. Panel C shows a ribbon representation of the backbone topology of Pdcd4 MA-3_C, in the same orientation as panels A,B,D and E.

Calculation of the electrostatic surface of Pdcd4 MA-3_C revealed that the eIF4A interaction site is predominantly negatively charged, as shown in figure 3.21 panel D, which is consistent with the high number of acidic residues in this region. Interestingly, several of these acidic residues are conserved in the other MA-3 domains known to bind eIF4A, and in the case of eIF4G MA-3, they have been shown to form part of a similar negatively patch on the surface of the domain. The structure of the N-terminal domain of human eIF4A, which is identical in sequence to mouse eIF4A has recently been deposited (PDB code 2G9N). This reveals that the surface of the N-terminal domain of eIF4A is characterised by a number of large basic patches, located in and around the regions involved in ATP and RNA binding, as shown in figure 3.22 (Pause *et al.*, 1993; Pause & Sonenberg, 1992; Svitkin *et al.*, 2001; Tanner *et al.*, 2003). It therefore seems very likely that Pdcd4 MA-3_C will bind to one or more of these basic regions on eIF4A, which is supported by recent mutagenesis data (Zakowicz *et al.*, 2005).

Pdcd4 MA-3_C has recently been shown to directly compete with eIF4G MA-3 for binding to eIF4A, which suggests that both these MA-3 domains bind to the same/similar surface of the N-terminal domain of eIF4A (Laronde-Leblanc *et al.*, 2006). Interestingly, analysis of the amino acid sequence of the MA-3 domains of Pdcd4 and eIF4G suggests that the secondary and tertiary structure of Pdcd4 MA-3_M might differ slightly from the structures that have been solved for Pdcd4 MA-3_C and eIF4G MA-3 (discussed in section 3.4.4). These minor structural differences would be consistant with the hypothesis that Pdcd4 MA-3_M binds to a distinct interaction site on the C-terminal domain of eIF4A. eIF4Gm has previously been shown to bind the C-terminal domain of eIF4A at a surface adjacent to the RNA-, ATP- and eIF4A NTD interaction sites (Oberer *et al.*, 2005). However, it seems unlikely that Pdcd4 MA-3_M binds to the same region, as previous competition assays have shown no evidence for any competition between the domains for binding to eIF4A (Yang

et al., 2003). The surface of the C-terminal domain of eIF4A also contains a number of large positively charged patches, given the sequence conservation of the negatively charged eIF4A binding surface of the MA-3 domains (discussed below), it seems like that Pdcd4 MA-3_M will interact with one of these regions.

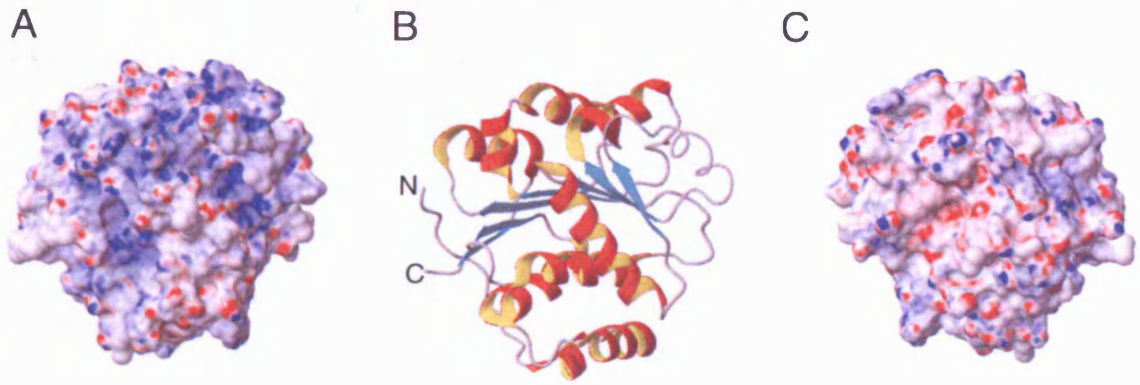


Figure 3.22 Structure and properties of the isolated N-terminal domain of human eIF4A. Panels A and C show a surface view of the N-terminal domain of human eIF4A (PDB code 2G9N), which is coloured according to electrostatic potential, with areas of significant negative charge in red, positive charge in blue and neutral in white. The electrostatic potential was calculated using MOLMOL (Koradi *et al.*, 1996), with the threshold for depicting significant areas of charge chosen to obtain a neutral representation for the aromatic ring of F52. Calculations were done with the assumption that all histidine residues were uncharged. The structure in panel C is rotated by 180° about the Y axis from the view shown in panel A. Panel B shows a ribbon representation of the backbone topology of the N-terminal domain in the same orientation as panel A.

The degree of amino acid sequence conservation for surface exposed residues in the MA-3 domains of Pdcd4 and eIF4G, is illustrated by the space-filled views of Pdcd4 MA-3_C shown in figure 3.21 panels E and F. There is clearly significantly more conservation in

the region shown here to form the interface with eIF4A, which suggests that all three MA-3 domains share a common eIF4A binding site. Interestingly, one of the strictly conserved residues, D414, is substituted by a lysine in the MA-3 domain of the eIF4G homologue DAP-5/NAT1/p97, which is unable to bind eIF4A (Imataka & Sonenberg, 1997; Yang *et al.*, 2004).

Recent mutagenesis studies of Pdcd4 have shown that the substitution of D414 with lysine, or D418 with alanine, results in a dramatic fall in eIF4A binding (Yang *et al.*, 2004). These two aspartate residues are located adjacent to each other on the surface of Pdcd4 MA-3_C and lie at the heart of the eIF4A binding site identified by the NMR studies reported here, as highlighted in figure 3.21B. The equivalent mutations in Pdcd4 MA-3_M and eIF4G MA-3 also result in dramatic falls in eIF4A binding, which strongly supports the suggestion of a conserved eIF4A binding site (Yang *et al.*, 2004). The Pdcd4 D418A variant was shown to retain full $eIF4G_M$ binding activity, which clearly suggests that Pdcd4 interacts with $eIF4G_M$ through a site distinct from that involved in eIF4A binding (Yang *et al.*, 2003).

Similar site-directed mutagenesis studies of the interaction of eIF4G (residues 1080-1600, $eIF4G_C$) with eIF4A, suggested that a cluster formed by R1281 and four surrounding glutamate residues on the MA-3 domain formed part of the probable binding site for eIF4A (Bellsolell *et al.*, 2006). We were not able to obtain chemical shift perturbation data for the equivalent residue of Pdcd4 MA-3_C (Y365), however, it is located within the N-terminal half of α3. This region showed significant shifts in signals on eIF4A binding, which were attributed to localised conformational changes, rather than direct contact with eIF4A. As discussed previously, the eIF4A binding site identified on Pdcd4 by NMR is largely conserved on eIF4G (see figure 3.21E), which strongly implies a common binding site for eIF4A on both MA-3 domains. In contrast, neither R1281, nor the surrounding

glutatmate residues are conserved in Pdcd4 MA-3_C. The study of eIF4G_C also suggested that the loop between helices 5 and 6 of eIF4G MA-3 was not required for eIF4A binding (Bellsolell *et al.*, 2006). The work reported here clearly shows that the equivalent loop in Pdcd4 MA-3_C forms a major part of the interface with eIF4A, which is supported by previous mutagenesis studies of Pdcd4 and eIF4G (Yang *et al.*, 2004; Yang *et al.*, 2003). Given the structural similarity of the two MA-3 domains, the level of sequence conservation in this region and the earlier mutational data, it seems very likely that the α5 - α6 loop is an important element of the eIF4A binding site in both Pdcd4 MA-3_C and eIF4G MA-3.

3.4.6 Conclusions

We have shown that Pdcd4 MA-3_C is composed of 3 layers of helix-turn-helix hairpins capped by a single helix. The structure shows close structural homology to the atypical HEAT repeats found in many eukaryotic initiation factors. The interaction data reported in this chapter unambiguously identifies the binding site for eIF4A on Pdcd4 MA-3_C. This binding site primarily involves residues located on a conserved surface region encompassing the loop connecting α5 and α6, and the turn linking α3 and α4. This site is strongly conserved in other MA-3 domains known to interact with eIF4A, including the preceding domain of Pdcd4, suggesting a common mode of binding. In addition, the N-terminal domain of eIF4A is shown to be sufficient for binding to Pdcd4 MA-3_C, which strongly implies that in complexes formed by the intact proteins the C-terminal domain of eIF4A will bind to Pdcd4 MA-3_M, resulting in an inhibitory complex with a head to tail arrangement of the proteins. Analysis of the sequence conservation and features of NMR spectra obtained from the single domain and tandem MA-3 region, strongly suggests that the tandem MA-3 region is composed of two equivalent domains connected by a somewhat flexible linker.

CHAPTER FOUR

CONCLUSIONS

The regulation of eukaryotic gene expression involves the formation of a large number of multi-subunit protein-protein, protein-DNA and protein-RNA complexes, for example the transcription and 43S pre-initiation complexes and enhanceosome complexes. In the case of the enhanceosome different combinations of regulators are recruited depending on the promoter and cell type (reviewed in Perissi *et al.* 2005; Rosenfeld *et al.* 2006). Central to many of these complexes are scaffold proteins, such as the transcriptional coactivators CBP and p300, and the translation initiation factor eIF4G, which function to recruit specific combinations of cofactors that are required to correctly regulate gene expression.

Many proteins involved in the regulation of eukaryotic gene expression have been shown to contain large (>50 residues) unstructured/disordered regions, which lack stable secondary or tertiary structure (Dunker *et al.* 2002; Uversky 2002). These regions are characterised by a low content of bulky hydrophobic residues and a high proportion of polar and charged residues. The length, composition and in some cases the sequences of these regions are often highly conserved between species and homologous proteins, suggesting that they may be functionally important. Many of these regions have been shown to fold upon binding to their biological targets, or else form linker regions between functional domains (reviewed in Wright *et al.* 1999; Dyson *et al.* 2005). These flexible linkers are necessary for the correct assembly of large complexes. For example, several of the interaction domains of CBP/p300 are separated by long linkers (~150-400 residues) that allow the domains to simultaneously form interactions with several transcriptional regulators that are bound over large promoter/enhancer regions (reviewed in Dyson et al. 2005). The nuclear cap binding protein CBP80 is composed of three atypical HEAT

domains that are linked by extended linkers (Mazza *et al.* 2001). However, unlike the independent domains of CBP/p300 that are generally believed to behave like 'beads on a string', the CBP80 domains interact to form a compact structure. The two linker regions of CBP80 also interact extensively with the domains, contributing to the stability of the protein. The scaffold protein eIF4G is also believed to form a similar compact structure, although the role and positioning of the significantly longer linker regions is unknown (Marintchev *et al.* 2004).

As previously mentioned, many of these unstructured/disordered proteins and domains undergo coupled folding and binding reactions. For example, the activation domains of the transcription factors c-Jun and p53 and the N-terminal region of the viral protein Tat fold upon binding to the KIX domain of CBP/p300, whilst the CBP SID and p160 AD1 domains undergo mutual synergistic folding (Radhakrishnan *et al.* 1997; Campbell *et al.* 2002; Demarest *et al.* 2002; Vendel *et al.* 2003). Coupled folding and binding has an entropic cost that needs to be countered by a favourable enthalpic contribution (reviewed in Wright et al. 1999; Dyson et al. 2005). In the case of the interaction between the KID domain of CREB and the KIX domain of CBP this enthalpic compensation comes largely from the interactions formed with the phosphate group bound to S133 of CREB KID (Radhakrishnan et al. 1997).

The unstructured/disordered nature of these proteins/domains has several potential advantages, including the fact that it provides potential mechanisms by which the proteins can be easily regulated. Coupled folding and binding of unstructured/disordered proteins to their targets is often regulated by posttranslational modifications, for example, only the phosphorylated form of the KID domain of CREB is able to bind the CBP KIX domain (Radhakrishnan et al. 1997; Dyson et al. 2005). In addition, it has been suggested that the

targeted degradation of these relative unstable proteins by the ubiquitin-proteasome pathway could allow their cellular levels to tightly regulated (Salghetti *et al.* 2001; Dyson et al. 2005).

Some of the proteins/domains that undergo coupled folding and binding contain disproportionately large hydrophobic interfaces that they are unable to bury in the absence of binding partners. For example, the SRC1 AD1 domain folds upon binding to CBP SID to form an 'L' shaped helical domain. In the absence of CBP the interaction surface would be solvent exposed. Similar results were observed for the AD1 domain of ACTR (Demarest *et al.* 2002). In addition the SRC1 and ACTR AD1 domains contain very few/no long range intramolecular interactions, instead their tertiary structures are stabilised by interactions with CBP SID. The molten globule CBP SID domain is also unable to sufficiently bury its hydrophobic rich motifs in the absence of a binding partner (Demarest et al. 2002; Demarest *et al.* 2004). The extensive interdomain interfaces present in these domains allow for the formation of complexes with high specificity, while the coupled folding and binding produces complexes with relatively low affinity (reviewed in Dyson et al. 2005). This mechanism for the formation of reversible complexes may be an important feature in the regulation of gene expression. Interestingly, NMR data obtained for the CBP SID / SRC1 AD1 complex suggested that the SRC1 AD1 domain has very limited conformational stability, which may facilitate the formation of a transient complex. In the absence of coupled folding and binding these proteins/domains would have to be several times larger in order to stabilise their tertiary structure in isolation.

A final advantage of coupled folding and binding is the ability of some unstructured/disordered proteins to interact with several unrelated proteins. For example, as discussed in chapter 2, the SID domain of CBP folds upon binding into at least two

distinct conformations (Demarest et al. 2002; Qin *et al.* 2005; Waters *et al.* 2006). The first conformation was observed in the mutually folded CBP SID / p160 AD1 domain complexes, whilst the second conformation was observed in the complex formed with the natively folded IRF-3 IAD domain. This structural plasticity may explain the ability of the CBP SID domain to interact with a number of diverse proteins that contain very little or no sequence homology. Whether the tertiary structure of the SID domain is restricted to just these two reported folds remains to be answered. These observations highlight the importance of detailed structural analysis of multiple complexes containing proteins/domains such as CBP SID that have the potential to form multiple conformations upon binding to different unrelated proteins. The functional significance of the different conformations observed for the CBP SID domain may extend beyond its ability to interact with these unrelated families of transcriptional regulators. The different CBP SID conformations formed upon binding could potentially influence the subsequent binding of additional transcriptional regulators and as a result the transcriptional outcome.

Many folded proteins/domains undergo localised folding upon binding, for example as observed for eIF4E upon binding to eIF4G (Gross *et al.* 2003). As reported in chapter three of this thesis, the unstructured 13 residue loop between α5 and α6 of Pdcd4 MA-3_C is believed to fold into a helical conformation upon binding to the N-terminal domain of eIF4A. The additional mobility of this loop may facilitate the recruitment of eIF4A or other partner proteins. In addition, short linker regions (~15 residues) have also been identified between the MA-3 domains of Pdcd4, and N- and C-terminal domains of eIF4A (Caruthers *et al.* 2000; Waters *et al.* 2007). These linkers allow significant flexibility in the positioning and orientation of the domains, facilitating interactions with partner proteins, and in the case of eIF4A with RNA and ATP.

As previously mentioned the enhanceosome complexes selectively recruit different combinations of regulators depending on the context. It is possible that the ability of the domains of scaffold proteins to form different conformations upon binding to distinct transcriptional regulators could result in the selective ordered recruitment of cofactors to the enhanceosome complex. In order to understand the molecular basis for the combinatorial specificity observed in the enhanceosome we need to study large higher order complexes as a whole and not just focus on the individual interactions. Structural studies of such complexes would need to involve a combination of high resolution structural data obtained by X ray crystallography and NMR, and low resolution data obtained by techniques such as electron microscopy. For example, as used to develop the model of the transcription pre-initiation complex shown in figure 1.3 (Boeger *et al.* 2005). Similarly, detailed structural analysis of other large multisubunit complexes, such as the 43S translation pre-initiation complex would provide important insights into the recruitment of transcriptional/translational regulators to these complexes, as well as increasing our understanding of the complexes function. The work described in this thesis represents a significant step towards the ultimate goal of obtaining detailed structural information for the large dynamic complexes involved in the regulation of eukaryotic gene expression.

APPENDICES

A.1 Culture Media

LB Broth

Per litre, 10 g tryptone, 5 g yeast extract and 10 g NaCl. Make up to volume with deionised H_2O and autoclave. When cool add the desired antibiotic.

Minimal Media

Minimal media allowing the introduction of selective nitrogen or carbon isotopes into expressed proteins was prepared as follows;

Per litre culture,

$(NH_4)_2SO_4$ (1)	1.0 g	$MnCl_2$	16.0 mg
Na_2HPO_4	6.8 g	$FeCl_3$	5.0 mg
KH_2PO_4	3.0 g	$ZnCl_2$	0.5 mg
NaCl	0.5 g	$CuCl_2$	0.1 mg
Na_2SO_4	0.3 mM	$CoCl_2$	0.1 mg
EDTA	50.0 mg	H_3BO_3	0.1 mg

were dissolved in dH_2O and autoclaved. When cool the following were added aseptically:

$MgSO_4$	1.0 mM	Thiamine	1.0 mg
$CaCl_2$	0.3 mM	Glucose (2)	4.0 g
d-Biotin	1.0 mg	Antibiotic	

(1) For ^{15}N labelled and ^{15}N ^{13}C labelled samples the ammonium sulphate was replaced with 0.6 g/l $(^{15}NH_4)_2SO_4$ (^{15}N 99% - ISOTEC, Inc.).

(2) For ^{13}C labelled and ^{15}N ^{13}C labelled samples the glucose was replaced with 2 g/l D-glucose U-$^{13}C_6$ (^{13}C 99% +, Cambridge Isotope Laboratories, Inc.).

A.2 TALOS Based Torsion Angle Constraints used for Structural Calculations

A.2.1 CBP SID Torsion Angle Constraints

Residue		Angle	Constraint Range	
2064	SER	PHI	-118	-50
2064	SER	PSI	90	182
2066	SER	PHI	-100	-20
2066	SER	PSI	-65	-13
2067	ALA	PHI	-90	-30
2067	ALA	PSI	-90	-20
2068	LEU	PHI	-90	-30
2068	LEU	PSI	-56	-16
2069	GLN	PHI	-75	-55
2069	GLN	PSI	-53	-29
2070	ASP	PHI	-77	-49
2070	ASP	PSI	-50	-30
2071	LEU	PHI	-82	-54
2071	LEU	PSI	-90	-25
2072	LEU	PHI	-72	-52
2072	LEU	PSI	-54	-26
2073	ARG	PHI	-75	-55
2073	ARG	PSI	-55	-19
2074	THR	PHI	-84	-52
2074	THR	PSI	-57	-25
2075	LEU	PHI	-89	-49
2075	LEU	PSI	-53	-17
2076	LYS	PHI	-109	-53
2076	LYS	PSI	-62	22
2082	GLN	PHI	-75	-55
2082	GLN	PSI	-90	-20
2083	GLN	PHI	-90	-30
2083	GLN	PSI	-60	-15
2084	GLN	PHI	-90	-30
2084	GLN	PSI	-56	-28
2085	GLN	PHI	-75	-55
2085	GLN	PSI	-51	-23
2086	GLN	PHI	-83	-59
2086	GLN	PSI	-52	-24
2087	VAL	PHI	-76	-56
2087	VAL	PSI	-55	-35
2088	LEU	PHI	-68	-48
2088	LEU	PSI	-52	-32
2089	ASN	PHI	-74	-54
2089	ASN	PSI	-90	-30
2090	ILE	PHI	-78	-54
2090	ILE	PSI	-90	-30
2091	LEU	PHI	-90	-30

2091	LEU	PSI	-54	-30
2092	LYS	PHI	-90	-30
2092	LYS	PSI	-51	-27
2096	GLN	PHI	-84	-48
2096	GLN	PSI	-62	-18
2097	LEU	PHI	-84	-52
2097	LEU	PSI	-54	-22
2098	MET	PHI	-71	-51
2098	MET	PSI	-55	-27
2099	ALA	PHI	-70	-50
2099	ALA	PSI	-59	-27
2100	ALA	PHI	-80	-56
2100	ALA	PSI	-60	-15
2101	PHE	PHI	-75	-55
2101	PHE	PSI	-64	-24
2102	ILE	PHI	-90	-30
2102	ILE	PSI	-55	-31
2103	LYS	PHI	-74	-54
2103	LYS	PSI	-61	-29
2104	GLN	PHI	-72	-52
2104	GLN	PSI	-68	-12
2105	ARG	PHI	-97	-47
2105	ARG	PSI	-75	24
2106	THR	PHI	-103	-39
2106	THR	PSI	-54	-12
2107	ALA	PHI	-78	-53
2107	ALA	PSI	-62	-5
2108	LYS	PHI	-106	-39
2108	LYS	PSI	-73	14
2109	TYR	PHI	-108	-40
2109	TYR	PSI	-88	36
2110	VAL	PHI	-112	-36
2110	VAL	PSI	-69	3
2111	ALA	PHI	-107	-51
2111	ALA	PSI	-74	42
2113	GLN	PHI	-126	-46
2113	GLN	PSI	80	192

A.2.2 SRC1 AD1 Torsion Angle Constraints

Residue	Angle	Constraint Range	
929 GLU	PHI	-90	-30
929 GLU	PSI	-90	-20
930 LYS	PHI	-84	-48
930 LYS	PSI	-61	-17
931 ALA	PHI	-77	-57
931 ALA	PSI	-57	-17
932 LEU	PHI	-90	-30
932 LEU	PSI	-60	-20
933 LEU	PHI	-77	-49
933 LEU	PSI	-54	-14
934 GLU	PHI	-90	-30
934 GLU	PSI	-48	-28
935 GLN	PHI	-77	-57
935 GLN	PSI	-53	-29
936 LEU	PHI	-84	-56
936 LEU	PSI	-58	-22
937 VAL	PHI	-74	-50
937 VAL	PSI	-54	-34
938 SER	PHI	-72	-52
938 SER	PSI	-53	-29
939 PHE	PHI	-74	-54
939 PHE	PSI	-57	-33
940 LEU	PHI	-72	-52
940 LEU	PSI	-51	-31
941 SER	PHI	-85	-53
941 SER	PSI	-54	-18
943 LYS	PHI	-154	-42
943 LYS	PSI	56	196
945 GLU	PHI	-70	-46
945 GLU	PSI	-56	-12
946 THR	PHI	-79	-51
946 THR	PSI	-56	-24
947 GLU	PHI	-79	-55
947 GLU	PSI	-90	-5
948 LEU	PHI	-90	-30
948 LEU	PSI	-55	-15
949 ALA	PHI	-79	-51
950 GLU	PHI	-90	-30
950 GLU	PSI	-56	-24
951 LEU	PHI	-81	-53
951 LEU	PSI	-52	-32
952 ASP	PHI	-79	-55
952 ASP	PSI	-90	-16
953 ARG	PHI	-90	-30
953 ARG	PSI	-64	-12
954 ALA	PHI	-90	-30
954 ALA	PSI	-62	-2

955	LEU	PHI	-110	-74
955	LEU	PSI	-30	26
956	GLY	PHI	64	104
956	GLY	PSI	-90	33
957	ILE	PSI	-70	20
958	ASP	PHI	-94	-46
958	ASP	PSI	-72	24
959	LYS	PHI	-87	-47
959	LYS	PSI	-66	-6
960	LEU	PHI	-100	-40
960	LEU	PSI	-63	-7
961	VAL	PHI	-113	-45
961	VAL	PSI	-90	-5
962	GLN	PHI	-107	-30
962	GLN	PSI	-45	19

A.2.3 Pdcd4 MA-3$_C$ Torsion Angle Constraints

Residue		Angle	Constraint Range	
326	HIS	PHI	-78	-46
326	HIS	PSI	-53	-29
327	LEU	PHI	-75	-55
327	LEU	PSI	-49	-28
328	VAL	PHI	-90	-46
328	VAL	PSI	-54	-34
329	LYS	PHI	-72	-50
329	LYS	PSI	-48	-28
330	GLU	PHI	-81	-50
330	GLU	PSI	-55	-26
331	ILE	PHI	-76	-56
331	ILE	PSI	-53	-29
332	ASP	PHI	-71	-50
332	ASP	PSI	-53	-33
333	MET	PHI	-76	-55
333	MET	PSI	-53	-30
334	LEU	PHI	-79	-53
334	LEU	PSI	-56	-24
335	LEU	PHI	-81	-54
335	LEU	PSI	-51	-29
336	LYS	PHI	-80	-50
336	LYS	PSI	-58	-15
337	GLU	PHI	-79	-58
337	GLU	PSI	-54	-24
338	TYR	PHI	-76	-52
338	TYR	PSI	-52	-32
339	LEU	PHI	-82	-51
339	LEU	PSI	-58	-18
340	LEU	PHI	-87	-59
340	LEU	PSI	-53	-5
344	ILE	PHI	-71	-47
344	ILE	PSI	-64	-20
345	SER	PHI	-76	-48
345	SER	PSI	-54	-26
346	GLU	PHI	-73	-53
346	GLU	PSI	-63	-23
347	ALA	PHI	-77	-52
347	ALA	PSI	-55	-25
348	GLU	PHI	-75	-55
348	GLU	PSI	-50	-30
349	HIS	PHI	-73	-53
349	HIS	PSI	-56	-32
350	CYS	PHI	-72	-52
350	CYS	PSI	-55	-35
351	LEU	PHI	-74	-54
351	LEU	PSI	-54	-29
352	LYS	PHI	-77	-48

352	LYS	PSI	-51	-26
353	GLU	PHI	-82	-50
353	GLU	PSI	-58	-14
361	HIS	PHI	-77	-49
361	HIS	PSI	-51	-31
362	GLU	PHI	-75	-55
362	GLU	PSI	-57	-20
363	LEU	PHI	-77	-57
363	LEU	PSI	-52	-23
364	VAL	PHI	-74	-53
364	VAL	PSI	-56	-36
365	TYR	PHI	-77	-48
365	TYR	PSI	-61	-28
366	GLU	PHI	-71	-51
366	GLU	PSI	-52	-29
367	ALA	PHI	-75	-55
367	ALA	PSI	-58	-22
368	ILE	PHI	-74	-54
368	ILE	PSI	-53	-33
369	VAL	PHI	-75	-54
369	VAL	PSI	-54	-34
370	MET	PHI	-71	-51
370	MET	PSI	-54	-21
371	VAL	PHI	-74	-54
371	VAL	PSI	-53	-33
372	LEU	PHI	-76	-45
372	LEU	PSI	-60	-23
373	GLU	PHI	-91	-51
373	GLU	PSI	-63	-11
378	SER	PHI	-73	-49
378	SER	PSI	-56	-16
379	ALA	PHI	-84	-56
379	ALA	PSI	-53	-29
380	PHE	PHI	-74	-54
380	PHE	PSI	-51	-30
381	LYS	PHI	-75	-48
381	LYS	PSI	-55	-30
382	MET	PHI	-72	-52
382	MET	PSI	-61	-21
383	ILE	PHI	-77	-55
383	ILE	PSI	-52	-27
384	LEU	PHI	-72	-52
384	LEU	PSI	-56	-33
385	ASP	PHI	-74	-54
385	ASP	PSI	-47	-27
386	LEU	PHI	-74	-54
386	LEU	PSI	-55	-35
387	LEU	PHI	-72	-52
387	LEU	PSI	-56	-31
388	LYS	PHI	-71	-51

388	LYS	PSI	-50	-30
389	SER	PHI	-73	-53
389	SER	PSI	-57	-32
390	LEU	PHI	-72	-47
390	LEU	PSI	-60	-25
391	TRP	PHI	-72	-51
391	TRP	PSI	-54	-32
392	LYS	PHI	-83	-47
392	LYS	PSI	-52	-4
393	SER	PHI	-127	-55
393	SER	PSI	-17	11
398	ILE	PHI	-70	-50
398	ILE	PSI	-54	-22
399	ASP	PHI	-78	-54
399	ASP	PSI	-50	-26
400	GLN	PHI	-76	-56
400	GLN	PSI	-57	-31
401	MET	PHI	-73	-53
401	MET	PSI	-52	-32
402	LYS	PHI	-72	-52
402	LYS	PSI	-54	-32
403	ARG	PHI	-72	-52
403	ARG	PSI	-57	-29
404	GLY	PHI	-77	-54
404	GLY	PSI	-63	-25
405	TYR	PHI	-76	-49
405	TYR	PSI	-56	-36
406	GLU	PHI	-68	-48
406	GLU	PSI	-53	-33
407	ARG	PHI	-75	-55
407	ARG	PSI	-58	-26
408	ILE	PHI	-97	-37
408	ILE	PSI	-62	10
414	ASP	PHI	-103	-39
414	ASP	PSI	-73	11
415	ILE	PHI	-100	-44
415	ILE	PSI	-59	-23
422	SER	PHI	-72	-52
422	SER	PSI	-54	-22
423	TYR	PHI	-73	-53
423	TYR	PSI	-57	-36
424	SER	PHI	-75	-55
424	SER	PSI	-49	-26
425	VAL	PHI	-80	-54
425	VAL	PSI	-53	-33
426	LEU	PHI	-71	-51
426	LEU	PSI	-52	-25
427	GLU	PHI	-75	-55
427	GLU	PSI	-50	-29
428	ARG	PHI	-76	-56

428	ARG	PSI	-51	-31
429	PHE	PHI	-75	-55
429	PHE	PSI	-55	-35
430	VAL	PHI	-73	-53
430	VAL	PSI	-56	-32
431	GLU	PHI	-77	-52
431	GLU	PSI	-57	-15
432	GLU	PHI	-84	-53
432	GLU	PSI	-57	-28
433	CYS	PHI	-74	-54
433	CYS	PSI	-50	-29
434	PHE	PHI	-76	-56
434	PHE	PSI	-53	-33
435	GLN	PHI	-82	-49
435	GLN	PSI	-50	-25
436	ALA	PHI	-125	-57
436	ALA	PSI	-29	27
441	LYS	PHI	-67	-47
441	LYS	PSI	-58	-22
442	GLN	PHI	-72	-49
442	GLN	PSI	-57	-21
443	LEU	PHI	-75	-55
443	LEU	PSI	-52	-27
444	ARG	PHI	-83	-51
444	ARG	PSI	-55	-17
445	ASP	PHI	-81	-49
445	ASP	PSI	-53	-25

A.3 Chemical Shift Index Data

A.3.1 Sequence Specific Assignments Obtained for the CBP SID Domain

2059	PRO	CA	62.4
		CB	34.6
		CG	24.6
		CD	50.2
		HA	4.72
		HB2	2.43
		HB3	2.17
		HG2	1.96
		HG3	1.86
		HD2	3.59
		HD3	3.56
2060	ASN	CA	53.5
		CB	39.0
		HA	4.76
		HB2	2.88
		HB3	2.80
		HD21	7.60
		HD22	6.94
		ND2	112.7
2061	ARG	CA	55.9
		CB	30.7
		CG	27.2
		CD	43.3
		HA	4.43
		HB2	1.92
		HB3	1.78
		HG2	1.66
		HG3	1.65
		HD2	3.22
		HD3	3.22
2062	SER	CA	58.3
		CB	64.2
		H	8.40
		HA	4.55
		HB2	3.89
		HB3	3.86
		N	117.6
2063	ILE	CA	60.9
		CB	39.3
		CG1	28.0
		CG2	18.0
		CD1	13.5
		H	8.46
		HA	4.08
		HB	1.75
		HG12	1.59
		HG13	0.95
		HG2	0.87
		HD1	0.82
		N	123.5
2064	SER	CA	55.2
		CB	63.4
		H	8.46
		HA	4.80
		HB2	4.07
2064	SER	HB3	3.89
		N	122.6
2065	PRO	CA	65.0
		CB	32.0
		CG	27.7
		CD	51.0
		HA	4.29
		HB2	2.34
		HB3	2.02
		HG2	2.21
		HG3	2.02
		HD2	4.11
		HD3	3.96
2066	SER	CA	60.2
		CB	62.4
		H	8.08
		HA	4.26
		HB2	3.93
		HB3	3.85
		N	113.7
2067	ALA	CA	54.8
		CB	18.6
		H	7.83
		HA	4.24
		HB	1.42
		N	126.3
2068	LEU	CA	58.3
		CB	41.5
		CD1	25.9
		CD2	25.9
		H	7.86
		HA	3.93
		HB2	1.77
		HB3	1.61
		HD1	0.90
		HD2	0.86
		N	118.3
2069	GLN	CA	59.0
		CB	28.0
		CG	33.9
		H	8.16
		HA	3.93
		HB2	2.18
		HB3	2.12
		HG2	2.43
		HG3	2.43
		N	118.1
2070	ASP	CA	57.2
		CB	40.1
		H	8.27
		HA	4.41
		HB2	2.84
		HB3	2.60
		N	121.2

2071	LEU	CA	58.2
		CB	41.1
		CD1	24.3
		H	8.60
		HA	3.94
		HD1	0.92
		N	123.5
2072	LEU	CA	58.4
		CB	41.1
		CG	27.1
		CD1	25.3
		CD2	22.9
		H	8.42
		HA	3.64
		HB2	1.76
		HB3	1.27
		HG	1.63
		HD1	0.77
		HD2	0.58
		N	119.3
2073	ARG	CA	59.3
		CB	30.2
		CG	27.7
		CD	43.0
		H	8.05
		HA	3.90
		HB2	1.93
		HB3	1.93
		HG2	1.73
		HG3	1.58
		HD2	3.25
		HD3	3.23
		N	117.6
2074	THR	CA	66.5
		CB	68.4
		CG2	23.2
		H	7.99
		HA	3.94
		HB	4.16
		HG2	1.24
		N	116.5
2075	LEU	CA	57.1
		CB	42.1
		CG	26.9
		CD1	26.2
		CD2	26.4
		H	8.55
		HA	3.94
		HB2	1.78
		HB3	1.39
		HG	1.77
		HD1	0.82
		HD2	0.79
		N	122.6
2076	LYS	CA	57.0
		CB	32.7
		CG	25.9
		CD	29.3
		CE	41.9
		H	7.39
		HA	4.31
		HB2	2.01
		HB3	1.76
		HG2	1.51
2076	LYS	HG3	1.43
		HD2	1.62
		HD3	1.61
		HE2	2.90
		HE3	2.79
		N	116.5
2077	SER	CA	58.0
		CB	62.7
		H	7.41
		HA	4.74
		HB2	4.10
		HB3	4.02
		N	117.9
2078	PRO	CA	63.0
		CB	32.0
		CG	27.7
		CD	50.5
		HA	4.36
		HB2	2.36
		HB3	1.83
		HG2	2.11
		HG3	2.10
		HD2	3.91
		HD3	3.63
2079	SER	CA	59.0
		CB	63.1
		H	8.45
		HA	4.30
		HB2	3.90
		HB3	3.84
		N	117.4
2080	SER	CA	56.9
		CB	63.0
		H	7.84
		HA	4.89
		HB2	4.24
		HB3	4.12
		N	119.8
2081	PRO	CA	65.5
		CB	31.8
		CG	28.0
		CD	50.5
		HA	4.39
		HB2	2.46
		HB3	1.95
		HG2	2.21
		HG3	2.06
		HD2	3.92
		HD3	3.92
2082	GLN	CA	59.0
		CB	27.7
		CG	34.2
		H	8.24
		HA	4.11
		HB2	2.11
		HB3	2.02
		HG2	2.48
		HG3	2.41
		N	117.2
2083	GLN	CA	59.1
		CB	27.1
		H	8.05

2083	GLN	N	124.4
2084	GLN	CA	59.3
		CB	28.0
		CG	34.3
		H	8.33
		HA	3.73
		HB2	2.28
		HB3	1.95
		HG2	2.59
		HG3	2.27
		N	119.5
2085	GLN	CA	58.3
		CB	28.0
		CG	33.9
		H	7.89
		HA	3.93
		HB2	2.17
		HB3	2.17
		HG2	2.44
		HG3	2.43
		N	117.9
2086	GLN	CA	59.1
		CB	28.3
		CG	33.9
		H	7.97
		HA	4.13
		HB2	2.30
		HB3	2.11
		HG2	2.51
		HG3	2.41
		N	119.8
2087	VAL	CA	67.4
		CB	31.2
		CG1	23.4
		CG2	21.5
		H	8.14
		HA	3.31
		HB	2.24
		HG1	0.96
		HG2	0.66
		N	120.0
2088	LEU	CA	58.4
		CB	41.1
		CD1	24.5
		CD2	24.3
		H	8.00
		HA	3.76
		HB2	1.75
		HB3	1.59
		HG	1.68
		HD1	0.85
		HD2	0.83
		N	119.0
2089	ASN	CA	56.1
		CB	37.9
		H	8.34
		HA	4.37
		HB2	2.93
		HB3	2.82
		HD21	7.55
		HD22	6.96
		N	117.2
		ND2	111.5
2090	ILE	CA	65.5
		CB	38.2
		CG1	28.8
		CG2	18.0
		CD1	14.0
		H	8.03
		HA	3.69
		HB	1.86
		HG12	1.03
		HG13	1.02
		HG2	0.86
		HD1	0.73
		N	122.3
2091	LEU	CA	58.4
		CB	40.9
		CG	25.9
		CD1	22.7
		CD2	25.9
		H	8.27
		HA	3.89
		HB2	1.87
		HB3	1.34
		HG	1.78
		HD1	0.55
		HD2	0.25
		N	120.0
2092	LYS	CA	59.0
		CB	32.3
		CG	25.9
		CD	29.3
		CE	41.8
		H	8.67
		HA	3.94
		HB2	1.87
		HB3	1.87
		HG2	1.61
		HG3	1.40
		HD2	1.64
		HD3	1.65
		HE2	2.90
		HE3	2.90
		N	116.5
2093	SER	CA	59.4
		CB	64.2
		H	7.49
		HA	4.50
		HB2	4.00
		HB3	4.00
		N	111.8
2094	ASN	CA	51.5
		CB	41.4
		H	7.36
		HA	5.33
		HB2	2.79
		HB3	2.59
		N	118.8
2095	PRO	CA	65.8
		CB	32.3
		CG	27.5
		CD	50.5
		HA	4.41
		HB2	2.42
		HB3	2.00
		HG2	2.16

2095	PRO	HG3	2.07
		HD2	3.92
		HD3	3.62
2096	GLN	CA	59.0
		CB	28.3
		CG	34.2
		H	8.97
		HA	4.21
		HB2	2.11
		HB3	2.05
		HG2	2.47
		HG3	2.42
		N	118.3
2097	LEU	CA	56.9
		CB	41.7
		CG	26.9
		CD2	24.8
		H	7.85
		HA	4.21
		HB2	1.77
		HG	1.66
		HD1	0.87
		HD2	0.81
		N	121.9
2098	MET	H	8.21
		N	119.8
2099	ALA	CA	55.2
		CB	18.1
		H	8.17
		HA	4.04
		HB	1.51
		N	120.0
2100	ALA	CA	54.9
		CB	18.1
		H	7.55
		HA	4.13
		HB	1.52
		N	120.9
2101	PHE	CA	62.3
		CB	40.2
		H	8.58
		HA	3.87
		HB2	3.44
		HB3	2.98
		HD1	6.98
		HD2	6.98
		HE1	7.20
		HE2	7.20
		N	120.5
2102	ILE	CA	65.8
		CB	37.9
		CG1	29.9
		CG2	17.5
		CD1	13.8
		H	8.80
		HA	3.46
		HB	1.96
		HG12	1.13
		HG13	1.13
		HG2	0.85
		HD1	0.91
		N	118.8
2103	LYS	CA	59.9
		CB	32.2
		CG	25.1
		CD	29.3
		CE	41.9
		H	8.14
		HA	3.93
		HB2	1.94
		HB3	1.87
		HG2	1.52
		HG3	1.38
		HD2	1.68
		HD3	1.66
		HE2	2.94
		HE3	2.94
		N	121.6
2104	GLN	CA	58.3
		CB	27.1
		CG	33.9
		H	7.66
		HA	4.04
		HB2	2.11
		HB3	1.88
		HG2	2.46
		HG3	2.22
		N	117.6
2105	ARG	CA	56.4
		CB	28.8
		CG	25.1
		CD	41.9
		H	8.14
		HA	4.30
		HB2	1.66
		HB3	1.67
		HG2	1.49
		HG3	1.42
		HD2	2.97
		HD3	2.97
		N	117.4
2106	THR	CA	64.4
		CB	69.3
		CG2	21.8
		H	8.18
		HA	4.11
		HB	4.31
		HG2	1.22
		N	110.6
2107	ALA	CA	54.5
		CB	18.0
		H	7.46
		HA	4.16
		HB	1.46
		N	124.2
2108	LYS	CA	56.9
		CB	31.8
		CG	24.3
		CD	28.5
		CE	41.8
		H	7.73
		HA	4.11
		HB2	1.72
		HB3	1.72
		HG2	1.26
		HG3	1.20

2108	LYS	HD2	1.55
		HD3	1.55
		HE2	2.90
		HE3	2.90
		N	117.4
2109	TYR	CA	59.3
		CB	39.0
		H	7.84
		HA	4.40
		HB2	3.12
		HB3	2.94
		HD1	7.11
		HD2	7.11
		HE1	6.77
		HE2	6.77
		N	119.8
2110	VAL	CA	63.0
		CB	32.5
		CG1	21.3
		CG2	21.3
		H	7.88
		HA	3.96
		HB	2.10
		HG1	0.99
		HG2	0.94
		N	120.2
2111	ALA	CA	52.9
		CB	18.9
		H	8.02
		HA	4.25
		HB	1.41
		N	125.6
2112	ASN	CA	52.9
		CB	39.0
		H	8.09
		HA	4.69
		HB2	2.82
		HB3	2.72
		HD21	7.63
		HD22	6.90
		N	117.2
		ND2	113.0
2113	GLN	CA	53.7
		CB	28.8
		CG	33.4
		H	8.06
		HA	4.58
		HB2	2.09
		HB3	1.95
		HG2	2.36
		HG3	2.36
		N	121.6
2114	PRO	CA	63.6
		CB	32.0
		CG	27.5
		CD	50.5
		HA	4.41
		HB2	2.30
		HB3	1.94
		HG2	2.06
		HG3	2.01
		HD2	3.78
		HD3	3.66
2115	GLY	CA	45.1
		H	8.53
		HA2	3.96
		HA3	3.96
		N	109.7
2116	MET	H	8.08
		HB2	2.13
		HB3	2.02
		HG2	2.62
		HG3	2.51
		N	120.2

A.3.2 Sequence Specific Assignments Obtained for the SRC1 AD1 Domain

920	PRO	CA	63.0
		CB	32.3
		CG	27.1
		CD	49.6
		HA	4.53
		HB2	2.31
		HB3	1.98
		HG2	2.04
		HG3	2.03
		HD2	3.66
		HD3	3.66
921	THR	CA	61.8
		CB	69.8
		CG2	21.5
		H	8.37
		HA	4.42
		HB	4.25
		HG2	1.23
		N	114.6
922	THR	CA	61.5
		CB	70.1
		CG2	21.5
		H	8.14
		HA	4.42
		HB	4.24
		HG2	1.20
		N	117.2
923	VAL	CA	62.6
		CB	32.6
		CG1	21.0
		CG2	21.0
		H	8.27
		HA	4.11
		HB	2.07
		HG1	0.93
		HG2	0.93
		N	123.0
924	GLU	CA	56.9
		CB	30.2
		CG	36.3
		H	8.54
		HA	4.27
		HB2	2.05
		HB3	1.97
		HG2	2.27
		HG3	2.27
		N	125.1
925	GLY	CA	45.3
		H	8.47
		HA2	3.96
		HA3	3.96
		N	110.6
926	ARG	CA	56.2
		CB	30.7
		CG	26.9
		CD	43.3
		H	8.19
		HA	4.34
		HB2	1.89
926	ARG	HB3	1.78
		HG2	1.60
		HG3	1.60
		HD2	3.20
		HD3	3.20
		N	120.9
927	ASN	CA	53.5
		CB	38.7
		HA	4.72
		HB2	2.85
		HB3	2.78
		HD21	7.69
		HD22	7.01
		ND2	113.4
928	ASP	CA	54.8
		CB	40.9
		H	8.26
		HA	4.58
		HB2	2.74
		HB3	2.70
		N	121.4
929	GLU	CA	59.6
		CB	29.6
		CG	36.9
		H	8.53
		HA	3.94
		HB2	2.08
		HB3	2.09
		HG2	2.29
		HG3	2.25
		N	121.9
930	LYS	CA	59.3
		CB	32.3
		CG	25.1
		CD	29.0
		CE	41.9
		H	8.06
		HA	4.01
		HB2	1.88
		HB3	1.88
		HG2	1.52
		HG3	1.41
		HD2	1.69
		HD3	1.70
		HE2	3.00
		HE3	3.00
		N	119.5
931	ALA	CA	54.5
		CB	18.4
		H	7.93
		HA	4.21
		HB	1.49
		N	122.1
932	LEU	CA	57.4
		CB	41.9
		CG	27.1
		CD1	25.9
		CD2	25.5
		H	8.02
		HA	4.06

932	LEU	HB2	1.86
		HB3	1.46
		HG	1.75
		HD1	0.87
		HD2	0.85
		N	119.8
933	LEU	CA	58.0
		CB	41.4
		CG	26.9
		CD1	24.8
		CD2	24.3
		H	8.22
		HA	4.01
		HB2	1.83
		HB3	1.66
		HG	1.75
		HD1	0.90
		HD2	0.86
		N	120.0
934	GLU	CA	59.3
		CB	29.0
		CG	36.5
		H	8.03
		HA	3.99
		HB2	2.20
		HB3	2.08
		HG2	2.46
		HG3	2.26
		N	118.8
935	GLN	CA	59.1
		CB	28.5
		CG	33.9
		H	7.96
		HA	4.13
		HB2	2.15
		HB3	2.11
		HG2	2.51
		HG3	2.41
		N	119.5
936	LEU	CA	58.3
		CB	41.4
		H	8.19
		HA	4.15
		HB2	2.15
		HB3	1.76
		HD1	0.93
		HD2	0.87
		N	121.9
937	VAL	CA	67.4
		CB	31.5
		CG1	22.9
		CG2	21.3
		H	8.38
		HA	3.49
		HB	2.16
		HG1	1.04
		HG2	0.93
		N	119.5
938	SER	CA	61.5
		CB	62.7
		H	8.32
		HA	4.21
		HB2	4.00
938	SER	HB3	3.98
		N	116.5
939	PHE	CA	61.2
		CB	39.3
		H	8.19
		HA	4.26
		HB2	3.38
		HB3	3.23
		HD1	7.20
		HD2	7.20
		HE1	7.26
		HE2	7.26
		N	122.8
940	LEU	CA	57.1
		CB	41.8
		CG	26.9
		CD1	22.9
		CD2	25.5
		H	8.45
		HA	3.81
		HB2	1.88
		HB3	1.39
		HG	2.03
		HD1	0.86
		HD2	0.66
		N	118.6
941	SER	CA	60.7
		CB	62.8
		H	8.09
		HA	4.22
		HB2	3.99
		HB3	3.95
		N	113.9
942	GLY	CA	44.8
		H	7.47
		N	107.8
943	LYS	CA	54.3
		CB	33.4
		CG	24.5
		CD	28.5
		CE	41.9
		H	7.13
		HA	4.36
		HB2	1.74
		HB3	1.68
		HG2	1.30
		HG3	1.26
		HD2	1.52
		HD3	1.48
		HE2	2.77
		HE3	2.68
		N	121.2
944	ASP	CA	54.0
		CB	42.1
		H	8.73
		HA	4.54
		HB2	2.84
		HB3	2.71
		N	123.3
945	GLU	CA	59.3
		CB	29.6
		CG	36.2

945	GLU	H	8.84
		HA	4.03
		HB2	2.08
		HB3	2.08
		HG2	2.33
		HG3	2.33
		N	123.3
946	THR	CA	65.5
		CB	68.4
		CG2	21.8
		H	8.45
		HA	4.10
		HB	4.28
		HG2	1.30
		N	114.6
947	GLU	CA	58.6
		CB	30.2
		CG	36.8
		H	7.91
		HA	4.22
		HB2	2.27
		HB3	2.08
		HG2	2.46
		HG3	2.26
		N	122.3
948	LEU	CA	55.0
		CB	40.6
		H	8.26
		N	121.2
949	ALA	CA	54.9
		CB	18.1
		H	7.98
		HA	4.25
		HB	1.56
		N	122.1
950	GLU	CA	59.0
		CB	28.7
		CG	36.3
		H	7.94
		HA	4.12
		HB2	2.16
		HB3	2.09
		HG2	2.42
		HG3	2.32
		N	117.6
951	LEU	CA	58.0
		CB	41.4
		CG	27.1
		CD1	24.8
		CD2	24.8
		H	7.77
		HA	4.14
		HB2	2.02
		HB3	1.78
		HG	1.84
		HD1	0.97
		HD2	0.90
		N	123.3
952	ASP	CA	57.3
		CB	39.9
		H	8.71
		N	120.2
953	ARG	CA	58.3
		CB	29.9
		CG	27.7
		CD	43.4
		H	7.86
		HA	4.16
		HB2	1.91
		HB3	1.91
		HG2	1.59
		HG3	1.59
		HD2	3.26
		HD3	3.23
		N	119.0
954	ALA	CA	54.6
		CB	18.6
		H	8.08
		HA	4.16
		HB	1.54
		N	124.2
955	LEU	CA	54.9
		CB	42.7
		CG	26.9
		CD1	23.4
		CD2	26.2
		H	8.06
		HA	4.26
		HB2	1.78
		HB3	1.58
		HG	1.79
		HD1	0.78
		HD2	0.72
		N	115.5
956	GLY	CA	45.9
		H	7.80
		HA2	4.03
		HA3	4.03
		N	107.6
957	ILE	CA	62.3
		CB	38.2
		CG1	28.0
		CG2	17.3
		CD1	13.5
		H	8.19
		HA	3.96
		HB	1.92
		HG12	1.50
		HG13	1.50
		HG2	0.76
		HD1	0.94
		N	117.2
958	ASP	CA	56.1
		CB	39.0
		H	8.69
		HA	3.98
		HB2	2.57
		HB3	2.47
		N	119.3
959	LYS	CA	57.1
		CB	32.3
		CG	24.8
		CD	28.8
		CE	41.9
		H	7.61

959	LYS	HA	4.11
		HB2	1.85
		HB3	1.80
		HG2	1.46
		HG3	1.38
		HD2	1.66
		HD3	1.64
		HE2	2.99
		HE3	2.99
		N	118.8
960	LEU	CA	56.7
		CB	42.2
		CD1	24.9
		CD2	23.7
		H	7.63
		HA	4.15
		HB2	1.76
		HB3	1.57
		HD1	0.88
		HD2	0.83
		N	120.9
961	VAL	CA	63.7
		CB	32.1
		CG1	20.9
		CG2	21.3
		H	7.79
		HA	3.82
		HB	2.01
		HG1	0.83
		HG2	0.80
		N	117.4
962	GLN	CA	56.7
		CB	28.8
		CG	33.7
		H	7.93
		HA	4.16
		HB2	2.13
		HB3	2.01
		HG2	2.41
		HG3	2.35
		N	120.2
963	GLY	CA	45.4
		H	8.09
		HA2	3.99
		N	108.5
964	GLY	CA	45.1
		H	8.15
		HA2	3.99
		N	108.7
965	GLY	CA	45.2
		H	8.34
		HA2	3.96
		N	109.0
966	LEU	CA	54.9
		CB	42.1
		H	8.11
		HA	4.35
		HB2	1.62
		HD1	0.92
		HD2	0.87
		N	121.2
967	ASP	CA	54.5
		CB	40.9
		H	8.36
		HA	4.58
		HB2	2.73
		HB3	2.64
		N	120.9
968	VAL	CA	63.7
		CB	32.1
		CG1	21.0
		CG2	21.3
		H	7.86
		HA	3.95
		HB	2.14
		HG1	0.96
		HG2	0.93
		N	119.3
969	LEU	CA	55.9
		CB	41.9
		CD1	25.1
		CD2	23.7
		H	8.10
		HA	4.22
		HB2	1.65
		HD1	0.90
		HD2	0.84
		N	122.6
970	SER	CA	59.4
		CB	63
		H	8.07
		HA	4.29
		HB2	3.92
		HB3	3.84
		N	115.3

A.3.3 Sequence Specific Assignments Obtained for the Pdcd4 MA-3$_C$ Domain

315	PRO	CA	63.2
		CB	32.3
		CG	27.1
		CD	49.7
		HA	4.47
		HB2	2.31
		HB3	1.94
		HG2	2.00
		HG3	2.00
		HD2	3.57
		HD3	3.57
316	LEU	CA	55.4
		CB	42.3
		CG	27.1
		CD1	24.8
		CD2	23.4
		H	8.56
		HA	4.35
		HB2	1.66
		HB3	1.59
		HG	1.65
		HD1	0.92
		HD2	0.87
		N	122.4
317	GLY	CA	45.4
		H	8.43
		HA2	4.01
		HA3	4.01
		N	109.9
318	SER	CA	58.6
		CB	63.9
		H	8.33
		HA	4.46
		HB2	3.91
		HB3	3.86
		N	115.9
319	GLY	CA	45.6
		H	8.57
		HA2	4.00
		HA3	4.00
		N	111.1
320	GLY	CA	45.3
		H	8.29
		HA2	3.97
		HA3	3.97
		N	108.7
321	GLN	CA	55.7
		CB	29.4
		CG	33.7
		H	8.27
		HA	4.34
		HB2	2.10
		HB3	1.96
		HG2	2.34
		HE21	7.55
		HE22	6.85
		HG3	2.34
		N	119.8
		NE2	112.6
322	GLN	CA	53.8
		CB	28.7
		CG	33.4
		H	8.45
		HA	4.60
		HB2	2.07
		HB3	1.91
		HG2	2.38
		HG3	2.38
		HE21	7.52
		HE22	6.86
		N	122.8
		NE2	112.5
323	PRO	CA	63.0
		CB	32.0
		CG	27.4
		CD	50.6
		HA	4.45
		HB2	2.26
		HB3	1.90
		HG2	1.99
		HG3	1.99
		HD2	3.78
		HD3	3.63
324	VAL	CA	62.8
		CB	32.5
		CG1	20.6
		CG2	21.0
		H	8.28
		HA	3.97
		HB	2.00
		HG1	0.92
		HG2	0.87
		N	120.2
325	ASN	CA	53.7
		CB	38.6
		H	8.41
		HA	4.60
		HB2	2.79
		HB3	2.79
		HD21	7.69
		HD22	6.73
		N	121.4
		ND2	112.5
326	HIS	CA	57.8
		CB	29.9
		HA	4.43
		HB2	3.14
		HB3	3.14
		HD2	7.05
		HE1	7.93
327	LEU	CA	56.6
		CB	42.0
		CG	26.8
		CD1	24.8
		CD2	23.7
		H	7.83
		HA	4.09
		HB2	1.52
		HB3	1.32

327	LEU	HG	1.31
		HD1	0.73
		HD2	0.72
		N	120.9
328	VAL	CA	65.2
		CB	31.4
		CG1	21.1
		CG2	21.9
		H	7.89
		HA	3.69
		HB	2.01
		HG1	0.87
		HG2	0.83
		N	119.6
329	LYS	CA	59.4
		CB	31.9
		CG	24.7
		CD	29.0
		HA	4.04
		HB2	1.89
		HB3	1.89
		HG2	1.50
		HG3	1.41
		HD2	1.65
		HD3	1.65
		HE2	2.97
		HE3	2.97
330	GLU	CA	59.8
		CB	29.1
		CG	36.3
		HA	4.07
		HB2	2.05
		HB3	2.05
		HG2	2.35
		HG3	2.27
331	ILE	CA	64.2
		CB	36.8
		CG1	29.6
		CG2	17.5
		CD1	11.9
		H	7.96
		HA	3.59
		HB	2.08
		HG12	1.57
		HG13	1.01
		HG2	0.76
		HD1	0.67
		N	121.6
332	ASP	CA	58.0
		CB	41.6
		H	8.33
		HA	4.33
		HB2	2.81
		HB3	2.69
		N	121.2
333	MET	CA	58.8
		CB	32.5
		CG	32.0
		CE	16.8
		H	8.09
		HA	4.09
		HB2	2.16
		HB3	2.16
333	MET	HG2	2.76
		HG3	2.61
		HE	2.08
		N	116.0
334	LEU	CA	58.9
		CB	42.1
		CG	26.7
		CD1	26.9
		CD2	24.9
		H	7.80
		HA	4.13
		HB2	1.93
		HB3	1.93
		HG	2.08
		HD1	0.80
		HD2	0.73
		N	122.0
335	LEU	CA	58.1
		CB	42.0
		CG	27.1
		CD1	25.8
		CD2	22.4
		H	8.33
		HA	3.99
		HB2	2.10
		HB3	1.46
		HG	1.73
		HD1	0.88
		HD2	0.64
		N	122.8
336	LYS	CA	60.5
		CB	32.4
		CG	26.8
		CD	29.0
		CE	41.8
		H	8.51
		HA	3.85
		HB2	1.93
		HB3	1.81
		HG2	1.61
		HG3	1.36
		HD2	1.57
		HD3	1.57
		HE2	2.94
		HE3	2.94
		N	119.6
337	GLU	CA	59.0
		CB	28.7
		CG	36.1
		H	8.37
		HA	4.10
		HB2	2.15
		HB3	2.07
		HG2	2.39
		HG3	2.39
		N	118.9
338	TYR	CA	60.4
		CB	37.5
		H	8.11
		HA	4.66
		HB2	3.19
		HB3	3.14
		HD1	6.85
		HD2	6.85

338	TYR	HE1	6.38
		HE2	6.38
		N	122.1
339	LEU	CA	57.1
		CB	41.1
		CG	27.3
		CD1	26.1
		CD2	23.4
		H	8.26
		HA	3.70
		HB2	1.80
		HB3	1.37
		HG	1.91
		HD1	0.90
		HD2	0.87
		N	119.2
340	LEU	CA	56.5
		CB	42.8
		CG	26.6
		CD1	24.8
		CD2	24.8
		H	7.24
		HA	4.22
		HB2	1.83
		HB3	1.57
		HG	1.71
		HD1	0.87
		HD2	0.86
		N	116.7
341	SER	CA	59.2
		CB	65.2
		H	8.20
		HA	4.45
		HB2	3.98
		HB3	3.98
		N	114.0
342	GLY	CA	46.3
		H	9.44
		HA2	3.97
		HA3	3.17
		N	114.4
343	ASP	CA	53.8
		CB	40.5
		H	8.22
		HA	4.80
		HB2	3.12
		HB3	2.40
		N	119.1
344	ILE	CA	65.5
		CB	38.6
		CG1	30.6
		CG2	16.8
		CD1	13.5
		H	8.63
		HA	3.57
		HB	1.92
		HG12	2.04
		HG13	1.08
		HG2	0.94
		HD1	1.00
		N	130.0
345	SER	CA	61.7
		CB	62.5
		H	8.21
		HA	4.31
		HB2	4.00
		HB3	4.00
		N	116.0
346	GLU	CA	58.3
		CB	29.2
		CG	35.4
		H	8.08
		HA	4.28
		HB2	1.99
		HB3	1.89
		HG2	2.36
		HG3	2.26
		N	123.9
347	ALA	CA	55.6
		CB	18.3
		H	7.51
		HA	4.23
		HB	1.24
		N	123.3
348	GLU	CA	60.3
		CB	29.9
		CG	36.8
		H	8.50
		HA	3.59
		HB2	2.05
		HB3	2.05
		HG2	2.09
		HG3	2.09
		N	116.9
349	HIS	CA	59.5
		CB	28.8
		H	8.30
		HA	4.26
		HB2	3.35
		HB3	3.35
		HD2	7.28
		HE1	8.31
		N	117.9
350	CYS	CA	62.9
		CB	26.6
		H	8.68
		HA	3.95
		HB2	3.12
		HB3	2.84
		N	118.4
351	LEU	CA	57.7
		CB	42.0
		CG	26.6
		CD1	22.3
		CD2	26.0
		H	8.24
		HA	3.89
		HB2	1.87
		HB3	1.35
		HG	1.30
		HD1	0.69
		HD2	0.26
		N	119.7

352	LYS	CA	60.2
		CB	31.9
		CG	26.0
		CD	29.3
		CE	41.8
		H	8.45
		HA	4.05
		HB2	1.93
		HB3	1.93
		HG2	1.70
		HG3	1.47
		HD2	1.70
		HD3	1.70
		HE2	2.99
		HE3	2.99
		N	120.5
353	GLU	CA	58.4
		CB	29.5
		CG	36.7
		H	8.00
		HA	4.04
		HB2	2.04
		HB3	1.97
		HG2	2.12
		HG3	2.12
		N	118.9
354	LEU	CA	56.7
		CB	43.1
		CG	26.1
		CD1	22.6
		CD2	25.5
		H	7.48
		HA	4.08
		HB2	2.08
		HB3	1.47
		HG	2.04
		HD1	0.81
		HD2	0.77
		N	117.9
355	GLU	CA	57.0
		CB	27.9
		CG	36.8
		H	7.53
		HA	3.88
		HB2	2.25
		HB3	2.10
		HG2	2.29
		HG3	2.29
		N	113.0
356	VAL	CA	57.4
		CB	32.4
		CG1	18.7
		CG2	21.6
		HA	4.78
		HB	1.96
		HG1	0.97
		HG2	0.94
357	PRO	CA	64.5
		CB	30.5
		CG	27.7
		CD	49.7
		HA	2.39
		HB2	1.85
		HB3	1.82
357	PRO	HG2	2.03
		HG3	2.03
		HD2	3.67
		HD3	3.35
358	HIS	CA	57.0
		CB	29.2
		H	8.11
		HA	4.53
		HB2	3.11
		HB3	2.94
		N	114.3
359	PHE	CA	55.7
		CB	40.8
		H	8.29
		HA	4.96
		HB2	3.24
		HB3	2.51
		HD1	6.99
		HD2	6.99
		HE1	7.35
		HE2	7.35
		HZ	7.29
		N	119.5
360	HIS	CA	58.5
		CB	28.6
		HA	4.38
		HB2	3.31
		HB3	3.08
		HD2	6.45
		HE1	7.76
361	HIS	CA	59.1
		CB	31.4
		H	9.27
		HA	4.28
		HB2	3.18
		HB3	2.92
		N	118.7
362	GLU	CA	57.8
		CB	28.1
		H	6.72
		HA	4.13
		HB2	1.71
		HB3	1.40
		N	121.6
363	LEU	CA	58.6
		CB	40.4
		CG	27.5
		CD1	26.5
		CD2	24.8
		H	7.05
		HA	3.91
		HB2	2.11
		HB3	1.41
		HG	1.57
		HD1	0.77
		HD2	0.62
		N	119.8
364	VAL	CA	67.6
		CB	32.3
		CG1	23.3
		CG2	21.7
		H	7.40

364	VAL	HA	3.16
		HB	2.22
		HG1	1.15
		HG2	0.98
		N	118.2
365	TYR	CA	61.7
		CB	39.4
		H	7.92
		HA	3.73
		HB2	3.12
		HB3	2.94
		HD1	6.89
		HD2	6.89
		HE1	6.51
		HE2	6.51
		N	116.8
366	GLU	CA	59.2
		CB	28.6
		CG	35.3
		H	8.52
		HA	3.46
		HB2	1.98
		HB3	1.85
		HG2	2.83
		HG3	2.50
		N	116.3
367	ALA	CA	55.2
		CB	17.4
		H	8.56
		HA	3.84
		HB	1.40
		N	121.4
368	ILE	CA	65.9
		CB	36.8
		CG1	28.9
		CG2	17.3
		CD1	12.8
		H	7.67
		HA	2.91
		HB	1.39
		HG12	1.72
		HG13	0.26
		HG2	0.39
		HD1	-0.05
		N	117.2
369	VAL	CA	68.4
		CB	30.8
		CG1	21.6
		CG2	23.2
		H	8.03
		HA	3.06
		HB	1.72
		HG1	0.66
		HG2	0.29
		N	121.4
370	MET	CA	59.8
		CB	31.7
		CG	33.8
		CE	18.9
		H	7.71
		HA	3.96
		HB2	2.25
		HB3	1.78
370	MET	HG2	2.81
		HG3	2.42
		HE	2.07
		N	116.3
371	VAL	CA	67.1
		CB	31.1
		CG1	23.0
		CG2	21.3
		H	7.41
		HA	3.22
		HB	1.95
		HG1	0.55
		HG2	0.53
		N	119.5
372	LEU	CA	57.8
		CB	40.8
		CD1	26.1
		CD2	22.1
		H	8.11
		HA	3.89
		HB2	1.95
		HB3	1.28
		HD1	0.73
		HD2	0.74
		N	119.5
373	GLU	CA	56.9
		CB	30.3
		CG	36.5
		H	8.09
		HA	4.16
		HB2	2.04
		HB3	1.90
		HG2	2.41
		HG3	2.12
		N	113.9
374	SER	CA	59.2
		CB	64.5
		H	7.42
		HA	4.46
		HB2	4.08
		HB3	4.08
		N	115.3
375	THR	CA	61.7
		CB	70.2
		CG2	21.3
		H	8.52
		HA	4.57
		HB	4.28
		HG2	1.25
		N	115.0
376	GLY	CA	45.1
		H	8.50
		HA2	4.15
		HA3	4.15
		N	111.4
377	GLU	CA	57.2
		CB	31.1
		CG	36.2
		H	8.65
		HA	4.44
		HB2	2.14
		HB3	1.95

377	GLU	HG2	2.23
		HG3	2.23
		N	119.1
378	SER	CA	61.6
		CB	63.0
		H	8.27
		HA	4.32
		HB2	4.04
		HB3	4.04
		N	115.5
379	ALA	CA	55.9
		CB	17.5
		H	8.55
		HA	3.99
		HB	1.43
		N	124.5
380	PHE	CA	61.4
		CB	39.2
		H	8.28
		HA	3.97
		HB2	3.48
		HB3	3.17
		HD1	7.18
		HD2	7.18
		HE1	7.44
		HE2	7.44
		HZ	7.10
		N	118.2
381	LYS	CA	59.4
		CB	32.7
		CG	25.3
		CD	28.9
		CE	42.0
		H	8.02
		HA	4.17
		HB2	2.08
		HB3	2.08
		HG2	1.67
		HG3	1.59
		HD2	1.81
		HD3	1.81
		HE2	3.03
		HE3	3.03
		N	117.4
382	MET	CA	59.7
		CB	33.1
		CG	33.1
		CE	17.5
		H	8.68
		HA	4.40
		HB2	2.03
		HB3	2.03
		HG2	2.95
		HG3	2.51
		HE	1.80
		N	116.5
383	ILE	CA	62.6
		CB	34.1
		CG2	18.6
		CD1	9.3
		H	9.38
		HA	4.05
		HB	2.12
383	ILE	HG12	1.72
		HG13	1.51
		HG2	0.74
		HD1	0.83
		N	124.7
384	LEU	CA	59.8
		CB	41.5
		CG	27.6
		CD1	26.4
		CD2	24.6
		H	8.26
		HA	3.95
		HB2	2.07
		HB3	1.62
		HG	1.46
		HD1	0.84
		HD2	0.55
		N	123.9
385	ASP	CA	57.8
		CB	40.0
		H	8.36
		HA	4.43
		HB2	2.83
		HB3	2.61
		N	117.0
386	LEU	CA	58.5
		CB	41.4
		CG	26.6
		CD1	25.1
		CD2	22.8
		H	7.85
		HA	3.92
		HB2	1.96
		HB3	1.25
		HG	1.15
		HD1	0.32
		HD2	0.23
		N	122.2
387	LEU	CA	58.5
		CB	41.2
		CG	26.8
		CD1	25.8
		CD2	22.0
		H	8.63
		HA	3.87
		HB2	1.95
		HB3	1.28
		HG	1.94
		HD1	0.96
		HD2	0.70
		N	117.6
388	LYS	CA	60.7
		CB	32.7
		CG	25.4
		CD	29.8
		CE	41.3
		H	8.74
		HA	4.03
		HB2	2.01
		HB3	1.93
		HG2	1.68
		HG3	1.42
		HD2	1.71
		HD3	1.71

388	LYS	HE2	2.97
		HE3	2.97
		N	117.2
389	SER	CA	62.4
		CB	63.4
		H	7.85
		HA	4.25
		HB2	4.03
		HB3	4.03
		N	114.2
390	LEU	CA	57.3
		CB	41.7
		CG	26.8
		CD1	22.5
		CD2	26.0
		H	8.59
		HA	4.18
		HB2	1.83
		HB3	1.21
		HG	1.94
		HD1	0.69
		HD2	0.56
		N	119.7
391	TRP	CA	60.1
		CB	29.5
		H	8.52
		HA	4.73
		HB2	3.54
		HB3	3.24
		HD1	7.12
		HE1	9.74
		HE3	7.56
		HH2	7.17
		HZ2	7.44
		HZ3	6.96
		N	120.7
		NE1	128.8
392	LYS	CA	59.5
		CB	32.1
		CG	25.4
		CD	29.0
		CE	41.9
		H	8.56
		HA	3.75
		HB2	2.14
		HB3	2.07
		HG2	1.74
		HG3	1.60
		HD2	1.76
		HD3	1.76
		HE2	3.02
		HE3	3.02
		N	122.2
393	SER	CA	58.0
		CB	63.7
		H	8.10
		HA	4.43
		HB2	4.18
		HB3	4.18
		N	111.7
394	SER	CA	59.5
		CB	61.4
		H	7.79
394	SER	HA	4.07
		HB2	4.04
		HB3	4.04
		N	114.1
395	THR	CA	66.0
		CB	68.2
		CG2	23.1
		H	8.28
		HA	4.04
		HB	4.06
		HG2	1.16
		N	117.8
396	ILE	CA	59.3
		CB	39.6
		CG1	27.4
		CG2	18.2
		CD1	13.9
		H	7.06
		HA	4.98
		HB	1.60
		HG12	1.40
		HG13	0.89
		HG2	1.06
		HD1	0.73
		N	117.1
397	THR	CA	60.6
		CB	71.5
		CG2	21.8
		HA	4.48
		HB	4.73
		HG2	1.28
398	ILE	CA	65.0
		CB	38.3
		CG1	29.2
		CG2	17.0
		CD1	13.7
		H	9.02
		HA	4.12
		HB	2.11
		HG12	1.77
		HG13	1.48
		HG2	1.13
		HD1	1.06
		N	120.8
399	ASP	CA	57.4
		CB	41.1
		H	8.41
		HA	4.47
		HB2	2.65
		HB3	2.65
		N	118.9
400	GLN	CA	58.7
		CB	28.7
		H	7.51
		HA	4.05
		HB2	2.50
		HB3	2.50
		N	119.7
401	MET	CA	58.4
		CB	32.8
		CG	32.5
		CE	17.9

401	MET	H	8.35
		HA	4.31
		HB2	2.29
		HB3	2.29
		HG2	2.66
		HG3	2.60
		HE	1.98
		N	118.8
402	LYS	CA	59.9
		CB	32.8
		CG	24.6
		CD	29.8
		CE	41.7
		H	8.45
		HA	3.95
		HB2	1.98
		HB3	1.98
		HG2	1.46
		HG3	1.33
		HD2	1.71
		HD3	1.71
		HE2	2.97
		HE3	2.97
		N	118.4
403	ARG	CA	59.0
		CB	30.3
		CG	27.4
		CD	43.3
		H	7.71
		HA	4.18
		HB2	1.82
		HB3	1.82
		HG2	1.65
		HG3	1.40
		HD2	2.96
		HD3	2.76
		N	116.3
404	GLY	CA	48.2
		H	8.16
		HA2	3.99
		HA3	3.75
		N	106.3
405	TYR	CA	62.6
		CB	38.4
		H	8.19
		HA	3.67
		HB2	2.92
		HB3	2.17
		HD1	7.18
		HD2	7.18
		HE1	6.60
		HE2	6.60
		N	119.8
406	GLU	CA	60.2
		CB	29.5
		CG	37.6
		H	8.64
		HA	3.67
		HB2	2.16
		HB3	1.93
		HG2	2.55
		HG3	2.14
		N	116.3
407	ARG	CA	60.0
		CB	30.7
		CG	27.1
		CD	44.0
		H	7.67
		HA	4.15
		HB2	1.95
		HB3	1.95
		HG2	1.35
		HG3	1.35
		HD2	2.94
		HD3	2.47
		HE	6.71
		N	117.9
		NE	83.8
408	ILE	CA	62.0
		CB	35.7
		CG1	27.5
		CG2	18.8
		CD1	10.9
		H	7.43
		HA	4.02
		HB	2.21
		HG12	1.58
		HG13	1.18
		HG2	0.94
		HD1	0.50
		N	119.8
409	TYR	CA	57.5
		CB	35.8
		H	8.85
		HA	4.38
		HB2	3.42
		HB3	3.23
		HD1	7.00
		HD2	7.00
		HE1	6.87
		HE2	6.87
		N	119.7
410	ASN	CA	55.5
		CB	39.5
		H	7.94
		HA	4.65
		HB2	3.02
		HB3	2.97
		HD21	7.69
		HD22	7.02
		N	115.1
		ND2	113.1
411	GLU	CA	57.0
		CB	31.3
		CG	36.6
		H	7.96
		HA	4.57
		HB2	2.42
		HB3	2.05
		HG2	2.70
		HG3	2.43
		N	117.5
412	ILE	CA	63.0
		CB	38.8
		CG1	29.0
		CG2	18.8
		CD1	15.1

412	ILE	H	8.27
		HA	4.26
		HB	2.01
		HG12	1.12
		HG13	1.12
		HG2	1.04
		HD1	0.68
		N	122.1
413	PRO	CA	66.2
		CB	30.9
		CG	28.3
		CD	51.2
		HA	4.44
		HB2	2.38
		HB3	1.84
		HG2	2.09
		HG3	1.99
		HD2	3.82
		HD3	3.51
414	ASP	CA	56.0
		CB	41.2
		H	7.58
		HA	4.65
		HB2	2.81
		HB3	2.77
		N	115.9
415	ILE	CA	64.2
		CB	38.9
		CG1	28.4
		CG2	16.1
		CD1	13.5
		H	8.00
		HA	3.93
		HB	1.79
		HG12	1.68
		HG13	1.15
		HG2	0.91
		HD1	0.78
		N	122.9
416	ASN	CA	54.7
		CB	39.9
		H	8.58
		HA	4.63
		HB2	2.77
		HB3	2.62
		HD21	7.66
		HD22	6.91
		N	117.6
		ND2	114.0
417	LEU	CA	56.3
		CB	42.0
		CG	26.6
		CD1	25.3
		CD2	22.8
		H	7.40
		HA	4.21
		HB2	1.83
		HB3	1.56
		HG	1.90
		HD1	0.98
		HD2	0.91
		N	117.6
418	ASP	CA	55.1
		CB	42.2
		HA	4.73
		HB2	2.78
		HB3	2.64
419	VAL	CA	62.1
		CB	32.8
		CG1	21.2
		CG2	20.3
		HA	4.12
		HB	2.09
		HG1	0.95
		HG2	0.95
420	PRO	CA	62.3
		CB	34.7
		CG	24.7
		CD	50.3
		HA	4.56
		HB2	2.39
		HB3	2.14
		HG2	1.94
		HG3	1.82
		HD2	3.58
		HD3	3.58
421	HIS	CA	57.0
		CB	28.7
		H	8.65
		HA	4.80
		HB2	3.48
		HB3	3.27
		HD2	7.09
		HE1	8.14
		N	123.1
422	SER	CA	62.7
		CB	63.0
		HA	4.13
		HB2	3.86
		HB3	3.86
423	TYR	CA	62.7
		CB	37.5
		H	8.32
		HA	4.09
		HB2	3.01
		HB3	3.01
		HD1	7.11
		HD2	7.11
		HE1	6.70
		HE2	6.70
		N	121.1
424	SER	CA	61.6
		CB	62.8
		H	7.83
		HA	4.30
		HB2	3.98
		HB3	3.98
		N	114.0
425	VAL	CA	66.7
		CB	31.6
		CG1	22.3
		CG2	22.2
		H	7.79
		HA	3.82
		HB	2.11

425	VAL	HG1	1.12
		HG2	1.10
		N	122.9
426	LEU	CA	58.5
		CB	41.1
		CG	27.0
		CD1	27.0
		CD2	23.7
		H	8.41
		HA	4.03
		HB2	2.10
		HB3	1.53
		HG	2.00
		HD1	1.05
		HD2	1.05
		N	120.3
427	GLU	CA	59.5
		CB	28.4
		CG	35.9
		H	8.46
		HA	4.19
		HB2	2.24
		HB3	2.07
		HG2	2.50
		HG3	2.50
		N	116.8
428	ARG	CA	59.6
		CB	30.9
		CG	28.2
		CD	43.6
		H	7.85
		HA	4.28
		HB2	2.22
		HB3	2.22
		HG2	2.00
		HG3	1.81
		HD2	3.24
		HD3	3.16
		HE	7.13
		N	119.6
		NE	83.9
429	PHE	CA	63.0
		CB	39.4
		H	9.18
		HA	4.03
		HB2	3.17
		HB3	3.08
		HD1	6.23
		HD2	6.23
		HE1	6.97
		HE2	6.97
		HZ	6.72
		N	122.3
430	VAL	CA	67.6
		CB	31.8
		CG1	24.2
		CG2	22.4
		H	9.65
		HA	3.59
		HB	2.44
		HG1	1.53
		HG2	1.17
		N	121.1
431	GLU	CA	59.6
		CB	28.7
		CG	34.9
		H	8.23
		HA	4.34
		HB2	2.25
		HB3	2.25
		HG2	2.53
		HG3	2.29
		N	121.9
432	GLU	CA	59.8
		CB	28.7
		CG	36.4
		H	8.30
		HA	4.08
		HB2	2.16
		HB3	2.03
		HG2	2.59
		HG3	2.54
		N	118.8
433	CYS	CA	64.0
		CB	26.2
		H	8.31
		HA	3.70
		HB2	3.13
		HB3	2.11
		N	119.1
434	PHE	CA	60.5
		CB	39.0
		H	8.66
		HA	4.24
		HB2	3.27
		HB3	2.86
		HD1	7.18
		HD2	7.18
		HE1	7.28
		HE2	7.28
		N	122.6
435	GLN	CA	58.4
		CB	27.5
		CG	33.1
		H	8.91
		HA	3.58
		HB2	2.16
		HB3	1.98
		HG2	2.62
		HG3	2.58
		HE21	8.04
		HE22	6.85
		N	121.6
		NE2	112.8
436	ALA	CA	52.2
		CB	19.0
		H	7.51
		HA	4.28
		HB	1.42
		N	117.9
437	GLY	CA	45.2
		H	7.63
		HA2	3.95
		HA3	3.77
		N	106.0

438	ILE	CA	61.6
		CB	38.0
		CG2	17.7
		CD1	14.2
		H	7.18
		HA	2.79
		HB	1.23
		HG12	0.95
		HG13	0.82
		HG2	0.37
		HD1	0.64
		N	111.6
439	ILE	CA	58.6
		CB	42.1
		CG1	24.6
		CG2	18.1
		CD1	14.2
		H	6.46
		HA	4.30
		HB	1.58
		HG12	1.24
		HG13	0.32
		HG2	0.19
		HD1	0.53
		N	107.2
440	SER	CA	56.7
		CB	65.6
		H	7.94
		HA	4.49
		HB2	4.38
		HB3	4.04
		N	116.7
441	LYS	CA	59.1
		CB	32.3
		CG	24.7
		CD	28.4
		CE	42.0
		H	9.05
		HA	3.31
		HB2	1.94
		HB3	1.73
		HG2	1.37
		HG3	1.02
		HD2	1.63
		HD3	1.43
		HE2	2.95
		HE3	2.90
		N	121.8
442	GLN	CA	59.4
		CB	28.0
		CG	34.0
		H	8.14
		HA	3.97
		HB2	2.05
		HB3	1.93
		HG2	2.40
		HG3	2.36
		HE21	7.57
		HE22	6.94
		N	116.6
		NE2	112.4
443	LEU	CA	57.1
		CB	42.3
		CG	27.0
443	LEU	CD1	24.2
		CD2	25.2
		H	7.63
		HA	4.22
		HB2	1.50
		HB3	1.50
		HG	1.47
		HD1	1.00
		HD2	0.73
		N	119.9
444	ARG	CA	58.9
		CB	28.8
		CG	25.8
		CD	42.9
		H	8.09
		HA	3.71
		HB2	1.59
		HB3	1.40
		HG2	1.30
		HG3	1.27
		HD2	2.67
		HD3	2.52
		HE	7.85
		N	123.6
		NE	84.2
445	ASP	CA	56.3
		CB	39.8
		H	8.67
		HA	4.28
		HB2	2.60
		HB3	2.60
		N	120.1
446	LEU	CA	54.7
		CB	43.4
		CG	26.7
		CD1	25.1
		CD2	23.0
		H	7.09
		HA	4.10
		HB2	1.89
		HB3	1.72
		HG	1.86
		HD1	0.97
		HD2	0.91
		N	119.6
447	CYS	CA	57.0
		CB	27.3
		H	7.13
		HA	3.06
		HB2	3.08
		HB3	2.79
		N	122.3
448	PRO	CA	64.5
		CB	32.7
		CG	27.6
		CD	51.2
		HA	4.20
		HB2	2.29
		HB3	1.90
		HG2	1.97
		HG3	1.97
		HD2	3.56
		HD3	3.22

REFERENCES

Abramson RD, Dever TE, Lawson TG, Ray BK, Thach RE and Merrick WC. (1987). The ATP-dependent interaction of eukaryotic initiation-factors with messenger-RNA. *J. Biol. Chem.*, **262:** 3826-3832.

Afonja O, Juste D, Das S, Matsuhashi S and Samuels H. (2004). Induction of PDCD4 tumor suppressor gene expression by RAR agonists, antiestrogen and HER-2/neu antagonist in breast cancer cells. Evidence for a role in apoptosis. *Oncogene*, **23:** 8135-8145.

Agalioti T, Chen GY and Thanos D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. *Cell*, **111:** 381-392.

Ait-Si-Ali S, Ramirez S, Barre FX, Dkhissi F, Magnaghi-Jaulin L, Girault JA, *et al.* (1998). Histone acetyltransferase activity of CBP is controlled by cycle-dependent kinases and oncoprotein E1A. *Nature*, **396:** 184-186.

Ali I and Jackson R. (2001). The translation of capped mRNAs has an absolute requirement for the central domain of eIF4G but not for the cap-binding initiation factor eIF4E. *Cold Spring Harb. Symp. Quant. Biol.*, **66:** 377-387.

Andersen CBF, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CLP, *et al.* (2006). Structure of the cxon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*, **313:** 1968-1972.

Andrade M and Bork P. (1995). HEAT repeats in the huntingtons-disease protein. *Nat. Genet.*, **11:** 115-116.

Anzick S, Kononen J, Walker R, Azorsa D, Tanner M, Guan X, *et al.* (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science*, **277:** 965-968.

Aranda A and Pascual A. (2001). Nuclear hormone receptors and gene expression. *Physiol. Rev.*, **81:** 1269-1304.

Arany Z, Huang LE, Eckner R, Bhattacharya S, Jiang C, Goldberg MA, *et al.* (1996). An essential role for p300/CBP in the cellular response to hypoxia. *Proc. Natl. Acad. Sci. USA*, **93:** 12969-73.

Auger AP, Tetel MJ and McCarthy MM. (2000). Steroid receptor coactivator-1 (SRC-1) mediates the development of sex-specific brain morphology and behavior. *Proc. Natl. Acad. Sci. USA*, **97:** 7551-7555.

Avantaggiati ML, Ogryzko V, Gardner K, Giordano A, Levine AS and Kelly K. (1997). Recruitment of p300/CBP in p53-dependent signal pathways. *Cell*, **89:** 1175-84.

Baek SH and Rosenfeld MG. (2004). Nuclear receptor coregulators: their modification codes and regulatory mechanism by translocation. *Biochem. Biophys. Res. Commun.*, **319:** 707-714.

Banerjee AC, Recupero AJ, Mal A, Piotrkowski AM, Wang DM and Harter ML. (1994). The adenovirus E1A 289R and 243R proteins inhibit the phosphorylation of p300. *Oncogene*, **9:** 1733-1737.

Bannister AJ and Kouzarides T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature*, **384:** 641-643.

Bannister AJ, Oehler T, Wilhelm D, Angel P and Kouzarides T. (1995). Stimulation of c-Jun activity by CBP: c-Jun residues Ser63/73 are required for CBP induced stimulation in vivo and CBP binding in vitro. *Oncogene*, **11:** 2509-2514.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, *et al.* (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, **410:** 120-124.

Bartels C, Xia TH, Billeter M, Guntert P and Wuthrich K. (1995). The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J. Biomol. NMR*, **6:** 1-10.

Bauer C, Diesinger I, Brass N, Steinhart H, Iro H and Meese E. (2001). Translation initiation factor eIF-4G is immunogenic, overexpressed, and amplified in patients with squamous cell lung carcinoma. *Cancer*, **92:** 822-829.

Bautista S, Valles H, Walker RL, Anzick S, Zeillinger R, Meltzer P, *et al.* (1998). In breast cancer, amplification of the steroid receptor coactivator gene AIB1 is correlated with estrogen and progesterone receptor positivity. *Clin. Cancer Res.*, **4:** 2925-2929.

Bax A, Clore GM and Gronenborn AM. (1990). H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13-enriched proteins. *J. Magn. Reson.*, **88:** 425-431.

Bellsolell L, Cho-Park PF, Poulin F, Sonenberg N and Burley SK. (2006). Two structurally atypical HEAT domains in the C-terminal portion of human eIF4G support binding to eIF4A and Mnk1. *Structure*, **14:** 913-23.

Belsham GJ and Sonenberg N. (1996). RNA-Protein interactions in regulation of picornavirus RNA translation. *Microbiol. Rev.*, **60:** 499-&.

Bevan C, Hoare S, Claessens F, Heery D and Parker M. (1999). The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Mol. Cell. Biol.*, **19:** 8383-8392.

Bex F, Yin MJ, Burny A and Gaynor RB. (1998). Differential transcriptional activation by

human T-cell leukemia virus type 1 Tax mutants is mediated by distinct interactions with CREB binding protein and p300. *Mol. Cell. Biol.*, **18:** 2392-2405.

Bhattacharya S, Michels CL, Leung MK, Arany ZP, Kung AL and Livingston DM. (1999). Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1. *Genes Dev*, **13:** 64-75.

Bitomsky N, Bohm M and Klempnauer K. (2004). Transformation suppressor protein Pdcd4 interferes with JNK-mediated phosphorylation of c-Jun and recruitment of the coactivator p300 by c-Jun. *Oncogene*, **23:** 7484-7493.

Bodenhausen G and Ruben DJ. (1980). Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69:** 185-189.

Boeger H, Bushnell DA, Davis R, Griesenbeck J, Lorch Y, Strattan JS, *et al.* (2005). Structural basis of eukaryotic gene transcription. *FEBS Lett.*, **579:** 899-903.

Boesen T, Mohammad SS, Pavitt GD and Andersen GR. (2004). Structure of the catalytic fragment of translation initiation factor 2B and identification of a critically important catalytic residue. *J. Biol. Chem.*, **279:** 10584-10592.

Bohm M, Sawicka K, Siebrasse J, Brehmer-Fastnacht A, Peters R and Klempnauer K. (2003). The transformation suppressor protein Pdcd4 shuttles between nucleus and cytoplasm and binds RNA. *Oncogene*, **22:** 4905-4910.

Bono F, Ebert J, Lorentzen E and Conti E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*, **126:** 713-725.

Bordoli L, Husser S, Luthi U, Netsch M, Osmani H and Eckner R. (2001). Functional analysis of the p300 acetyltransferase domain: the PHD finger of p300 but not of CBP is dispensable for enzymatic activity. *Nucl. Acids Res.*, **29:** 4462-71.

Borman AM, Michel YM and Kean KM. (2000). Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5 '-end. *Nucl. Acids Res.*, **28:** 4068-4075.

Bourguet W, Ruff M, Chambon P, Gronemeyer H and Moras D. (1995). Crystal-structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature*, **375:** 377-382.

Braunschweiler L and Ernst RR. (1983). Coherence transfer by isotropic mixing - application to proton correlation spectroscopy. *J. Magn. Reson.*, **53:** 521-528.

Brzozowski AM, Pike ACW, Dauter Z, Hubbard RE, Bonn T, Engstrom O, *et al.* (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature*, **389:** 753-758.

Bunone G, Briand PA, Miksicek RJ and Picard D. (1996). Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J.*, **15:** 2174-2183.

Bushnell DA, Westover KD, Davis RE and Kornberg RD. (2004). Structural basis of transcription: An RNA polymerase II-TFIIB cocrystal at 4.5 angstroms. *Science*, **303:** 983-988.

Campbell KM and Lumb KJ. (2002). Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry*, **41:** 13956-13964.

Cao R, Wang LJ, Wang HB, Xia L, Erdjument-Bromage H, Tempst P, *et al.* (2002). Role of histone H3 lysine 27 methylation in polycomb-group silencing. *Science*, **298:** 1039-1043.

Carapeti M, Aguiar RCT, Goldman JM and Cross NCP. (1998). A novel fusion between MOZ and the nuclear receptor coactivator TIF2 in acute myeloid leukemia. *Blood*, **91:** 3127-3133.

Caron C, Col E and Khochbin S. (2003). The viral control of cellular acetylation signaling. *Bioessays*, **25:** 58-65.

Caron S, Charon M, Cramer E, Sonenberg N and Dusanter-Fourt I. (2004). Selective modification of eukaryotic initiation factor 4F (eIF4F) at the onset of cell differentiation: Recruitment of eIF4GII and long-lasting phosphorylation of eIF4E. *Mol. Cell. Biol.*, **24:** 4920-4928.

Carr M, Bloemink M, Dentten E, Whelan A, Gordon S, Kelly G, *et al.* (2003). Solution structure of the Mycobacterium tuberculosis complex protein MPB70 - From tuberculosis pathogenesis to inherited human corneal disease. *J. Biol. Chem.*, **278:** 43736-43743.

Caruthers J, Johnson E and McKay D. (2000). Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl. Acad. Sci. USA*, **97:** 13080-13085.

Cavanagh J, Fairbrother WJ, Palmer AG and Skelton NJ. (1995). *Protein NMR Spectroscopy: Principles and Practice: Section 5.6: Chemical exchange effects in NMR spectroscopy*. Academic Press Inc.: London, pp 290-299.

Chakravarti D, LaMorte VJ, Nelson MC, Nakajima T, Schulman IG, Juguilon H, *et al.* (1996). Role of CBP/p300 in nuclear receptor signalling. *Nature*, **383:** 99-103.

Chan HM and La Thangue NB. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.*, **114:** 2363-2373.

Chawla S, Hardingham GE, Quinn DR and Bading H. (1998). CBP: A signal-regulated transcriptional coactivator controlled by nuclear calcium and CaM kinase IV. *Science*,

281: 1505-1509.

Chen D, Ma H, Hong H, Koh S, Huang S, Schurter B, *et al.* (1999a). Regulation of transcription by a protein methyltransferase. *Science*, **284:** 2174-2177.

Chen H, Lin R, Schiltz R, Chakravarti D, Nash A, Nagy L, *et al.* (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell*, **90:** 569-580.

Chen HW, Lin RJ, Xie W, Wilpitz D and Evans RM. (1999b). Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell*, **98:** 675-686.

Chen S, Dowhan D, Hosking B and Muscat G. (2000). The steroid receptor coactivator, GRIP-1, is necessary for MEF-2C-dependent gene expression and skeletal muscle differentiation. *Gene Dev.*, **14:** 1209-1228.

Chen Y, Knosel T, Kristiansen G, Pietas A, Garber M, Matsuhashi S, *et al.* (2003). Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis. *J. Pathol.*, **200:** 640-646.

Cheung P, Allis CD and Sassone-Corsi P. (2000). Signaling to chromatin through histone modifications. *Cell*, **103:** 263-271.

Chiba H, Muramatsu M, Nomoto A and Kato H. (1994). 2 human homologs of Saccharomyces-cerevisiae SW12/SNF2 and drosophila-brahma are transcriptional coactivators cooperating with the estrogen-receptor and the retinoic acid receptor. *Nucl. Acids Res.*, **22:** 1815-1820.

Chrivia JC, Kwok RPS, Lamb N, Hagiwara M, Montminy MR and Goodman RH. (1993). Phosphorylated CREB binds specifically to the nuclear-protein CBP. *Nature*, **365:** 855-859.

Cmarik J, Min H, Hegamyer G, Zhan S, Kulesz-Martin M, Yoshinaga H, *et al.* (1999). Differentially expressed protein Pdcd4 inhibits tumor promoter-induced neoplastic transformation. *Proc. Natl. Acad. Sci. USA*, **96:** 14037-14042.

Collins HM, Kindle KB, Matsuda S, Ryan C, Troke PJF, Kalkhoven E, *et al.* (2006). MOZ-TIF2 alters cofactor recruitment and histone modification at the RAR beta 2 promoter - Differential effects of MOZ fusion proteins on CBP- AND MOZ-dependent activators. *J. Biol. Chem.*, **281:** 17124-17133.

Cornilescu G, Delaglio F and Bax A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13:** 289-302.

Coulthard VH, Matsuda S and Heery DM. (2003). An extended LXXLL motif sequence

determines the nuclear receptor binding specificity of TRAP220. *J. Biol. Chem.*, **278:** 10942-10951.

Cuff JA, Clamp ME, Siddiqui AS, Finlay M and Barton GJ. (1998). JPRED: a consensus secondary structure prediction server. *Bioinformatics*, **14:** 892-3.

Dai P, Akimaru H, Tanaka Y, Hou DX, Yasukawa T, Kaneilshii C, *et al.* (1996). CBP as a transcriptional coactivator of c-Myb. *Genes Dev*, **10:** 528-540.

Davie JR and Murphy LC. (1990). Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. *Biochemistry*, **29:** 4752-4757.

De Benedetti A and Harris AL. (1999). eIF4E expression in tumors: its possible role in progression of malignancies. *Int. J. Biochem. Cell Biol.*, **31:** 59-72.

De Guzman RN, Liu HY, Martinez-Yamout M, Dyson HJ and Wright PE. (2000). Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP. *J. Mol. Biol.*, **303:** 243-253.

De Guzman RN, Martinez-Yamout MA, Dyson HJ and Wright PE. (2004). Interaction of the TAZ1 domain of the CREB-binding protein with the activation domain of CITED2: regulation by competition between intrinsically unstructured ligands for non-identical binding sites. *J. Biol. Chem.*, **279:** 3042-9.

De Guzman RN, Wojciak JM, Martinez-Yamout MA, Dyson HJ and Wright PE. (2005). CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry*, **44:** 490-497.

Delage-Mourroux R, Martini PGV, Choi I, Kraichely DM, Hoeksema J and Katzenellenbogen BS. (2000). Analysis of estrogen receptor interaction with a repressor of estrogen receptor activity (REA) and the regulation of estrogen receptor transcriptional activity by REA. *J. Biol. Chem.*, **275:** 35848-35856.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J and Bax A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR*, **6:** 277-293.

Demarest SJ, Deechongkit S, Dyson HJ, Evans RM and Wright PE. (2004). Packing, specificity, and mutability at the binding interface between the p160 coactivator and CREB-binding protein. *Protein Sci.*, **13:** 203-210.

Demarest SJ, Martinez-Yamout M, Chung J, Chen HW, Xu W, Dyson HJ, *et al.* (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature*, **415:** 549-553.

Derijard B, Hibi M, Wu IH, Barrett T, Su B, Deng TL, *et al.* (1994). JNK1 - a protein-kinase stimulated by UV-light and Ha-Ras that binds and phosphorylates the c-Jun

activation domain. *Cell*, **76:** 1025-1037.

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK and Zhou MM. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, **399:** 491-496.

Dorrello NV, Peschiaroli A, Guardavaccaro D, Colburn NH, Sherman NE and Pagano M. (2006). S6K1- and betaTRCP-mediated degradation of PDCD4 promotes protein translation and cell growth. *Science*, **314:** 467-71.

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM and Obradovic Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, **41:** 6573-6582.

Dutertre M and Smith CL. (2003). Ligand-independent interactions of p160/steroid receptor coactivators and CREB-binding protein (CBP) with estrogen receptor-alpha: Regulation by phosphorylation sites in the A/B region depends on other receptor domains. *Mol. Endocrinol.*, **17:** 1296-1314.

Dyson HJ and Wright PE. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6:** 197-208.

Eberle J, Krasagakis K and Orfanos C. (1997). Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells in vitro. *Int. J. Cancer*, **71:** 396-401.

Eckner R, Ewen ME, Newsome D, Gerdes M, Decaprio JA, Lawrence JB, *et al.* (1994). Molecular-cloning and functional analysis of the adenovirus E1A-associated 300-kDa protein (p300) reveals a protein with properties of a transcriptional adapter. *Genes Dev*, **8:** 869-884.

Eckner R, Ludlow JW, Lill NL, Oldread E, Arany Z, Modjtahedi N, *et al.* (1996). Association of p300 and CBP with simian virus 40 large T antigen. *Mol. Cell. Biol.*, **16:** 3454-3464.

Eferl R and Wagner EF. (2003). AP-1: A double-edged sword in tumorigenesis. *Nature Reviews Cancer*, **3:** 859-868.

Feeney J, Bauer CJ, Frenkiel TA, Birdsall B, Carr MD, Roberts GCK, *et al.* (1991). Deceptively simple exchange effects in homonuclear Hartmann-Hahn (HOHAHA) spectra of protein-ligand complexes. *J. Magn. Reson.*, **91:** 607-613.

Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, *et al.* (2006). Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34:** D247-51.

Fischle W, Wang YM and Allis CD. (2003). Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.*, **15:** 172-183.

Freedman SJ, Sun ZY, Poy F, Kung AL, Livingston DM, Wagner G, *et al.* (2002). Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *Proc.*

Natl. Acad. Sci. USA, **99:** 5367-72.

Gehin M, Mark M, Dennefeld C, Dierich A, Gronemeyer H and Chambon P. (2002). The function of TIF2/GRIP1 in mouse reproduction is distinct from those of SRC-1 and p/CIP. *Mol. Cell. Biol.*, **22:** 5923-5937.

Geiger JH, Hahn S, Lee S and Sigler PB. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science*, **272:** 830-836.

Giguere V. (1999). Orphan nuclear receptors: From gene to function. *Endocrinol. Rev.*, **20:** 689-725.

Gingras AC, Raught B and Sonenberg N. (1999). eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. *Annual Review of Biochemistry*, **68:** 913-963.

Giordano A and Avantaggiati ML. (1999). p300 and CBP: Partners for life and death. *J. Cell. Physiol.*, **181:** 218-230.

Glass CK and Rosenfeld MG. (2000). The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev*, **14:** 121-141.

Goddard TD, Kneller DG and. (unpublished). SPARKY 3. *University of California, San Francisco*: http://www.cgl.ucsf.edu/home/sparky/.

Goke A, Goke R, Knolle A, Trusheim H, Schmidt H, Wilmen A, *et al.* (2002). DUG is a novel homologue of translation initiation factor 4G that binds eIF4A. *Biochem. Biophys. Res. Commun.*, **297:** 78-82.

Goke R, Barth P, Schmidt A, Samans B and Lankat-Buttgereit B. (2004). Programmed cell death protein 4 suppresses CDK1/cdc2 via induction of p21(Waf1/Cip1). *Am. J. Physiol. Cell Physiol.*, **287:** C1541-C1546.

Goodman RH and Smolik S. (2000). CBP/p300 in cell growth, transformation, and development. *Genes Dev*, **14:** 1553-1577.

Gorbalenya AE and Koonin EV. (1993). Helicases - amino-acid-sequence comparisons and structure-function-relationships. *Curr. Opin. Cell Biol.*, **3:** 419-429.

Groft CM and Burley SK. (2002). Recognition of eIF4G by rotavirus NSP3 reveals a basis for mRNA circularization. *Mol. Cell*, **9:** 1273-1283.

Gronemeyer H, Gustafsson JA and Laudet V. (2004). Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.*, **3:** 950-964.

Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JEG, *et al.* (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell*, **115:** 739-750.

Groves MR and Barford D. (1999). Topological characteristics of helical repeat proteins.

Curr. Opin. Struct. Biol., **9:** 383-9.

Grzesiek S and Bax A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. *J. Magn. Reson.*, **96:** 432-440.

Grzesiek S and Bax A. (1993). Amino-acid type determination in the sequential assignment procedure of uniformly C-13/N-15-enriched proteins. *J. Biomol. NMR*, **3:** 185-204.

Guntert P, Mumenthaler C and Wuthrich K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273:** 283-298.

Guntert P and Wuthrich K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR*, **1:** 447-56.

Haghighat A, Mader S, Pause A and Sonenberg N. (1995). Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. *EMBO J.*, **14:** 5701-9.

Haghighat A and Sonenberg N. (1997). eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J. Biol. Chem.*, **272:** 21677-21680.

Hayashi Y, Ohmori S, Ito T and Seo H. (1997). A splicing variant of Steroid Receptor Coactivator-1 (SRC-1E): The major isoform of SRC-1 to mediate thyroid hormone action. *Biochem. Biophys. Res. Commun.*, **236:** 83-87.

Hebbes TR, Thorne AW and Cranerobinson C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.*, **7:** 1395-1402.

Heery D, Kalkhoven E, Hoare S and Parker M. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptor. *NATURE*, **387:** 733-736.

Heery DM. (Unpublished data).

Heery DM, Hoare S, Hussain S, Parker MG and Sheppard H. (2001). Core LXXLL motif sequences in CREB-binding protein, SRC1, and RIP140 define affinity and selectivity for steroid and retinoid receptors. *J. Biol. Chem.*, **276:** 6695-6702.

Hernandez G and Vazquez-Pianzola P. (2005). Functional diversity of the eukaryotic translation initiation factors belonging to eIF4 families. *Mech. Dev.*, **122:** 865-876.

Herrmann T, Guntert P and Wuthrich K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319:** 209-227.

Hershey JWB, and Merrick, W. C. (2000a). *Translational Control of Gene Expression.: Pathway and mechanism of initiation of protein synthesis.* Sonenberg N, Hershey, J. W. B., and Mathews M. B. (ed.). Cold Spring Harbor Laboratory Press, Cold Spring

Harbor: New York, pp 33-88.

Hershey JWB, and Miyamoto, S. (2000b). *Translational Control of Gene Expression.: Translational control and Cancer.* Sonenberg N, Hershey, J. W. B., and Mathews M. B. (ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor: New York, pp 637-654.

Hibi M, Lin AN, Smeal T, Minden A and Karin M. (1993). Identification of an oncoprotein-responsive and UV-responsive protein-kinase that binds and potentiates the c-Jun activation domain. *Genes Dev*, **7:** 2135-2148.

Hollenberg SM and Evans RM. (1988). Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. *Cell*, **55:** 899-906.

Holm L and Sander C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233:** 123-38.

Hong H, Kohli K, Trivedi A, Johnson D and Stallcup M. (1996). GRIP1, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone binding domains of steroid receptors. *Proc. Natl. Acad. Sci. USA*, **93:** 4948-4952.

Hong L, Schroth GP, Matthews HR, Yau P and Bradbury EM. (1993). Studies of the DNA-binding properties of histone H4 amino terminus - Thermal-denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 tail to DNA. *J. Biol. Chem.*, **268:** 305-314.

Horvai AE, Xu L, Korzus E, Brard G, Kalafus D, Mullen TM, *et al.* (1997). Nuclear integration of JAK/STAT and Ras/AP-1 signaling by CBP and p300. *Proc. Natl. Acad. Sci. USA*, **94:** 1074-9.

Hottiger MO, Felzien LK and Nabel GJ. (1998). Modulation of cytokine-induced HIV gene expression by competitive binding of transcription factors to the coactivator p300. *EMBO J.*, **17:** 3124-34.

Hu SC, Chrivia J and Ghosh A. (1999). Regulation of CBP-mediated transcription by neuronal calcium signaling. *Neuron*, **22:** 799-808.

Imataka H, Gradi A and Sonenberg N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J.*, **17:** 7480-7489.

Imataka H and Sonenberg N. (1997). Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol. Cell. Biol.*, **17:** 6940-6947.

Iyer NG, Ozdag H and Caldas C. (2004). p300/CBP and cancer. *Oncogene*, **23:** 4225-4231.

Jackson RJ. (2000). *Transcriptional Control of Gene Expression.: Comparative view of*

initiation site selection mechanisms Sonenberg N, Hershey, J. W. B., and Mathews M. B. (ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor: New York, pp 127-184.

Janknecht R and Nordheim A. (1996). MAP kinase-dependent transcriptional coactivation by Elk-1 and its cofactor CBP. *Biochem. Biophys. Res. Commun.*, **228:** 831-837.

Jansen A, Camalier C, Stark C and Colburn N. (2004). Characterization of programmed cell death 4 in multiple human cancers reveals a novel enhancer of drug sensitivity. *Mol. Cancer Ther.*, **3:** 103-110.

Jansen AP, Camalier CE and Colburn NH. (2005). Epidermal expression of the translation inhibitor programmed cell death 4 suppresses tumorigenesis. *Cancer Res.*, **65:** 6034-6041.

Jenuwein T and Allis CD. (2001). Translating the histone code. *Science*, **293:** 1074-1080.

Jepsen K, Hermanson O, Onami TM, Gleiberman AS, Lunyak V, McEvilly RJ, *et al.* (2000). Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell*, **102:** 753-763.

Jepsen K and Rosenfeld MG. (2002). Biological roles and mechanistic actions of co-repressor complexes. *J. Cell Sci.*, **115:** 689-698.

Jin H, Kim T, Hwang S, Chang S, Kim H, Anderson H, *et al.* (2006). Aerosol delivery of urocanic acid-modified chitosan/programmed cell death 4 complex regulated apoptosis, cell cycle, and angiogenesis in lungs of K-ras null mice. *Mol. Cancer Ther.*, **5:** 1041-1049.

Jordan VC. (2003). Antiestrogens and selective estrogen receptor modulators as multifunctional medicines. 2. Clinical considerations and new agents. *J. Med. Chem.*, **46:** 1081-1111.

Kalkhoven E. (2004). CBP and p300: HATs for different occasions. *Biochem. Pharmacol.*, **68:** 1145-1155.

Kalkhoven E, Teunissen H, Houweling A, Verrijzer CP and Zantema A. (2002). The PHD type zinc finger is an integral part of the CBP acetyltransferase domain. *Mol. Cell. Biol.*, **22:** 1961-70.

Kalkhoven E, Valentine J, Heery D and Parker M. (1998). Isoforms of steroid receptor co-activator 1 differ in their ability to potentiate transcription by the oestrogen receptor. *EMBO J.*, **17:** 232-243.

Kamei Y, Xu L, Heinzel T, Torchia J, Kurokawa R, Gloss B, *et al.* (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell*, **85:** 403-414.

Kang MJ, Ahn HS, Lee JY, Matsuhashi S and Park WY. (2002a). Up-regulation of PDCD4 in senescent human diploid fibroblasts. *Biochem. Biophys. Res. Commun.*, **293:** 617-621.

Kang ZG, Pirskanen A, Janne OA and Palvimo JJ. (2002b). Involvement of proteasome in the dynamic assembly of the androgen receptor transcription complex. *Journal of Biological Chemistry*, **277:** 48366-48371.

Kato S, Endoh H, Masuhiro Y, Kitamoto T, Uchiyama S, Sasaki H, *et al.* (1995). Activation of the estrogen-receptor through phosphorylation by mitogen-activated protein-kinase. *Science*, **270:** 1491-1494.

Kawasaki H, Eckner R, Yao TP, Taira K, Chiu R, Livingston DM, *et al.* (1998). Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. *Nature*, **393:** 284-289.

Kim M, Hsiao S and Kraus W. (2001). A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation. *EMBO J.*, **20:** 6084-6094.

Kimura A, Matsubara K and Horikoshi M. (2005). A decade of histone acetylation: Marking eukaryotic chromosomes with specific codes. *J. Biochem.*, **138:** 647-662.

Kindle KB, Troke PJF, Collins HM, Matsuda S, Bossi D, Bellodi C, *et al.* (2005). MOZ-TIF2 inhibits transcription by nuclear receptors and p53 by impairment of CBP function. *Mol. Cell. Biol.*, **25:** 988-1002.

Kishimoto M, Kohno T, Okudela K, Otsuka A, Sasaki H, Tanabe C, *et al.* (2005). Mutations and deletions of the CBP gene in human lung cancer. *Clin. Cancer Res.*, **11:** 512-519.

Koh S, Chen D, Lee Y and Stallcup M. (2001). Synergistic enhancement of nuclear receptor function by p160 coactivators and two coactivators with protein methyltransferase activities. *J. Biol. Chem.*, **276:** 1089-1098.

Koradi R, Billeter M and Wuthrich K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14:** 51-55.

Kurdistani SK and Grunstein M. (2003). Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell Biol.*, **4:** 276-284.

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P and Reinberg D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev*, **16:** 2893-2905.

Lamphear BJ, Kirchweger R, Skern T and Rhoads RE. (1995). Mapping of functional domains in eukaryotic protein-synthesis initiation-factor 4G (eIF4G) with picornaviral proteases - implications for cap-dependent and cap-independent translational initiation.

J. Biol. Chem., **270:** 21975-21983.

Lankat-Buttgereit B, Gregel C, Knolle A, Hasilik A, Arnold R and Goke R. (2004). Pdcd4 inhibits growth of tumor cells by suppression of carbonic anhydrase type II. *Mol. Cell. Endocrinol.*, **214:** 149-153.

LaRonde-LeBlanc N, Santhanam AN, Baker AR, Wlodawer A and Colburn NH. (2007). Structural basis for inhibition of translation by the tumor suppressor Pdcd4. *Mol. Cell. Biol.*, **27:** 147-156.

Laudet V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.*, **19:** 207-226.

Lee CW, Sorensen TS, Shikama N and La Thangue NB. (1998a). Functional interplay between p53 and E2F through co-activator p300. *Oncogene*, **16:** 2695-710.

Lee JS, See RH, Deng TL and Shi Y. (1996). Adenovirus E1A downregulates cJun- and JunB-mediated transcription by targeting their coactivator p300. *Mol. Cell. Biol.*, **16:** 4312-4326.

Lee SK, Kim HJ, Na SY, Kim TS, Choi HS, Im SY, *et al.* (1998b). Steroid receptor coactivator-1 coactivates activating protein-1-mediated transactivations through interaction with the c-Jun and c-Fos subunits. *J. Biol. Chem.*, **273:** 16651-16654.

Lee TI and Young RA. (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.*, **34:** 77-137.

Lees JA, Fawell SE and Parker MG. (1989). Identification of 2 transactivation domains in the mouse estrogen-receptor. *Nucl. Acids Res.*, **17:** 5477-5488.

Legge GB, Martinez-Yamout MA, Hambly DM, Trinh T, Lee BM, Dyson HJ, *et al.* (2004). ZZ domain of CBP: An unusual zinc finger fold in a protein interaction module. *J. Mol. Biol.*, **343:** 1081-1093.

Legoff P, Montano MR, Schodin DJ and Katzenellenbogen BS. (1994). Phosphorylation of the human estrogen-receptor - identification of hormone-regulated sites and examination of their influence on transcriptional activity. *J. Biol. Chem.*, **269:** 4458-4466.

Lemon B and Tjian R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev*, **14:** 2551-2569.

Li QY, Imataka H, Morino S, Rogers GW, Richter-Cook NJ, Merrick WC, *et al.* (1999). Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell. Biol.*, **19:** 7336-7346.

Lin CH, Hare BJ, Wagner G, Harrison SC, Maniatis T and Fraenkel E. (2001). A small domain of CBP/p300 binds diverse proteins: Solution structure and functional studies.

Mol. Cell, **8:** 581-590.

Linder P, Lasko P, Ashburner M, Leroy P, Nielsen P, Nishi K, *et al.* (1989). Birth of the D-E-A-D box. *Nature*, **337:** 121-122.

Litterst CM and Pfitzner E. (2001). Transcriptional activation by STAT6 requires the direct interaction with NCoA-1. *J. Biol. Chem.*, **276:** 45713-45721.

Litterst CM and Pfitzner E. (2002). An LXXLL motif in the transactivation domain of STAT6 mediates recruitment of NCoA-1/SRC-1. *J. Biol. Chem.*, **277:** 36052-36060.

Liu Z, Wong J, Tsai S, Tsai M and O'Malley B. (2001). Sequential recruitment of steroid receptor coactivator-1 (SRC-1) and p300 enhances progesterone receptor-dependent initiation and reinitiation of transcription from chromatin. *Proc. Natl. Acad. Sci. USA*, **98:** 12426-12431.

Livengood JA, Scoggin KE, Van Orden K, McBryant SJ, Edayathumangalam RS, Laybourn PJ, *et al.* (2002). p53 Transcriptional activity is mediated through the SRC1-interacting domain of CBP/p300. *J. Biol. Chem.*, **277:** 9054-61.

Lonard DM and O'Malley BW. (2005). Expanding functional diversity of the coactivators. *Trends Biochem. Sci.*, **30:** 126-132.

Lucey M, Chen D, Lopez-Garcia J, Hart S, Phoenix F, Al-Jehani R, *et al.* (2005). T: G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif. *Nucl. Acids Res.*, **33:** 6393-6404.

Luger K, Mader A, Richmond R, Sargent D and Richmond T. (1997). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature*, **389:** 251-260.

Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR and Sigler PB. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, **352:** 497-505.

Ma H, Hong H, Huang S, Irvine R, Webb P, Kushner P, *et al.* (1999). Multiple signal input and output domains of the 160-kilodalton nuclear receptor coactivator proteins. *Mol. Cell. Biol.*, **19:** 6164-6173.

Macura S and Ernst RR. (1980). Elucidation of cross relaxation in liquids by two-dimensional NMR-spectroscopy. *Mol. Phys.*, **41:** 95-117.

Mader S, Lee H, Pause A and Sonenberg N. (1995). The translation initiation-factor eIF-4E binds to a common motif shared by the translation factor eIF-4-Gamma and the translational repressors 4E-binding proteins. *Mol. Cell. Biol.*, **15:** 4990-4997.

Marcotrigiano J, Gingras AC, Sonenberg N and Burley SK. (1999). Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol.*

Cell, **3:** 707-716.

Marcotrigiano J, Lomakin I, Sonenberg N, Pestova T, Hellen C and Burley S. (2001). A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol. Cell*, **7:** 193-203.

Margueron R, Trojer P and Reinberg D. (2005). The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.*, **15:** 163-176.

Marintchev A and Wagner G. (2004). Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys*, **37:** 197-284.

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, *et al.* (1989a). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*, **28:** 6150-6.

Marion D, Kay LE, Sparks SW, Torchia DA and Bax A. (1989b). 3-Dimensional heteronuclear NMR of N-15-labeled proteins. *J. Am. Chem. Soc.*, **111:** 1515-1517.

Matsuda S, Harries JC, Viskaduraki M, Troke PJF, Kindle KB, Ryan C, *et al.* (2004). A conserved alpha-helical motif mediates the binding of diverse nuclear proteins to the SRC1 interaction domain of CBP. *J. Biol. Chem.*, **279:** 14055-14064.

Mazza C, Ohno M, Segref A, Mattaj IW and Cusack S. (2001). Crystal structure of the human nuclear cap binding complex. *Mol. Cell*, **8:** 383-96.

McInerney E, Rose D, Flynn S, Westin S, Mullen T, Krones A, *et al.* (1998). Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes Dev*, **12:** 3357-3368.

McKenna NJ and O'Malley BW. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell*, **108:** 465-474.

McManus KJ and Hendzel MJ. (2003). Quantitative analysis of CBP- and p300-induced histone acetylations in vivo using native chromatin. *Mol. Cell. Biol.*, **23:** 7611-7627.

Mellor J. (2006). Dynamic nucleosomes and gene transcription. *Trends Genet.*, **22:** 320-329.

Metivier R, Penot G, Hubner MR, Reid G, Brand H, Kos M, *et al.* (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, **115:** 751-763.

Metivier R, Reid G and Gannon F. (2006). Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.*, **7:** 161-167.

Miller RW and Rubinstein JH. (1995). Tumors in Rubinstein-Taybi Syndrome. *Am. J.*

Med. Genet., **56:** 112-115.

Mink S, Haenig B and Klempnauer KH. (1997). Interaction and functional collaboration of p300 and C/EBP beta. *Mol. Cell. Biol.*, **17:** 6609-6617.

Morino S, Imataka H, Svitkin Y, Pestova T and Sonenberg N. (2000). Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol. Cell. Biol.*, **20:** 468-477.

Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, Sachchidanand, *et al.* (2004). Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell*, **13:** 251-263.

Muskett F, Frenkiel T, Feeney J, Freedman R, Carr M and Williamson R. (1998). High resolution structure of the N-terminal domain of tissue inhibitor of metalloproteinases-2 and characterization of its interaction site with matrix metalloproteinase-3. *J. Biol. Chem.*, **273:** 21736-21743.

Nagaich AK, Walker DA, Wolford R and Hager GL. (2004). Rapid periodic binding and displacement of the glucocorticoid receptor during chromatin remodeling. *Molecular Cell*, **14:** 163-174.

Nagy L and Schwabe JWR. (2004). Mechanism of the nuclear receptor molecular switch. *Trends Biochem Sci.*, **29:** 317-324.

Nakajima T, Fukamizu A, Takahashi J, Gage FH, Fisher T, Blenis J, *et al.* (1996). The signal-dependent coactivator CBP is a nuclear target for pp90RSK. *Cell*, **86:** 465-74.

Nielsen PJ and Trachsel H. (1988). The mouse protein-synthesis initiation factor-4A gene family includes 2 related functional genes which are differentially expressed. *EMBO J.*, **7:** 2097-2105.

Nightingale KP, O'Neill LP and Turner BM. (2006). Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr. Opin. Genet. Dev.*, **16:** 125-136.

Nikolov DB and Burley SK. (1997). RNA polymerase II transcription initiation: A structural view. *Proc. Natl. Acad. Sci. USA*, **94:** 15-22.

Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, *et al.* (1995). Crystal-structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, **377:** 119-128.

Nishihara E, Yoshida-Komiya H, Chan C, Liao L, Davis R, O'Malley B, *et al.* (2003). SRC-1 null mice exhibit moderate motor dysfunction and delayed development of cerebellar Purkinje cells. *J. Neurosci.*, **23:** 213-222.

Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, *et al.* (1998).

Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature*, **395:** 137-143.

Nowak S and Corces V. (2004). Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet.*, **20:** 214-220.

Oberer M, Marintchev A and Wagner G. (2005). Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev*, **19:** 2212-2223.

Ogawa S, Lozach J, Jepsen K, Sawka-Verhelle D, Perissi V, Sasik R, *et al.* (2004). A nuclear receptor corepressor transcriptional checkpoint controlling activator protein 1-dependent gene networks required for macrophage activation. *Proc. Natl. Acad. Sci. USA*, **101:** 14461-14466.

Ogryzko VV, Schiltz RL, Russanova V, Howard BH and Nakatani Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, **87:** 953-959.

Onate S, Tsai S, Tsai M and O Malley B. (1995). Sequence and characterization of a coactivator for the steroid-hormone receptor superfamily. *Science*, **270:** 1354-1357.

Palamarchuk A, Efanov A, Maximov V, Aqeilan R, Croce C and Pekarsky Y. (2005). Akt phosphorylates and regulates Pdcd4 tumor suppressor protein. *Cancer Res.*, **65:** 11282-11286.

Pause A, Methot N and Sonenberg N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic tranlsation initiation factor-4A is required for RNA-binding and ATP hydrolysis. *Mol. Cell. Biol.*, **13:** 6789-6798.

Pause A, Methot N, Svitkin Y, Merrick WC and Sonenberg N. (1994). Dominant-negative mutants of mammalian translation initiation-factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation. *EMBO J.*, **13:** 1205-1215.

Pause A and Sonenberg N. (1992). Mutational analysis of a DEAD box RNA helicase - The mammalian translation initiation-factor eIF4A. *EMBO J.*, **11:** 2643-2654.

Perissi V and Rosenfeld MG. (2005). Controlling nuclear receptors: The circular logic of cofactor cycles. *Nature Reviews Molecular Cell Biology*, **6:** 542-554.

Perkins ND, Felzien LK, Betts JC, Leung KY, Beach DH and Nabel GJ. (1997). Regulation of NF-kappa B by cyclin-dependent kinases associated with the p300 coactivator. *Science*, **275:** 523-527.

Piotto M, Saudek V and Sklenar V. (1992). Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2:** 661-665.

Polesskaya A, Naguibneva I, Duquet A, Bengal E, Robin P and Harel-Bellan A. (2001).

Interaction between acetylated MyoD and the bromodomain of CBP and/or p300. *Mol. Cell. Biol.*, **21:** 5312-20.

Ponting CP. (2000). Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem. Sci.*, **25:** 423-6.

Preiss T and Hentze MW. (2003). Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays*, **25:** 1201-1211.

Prevot D, Darlix JL and Ohlmann T. (2003). Conducting the initiation of protein synthesis: the role of eIF4G. *Biol. Cell*, **95:** 141-156.

Ptashne M and Gann A. (1997). Transcriptional activation by recruitment. *Nature*, **386:** 569-577.

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

Pyronnet S, Imataka H, Gingras AC, Fukunaga R, Hunter T and Sonenberg N. (1999). Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. *EMBO J.*, **18:** 270-279.

Qin BY, Liu C, Srinath H, Lam SS, Correia JJ, Derynck R, *et al.* (2005). Crystal structure of IRF-3 in complex with CBP. *Structure*, **13:** 1269-1277.

Radhakrishnan I, PerezAlvarado GC, Parker D, Dyson HJ, Montminy MR and Wright PE. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell*, **91:** 741-752.

Ragvin A, Valvatne H, Erdal S, Arskog V, Tufteland KR, Breen K, *et al.* (2004). Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J Mol Biol*, **337:** 773-88.

Razeto A, Ramakrishnan V, Litterst CM, Giller K, Griesinger C, Carlomagno T, *et al.* (2004). Structure of the NCoA-1/SRC-1 PAS-B domain bound to the LXXLL motif of the STAT6 transactivation domain. *J. Mol. Biol.*, **336:** 319-329.

Reid G, Hubner MR, Metivier R, Brand H, Denger S, Manu D, *et al.* (2003). Cyclic, proteasome-mediated turnover of unliganded and liganded ER alpha on responsive promoters is an integral feature of estrogen signaling. *Mol. Cell*, **11:** 695-707.

Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H, *et al.* (1995). Crystal-structure of the RAR-gamma ligand-binding domain bound to all-trans-retinoic acid. *Nature*, **378:** 681-689.

Richter-Cook NJ, Dever TE, Hensold JO and Merrick WC. (1998). Purification and characterization of a new eukaryotic protein translation factor - Eukaryotic initiation factor 4H. *J. Biol. Chem.*, **273:** 7579-7587.

RochetteEgly C, Adam S, Rossignol M, Egly JM and Chambon P. (1997). Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7. *Cell*, **90:** 97-107.

Rogatsky I, Trowbridge JM and Garabedian MJ. (1999). Potentiation of human estrogen receptor alpha transcriptional activation through phosphorylation of serines 104 and 106 by the cyclin A-CDK2 complex. *J. Biol. Chem.*, **274:** 22296-22302.

Rogers GW, Komar AA and Merrick WC. (2002). eIF4A: The godfather of the DEAD box helicases. *Prog. Nucleic Acid Res. Mol. Biol.*, **72:** 307-331.

Rogers GW, Richter NJ, Lima WF and Merrick WC. (2001). Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.*, **276:** 30914-30922.

Rosenfeld MG, Lunyak VV and Glass CK. (2006). Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptionaf response. *Genes Dev*, **20:** 1405-1428.

Rowan BG, Garrison N, Weigel NL and O'Malley BW. (2000). 8-bromo-cyclic AMP induces phosphorylation of two sites in SRC-1 that facilitate ligand-independent activation of the chicken progesterone receptor and are critical for functional cooperation between SRC-1 and CREB binding protein. *Mol. Cell. Biol.*, **20:** 8720-8730.

Rozen F, Edery I, Meerovitch K, Dever TE, Merrick WC and Sonenberg N. (1990). Bidirectional Rna Helicase Activity of eukaryotic translation initiation factor-4a and factor-4F. *Mol. Cell. Biol.*, **10:** 1134-1144.

Rubinstein JH and Taybi H. (1963). Broad thumbs and toes and facial abnormalities. A possible mental retardation syndrome. *Am. J. Dis. Child.*, **105:** 588-608.

Ryan CM, Harries JC, Kindle KB, Collins HM and Heery DM. (2006). Functional interaction of CREB binding protein (CBP) with nuclear transport proteins and modulation by HDAC inhibitors. *Cell Cycle*, **5:** 2146-2152.

Salghetti SE, Caudy AA, Chenoweth JG and Tansey WP. (2001). Regulation of transcriptional activation domain function by ubiquitin. *Science*, **293:** 1651-1653.

Santos-Rosa H and Caldas C. (2005). Chromatin modifier enzymes, the histone code and cancer. *Eur. J. Cancer*, **41:** 2381-2402.

Scheper GC and Proud CG. (2002). Does phosphorylation of the cap-binding protein eIF4E play a role in translation initiation? *Eur. J. Biochem.*, **269:** 5350-5359.

Schwabe JWR, Chapman L, Finch JT and Rhodes D. (1993). The crystal-structure of the estrogen-receptor DNA-binding domain bound to DNA - How receptors discriminate between their response elements. *Cell*, **75:** 567-578.

Scoggin KE, Ulloa A and Nyborg JK. (2001). The oncoprotein Tax binds the SRC-1-interacting domain of CBP/p300 to mediate transcriptional activation. *Mol Cell Biol*, **21:** 5520-30.

Shang YF, Hu X, DiRenzo J, Lazar MA and Brown M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell*, **103:** 843-852.

Sharma D and Fondell JD. (2002). Ordered recruitment of histone acetyltransferases and the TRAP/Mediator complex to thyroid hormone-responsive promoters in vivo. *Proc. Natl. Acad. Sci. USA*, **99:** 7934-7939.

Shaulian E and Karin M. (2002). AP-1 as a regulator of cell life and death. *Nat. Cell Biol.*, **4:** E131-E136.

Sheppard H, Harries J, Hussain S, Bevan C and Heery D. (2001). Analysis of the steroid receptor coactivator 1 (SRC1)-CREB binding protein interaction interface and its importance for the function of SRC1. *Mol. Cell. Biol.*, **21:** 39-50.

Sheppard H, Matsuda S, Harries J, Kindle K and Heery D. (2003). Transcriptional activation by estrogen receptor (ER alpha) and steroid receptor coactivator (SRC1) involves distinct mechanisms in yeast and mammalian cells. *J. Mol. Endocrinol.*, **30:** 411-422.

Shiau A, Barstad D, Loria P, Cheng L, Kushner P, Agard D, *et al.* (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*, **95:** 927-937.

Shibahara K, Asano M, Ishida Y, Aoki T, Koike T and Honjo T. (1995). Isolation of a novel mouse gene MA-3 that is induced upon programmed cell death. *GENE*, **166:** 297-301.

Smeal T, Binetruy B, Mercola DA, Birrer M and Karin M. (1991). Oncogenic and transcriptional cooperation with Ha-Ras requires phosphorylation of c-Jun on serine-63 and serine-73. *Nature*, **354:** 494-496.

Spencer T, Jenster G, Burcin M, Allis C, Zhou J, Mizzen C, *et al.* (1997). Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature*, **389:** 194-198.

Spera S and Bax A. (1991). Empirical correlation between protein backbone conformation and C-alpha and C-beta C-13 nuclear-magnetic-resonance chemical-shifts. *J. Am. Chem. Soc.*, **113:** 5490-5492.

Spiegelman BM and Heinrich R. (2004). Biological control through regulated transcriptional coactivators. *Cell*, **119:** 157-167.

Sreerama N and Woody RW. (1994). Protein secondary structure from circular-dichroism spectroscopy - Combining variable selection principle and cluster-analysis with neural-

network, ridge-regression and self-consistent methods. *J. Mol. Biol.*, **242:** 497-507.

Sreerama N and Woody RW. (2000). Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, **287:** 252-260.

Stein RW, Corrigan M, Yaciuk P, Whelan J and Moran E. (1990). Analysis of E1A-mediated growth-regulation functions - Binding of the 300-kilodalton cellular-product correlates with E1A enhancer repression function and DNA synthesis-inducing activity. *J. Virol.*, **64:** 4421-4427.

Sterner D and Berger S. (2000). Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.*, **64:** 435-459.

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

Svitkin Y, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham G, *et al.* (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5 ' secondary structure. *RNA*, **7:** 382-394.

Taatjes DJ, Marr MT and Tjian R. (2004). Opinion - Regulatory diversity among metazoan co-activator complexes. *Nat. Rev. Mol. Cell Biol.*, **5:** 403-410.

Takeshita A, Yen P, Misiti S, Cardona G, Liu Y and Chin W. (1996). Molecular cloning and properties of a full-length putative thyroid hormone receptor coactivator. *Endocrin.*, **137:** 3594-3597.

Tan S, Hunziker Y, Sargent DF and Richmond TJ. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*, **381:** 127-134.

Tanner N, Cordin O, Banroques J, Doere M and Linder P. (2003). The Q motif: A newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell*, **11:** 127-138.

Tarun SZ and Sachs AB. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.*, **15:** 7168-7177.

Thomas MC and Chiang CM. (2006). The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.*, **41:** 105-178.

Thompson PR, Wang D, Wang L, Fulco M, Pediconi N, Zhang D, *et al.* (2004). Regulation of the p300 HAT domain via a novel activation loop. *Nat. Struct. Mol. Biol.*, **11:** 308-15.

Timmermann S, Lehrmann H, Polesskaya A and Harel-Bellan A. (2001). Histone acetylation and disease. *Cell. Mol. Life Sci.*, **58:** 728-736.

Tora L, White J, Brou C, Tasset D, Webster N, Scheer E, *et al.* (1989). The human

estrogen-receptor has 2 independent nonacidic transcriptional activation functions. *Cell*, **59:** 477-487.

Torchia J, Rose D, Inostroza J, Kamei Y, Westin S, Glass C, *et al.* (1997). The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature*, **387:** 677-684.

Triezenberg SJ. (1995). Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.*, **5:** 190-6.

Troke PJF, Kindle KB, Collins HM and Heery DM. (2006). MOZ fusion proteins in acute myeloid leukaemia. *Transcription*: 23-39.

Turner BM. (2000). Histone acetylation and an epigenetic code. *Bioessays*, **22:** 836-845.

Turner BM, Birley AJ and Lavender J. (1992). Histone-H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in drosophila polytene nuclei. *Cell*, **69:** 375-384.

Ugai H, Uchida K, Kawasaki H and Yokoyama KK. (1999). The coactivators p300 and CBP have different functions during the differentiation of F9 cells. *J. Mol. Med.*, **77:** 481-494.

Underhill C, Qutob MS, Yee SP and Torchia J. (2000). A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. *J. Biol. Chem.*, **275:** 40463-40470.

Uversky VN. (2002). Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.*, **11:** 739-756.

Vaisanen S, Dunlop TW, Sinkkonen L, Frank C and Carlberg C. (2005). Spatio-temporal activation of chromatin on the human CYP24 gene promoter in the presence of 1 alpha,25-dihydroxyvitamin D-3. *J. Mol. Biol.*, **350:** 65-77.

Vendel AC and Lumb KJ. (2003). Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator Tat. *Biochemistry*, **42:** 910-6.

Vendel AC, McBryant SJ and Lumb KJ. (2003). KIX-mediated assembly of the CBP-CREB-HTLV-1 tax coactivator-activator complex. *Biochemistry*, **42:** 12481-7.

Voegel J, Heine M, Tini M, Vivat V, Chambon P and Gronemeyer H. (1998). The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *EMBO J.*, **17:** 507-519.

Voegel J, Heine M, Zechel C, Chambon P and Gronemeyer H. (1996). TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. *EMBO J.*, **15:** 3667-3675.

Volpon L, Osborne MJ, Topisirovic I, Siddiqui N and Borden KL. (2006). Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands. *EMBO J.*, **25:** 5138-49.

Wachtel M, Dettling M, Koscielniak E, Stegmaier S, Treuner J, Simon-Klingenstein K, *et al.* (2004). Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. *Cancer Research*, **64:** 5539-5545.

Wang Z, Rose D, Hermanson O, Liu F, Herman T, Wu W, *et al.* (2000). Regulation of somatic growth by the p160 coactivator p/CIP. *Proc. Natl. Acad. Sci. USA*, **97:** 13549-13554.

Waskiewicz AJ, Johnson JC, Penn B, Mahalingam M, Kimball SR and Cooper JA. (1999). Phosphorylation of the cap-binding protein eukaryotic translation initiation factor 4E by protein kinase Mnk1 in vivo. *Mol. Cell. Biol.*, **19:** 1871-1880.

Waters L, Yue B, Veverka V, Renshaw P, Bramham J, Matsuda S, *et al.* (2006a). Structural diversity in p160/CREB-binding protein coactivator complexes. *J. Biol. Chem.*, **281:** 14787-14795.

Waters LC, Bohm M, Veverka V, Muskett FW, Frenkiel TA, Kelly GP, *et al.* (2006b). NMR assignment and secondary structure determination of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4. *J. Biomol. NMR.* **36:** S5:18.

Waters LC, Veverka V, Bohm M, Schmedt T, Choong PT, Muskett FW, *et al.* (2007). Structure of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4 and characterisation of its interaction with eIF4A. *Oncogene*: (Accepted).

Webb P, Nguyen P, Shinsako J, Anderson C, Feng W, Nguyen M, *et al.* (1998). Estrogen receptor activation function 1 works by binding p160 coactivator proteins. *Mol. Endocrinol.*, **12:** 1605-1618.

Webster NJG, Green S, Jin JR and Chambon P. (1988). The Hormone-Binding Domains of the Estrogen and Glucocorticoid Receptors Contain an Inducible Transcription Activation Function. *Cell*, **54:** 199-207.

Wei CC, Balasta ML, Ren JH and Goss DJ. (1998). Wheat germ poly(A) binding protein enhances the binding affinity of eukaryotic initiation factor 4F and (iso)4F for cap analogues. *Biochemistry*, **37:** 1910-1916.

Weiss R, Xu J, Ning G, Pohlenz J, O'Malley B and Refetoff S. (1999). Mice deficient in the steroid receptor co-activator 1 (SRC-1) are resistant to thyroid hormone. *EMBO J.*, **18:** 1900-1904.

White JH, Fernandes I, Mader S and Yang XJ. (2004). Corepressor recruitment by agonist-

bound nuclear receptors. *Vitam. Horm.*, **68:** 123-143.

Williamson R, Carr M, Frenkiel T, Feeney J and Freedman R. (1997). Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry*, **36:** 13882-13889.

Wishart DS and Sykes BD. (1994). The C-13 Chemical-Shift Index - a simple method for the identification of protein secondary structure using C-13 chemical-shift data. *J. Biomol. NMR*, **4:** 171-180.

Wishart DS, Sykes BD and Richards FM. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol*, **222:** 311-33.

Wishart DS, Sykes BD and Richards FM. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31:** 1647-51.

Wittekind M and Mueller L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J. Magn. Reson. B.*, **101:** 201-205.

Wright PE and Dyson HJ. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, **293:** 321-331.

Wu H, Hamamori Y, Xu J, Chang S, Saluna T, Chang M, *et al.* (2005). Nuclear hormone receptor coregulator GRIP1 suppresses, whereas SRC1A and p/CIP coactivate, by domain-specific binding of MyoD. *J. Biol. Chem.*, **280:** 3129-3137.

Wu RC, Qin J, Yi P, Wong JM, Tsai SY, Tsai MJ, *et al.* (2004). Selective phosphorylations of the SRC-3/AIB1 coactivator integrate genomic reponses to multiple cellular signaling pathways. *Mol. Cell*, **15:** 937-949.

Wu SY and Chiang CM. (1998). Properties of PC4 and an RNA polymerase II complex in directing activated and basal transcription in vitro. *J. Biol. Chem.*, **273:** 12492-12498.

Wuthrich K. (1986). *NMR or Proteins and Nucleic Acids.: Polypeptide Secondary Structures in Proteins by NMR.* Wuthrich K (ed.). John Wiley and Sons Inc.: New York, pp 117-129.

Xu J and Li Q. (2003). Review of the in vivo functions of the p160 steroid receptor coactivator family. *Mol. Endocrinol.*, **17:** 1681-1692.

Xu J, Liao L, Ning C, Yoshida-Komiya H, Deng C and O'Malley B. (2000). The steroid receptor coactivator SRC-3 (p/CIP/RAC3/AIB1/ACTR/TRAM-1) is required for normal growth, puberty, female reproductive function, and mammary gland development. *Proc. Natl. Acad. Sci. USA*, **97:** 6379-6384.

Xu J and O'Malley BW. (2002). Molecular mechanisms and cellular biology of the steroid receptor coactivator (SRC) family in steroid receptor function. *Rev. Endocr. Metab. Disord.*, **3:** 185-92.

Xu J, Qiu Y, DeMayo F, Tsai S, Tsai M and O'Malley B. (1998). Partial hormone resistance in mice with disruption of the steroid receptor coactivator-1 (SRC-1) gene. *Science*, **279:** 1922-1925.

Xu W, Chen HW, Du KY, Asahara H, Tini M, Emerson BM, *et al.* (2001). A transcriptional switch mediated by cofactor methylation. *Science*, **294:** 2507-2511.

Yamamoto Y, Singh CR, Marintchev A, Hall NS, Hannig EM, Wagner G, *et al.* (2005). The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G. *Proc. Natl. Acad. Sci. USA*, **102:** 16164-9.

Yang H, Cho M, Zakowicz H, Hegamyer G, Sonenberg N and Colburn N. (2004). A novel function of the MA-3 domains in transformation and translation suppressor Pdcd4 is essential for its binding to eukaryotic translation initiation factor 4A. *Mol. Cell. Biol.*, **24:** 3894-3906.

Yang H, Jansen A, Komar A, Zheng X, Merrick W, Costes S, *et al.* (2003a). The transformation suppressor Pdcd4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation. *Mol. Cell. Biol.*, **23:** 26-37.

Yang H, Jansen A, Nair R, Shibahara K, Verma A, Cmarik J, *et al.* (2001). A novel transformation suppressor, Pdcd4, inhibits AP-1 transactivation but not NF-kappa B or ODC transactivation. *Oncogene*, **20:** 669-676.

Yang H, Knies J, Stark C and Colburn N. (2003b). Pdcd4 suppresses tumor phenotype in JB6 cells by inhibiting AP-1 transactivation. *Oncogene*, **22:** 3712-3720.

Yang HS, Matthews CP, Clair T, Wang Q, Baker AR, Li CCH, *et al.* (2006). Tumorigenesis suppressor Pdcd4 down-regulates mitogen-activated protein kinase kinase kinase kinase 1 expression to suppress colon carcinoma cell invasion. *Mol. Cell. Biol.*, **26:** 1297-1306.

Yang XJ, Ogryzko VV, Nishikawa J, Howard BH and Nakatani Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature*, **382:** 319-324.

Yao TP, Ku G, Zhou ND, Scully R and Livingston DM. (1996). The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *Proc. Natl. Acad. Sci. USA*, **93:** 10626-10631.

Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, *et al.* (1998). Gene dosage-

dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell*, **93:** 361-372.

Yoderhill J, Pause A, Sonenberg N and Merrick WC. (1993). The p46 subunit of eukaryotic initiation-factor (eIF)-4F exchanges with eIF-4A. *J. Biol. Chem.*, **268:** 5566-5573.

Yoshinaga H, Matsuhashi S, Fujiyama C and Masaki Z. (1999). Novel human PDCD4(H731) gene expressed in proliferative cells is expressed in the small duct epithelial cells of the breast as revealed by an anti-H731 antibody. *Pathol. Int.*, **49:** 1067-1077.

Zakowicz H, Yang H, Stark C, Wlodawer A, Laronde-Leblanc N and Colburn N. (2005). Mutational analysis of the DEAD-box RNA helicase eIF4AII characterizes its interaction with transformation suppressor Pdcd4 and eIF4GI. *RNA*, **11:** 261-274.

Zhang H, Yi X, Sun X, Yin N, Shi B, Wu H, *et al.* (2004). Differential gene regulation by the SRC family of coactivators. *Genes Dev*, **18:** 1753-1765.

Zhang Y, Ohyashiki JH, Takaku T, Shimizu N and Ohyashiki K. (2006). Transcriptional profiling of Epstein-Barr virus (EBV) genes and host cellular genes in nasal NK/T-cell lymphoma and chronic active EBV infection. *Br. J. Cancer*, **94:** 599-608.

Zhang Y and Reinberg D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*, **15:** 2343-2360.

Zuiderweg ERP, Mcintosh LP, Dahlquist FW and Fesik SW. (1990). 3-Dimensional C-13-resolved proton NOE spectroscopy of uniformly C-13-labeled proteins for the NMR assignment and structure determination of larger molecules. *J. Magn. Reson.*, **86:** 210-216-216.

PUBLISHED PAPERS

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Carr, M. and Heery, D. M. (2006). Structural Diversity in p160/CREB-binding Protein Coactivator Complexes. *J. Biol. Chem.*, **281:** 14787-14795.

Waters, L. C., Bohm, M., Veverka, V., Muskett, F. W., Frenkiel, T. A., Kelly, G. P., Prescott, A., Dosanjh, N. S., Klempnauer, K. H. and Carr, M. D. (2006). NMR Assignment and Secondary Structure Determination of the C-terminal MA-3 Domain of the Tumour Suppressor Protein Pdcd4. *J. Biomol. NMR*, **36:** S5:18.

Waters, L. C., Veverka, V., Bohm, M., Schmedt, T., Choong, P. T., Muskett, F. W., Klempnauer, K. H. and Carr, M. D. (2007). Structure of the C-terminal MA-3 domain of the Tumour Suppressor Protein Pdcd4 and Characterisation of its Interaction with eIF4A. *Oncogene*, (Published Online).

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 21, pp. 14787–14795, May 26, 2006

Structural Diversity in p160/CREB-binding Protein Coactivator Complexes*

Received for publication, January 10, 2006, and in revised form, March 14, 2006 Published, JBC Papers in Press, March 15, 2006, DOI 10.1074/jbc.M600237200

Lorna Waters[‡1,2], **Baigong Yue**[§2], **Vaclav Veverka**[‡], **Philip Renshaw**[‡], **Janice Bramham**[‡], **Sachiko Matsuda**[‡], **Thomas Frenkiel**[¶], **Geoffrey Kelly**[¶], **Frederick Muskett**[‡], **Mark Carr**[‡3], **and David M. Heery**[§4]

From the [‡]*Department of Biochemistry, Henry Wellcome Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, United Kingdom,* [§]*School of Pharmacy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom, and* [¶]*Medical Research Council Biomedical NMR Centre, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom*

Ligand-induced transcription by nuclear receptors involves the recruitment of p160 coactivators such as steroid receptor coactivator 1 (SRC1), in complex with histone acetyltransferases such as CREB-binding protein (CBP) and p300. Here we describe the solution structure of a complex formed by the SRC1 interaction domain (SID) of CBP and the activation domain (AD1) of SRC1, both of which contain four helical regions (Cα1, Cα2, Cα3, and Cα3′ in CBP and Sα1, Sα2′, Sα2, and Sα3 in SRC1). A tight four-helix bundle is formed between Sα1, Cα1, Cα2, and Cα3 that is capped by Sα3. In contrast to the structure of the AD1 domain of the related p160 protein ACTR in complex with CBP SID, the sequences forming Sα2′ and Sα2 in SRC1 AD1 are not involved in the interface between the two domains but rather serve to position Sα3. Thus, although the CBP SID domain adopts a similar fold in complex with different p160 proteins, the topologies of the AD1 domains are strikingly different, a feature that is likely to contribute to functional specificity of these coactivator complexes.

The lysine acetyltransferase CBP[5] interacts with a large number of nuclear proteins, many of which are transcription factors (1, 2). This is achieved through direct or indirect protein-protein interactions that are mediated by distinct structural domains within the CBP protein such as the KIX, CRD, CH1, CH3, bromodomain and SID domains. The protein-binding domains of CBP display partial specificity, having both distinct and overlapping binding partner profiles, which contributes to the phenomena of synergy and cross-talk between transcription factors (1).

The recruitment of CBP to target gene promoters/enhancers facilitates acetylation of histone N-terminal tails, leading to chromatin remodeling and enhanced gene expression. This has been demonstrated for nuclear receptors, which activate transcription of their target genes in response to ligand binding (3, 4). Ligand-bound receptors undergo a conformational change that stimulates their interaction with cofactors that contain functional LXXLL motifs, such as the p160 coactivators SRC1, TIF2, and ACTR (5, 6). Studies of steroid-regulated gene promoters have revealed that p160s and HAT proteins are among the first cofactors recruited in response to ligand (7, 8). Such temporally ordered recruitment of coactivators to promoters/enhancers is crucial for the sequential chromatin modification and remodeling events preceding transcription (4).

The efficacy of nuclear receptor/cofactor interaction is influenced by a number of determinants, including the precise sequence and number of LXXLL motifs, the sequences flanking core motifs, and other distal sequences (9–11). CBP contains three LXXLL motifs, although they mediate only weak direct interactions with estrogen, androgen, and progesterone receptors (9, 10). Thus, at least in the case of the steroid receptors, efficient recruitment of CBP/p300 and associated factors is achieved indirectly via the p160 proteins (12).

A number of studies have shown that recruitment of CBP/p300 proteins is facilitated by the p160 activation domain AD1, which acts as a potent transcriptional activator in mammalian and yeast cells (3, 12–17). The AD1 domain docks with the SRC1 interaction domain (SID) of CBP, located within the sequence 2058–2130 (12), and this interaction is essential for ligand-dependent transcription mediated by steroid receptors (12, 18). In addition to binding p160s, the SID also facilitates interactions of CBP with activation domains in other nuclear proteins, including the transcription factors Ets1, Ets2, p53, and IRF3, and viral activators such as E1A, KSHV IRF1, and Tax (19–22). Indeed, it has been shown that competition between such proteins for binding to the SID contributes to the negative cross-talk observed between different signaling pathways (22). Similarly, binding of viral proteins to the SID and other CBP/p300 domains not only facilitates viral gene transcription but may also down-regulate expression of host defense genes, through exclusion of host factors from CBP-p300 complexes.

The AD1/SID interaction is also important in MOZ-TIF2-associated leukemogenesis. MOZ-TIF2 is a fusion protein expressed in acute myeloid leukemia blasts containing the *inv*8(p11;q13) translocation (23). This fusion protein contains the C-terminal sequence of TIF2, including the AD1 domain, facilitating its interaction with CBP/p300, as demonstrated by *in vitro* interactions and *in vivo* by using fluorescence resonance energy transfer experiments (24). As a consequence of this interaction, CBP is mislocalized from promyelocytic leukemia bodies, and cellular levels of CBP are depleted, leading to a reduced transcriptional

* This work was supported in part by Grant 054401/Z/98/B from the Wellcome Trust (to D. M. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2C52) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Supported by a Ph.D. studentship from the Biotechnology and Biological Sciences Research Council.

[2] Both authors should be considered as equal first authors.

[3] Supported by Wellcome Trust Grants 066047 and 063632. To whom correspondence may be addressed: Dept. of Biochemistry, Henry Wellcome Bldg., University of Leicester, Lancaster Rd., Leicester LE1 9HN, UK. Tel.: 44-116-229-7075; Fax: 44-116-229-7018; E-mail: mdc12@le.ac.uk.

[4] To whom correspondence may be addressed: School of Pharmacy, University of Nottingham, University Park, Nottingham NG7 2RD, UK. Tel.: 44-115-951-5087; Fax: 44-115-846-6249; E-mail: david.heery@nottingham.ac.uk.

[5] The abbreviations used are: CBP, CREB-binding protein; CREB, cAMP-response element-binding protein; SID, SRC1 interaction domain; AD1, activation domain; r.m.s.d., root mean standard deviation; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single quantum coherence; TOCSY, total correlation spectroscopy; HPLC, high pressure liquid chromatography; IRF, interferon regulatory factor; IAD, IRF association domain; SRC1, steroid receptor coactivator 1.

activity of CBP-dependent activators such as nuclear receptors and p53 (24). Consistent with this, AD1 integrity was found to be essential for transformation of hematopoietic progenitors by MOZ-TIF2 *in vitro* (24, 25) and for induction of acute myeloid leukemias by MOZ-TIF2 in mice (25). Therefore, understanding the structure of the CBP-p160 complexes has relevance to the etiology of acute myeloid leukemia.

Phenotypic differences in knock-out mice indicate that p160s have tissue-specific functions (reviewed in Ref. 26). Similarly, CBP and p300 proteins appear to have distinct roles *in vivo*, for example in hematopoiesis (27). Thus, the existence of different p160-CBP or p160-p300 complexes *in vivo* suggests they may have specific albeit partially redundant functions. For example, SRC1 and ACTR interact with thymine DNA glycosylase, whereas TIF2 does not. This is because of the presence of a tyrosine repeat motif (YXXY) in ACTR and SRC1, which is not conserved in the TIF2 sequence (28). Chromatin immunoprecipitation assays investigating cofactor recruitment at the *pS2* gene promoter have indicated that the presence of one p160 can exclude recruitment of others (7), although the molecular basis of this is unknown. Another study provided evidence that CBP-p160 complexes are functionally distinct, as it was observed that different p160 combinations can be detected on androgen receptor target gene promoters, possibly through the formation of specific p160 heterodimer pairs (29). Thus, determination of the structures of different complexes will be essential to understand how such selectivity is achieved. In this study, we describe a high resolution solution structure of the complex formed between the SID domain of CBP and the AD1 domain of SRC1. By comparison to the CBP SID/ACTR AD1 structure (30), we show that although the structure of CBP SID is strikingly similar in both complexes, p160 proteins adopt distinct conformations that are likely to be important for their different biological functions.

EXPERIMENTAL PROCEDURES

Protein Expression/Purification—To express the CBP/SRC1 interaction domain complex in *Escherichia coli*, we obtained a modified pET22b dual expression vector, containing sequences encoding the ACTR AD1 and CBP SID domains, preceded by independent ribosome-binding sites (a gift from Peter Wright, Scripps Institute). To generate the CBP SID/SRC1 AD1 expression vector, the ACTR AD1 sequence was removed by restriction digestion with NcoI and HindIII and replaced with an NcoI/HindIII-digested PCR fragment containing the coding region for residues 920–970 of SRC1 followed by a thrombin cleavage site (underlined) and a polyhistidine tag (KLVPRGSLEHHHHHH). The mouse CBP SID-(2059–2117) and human SRC1 AD1-(920–970) dual expression vector was transformed into *E. coli* strain B834, which was used to produce either unlabeled, ^{15}N-labeled, or ^{15}N/^{13}C-labeled samples of the CBP SID-SRC1 AD1 complex. For labeled samples, cells were grown in minimal media containing 0.6 g/liter [^{15}N]ammonium sulfate and 2 g/liter [^{13}C]glucose as the sole nitrogen and/or carbon sources. The medium was also supplemented with 50 mg/liter unlabeled methionine, as B834 is a methionine auxotroph. Transformants were grown at 37 °C, and expression of the two proteins forming the complex was induced in mid-log phase by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside. The cells were harvested 3 h post-induction, resuspended in 20 mM sodium phosphate, 100 mM sodium chloride, and 5 mM imidazole buffer (pH 7.0). Lysis was achieved by sonication. Insoluble material was removed by high speed centrifugation, and the cleared lysate was filtered through a 0.2-μm filter prior to chromatography. The histidine-tagged CBP SID-SRC1 AD1 complex was affinity-purified on a HiTrap chelating HP column (Amersham Biosciences) charged with Ni^{2+} and eluted with a linear gradient of imidazole (5–250 mM). Fractions containing the complex were pooled and subjected to a final purification by gel filtration on a Superdex 75 16/60 pre-packed column (Amersham Biosciences). Typical yields of the >95% pure complex were 10 mg/liter.

Reverse Phase HPLC—Reverse phase HPLC analysis was performed on a C4 column. The individual polypeptides were resolved by applying a two-step linear gradient of acetonitrile (1.6–80%), and the relative amount of each was quantified by absorbance at 215 nm. Eluted peaks were identified by mass spectrometry.

Circular Dichroism Spectroscopy—Far UV CD spectra were obtained from 0.4 mM samples of the CBP SID-SRC1 AD1 complex in a 20 mM sodium phosphate, 100 mM sodium chloride buffer (pH 7.0). Samples were placed in a 0.1-mm path length cell, and spectra were recorded from 180 to 250 nm at a resolution of 1 nm and a scan speed of 20 nm/min, with each spectrum representing the average of 10 accumulations. Prior to analysis, CD spectra were corrected for buffer absorbance, and the raw data were converted to molar CD per residue.

NMR Spectroscopy—NMR spectra were acquired from 0.35-ml samples of 1.5 mM CBP SID-SRC1 AD1 complex in a 20 mM sodium phosphate, 100 mM sodium chloride, 10 μM EDTA, and 0.02% (w/v) sodium azide buffer (pH 7.0), containing either 10% D_2O, 90% H_2O, or 100% D_2O as appropriate. All NMR data were acquired at 25 °C on either an 800-MHz Varian Inova or a 600-MHz Bruker Avance spectrometer. The two-dimensional and three-dimensional spectra recorded to obtain sequence-specific assignments for CBP SID and SRC1 AD1 in complex were as follows: ^{1}H TOCSY with mixing times of 40 and 55 ms (31) and NOESY with an NOE mixing time of 100 ms (32); ^{15}N/^{1}H HSQC; TOCSY-HSQC with a mixing time of 50 ms and NOESY-HSQC with an NOE mixing time of 100 ms (33); ^{13}C/^{1}H HCCH-TOCSY with a mixing time of 20 ms (34), HMQC-NOESY with an NOE mixing time of 100 ms (35); and ^{15}N/^{13}C/^{1}H HNCACB (36) and CBCA(CO)NH (37). Typical acquisition times in F_1 and F_2 for the three-dimensional experiments were 11–13 ms for ^{15}N, 7.5–9.5 ms for ^{13}C, and 15 ms for ^{1}H, and an acquisition time of 75 ms in F_3 (^{1}H). The majority of the three-dimensional spectra were collected over ~88 h, two-dimensional ^{1}H experiments over 8.5–24 h, and ^{15}N/^{1}H HSQC spectra over about 30 min. Typical acquisition times in two-dimensional experiments were either 70 (^{15}N) or 35 ms (^{1}H) in F_1 and 250 ms in F_2 (^{1}H). The WATERGATE method (38) was used to suppress the water signal when required. The three-dimensional NMR data were processed using NMRPipe (39) with linear prediction used to extend the effective acquisition times by up to 1.5–2-fold in F_1 and F_2. The spectra were analyzed using the XEASY package (40).

Structure Calculations—A family of converged CBP SID/SRC1 AD1 structures was determined in a two-stage process using the program CYANA (41). Initially, the combined automated NOE assignment and structure determination protocol (CANDID) was used to automatically assign both the intra- and intermolecular NOE cross-peaks identified in three-dimensional ^{15}N- and ^{13}C-edited NOESY spectra. This approach provides a completely unbiased assignment of the NOE peaks, in which all peaks are evaluated as either intra- or intermolecular in origin. Subsequently, several cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce the final converged CBP SID/SRC1 AD1 structures (42). The input for the CANDID stage consisted of essentially complete ^{15}N, ^{13}C, and ^{1}H resonance assignments for the nonexchangeable groups in the CBP SID-SRC1 AD1 complex, two manually picked three-dimensional NOE peak lists corresponding to all NOEs involving amide protons (1179) and all NOEs between aliphatic protons (2371), and one manually picked two-dimensional NOE peak list corresponding to all NOEs involving aromatic side chain protons (127). In addition, the CANDID stage included

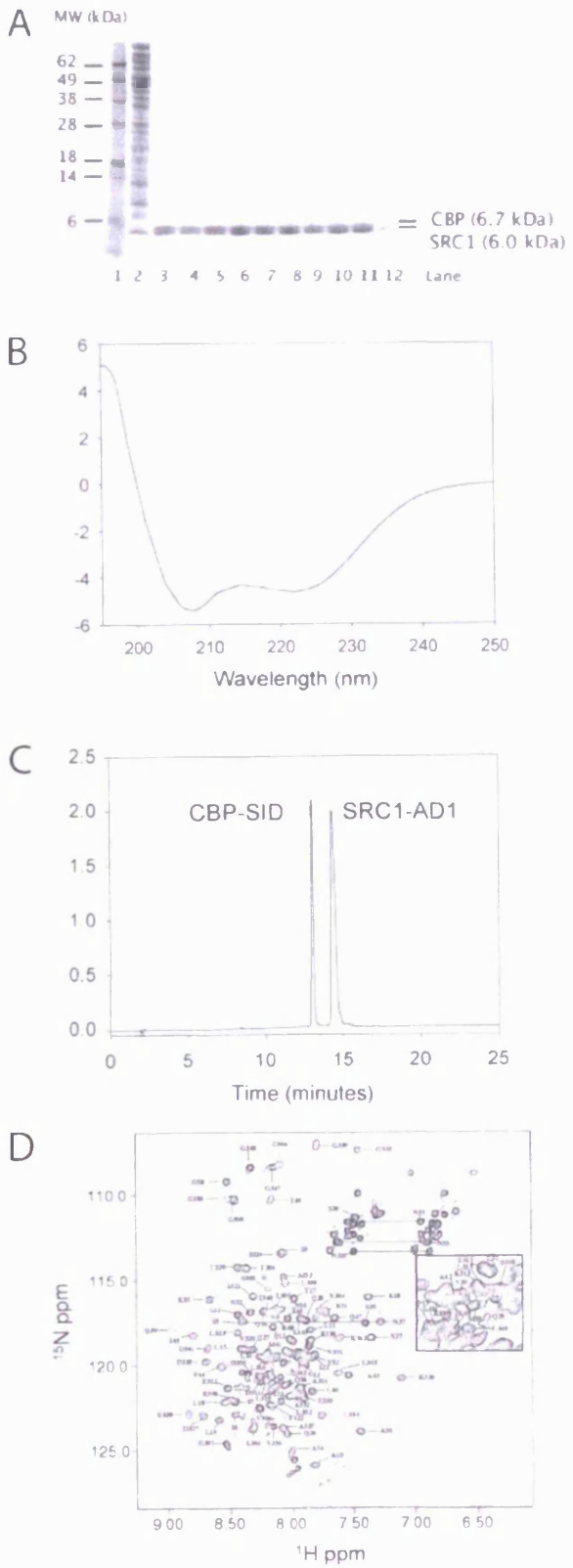


FIGURE 1. **Purification and characterization of the CBP SID-SRC1 AD1 complex.** *A*, 15% SDS-polyacrylamide gel showing separation of whole cell lysate from *E. coli* B834 expressing CBP SID-SRC1 AD1 complex (*lane 2*). The His_6-tagged complex was purified by affinity chromatography. *Lanes 3–12* show fractions containing the purified complex after gel filtration. Molecular weight markers are shown in *lane 1*. *B*, far UV circular dichroism spectrum acquired for the CBP SID-SRC1 AD1 complex. *C*, reverse phase HPLC of the purified CBP SID-SRC1 AD1 complex. The areas under the two peaks were used to calculate the ratio of the CBP SID domain to SRC1 AD1 domain in the purified complex. *D*, $^{15}N/^{1}H$ HSQC spectrum of the CBP SID-SRC1 AD1 complex. The assignments of the signals from backbone amide groups in both domains are indicated by residue type and number, with the overlapped region between 7.85 and 8.35 ppm in ^{1}H and 118.5 and 121.5 ppm in ^{15}N shown in the expanded region to the *right* of the complete spectrum. For clarity, residues are numbered according to their position in the domain. Thus, CBP residues Pro^{2059}–Gln^{2117} are numbered *P2–Q60*, and SRC1 residues Pro^{920}–Ser^{970} are numbered *P303–S353*, with the C-terminal linker to the His tag numbered 354–361. Assignments obtained for a number of side chain NH_2 groups are indicated on the spectrum.

80 torsion angle constraints for CBP SID and 64 torsion angle constraints for SRC1 AD1 obtained from analysis of chemical shift data using TALOS (43). Hydrogen bond constraints were also added for regions predicted to contain regular α-helical secondary structure by both the NOE and chemical shift data, which corresponds to residues Ala^{2067}–Leu^{2075}, Gln^{2082}–Lys^{2092}, and Ala^{2099}–Ala^{2107} of CBP SID and Glu^{929}–Ser^{941} of SRC1 AD1. The peak lists were prepared using XEASY, and the intensities of peaks were obtained using the "interactive integration" routine (40). CANDID calculations were carried out using the default parameter settings in CYANA 1.0.6, with the upper limit for NOE-derived distance constraints set at 5.5 Å and chemical shift tolerances set to 0.02 ppm (direct and indirect ^{1}H) and 0.3 ppm (^{15}N and ^{13}C). The final converged CBP SID/SRC1 AD1 structures were produced from 100 random starting coordinates using a standard torsion angle-based simulated annealing protocol combined with 5 cycles of redundant dihedral angle constraints (44, 45). The calculations were based upon 1855 nonredundant NOE-derived upper distance limits (maximum value 6.0 Å), assigned to unique pairs of protons using CANDID, 142 Φ and Ψ torsion angle constraints derived from TALOS, and 104 hydrogen bond constraints for helical residues. Analysis of the family of structures obtained was carried out using the programs CYANA and MOLMOL (41, 46)

Sequence-specific Assignments—Sequence-specific backbone resonance assignments (N, NH, Cα, and Cβ) were obtained for the CBP SID-SRC1 AD1 complex from the identification of intra- and inter-residue connectivities in HNCACB, CBCA(CO)NH, and $^{15}N/^{1}H$ NOESY-HSQC spectra. Assignments were then extended to the side chain signals using correlations observed primarily in $^{15}N/^{1}H$ TOCSY-HSQC and $^{13}C/^{1}H$ HCCH-TOCSY, with additional supporting evidence provided by $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ HMQC-NOESY spectra where required.

RESULTS AND DISCUSSION

Expression and Purification of the SRC1-CBP Complex—We previously used yeast two-hybrid and *in vitro* pulldown experiments to define the minimal sequences required for interaction of the CBP SID and SRC1 AD1 domains (12). A modified pET22b dual expression vector (30), containing sequences encoding the SRC1 AD1-(920–970) and CBP SID-(2059–2117) domains, was used to coexpress these polypeptides in *E. coli*. The addition of a polyhistidine tag at the C terminus of the AD1 domain facilitated purification of the complex, as described under "Experimental Procedures." A highly purified complex containing polypeptides of the expected size was obtained, as determined by SDS-PAGE (Fig. 1*A*). Far UV CD spectra acquired for the CBP SID-SRC1 AD1 complex indicated that the complex was predominantly helical, having characteristic negative ellipticity peaks at ~209 and 221 nm and a large positive peak at 195 nm (Fig. 1*B*). Analysis of the spectra using the CDPro software package (47) indicated that the complex con-

TABLE 1
NMR constraints and structural statistics for the CBP SID-SRC1 AD1 complex

(*a*) No. of constraints used in the final structural calculations		
Intraresidue NOEs	328	
Sequential NOEs ($i,i+1$)	514	
Medium range NOEs ($i,i \leq 4$)	643	
Intramolecular long range NOEs ($i,i \geq 5$)	166	
Intermolecular long range NOEs ($i,i \geq 5$)	108	
Torsion angles	142	(71Φ and 71Ψ)
Hydrogen bonding	104	
(*b*) Maximum and total constraint violations in 37 converged CBP SID/SRC1 AD1 structures		
Upper distance limits (Å)	0.40 ± 0.05	20.7 ± 1.84
Lower distance limits (Å)	0.24 ± 0.08	1.5 ± 0.24
van der Waals contacts (Å)	0.28 ± 0.04	14.8 ± 1.55
Torsion angle ranges (°)	3.44 ± 0.64	33.8 ± 6.53
Average CYANA target function (Å^2)	6.17 ± 0.95	
(*c*) Structural statistics for the family of converged CBP SID/ SRC1 AD1 structures		
Residues within allowed region of Ramachandran plot (%)		98
Backbone atom r.m.s.d. for structured region (residues 2063–2113 of CBP SID and 928–963 of SRC1 AD1)		0.49 ± 0.10 Å
Heavy atom r.m.s.d. for structured region (residues 2063–2113 of CBP SID and 928–963 of SRC1 AD1)		0.95 ± 0.09 Å

tained 54% (±7.3%) helical, 20% (±6.1%) turn, and 25% (±5.7%) random coil secondary structure.

Reverse phase HPLC analysis was performed to determine the stoichiometry of the purified complex. The eluted peaks were identified by mass spectrometry (data not shown). Quantification of the peptide bond absorbance from both peaks confirmed the formation of a 1:1 CBP SID-SRC1 AD1 complex (Fig. 1*C*).

Sequence-specific Assignments and Structure Calculation for the CBP-SRC1 Complex—Very comprehensive sequence-specific resonance assignments were obtained for the CBP SID-SRC1 AD1 complex despite the relatively poor dispersion observed in spectra, which is illustrated by the HSQC spectrum (Fig. 1*D*). For example, backbone amide assignments were obtained for all non-proline residues in the complex except as follows: Asn^{2060}, Arg^{2061}, and Gln^{2117} in CBP SID; Asn^{927} in SRC1 AD1 (96%); and for all Cα and Cβ signals apart from the two unlabeled methionines and two residues in CBP SID (Pro^{2059} and Gln^{2117}) (96%).

The CANDID protocol was effective in determining unique assignments for the NOEs identified in the three-dimensional ^{15}N- and ^{13}C-edited NOESY and the aromatic to aliphatic region of the two-dimensional NOESY. At the end of the final cycle, unique assignments were obtained for 89.8% (1059/1179) of the NOE peaks picked in the ^{15}N/^{1}H NOESY-HSQC spectra, 89.3% (2117/2371) in the ^{13}C/^{1}H HMQC-NOESY spectra, and 92.1% (117/127) in the NOESY spectrum, which produced 1759 nonredundant ^{1}H to ^{1}H upper distance limits. The final family of CBP SID-SRC1 AD1 complex structures was determined using a total of 2005 NMR-derived structural constraints (an average of 18.2 per residue), including 1759 NOE-based upper distance limits (328 intraresidue, 514 sequential ($i,i+1$), 643 medium range ($i,i \leq 4$), and 274 long range ($i,i \geq 5$) comprising 166 intramolecular NOEs and 108 intermolecular NOEs), 142 backbone torsion angle constraints (71Φ and 71Ψ), and 104 hydrogen bond constraints. Following the final round of CYANA calculations, 37 satisfactorily converged structures were obtained from 100 random starting structures. The converged structures contain no distance or van der Waals violation greater than 0.5 Å and no dihedral angle violations greater than 5°, with an average value for the CYANA target function of 6.17 ± 0.95Å^2. The sums of the violations for the upper distance limits, lower distance limits, van der Waals contacts, and torsion angle constraints were 20.7 ± 1.84 Å, 1.5 ± 0.24 Å, 14.8 ± 1.55 Å, and 33.8 ± 6.53°, respectively. Similarly, maximum violations for the converged structures were 0.40 ± 0.05 Å, 0.24 ± 0.08 Å, 0.28 ± 0.04 Å, and 3.44 ± 0.64°, respectively. The NMR constraints and structural statistics for the complex are summarized in Table 1. The family of converged CBP SID-SRC1 AD1 complex structures, together with the NMR constraints, have been deposited in the Protein Data Bank (code 2C52).

Structure of the CBP-SRC1 Complex—The solution structure of the CBP SID-SRC1 AD1 complex was determined to high precision, as evident from the superposition of the protein backbone of the family of converged structures (best fit for residues 2063–2113 of CBP SID and 928–963 of SRC1 AD1; see Fig. 2*A*) and reflected in low root mean standard deviation (r.m.s.d.) values to the mean structure for both the backbone (0.49 ± 0.10 Å) and all heavy atoms (0.95 ± 0.09 Å). The backbone topology of the complex is composed primarily of eight helices linked by turns and loops, including four α-helices in the CBP SID domain (Cα1, Ser^{2066}–Lys^{2076}; Cα2, Gln^{2082}–Ser^{2093}; Cα3, Pro^{2095}–Thr^{2106}, and Cα3′, Tyr^{2109}–Asn^{2112}) and three α-helices in the SRC1 AD1 domain (Sα1, Glu^{929}–Ser^{941}; Sα2′, Glu^{945}–Leu^{948}, and Sα3, Ile^{957}–Gln^{962}). SRC1 AD1 also contains a short 3_{10} helix (Sα2, Glu^{950}–Leu^{955}). The total helical content of the complex is just over 50%, in close agreement with the CD analysis (54% helical structure; see Fig. 1*B*).

The CBP SID and SRC1 AD1 domains are intimately associated in the complex, with Cα1, Cα2, Cα3, and Sα1 forming a four-helix bundle (Fig. 2*B*). The overall interface between the two domains is substantial, corresponding to a solvent-inaccessible surface of 1019 Å^2 on CBP SID and 1088 Å^2 on SRC1 AD1. Helices Sα2′ and Sα2 pack together on one side of the complex to stabilize the corner region of the "L"-shaped domain and are not involved in the interface with CBP SID. In contrast, Sα3 serves to cap the four-helix bundle region and forms contacts with Cα2, Cα3, Cα3′, and the two serine residues in the PSSP turn of CBP SID. The first two helices of CBP SID (Cα1 and Cα2) lie almost antiparallel to each other and are separated by a well defined five-residue turn consisting of the sequence SPSSP. The N-terminal portion of Cα2 contains a stretch of five glutamines preceding the more hydrophobic C-terminal half of the helix. Cα3 is inclined away from Cα1 and Cα2 exposing a hydrophobic groove between Cα1 and Cα3, which accommodates Sα1 of the SRC1 AD1 domain. Helices Cα3 and Cα3′ are separated by a short region of irregular structure and resemble a single bent helix. The stabilizing interactions between Cα1 and Cα2 and between Cα2 and Cα3 appear to involve no ionic or hydrogen bonds but rely on favorable van der Waals contacts, primarily involving residues Ala^{2067}, Leu^{2071}, and Thr^{2074} from Cα1; Gln^{2083}, Val^{2087}, Ile^{2090}, and Leu^{2091} from Cα2; and Leu^{2097}, Phe^{2101}, and Thr^{2106} from Cα3. Asn^{2094} also makes van der Waals interactions with both Cα2 and Cα3.

The SRC1 AD1 domain also appears to be primarily stabilized by van der Waals interactions, involving residues Ala^{931} and Gln^{935} from Sα1,

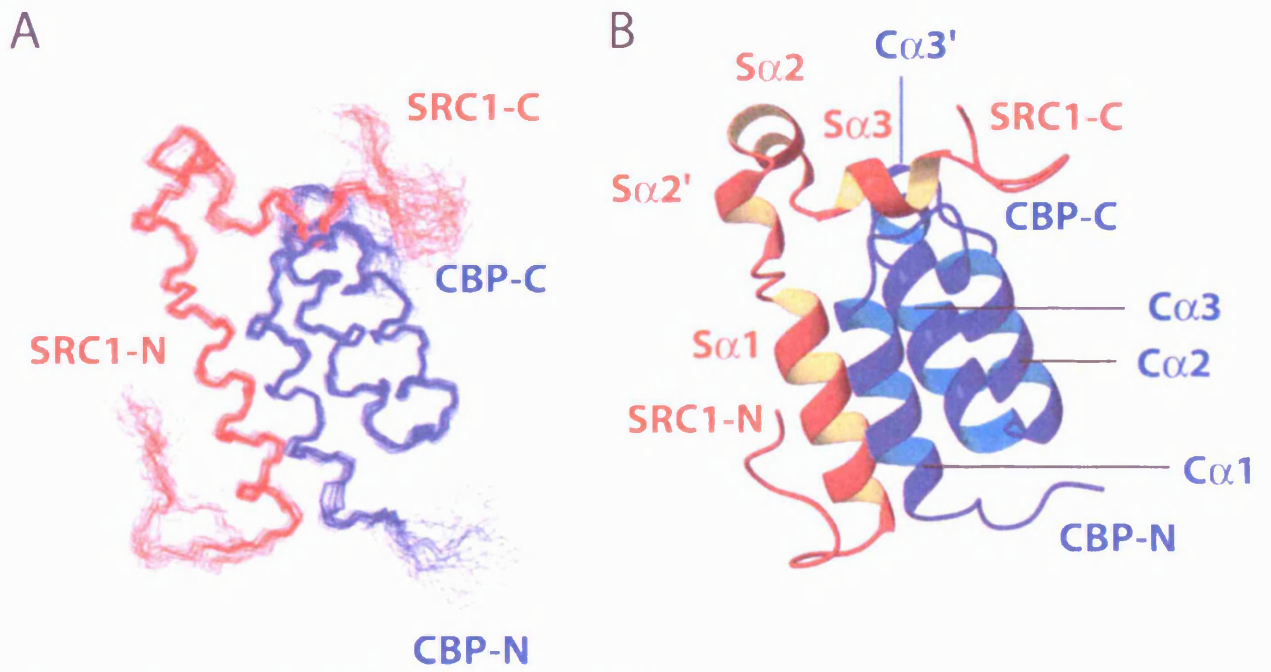


FIGURE 2. **Solution structure of the CBP SID-SRC1 AD1 complex.** *A*, best fit superimposition of the family of 37 converged NMR structures obtained for the CBP SID-SRC1 AD1 complex. The CBP SID domain is shown in *blue* and SRC1 AD1 in *red*. *B*, *schematic* (ribbon) representation of the backbone topology of the two polypeptides in the complex, using the same orientation as in *A*, and demonstrating the tight 4-helix bundle conformation. The four helices of CBP SID (Cα1, Cα2, Cα3, and Cα3′) and SRC1 AD1 (Sα1, Sα2′, Sα2, and Sα3) are indicated, along with the N and C termini of both domains.

Lys^{943} from the unstructured region between Sα1 and Sα2′, Thr^{946} from Sα2′, Ala^{949} that links Sα2′ to Sα2, and Gly^{956} that links Sα2 to Sα3. However, two potential salt bridges are formed between residues Glu^{945} and Lys^{959} and between Asp^{952} and Lys^{959}. The binding of Sα1 in the hydrophobic groove situated between Cα1 and Cα3 is stabilized by van der Waals interactions. The favorable contacts between the N-terminal helices of CBP SID and SRC1 AD1 mainly involve interactions between buried side chains found in the leucine-rich motifs of Cα1 (L*XX*LL*XX*L corresponding to Leu^{2068}, Leu^{2071}, Leu^{2072}, Leu^{2075}) and Sα1 (LL*XX*L*XX*FL corresponding to Leu^{932} and Leu^{936}). These motifs resemble the L*XX*LL and I/L*XX*I/H/L*XXX*IL motifs that mediate the interaction of coactivators and corepressors with nuclear receptors. Additional favorable van der Waals interactions are found between residues Gln^{935} and Phe^{939} from Sα1 and the leucine residues of Cα1. Similarly, stabilizing contacts occur between Sα1 and Cα3, involving the leucine-rich motif (Leu^{933}, Leu^{936}, and Leu^{940}) and residues Glu^{929} and Val^{937} of Sα1, and residues Gln^{2096}, Leu^{2097}, Ala^{2100}, Phe^{2101}, and Gln^{2104} of Cα3 in CBP. A potential salt bridge between Lys^{2076} and Asp^{944} may also stabilize this region of the complex.

The fourth helix of SRC1 AD1 (Sα3) fits into a second hydrophobic groove in CBP SID located between the well defined SPSSP turn linking Cα1 and Cα2 and the C terminus of Cα3. This interaction is stabilized by hydrophobic contacts involving Leu^{955}, Ile^{957}, Leu^{960}, Val^{961}, Gln^{962}, and Gly^{964} of SRC1 AD1 and Ser^{2079}, Ser^{2080}, Gln^{2084}, Arg^{2105}, Lys^{2108}, and Tyr^{2109} of CBP SID. The five C-terminal residues of SRC1 AD1 appear to loop back toward Sα3 and Cα3′, leading to some van der Waals contacts between Leu^{969} of SRC1 AD1 and Tyr^{2109} of CBP SID. However, this loop is primarily stabilized by intramolecular van der Waals interactions involving Gln^{962} and residues within a C-terminal linker sequence originating from the expression vector. In the absence of the linker sequence, contacts between Leu^{969} and Tyr^{2109} may not occur, and the C-terminal region of SRC1 AD1, including Sα3, may make additional contacts within the complex, or with other domains/proteins.

Some features of the NMR data obtained for the CBP SID-SRC1 AD1 complex clearly show the presence of a second minor conformational state, which is in exchange with the structure reported here (Fig. 1*D*). This is clearly indicated by the presence of cross-peaks between signals from backbone amide groups in $^{15}N/^{1}H$ TOCSY-HSQC spectra, which arise through chemical exchange processes (48). This also results in exchange broadening of a significant number of backbone amide signals. Interestingly, only a few of the TOCSY exchange peaks involve residues in CBP SID, whereas almost all of the residues from Leu^{936}–Val^{961} in SRC1 AD1 are affected, which corresponds to the six C-terminal residues of Sα1 through to the end of Sα3. This suggests that the SRC1 AD1 domain has a degree of conformational instability, even when bound to CBP SID, which may facilitate the rapid dissociation of the complex. However, we do not exclude the possibility that other protein/protein interactions may stabilize complexes of the full-length proteins.

Comparison of CBP-SRC1 and CBP-ACTR Complexes—The solution structure of CBP SID in complex with the AD1 domain from ACTR has been reported previously (30). Comparison of the two complexes revealed that they have both conserved and distinct structural features (Fig. 3, *A* and *B*). The SID adopts similar secondary and tertiary structures, when bound to different AD1 domains (Fig. 3, *A* and *B*), as highlighted by the superposition of the SID polypeptide backbone atoms (residues Ile^{2063}, Pro^{2065}–Ala^{2067}, Asp^{2070}–Ser^{2079}, and Gln^{2085}–Lys^{2108}), which yields an r.m.s.d. value of 1.79 Å (Fig. 3*C*). The most notable difference between the CBP SID polypeptides in the two complexes is the conformation of the sequence composed of five glutamine residues (2082–2086). In the CBP-SRC1 complex this region forms the N terminus of an extended Cα2 helix (Fig. 3, *A* and *C*), whereas in the CBP-ACTR complex the three proximal glutamine residues were pro-

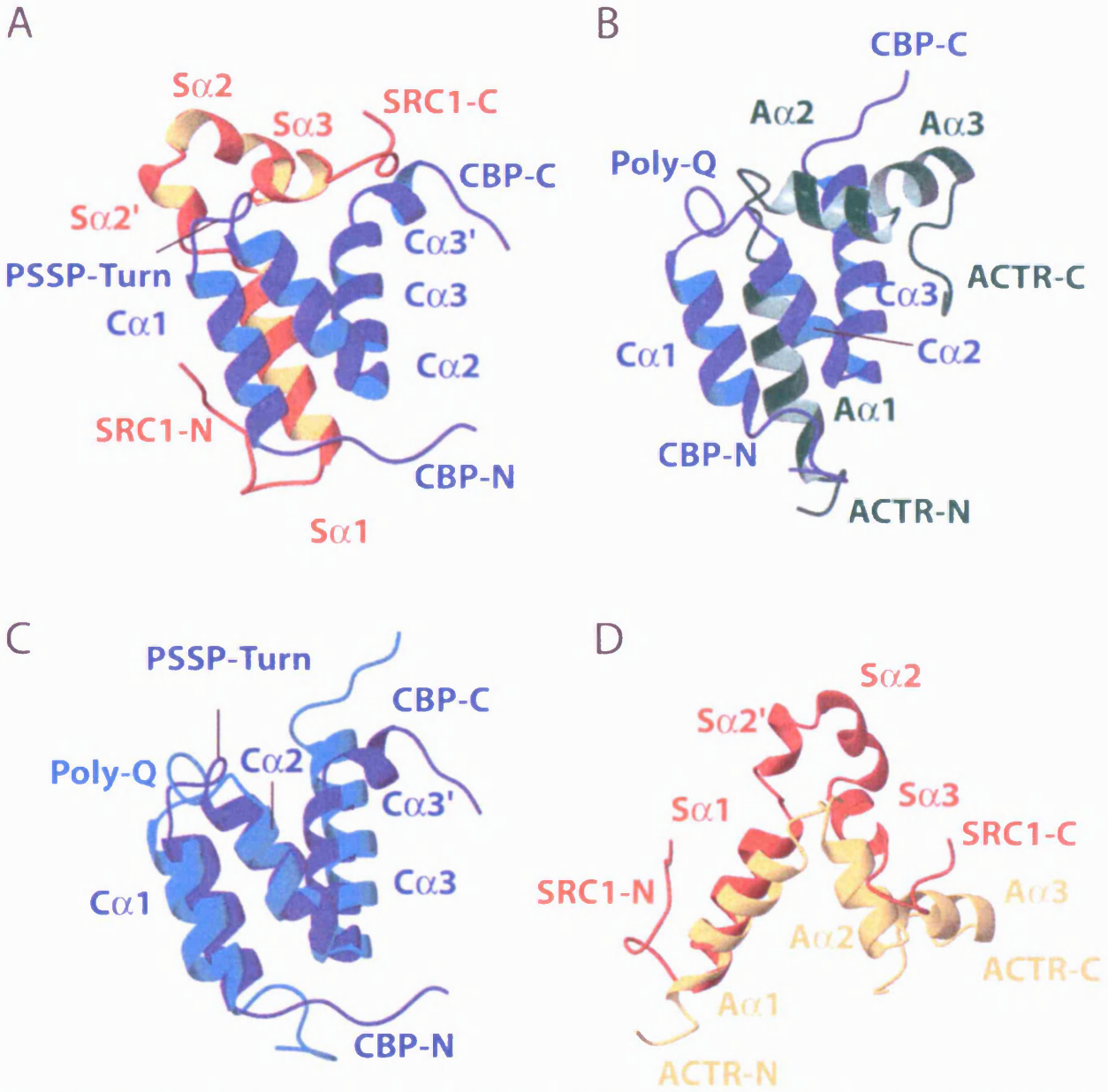


FIGURE 3. **Comparison of CBP SID/SRC1 AD1 and CBP SID-ACTR AD1 complexes.** *A* and *B*, equivalent views of the backbone topology of the CBP SID-SRC1 AD1 and CBP SID-ACTR AD1 complexes, respectively, which were obtained by superimposing the CBP SID domain from both complexes (residues Ile^{2063}, Pro^{2065}–Ala^{2067}, Asp^{2070}–Ser^{2079}, and Gln^{2085}–Lys^{2108}). The CBP SID domain is shown in *blue*; SRC1 AD1 is shown in *red* and ACTR AD1 in *green*. *C*, comparison of the backbone topologies of the two CBP SID domains, which were overlaid on the same residues as in *A* and *B*. The CBP SID domains from the SRC1 AD1 and ACTR AD1 complexes are shown in *blue* and *cyan*, respectively. *D*, comparison of the backbone folds of the AD1 domains of SRC1 (*red*) and ACTR (*yellow*), which were overlaid on residues Ile^{2063}, Pro^{2065}–Ala^{2067}, Asp^{2070}–Ser^{2079}, Gln^{2085}–Lys^{2108} of CBP-SID and residues Asp^{928}–Ser^{941} and Asp^{1044}–Ser^{1057} of SRC1 and ACTR, respectively.

posed to form part of a solvent-exposed loop (designated poly(Q); Fig. 3, *B* and *C*). This difference may be due to the paucity of NMR signal assignments for residues in this region in the CBP-ACTR complex (30), resulting in fewer NMR constraints on the conformation rather than a true structural difference. Another difference between the CBP SID domains is Cα3, which is 17 residues long (Pro^{2095}–Ala^{2111}) in the CBP/ACTR structure, although in the CBP-SRC1 complex this region forms two shorter helices (Cα3, Pro^{2095}–Thr^{2106}; Cα3′, Tyr^{2109}–Asn^{2112}) resembling a kinked helix. The Cα3′ helix is inclined away from the rest of the four-helix bundle, which is not observed in the ACTR-AD1 complex.

Structural similarity between p160 AD1 domains in the two complexes is confined to the N-terminal helices Aα1 and Sα1, which form four-helix bundles with the CBP SID (Fig. 3, *A* and *B*). The remaining sequences in the AD1 domains adopt strikingly distinct topologies as highlighted by superimposing the AD1 polypeptide structures (Fig. 3*D*). This is also reflected in the significant chemical shift differences for NMR signals from equivalent residues in the two AD1 domains and also

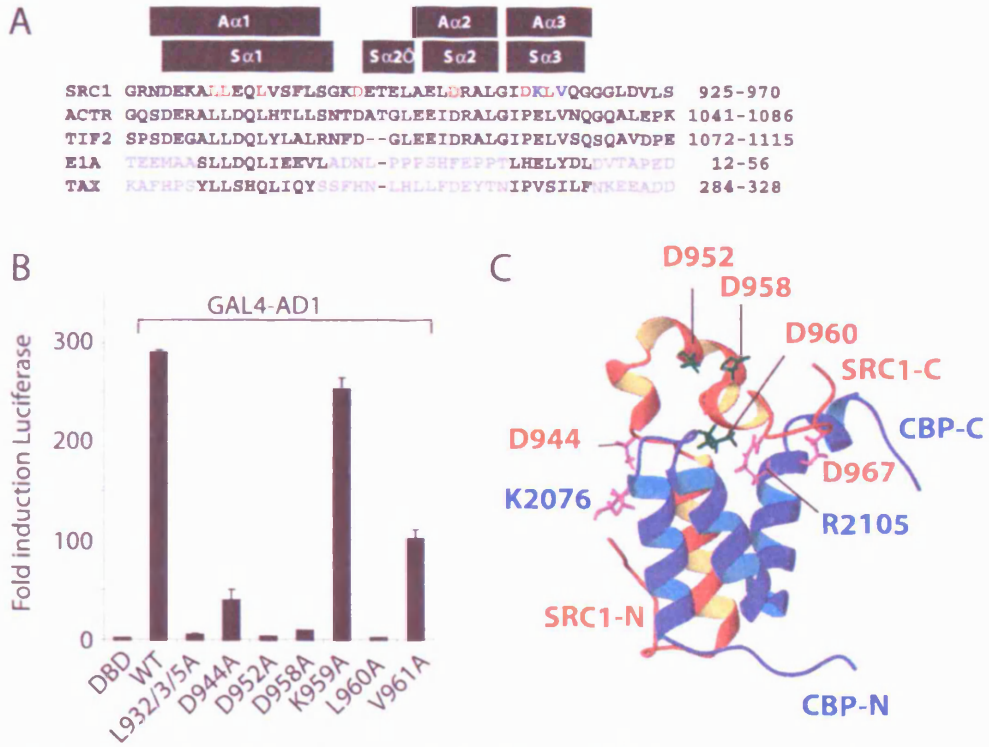


FIGURE 4. **Critical residues in SRC1 AD1 required for CBP binding.** *A*, sequence alignment of p160 AD1 domains and CBP-binding sequences from viral proteins E1A and TAX. For the viral proteins, low homology sequences are represented in *gray*, and regions showing homology to helices Sα1/Aα1 and Sα3/Aα3 are in *black*. SRC1 residues, which were subjected to replacement with alanine and which are required for CBP binding/transcriptional activity of AD1, are highlighted in *red*. SRC1 residues highlighted in *blue* had little effect on AD1 transcriptional activity when replaced with alanine (see *C*, below). The *black boxes* over the alignments indicate regions of helical secondary structure observed in complexes for SRC1 (Sα1, Sα2′, Sα2, and Sα3) and ACTR (Aα1, Aα2, and Aα3). *B*, reporter (luciferase) assay showing the relative transcriptional activities of GAL4-AD1 (SRC1) or mutants thereof containing alanine replacements for the residues indicated. Reporter activity is presented as fold induction over control (GAL4-DBD) and was normalized to a cotransfected β-galactosidase reporter. The data presented are averages from triplicates, and the *error bars* indicate the S.D. *DBD*, DNA binding domain; *WT*, wild type. *C*, ribbon representation of the backbone topology of CBP SID-SRC1 AD1 complex. The side chains of residues involved in two potential inter-molecular salt bridges (Lys^{2076} of CBP SID and Asp^{944} of SRC1 AD1; Arg^{2105} of CBP SID and Asp^{967} of SRC1 AD1) are highlighted in *purple*. Note that Arg^{2105} and Asp^{952} are spatially distant and cannot form a salt bridge equivalent to that observed between Arg^{2105} and Asp^{1086} in the CBP-ACTR complex. The side chains of residues Asp^{952}, Asp^{958}, and Asp^{960}, which are required for AD1 transcriptional activity, are represented in *green*.

for CBP SID residues that make distinct contacts with AD1 in the two complexes, in particular, residues in the Cα3/Cα3′ region (data not shown). Although SRC1 AD1 contains three further helical regions (residues Sα2′, Glu^{945}–Leu^{948}; Sα2, Glu^{950}–Ser^{954}; and Sα3, Ile^{957}–Gln^{962}), ACTR AD1 contains only two additional helices (Aα2, Leu^{1064}–Leu^{1071}, and Aα3, Ile^{1073}–Gln^{1079}) (Fig. 3*D*). These distinct topologies are somewhat unexpected given that SRC1 and ACTR share 53% sequence identity and 67% sequence similarity within the AD1 domain. In the CBP SID-ACTR AD1 complex, Aα2 and Aα3 wrap around Cα3 making extensive contacts with this helix, whereas Aα2 occupies a hydrophobic groove between the proposed poly(Q) loop and Cα3 (Fig. 3*B*). In the CBP SID-SRC1 AD1 complex, only the C-terminal residue (Leu^{955}) of Sα2 (corresponding to Leu^{1071} in Aα2) makes any contact with the SID domain (Fig. 3*A*). Instead, Sα3 fills the groove between the PSSP turn and Cα3, effectively capping the four-helix bundle (Fig. 3*A*). This contrasts with the position of corresponding helix of ACTR (Aα3), which packs against the adjacent hydrophobic face of Cα3 (Fig. 3*B*).

The CBP SID-ACTR AD1 complex was reported to contain a salt bridge between residues Asp^{1068} in Aα2 and Arg^{2105} in Cα3, coordinating a hydrogen-bonding network involving Asp^{1060} and Arg^{2105} and Asp^{1068} and Tyr^{2109}. This buried charged interaction has been proposed to be important for the specificity of the CBP SID/ACTR AD1 interaction, although it does not contribute directly to the stability of the complex (49). Both of these aspartate residues are conserved in SRC1 AD1 (Asp^{944} and Asp^{952}); however, they are spatially distant from the Cα3 helix. Asp^{944} is present in the loop between Sα1 and Sα2′, and Asp^{952} is contained within the Sα2 helix (Fig. 4*C*). However, it is possible that Asp^{944} forms a salt bridge with Lys^{2076} from Cα1. Similarly, an aspartate residue (Asp^{965}) near the C terminus of SRC1 AD1 may form a salt bridge with Arg^{2105}, which is involved in van der Waals contacts between Cα3 and Sα3. As would be expected from the tertiary structures of their respective complexes, the solvent-inaccessible surface areas of CBP SID (1681 Å^2) and ACTR AD1 (1839 Å^2) domains are greater than those observed for the complex of CBP SID (1019 Å^2) and SRC1 AD1 (1088 Å^2).

In summary, the backbone structure of the CBP SID and the first helix of AD1 are very similar in both complexes, as reflected in a backbone atom r.m.s.d. value of 2.05 Å (calculated by superimposition of residues Ile^{2063}, Pro^{2065}–Ala^{2067}, Asp^{2070}–Ser^{2079}, and Gln^{2085}–Lys^{2108} of CBP SID and residues Asp^{928}–Ser^{941} and Asp^{1044}–Ser^{1057} of SRC1 AD1 and ACTR-AD1, respectively). The remainder of the p160 AD1 domains adopt very distinct folds, although the C-terminal helices of AD1 appear to be essential to stabilize the complex.

Structural Flexibility of the CBP SID Permits Complex Formation with Multiple Partners—The NMR structure of the CBP SID in isolation (also termed IBiD because of its interaction with IRF3) revealed it to have significant α-helical content (19), although the conformation of the isolated CBP SID has been proposed to be consistent with that of a molten

globule (50). Isolated AD1 polypeptides have little if any intrinsic structure (50)[6] but appear to undergo induced folding in complex with CBP SID (30). An important question is whether short conserved structural motifs promote similar modes of binding of different proteins to CBP, or whether different complexes adopt drastically different conformations that influence the overall structure of CBP/p300 proteins, which in turn influences combinatorial complex formation.

This study reveals structural similarity within the α1 helix of two p160 AD1 domains, which contributes to a four-helix bundle in both the CBP-SRC1 and CBP-ACTR complexes. We have shown previously that the LL*XX*L*XXX*L sequence motif within α1 is conserved in other SID-binding proteins (see Fig. 4*A*), and that disruption of this sequence in Ets2 and E1A abrogates their binding to the CBP SID (22). Fusion of the SRC1 AD1 sequence to a GAL4 DNA binding domain permits very potent activation of a *GAL* reporter gene in mammalian cells, through recruitment of endogenous CBP and p300 (12). As shown in Fig. 4*B*, replacement of the Leu^{932}, Leu^{933}, and Leu^{936} residues in Sα1 with alanines resulted in loss of AD1 transcriptional activity. This highlights the requirement of the Sα1 for interaction with full-length CBP/p300 *in vivo*.

The C-terminal helix (α3) of the SRC1 and ACTR AD1 domains also plays a critical role in formation of the two complexes with CBP SID as discussed above. Interestingly, the sequence of the Sα3 helix of SRC1 AD1 (^{957}IDKLV^{961}) is divergent from that in ACTR and TIF2 (*IPELV*) (Fig. 4*A*). The IPELV sequence in ACTR and TIF2 resembles the LP*X*L motif that has been shown to mediate the interaction of CITED and HIF1α transcription factors with the CH1 domain of CBP (51). Mutagenesis of the SRC1 AD1 domain in the context of GAL4-AD1 revealed that the D958A or L960A mutations resulted in almost complete loss of reporter activation, whereas K959A retained almost full transactivation function (Fig. 4*B*). In addition, reporter activation by the V961A mutant was reduced by 70%, suggesting it is important but not essential for CBP/p300 recruitment (Fig. 4*B*). These results confirm the importance of Sα3 and key residues within that helix for CBP/p300 recruitment.

Replacement of the conserved aspartate residues (D944A and D952A) also resulted in loss of reporter activation by GAL4-AD1 (Fig. 4*B*). Thus, as suggested for the CBP-ACTR complex (30, 49), salt bridge formation may be important in stabilizing the higher order folding of CBP-p160 complexes. There is evidence to suggest that SRC1 sequences outside of the AD1 domain may influence the interaction with CBP. The signaling molecule 8-bromo-cAMP induces mitogen-activated protein kinase-dependent phosphorylation of SRC1 at Thr^{1179} and Ser^{1185}, which leads to increased ligand-independent transcriptional activity of the progesterone receptor and enhanced CBP recruitment (52, 53). Similarly, the interaction of ACTR with CBP was abrogated by mutation of residues outside of the AD1 domain, which are known to be phosphorylated *in vivo* (Thr^{24}, Ser^{543}, Ser^{857}, Ser^{860}, and Ser^{867}) (53). This may reflect binding between additional regions of SRC1 and CBP and/or conformational changes in SRC1 that effect the affinity of the AD1 region for CBP.

We have indicated previously that several CBP SID-binding proteins contain low level sequence homology in the regions of SRC1 AD1 corresponding to Sα1, and to a lesser extent Sα3, that may indicate conserved modes of binding to CBP. An alignment of the sequences of interaction domains of several CBP SID-binding proteins is shown in Fig. 4*A*. The transactivation domain of human T-cell leukemia Tax protein contains a short sequence (^{312}YTNIPISL^{319}) that is required for binding to CBP SID (21). This sequence may form an amphipathic α-helix analogous to Sα3 of SRC1 AD1 and play an equivalent role in stabilizing CBP/Tax interactions (Fig. 4*A*). The adenoviral oncoprotein E1A revealed also sequences essential for CBP binding, which may be functionally and structurally equivalent to SRC1 AD1 Sα1 and Sα3 (22) (Fig. 4*A*). IRF3 contains a sequence termed the IRF association domain (IAD), which mediates its interactions with CBP SID, and the crystal structure of this complex was recently reported (54). Little structural similarity is evident between the CBP-IRF3 complex and the CBP-p160 structures. Contacts between the IAD and SID are mediated by hydrophobic contacts involving Cα1, Cα2, and Cα3 in CBP SID with two amphipathic α-helices (H3 and H4) in the IAD containing I*XX*LI and L*XX*LV*XXX*V motifs, respectively. Interestingly, the secondary structure of the SID is similar to that observed in p160 complexes, and we note that the poly(Q) region of CBP is comprised within the Cα2 helix, as in the CBP-SRC1 complex. Thus the SID appears to be able to fold into distinct conformations to accommodate binding to different transcription factors.

In conclusion, the NMR structure reported here reveals both similar and diverse structural features in CBP SID-p160 AD1 complexes, which may reflect functional differences between p160 proteins. Structural versatility of the CBP SID is likely to underpin its ability to assemble different transcription factor-cofactor complexes in a promoter-dependent context.

Acknowledgments—We are grateful to Peter Wright for the gift of the dual expression plasmid pET22B. We thank Jonas Emsley and Paul McEwan for useful discussions and Sharad Mistry for technical assistance.

[6] J. Bramham and D. M. Heery, unpublished results.

REFERENCES

1. Goodman, R. H., and Smolik, S. (2000) *Genes Dev.* **14,** 1553–1577
2. Giordano, A., and Avantaggiati, M. L. (1999) *J. Cell. Physiol.* **181,** 218–230
3. Kamei, Y., Xu, L., Heinzel, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S. C., Heyman, R. A., Rose, D. W., Glass, C. K., and Rosenfeld, M. G. (1996) *Cell* **85,** 403–414
4. Perissi, V., and Rosenfeld, M. G. (2004) *Nat. Rev. Mol. Cell Biol.* **6,** 542–554
5. Heery, D. M., Kalkhoven, E., Hoare, S., and Parker, M. G. (1997) *Nature* **387,** 733–736
6. Torchia, J., Rose, D. W., Inostroza, J., Kamei, Y., Westin, S., Glass, C. K., and Rosenfeld, M. G. (1997) *Nature* **387,** 677–684
7. Metivier, R., Penot, G., Hubner, M. R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003) *Cell* **115,** 751–763
8. Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A., and Brown, M. (2000) *Cell* **103,** 843–852
9. Heery, D. M., Hoare, S., Hussain, S., Parker, M. G., and Sheppard, H. (2001) *J. Biol. Chem.* **276,** 6695–6702
10. Coulthard, V. H., Matsuda, S., and Heery, D. M. (2003) *J. Biol. Chem.* **278,** 10942–10951
11. Plevin, M. J., Mills, M. M., and Ikura, M. (2005) *Trends Biochem. Sci.* **30,** 66–69
12. Sheppard, H. M., Harries, J. C., Hussain, S., Bevan, C., and Heery, D. M. (2001) *Mol. Cell. Biol.* **21,** 39–50
13. Chen, H., Lin, R. J., Schiltz, R. L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M. L., Nakatani, Y., and Evans, R. M. (1997) *Cell* **90,** 569–580
14. Voegel, J. J., Heine, M. J., Tini, M., Vivat, V., Chambon, P., and Gronemeyer, H. (1998) *EMBO J.* **17,** 507–519
15. Kalkhoven, E., Valentine, J. E., Heery, D. M., and Parker, M. G. (1998) *EMBO J.* **17,** 232–243
16. McInerney, E. M., Rose, D. W., Flynn, S. E., Westin, S., Mullen, T. M., Krones, A., Inostroza, J., Torchia, J., Nolte, R. T., Assa-Munt, N., Milburn, M. V., Glass, C. K., and Rosenfeld, M. G. (1998) *Genes Dev.* **2,** 3357–3366
17. Sheppard, H. M., Matsuda, S., Harries, J. C., Kindle, K. B., and Heery, D. M. (2003) *J. Mol. Endocrinol.* **30,** 411–422
18. Kim, M. Y., Hsiao, S. J., and Kraus, W. L. (2001) *EMBO J.* **20,** 6084–6094
19. Lin, C. H., Hare, B. J., Wagner, G., Harrison, S. C., Maniatis, T., and Fraenkel, E. (2001) *Mol. Cell* **8,** 581–590
20. Livengood, J. A., Scoggin, K. E. S., Van Orden, K., McBryant, S. J., Edayathumangalam, R., Laybourn, P. J., and Nyborg, J. K. (2002) *J. Biol. Chem.* **277,** 9054–9061
21. Scoggin, K. E. S., Ulloa, A., and Nyborg, J. K. (2001) *Mol. Cell. Biol.* **21,** 5520–5530
22. Matsuda, S., Harries, J. C., Viskaduraki, M., Troke, P. J. F., Kindle, K. B., Ryan, C. M., and Heery, D. M. (2004) *J. Biol. Chem.* **279,** 14055–14064
23. Carapeti, M., Aguiar, R. C., Goldman, J. M., and Cross, N. C. (1998) *Blood* **1,** 3127–3133

24. Kindle, K. B., Troke, P. J., Collins, H. M., Matsuda, S., Bossi, D., Bellodi, C., Kalkhoven, E., Salomoni, P., Pelicci, P. G., Minucci, S., and Heery, D. M. (2005) *Mol. Cell. Biol.* **25,** 988–1002
25. Deguchi, K., Ayton, P. M., Carapeti, M., Kutok, J. L., Snyder, C. S., Williams, I. R., Cross, N. C., Glass, C. K., Cleary, M. L., and Gilliland, D. G. (2003) *Cancer Cell* **3,** 259–271
26. Xu, J., and Li, Q. (2003) *Mol. Endocrinol.* **17,** 1681–1692
27. Kung, A. L., Rebel, V. I., Bronson, R. T., Ch'ng, L.-E., Sieff, C. A., Livingston, D. M., and Yao, T.-P. (2000) *Genes Dev.* **14,** 272–277
28. Lucey, M. J., Chen, D., Phoenix, F., Lopez-Garcia, J., Hart, S. M., Al-Jehani, R., Alao, J. P., White, R., Kindle, K. B., Losson, R., Chambon, P., Parker, M. G., Schar, P., Heery, D. M., Buluwela, L., and Ali, S. (2005) *Nucleic Acids Res.* **33,** 6393–6404
29. Zhang, H., Yi, X., Sun, X., Yin, N., Shi, B., Wu, H., Wang, D., Wu, G., and Shang, Y. (2004) *Genes Dev.* **18,** 1753–1765
30. Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M., and Wright, P. E. (2002) *Nature* **415,** 549–553
31. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson.* **53,** 521–528
32. Macura, S., and Ernst, R. R. (1980) *Mol. Physiol.* **41,** 95–117
33. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc.* **111,** 1515–1517
34. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Magn. Reson.* **88,** 425–431
35. Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W., and Fesik, S. W. (1990) *J. Magn. Reson.* **86,** 210–216
36. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson.* **101,** 201–205
37. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR* **3,** 185–204
38. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* **2,** 661–665
39. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293
40. Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR* **5,** 1–10
41. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) *J. Mol. Biol.* **319,** 209–227
42. Guntert, P., and Wuthrich, K. (1991) *J. Biomol. NMR* **1,** 447–456
43. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302
44. Carr, M. D., Bloemink, M. J., Dentten, E., Whelan, A. O., Gordon, S. V., Kelly, G., Frenkiel, T. A., Hewinson, R. G., and Williamson, R. A. (2003) *J. Biol. Chem.* **278,** 43736–43743
45. Muskett, F. W., Frenkiel, T. A., Feeney, J., Freedman, R. B., Carr, M. D., and Williamson, R. A. (1998) *J. Biol. Chem.* **273,** 21736–21743
46. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics* **14,** 51–55
47. Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem.* **287,** 252–260
48. Feeney, J., Bauer, C. T., Frenkiel, T. A., Birdsall, B., Carr, M. D., Roberts, G. C. K., and Arnold, J. R. P. (1991) *J. Magn. Reson.* **91,** 607–613
49. Demarest, S. J., Deechongkit, S., Dyson, H. J., Evans, R. M., and Wright, P. E. (2004) *Protein Sci.* **13,** 203–210
50. Dyson, J. H., and Wright, P. E. (2005) *Nat. Rev. Mol. Cell Biol.* **6,** 197–208
51. Freedman, S. J., Sun, Z. Y., Kung, A. L., France, D. S., Wagner, G., and Eck, M. J. (2003) *Nat. Struct. Biol.* **10,** 504–512
52. Rowan, B. G., Garrison, N., Weigel, N. L., and O'Malley, B. W. (2000) *Mol. Cell. Biol.* **20,** 8720–8730
53. Wu, R. C., Qin, J., Yi, P., Wong, J., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. (2004) *Mol. Cell* **15,** 937–949
54. Qin, B. Y., Liu, C., Srinath, H., Lam, S. S., Correia, J. J., Derynck, R., and Lin, K. (2005) *Structure* **13,** 1269–1277

Journal of Biomolecular NMR (2006) 36:18

Letter to the Editor

NMR assignment and secondary structure determination of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4

DOI 10.1007/s10858-005-5887-6

Pdcd4 is a novel eukaryotic tumour suppressor protein, which is involved in the control of both transcription and translation (Cmarik et al., 1999). The protein contains two MA-3 domains which are involved in mediating specific protein–protein interactions with functional partners such as eIF4A (Yang et al., 2003). Here we report essentially complete backbone and sidechain ^{15}N, ^{13}C and ^{1}H assignments (97%) for the C-terminal MA-3 domain of murine Pdcd4. This reveals that Pdcd4 MA-3 contains seven helices (H326-S341, I344-E353, H361-E373, S378-K392, I398-I408, S422-A436 and K441-L446) linked by loops. BMRB accession number 6900.

References: Cmarik et al. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 14037–14042; Yang et al. (2003) *Mol. Cell. Biol.*, **23**, 26–37.

Lorna C. Waters[a], Maret Böhm[b], Vaclav Veverka[a], Frederick W. Muskett[a], Thomas A. Frenkiel[c], Geoffrey P. Kelly[c], Andrew Prescott[a], Nuvjeevan S. Dosanjh[a], Karl-Heinz Klempnauer[b] & Mark D. Carr[a,*]

[a]*Department of Biochemistry, Henry Wellcome Building, University of Leicester, Lancaster Road, Leicester, LE1 9HN, U.K.;* [b]*Institut für Biochemie, Westfälische-Wilhelms-Universität Münster, Wilhelm-Klemm-Str. 2, D-48149, Münster, Germany;* [c]*MRC Biomedical NMR Centre, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, U.K.*

*To whom correspondence should be addressed. E-mail: mdc12@le.ac.uk

Supplementary material is available in electronic format at http://dx.doi.org/10.1007/s10858-005-5887-6

Letter to the Editor: ^{15}N, ^{13}C and ^{1}H resonance assignments and secondary structure determination of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4.

Lorna C. Waters[a], Maret Böhm[b], Vaclav Veverka[a], Frederick W. Muskett[a], Thomas A. Frenkiel[c], Geoffrey P. Kelly[c], Andrew Prescott[a], Nuvjeevan S. Dosanjh[a], Karl-Heinz Klempnauer[b] and Mark D. Carr[a*]

[a]*Department of Biochemistry, Henry Wellcome Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, United Kingdom;* [b]*Institut für Biochemie, Westfälische-Wilhelms-Universität Münster, Wilhelm-Klemm-Str. 2, D-48149 Münster, Germany;* [c]*MRC Biomedical NMR Centre, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, United Kingdom.* [*]*Email: mdc12@le.ac.uk*

Key words: Pdcd4, MA-3 domain, NMR resonance assignments, secondary structure

Biological context

Pdcd4 is a novel eukaryotic regulatory protein, which was originally discovered in a screen for genes activated during apoptosis (Shibahara *et al.*, 1995). It was subsequently identified as a tumour suppressor in studies of a mouse keratinocyte model of tumour promotion, in which high levels of Pdcd4 rendered cells resistant to transformation by the tumour promoter TPA (Cmarik *et al.*, 1999). Recently, loss of Pdcd4 expression has been strongly implicated in the development and progression of both human lung cancer and aggressive malignant breast cancer (Chen *et al.*, 2003; Afonja *et al.*,

2004)

The 469 residue mouse Pdcd4 protein contains at least three domains, an N-terminal RNA-binding region (residues 1-157) and two MA3 domains (residues 157-275 and 319-449). MA3 domains have been found in over 85 eukaryotic proteins and appear to be involved in mediating specific protein-protein interactions. Surprisingly, despite their wide occurrence and functional importance there is no detailed published structural information reported for any MA3 domain or complex formed with an interacting protein.

The molecular functions of Pdcd4 are not fully understood, however, recent work indicates roles in the control of both transcription and translation, mediated via specific protein-protein and protein-RNA interactions (Yang *et al.*, 2003; 2004; Bitomsky *et al.*, 2004). The interaction with eIF4A is currently the best characterised and involves both MA3 domains in Pdcd4, with complex formation resulting in inhibition of the helicase activity of eIF4A and suppression of cap-dependent translation (Yang *et al.*, 2003; 2004).

Here we report the determination of essentially complete sequence-specific backbone and sidechain assignments for C-terminal MA3 domain of Pdcd4 and the positions of elements of regular secondary structure.

Methods and experiments

Unlabelled and uniformly ^{15}N and $^{15}N/^{13}C$ labelled samples of Pdcd4 MA-3 (319-449) were prepared from a pGex6P2 based *E. coli* expression vector.

The unlabelled samples were grown in LB medium and the labelled samples were grown in minimal media containing 0.6 g/l >99% ^{15}N ammonium sulphate and/or 2 g/l >99% ^{13}C glucose as the sole nitrogen and carbon sources. A ^{13}C labelled protein sample which was used for the $^{13}C/^{1}H$ NOESY-HSQC experiment was prepared as described above, but with the addition of non-isotopically labelled aromatic amino acids to the minimal media (His, Phe, Trp and Tyr at 50 mg/l). The recombinant glutathione-S-transferase – Pdcd4 MA-3 fusion protein was purified by affinity chromatography using s-linked glutathione agarose. The GST tag was then removed by cleavage with PreScission Protease and the two products separated by gel filtration on a superdex 75 16/60 column.

NMR spectra were acquired from 0.35 ml samples of 0.6 – 1.0 mM Pdcd4 MA-3 in a 25 mM sodium phosphate, 100 mM sodium chloride, 0.5 mM DTT, 10 µM EDTA and 0.02% (w/v) sodium azide buffer at pH 6.5, containing either 10% D_2O/90% H_2O or 100% D_2O as appropriate. All NMR data were acquired at 25°C on either 600 MHz Bruker Avance/DRX systems or an 800 MHz Varian Inova spectrometer. The 2D and 3D spectra recorded to obtain sequence specific assignments for Pdcd4 MA-3 were: ^{1}H-^{1}H TOCSY with mixing times of 40 and 60 ms, ^{1}H-^{1}H NOESY with an NOE mixing time of 100 ms; $^{15}N/^{1}H$ HSQC; TOCSY-HSQC with a mixing time of 55 ms; NOESY-HSQC with an NOE mixing time of 100 ms; $^{13}C/^{1}H$ HSQC; HCCH-TOCSY with a mixing time of 13 ms, NOESY-HSQC with an NOE mixing time of 100 ms and $^{15}N/^{13}C/^{1}H$ HNCACB and CBCA(CO)NH (Bax. 1994 and references therein). Typical acquisition times in F_1 and F_2 for the 3D experiments were

9.2-24 ms for ^{15}N, 6.6-10.0 ms for ^{13}C and 15-18 ms for ^{1}H, and with an acquisition time of 80 ms in F_3 (^{1}H). The majority of the 3D spectra were collected over approximately 72 hours, 2D ^{1}H experiments over 24 hours and $^{15}N/^{1}H$ and $^{13}C/^{1}H$ HSQC spectra over about 1 hour and 20 minutes respectively. Typical acquisition times in 2D experiments were either 60 ms (^{15}N), 10 ms (^{13}C) or 35-42 ms (^{1}H) in F_1 and 85-125 ms in F_2 (^{1}H). The WATERGATE method was used to suppress the water signal when required. The 3D NMR data were processed using either NMRPipe (Delaglio *et al.*, 1995) or Topspin (Bruker Biospin Ltd) with linear prediction used to extend the effective acquisition times by up to 2 fold in F_1 and F_2. The spectra were analysed using the Sparky package.

Pdcd4 gives rise to well-resolved spectra, as illustrated by the $^{15}N/^{1}H$ HSQC spectrum shown in Figure 1. This allowed essentially complete backbone resonance assignments to be made, which, together with the patterns of sequential and medium range NOEs involving backbone amides, was used to map the secondary structure. The data shows that Pdcd4 contains seven helices (residues K329-S341, I344-E353, E362-E373, A379-K392, I398-I408, V430-A436 and K441-L446).

Extent of assignments and data deposition

Backbone amide assignments were obtained for all residues in Pdcd4 except: H326, K329, E330, V356, H360, I397, D418, V419, S422 and S449 (92%) and for all Cα and Cβ signals apart from S449 (99%). In the case of Hα and Hβ signals, assignments were obtained for all residues except residue S449

(99%). For the remaining non-exchangeable side chain signals (^{13}C and ^{1}H) complete assignments were obtained apart from: K329 (Cε), H358 (Hε1, Hδ2), H361 (Hε1, Hδ2), L372 (Cγ, Hγ), I383 (Cγ1), Q400 (Cγ, Hγ2, Hγ3), I412 (Hγ12, Hγ13) I438 (Cγ1) and F434 (Hζ) (98%). The comprehensive ^{15}N, ^{13}C and ^{1}H resonance assignments obtained for Pdcd4 have been deposited at the BioMagResBank database. (accession number 6900).

Acknowledgements

This work was supported by grants from the Wellcome Trust and the Deutsche Krebshilfe. LW was supported by a BBSRC PhD studentship.

References

Afonja, O., Juste, D., Das, S., Matsuhashi, S. and Samuels, H.H. (2004) *Oncogene,* **23**, 8135-8145.

Bax, A. (1994) Curr. Opin. Struct. Biol., 4, 738–744.

Bitomsky, N., Böhm, M. and Klempnauer, K.-H. (2004) *Oncogene,* **23**, 7484-7493.

Chen, Y., Knosel, T., Kristiansen, G., Pietas, A., Garber, M.E., Matsuhashi, S., Ozaki, I. and Petersen, I. (2003) *J. Pathol.,* **200**, 640-646.

Cmarik, J.L., Min, H., Hegamyer, G., Zhan, S., Kulesz-Martin, M., Yoshinaga, H., Matsuhashi, S. and Colburn N.H. (1999) *Proc. Natl. Acad. Sci., USA* **96**, 14037-14042.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR.,* 6, 277-293.

Shibahara, K., Asano, M., Ishida, Y., Aoki, T., Koike, T. and Honjo, T. (1995) *Gene*, **166**, 297-301.

Yang, H.-S., Knies, J.L., Stark, C. and Colburn N.H. (2003) *Oncogene.* **20**, 669-676.

Yang, H.S., Jansen, A.P., Komar, A.A., Zheng, X., Merrick, W.C., Costes, S., Lockett, S.J., Sonenberg, N. and Colburn, N.H. (2003) *Mol. Cell. Biol.*, **23**, 26-37.

Yang, H.S., Cho, M.H., Zacowicz, H., Hegamyer, G., Sonenberg, N. and Colburn, N. (2004) *Mol. Cell. Biol.*, **24**, 3894-3906.

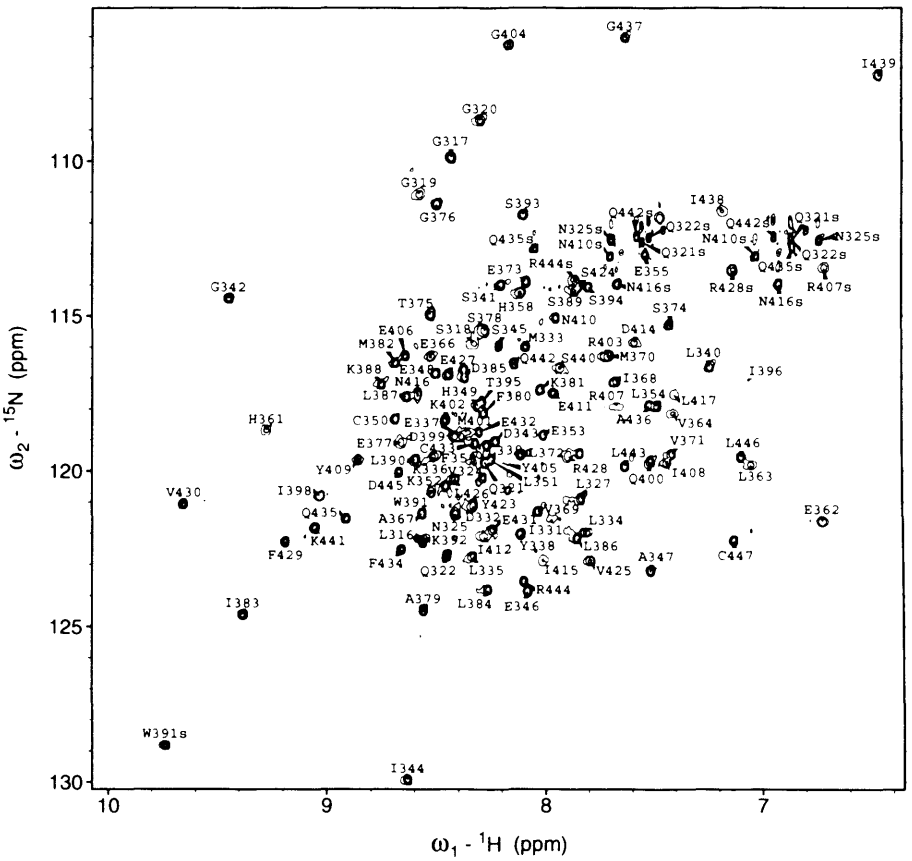


Figure 1: $^{15}N/^{1}H$ HSQC spectrum of the C-terminal MA-3 domain of Pdcd4. The assignments of the signals from backbone amide groups are indicated by residue type and number. In addition sidechain amide group assignments are labeled 's'.

Oncogene (2007), 1–10

www.nature.com/onc

ORIGINAL ARTICLE

Structure of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4 and characterization of its interaction with eIF4A

LC Waters[1], V Veverka[1], M Böhm[2], T Schmedt[2], PT Choong[1], FW Muskett[1], K-H Klempnauer[2] and MD Carr[1]

[1]*Department of Biochemistry, University of Leicester, Leicester, UK and* [2]*Institut für Biochemie, Westfälische-Wilhelms-Universität Münster, Münster, Germany*

Programmed cell death protein 4 (Pdcd4) is a novel tumour suppressor protein, which is involved in the control of eukaryotic transcription and translation. The regulation of translation involves specific interactions with eukaryotic initiation factor (eIF)4A and eIF4G, which are mediated via the two tandem MA-3 domains. We have determined the structure of the C-terminal MA-3 domain of Pdcd4 (Pdcd4 MA-3$_C$), characterized its interaction with eIF4A and compared the features of nuclear magnetic resonance (NMR) spectra obtained from the single domain and tandem MA-3 region. Pdcd4 MA-3$_C$ is composed of three layers of helix–turn–helix hairpins capped by a single helix and shows close structural homology to the atypical HEAT repeats found in many eIFs. The sequence conservation and NMR data strongly suggest that the tandem MA-3 region is composed of two equivalent domains connected by a somewhat flexible linker. Pdcd4 MA-3$_C$ was found to interact with the N-terminal domain of eIF4A through a conserved surface region encompassing the loop connecting α5 and α6 and the turn linking α3 and α4. This site is strongly conserved in other MA-3 domains known to interact with eIF4A, including the preceding domain of Pdcd4, suggesting a common mode of binding.

Oncogene advance online publication, 19 February 2007; doi:10.1038/sj.onc.1210305

Keywords: Pdcd4; MA-3; eIF4A; HEAT; NMR; translation

Introduction

Programmed cell death protein 4 (Pdcd4) is a novel, highly conserved, eukaryotic protein, which has recently been shown to play critical roles in the regulation of both transcription and translation (Yang *et al.*, 2003a, 2004; Bitomsky *et al.*, 2004). The protein was initially discovered in a screen for genes activated during apoptosis (Shibahara *et al.*, 1995) and then subsequently identified as a tumour suppressor in studies of a mouse keratinocyte model of tumour promotion, in which high levels of Pdcd4 were found to render cells resistant to transformation by the tumour promoter 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA) (Cmarik *et al.*, 1999). Pdcd4 has been shown to inhibit the activation of AP1-responsive promoters by c-Jun, providing a possible explanation for its ability to suppress TPA-induced transformation (Yang *et al.*, 2001, 2003b). Recently, loss of Pdcd4 expression has been strongly implicated in the development and progression of both human lung cancer and aggressive malignant breast cancer (Chen *et al.*, 2003; Afonja *et al.*, 2004). In addition, the expression of Pdcd4 has been shown to be reduced in many renal-, lung- and glia-derived tumours (Jansen *et al.*, 2004).

Pdcd4 has been shown to shuttle between the nucleus and the cytoplasm but under normal cell growth conditions it appears to be predominantly localized to the nucleus (Bohm *et al.*, 2003; Palamarchuk *et al.*, 2005). Recent work suggests that phosphorylation of Pdcd4 by Akt plays a key role in the nuclear translocation of Pdcd4, as well as significantly reducing the ability of both nuclear and cytoplasmic Pdcd4 to inhibit AP-1-dependent activation of gene expression (Palamarchuk *et al.*, 2005).

Human and mouse Pdcd4 are 469 residue proteins, which contain at least three domains, an N-terminal RNA-binding region (residues 1–157) and two MA3 domains (residues 157–284 and 319–449) (Ponting, 2000; Bohm *et al.*, 2003). To date, MA3 domains have been found in over 200 eukaryotic proteins and appear to be involved in mediating specific protein–protein interactions (http://www.sanger.ac.uk/Software/Pfam/) (Finn *et al.*, 2006). Characterization of the cellular functions of Pdcd4 is the focus of many ongoing investigations, however, recent work indicates essential roles in the control of both transcription and translation, mediated via specific protein–protein and protein–RNA interactions (Yang *et al.*, 2003a, 2004; Bitomsky *et al.*, 2004). The best characterized functional interaction is between the MA-3 domains of Pdcd4 and the eukaryotic initiation factors eIF4A and eIF4G, with complex

Correspondence: Dr M Carr, Department of Biochemistry, University of Leicester, Henry Wellcome Building, Lancaster Road, Leicester, Leicestershire, LE1 9HN, UK.
E-mail: mdc12@le.ac.uk
Received 13 October 2006; revised 19 December 2006; accepted 29 December 2006

formation resulting in the inhibition of cap-dependent translation (Yang *et al*., 2003a, 2004). eIF4A catalyses the unwinding of stable secondary structure in the 5′ untranslated region (UTR) of mRNAs, allowing the recruitment of the 40S ribosomal subunit to the 5′ cap of mRNA (Hershey and Merrick, 2000). The inherent helicase activity of eIF4A is strongly stimulated by binding to the scaffold protein eIF4G to form part of the eIF4F complex, which also includes eIF4E (Rozen *et al*., 1990).

Pdcd4 has also been shown to bind to c-Jun and Jun N-terminal kinase (JNK), thereby preventing JNK-dependent phosphorylation of the c-jun transactivation domain, as well as the interaction of c-Jun with the coactivator p300 (Bitomsky *et al*., 2004). This suggests a mechanism for the suppression of c-Jun activity by Pdcd4 and highlights its importance in transcriptional as well as translational control.

In this communication, we report the high-resolution solution structure of the C-terminal MA-3 domain of mouse Pdcd4 (Pdcd4 MA-3$_C$), characterization of its interaction with eIF4A and propose a model for the entire tandem MA-3 region based on sequence conservation and nuclear magnetic resonance (NMR) data.

Results

Structural calculations

The combined automated nuclear overhauser effect (NOE) assignment and structure determination protocol (CANDID) (Herrmann *et al*., 2002) proved very effective at determining unique assignments for the NOEs identified in the two- and three-dimensional nuclear overhauser effect spectroscopy-based spectra, with unique assignments obtained for 94% (3999/4254) of the NOE peaks, which produced 2545 non-redundant ^{1}H to ^{1}H upper distance limits. The final family of Pdcd4 MA-3$_C$ structures was determined using a total of 2787 NMR-derived structural constraints (an average of 21.3 per residue), including 2545 NOE-based upper distance limits, 174 backbone torsion angle constraints and 68 hydrogen bond constraints in regions of regular helical structure (Supplementary Table S1). Following the final round of CYANA calculations, 49 satisfactorily converged structures were obtained from 100 random starting structures. The converged structures contain no distance or van der Waals violation higher than 0.5 Å and no dihedral angle violations higher than 5°, with an average value for the CYANA target function of 1.46 ± 0.25 Å^2. The NMR constraints and structural statistics for Pdcd4 MA-3$_C$ are summarized in Supplementary Table S1. The family of converged Pdcd4-MA3 domain structures, together with the NMR constraints, have been deposited in the Protein Data Bank (accession code 2HM8).

The solution structure of Pdcd4 MA-3$_C$ is determined to high precision, which is clearly evident from the superposition of the protein backbone shown for the family of converged structures in Figure 1a (best fit for residues 327–356, 359–412 and 421–448) and is reflected in low root-mean squared deviation (RMSD) values to the mean structure for both the backbone and all heavy atoms of 0.49 ± 0.07 and 0.92 ± 0.07 Å, respectively.

Mapping of binding sites

The positions of signals from backbone amide groups in proteins are highly sensitive to changes in their local environment and have been used to localize the eIF4A-binding site on Pdcd4 MA-3$_C$. Typical ^{15}N/^{1}H heteronuclear single-quantum coherence (HSQC) spectra obtained from samples of ^{15}N-labelled Pdcd4 MA-3$_C$ (100 μM), in the presence and absence of an excess of full-length eIF4A (150 μM), are shown in Figure 2a. These spectra allowed backbone amide minimal shift values to be determined for nearly all the residues of bound Pdcd4 MA-3$_C$, which are summarized in the histogram shown in Figure 2b. Examination of the histogram clearly identifies 11 residues (H361, E362, E373, S374, T375, G376, N410, D414, I415, L416 and R428) in Pdcd4 MA-3$_C$ whose backbone amide signals undergo large chemical shift changes on binding of eIF4A. These residues are highlighted in red on the

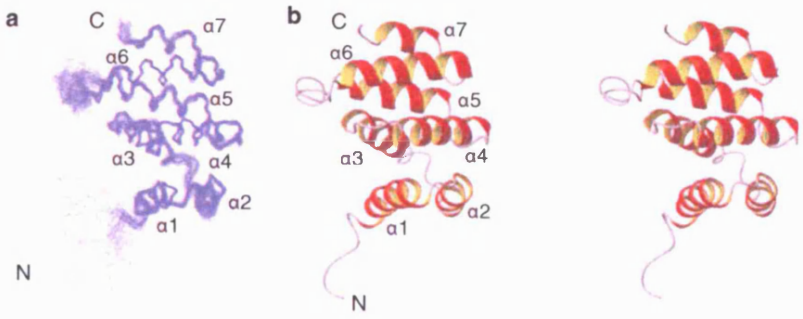


Figure 1 Solution structure of the Pdcd4 MA-3$_C$ domain. (**a**) Shows a best-fit superposition of the protein backbone for the family of 49 converged structures obtained. (**b**) Contains a stereo view of the ribbon representation of the backbone topology of Pdcd4 MA-3$_C$ based on the structure closest to the mean and is shown in the same orientation as in (**a**). The seven helical regions of Pdcd4 MA-3$_C$, α1, α2, α3, α4, α5, α6 and α7 are labelled, as are the N and C termini of the domain.

Pdcd4 MA-3$_C$ sequence shown in Supplementary Figure S1a.

Analogous chemical shift mapping experiments were also carried out for Pdcd4 MA-3$_C$ (100 μM) in the presence of a 5- and 10-fold molar excess of the isolated N-terminal domain of eIF4A. Similar changes in the HSQC spectra of Pdcd4 MA-3$_C$ were observed at both N-terminal domain eIF4A concentrations used, and are summarized for the fivefold molar excess by the histogram shown in Figure 2c. Analysis of the data clearly reveals the same groups of residues affected by binding of the isolated N-terminal domain of eIF4A as observed with the full-length protein, with the most substantial shifts (> 0.45 p.p.m.) seen for residues N325, H361, E362, E373, T375, G376, I408, I412, D414, I415, L417, R428 and V430. These residues are highlighted in red on the Pdcd4 MA-3$_C$ sequence shown in Supplementary Figure S1b. The side chain amide group of N416 was also significantly shifted by both full-length eIF4A and the isolated N-terminal domain. In contrast, no significant shifts were observed in ^{15}N/^{1}H HSQC spectra of Pdcd4 MA-3$_C$ (100 μM) acquired in the presence of a fivefold excess of the isolated C-terminal domain of eIF4A.

We were unable to use ^{15}N/^{1}H HSQC spectra to monitor the effects of eIF4A binding to a number of non-proline residues in Pdcd4 MA-3$_C$ (H326, V328, K329, E330, H349, V356, H360, Y365, K392, T395, I396, T397, M401, D418, V419, H421, S422 and S449) as it was not possible to identify the corresponding backbone amide peak in either the free or bound Pdcd4 MA-3$_C$ spectra. To try to establish whether any of these residues were located on the interface with eIF4A, ^{13}C/^{1}H HSQC spectra were acquired for Pdcd4 MA-3$_C$ (100 μM) in the presence and absence of the isolated N-terminal domain of eIF4A (400 μM). Analysis of the Cα-Hα region of the spectra provided reliable minimal shift information for 13 additional residues of Pdcd4 MA-3$_C$ (V328, V356, P557, K392, T395, I396, T397, M401, P413, D418, P420, S422 and P448) and substantial shifts were observed for residues I396, P413, D418 and S422, as shown in Figure 2c. In addition, several signals from methyl groups were perturbed significantly by eIF4A binding, in particular those from I331, I408 and I412, which also showed significant changes in their backbone amide signals.

Discussion

Structural features of the Pdcd4 MA-3$_C$ domain

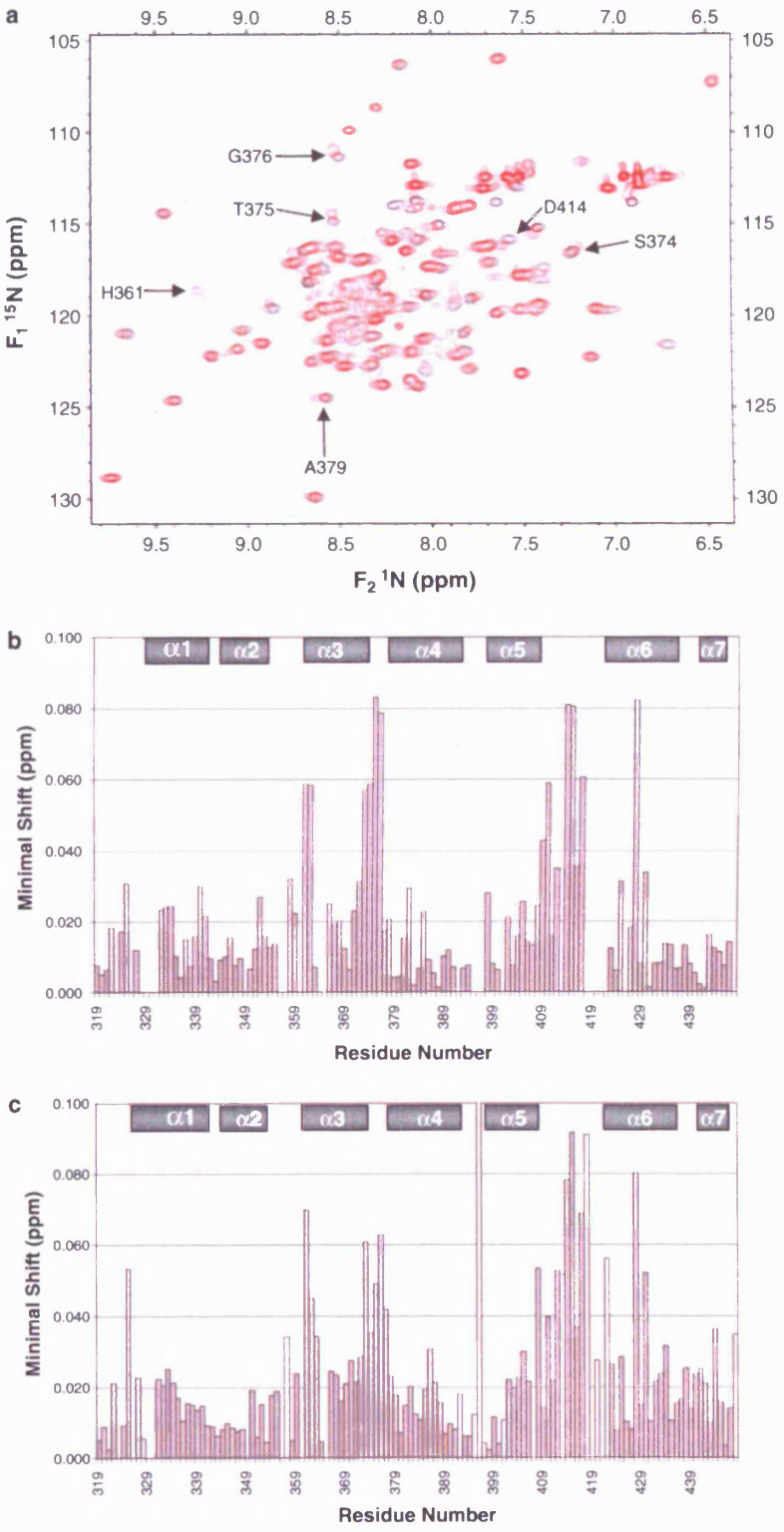
The backbone topology of the Pdcd4 MA-3$_C$ domain is illustrated by the ribbon diagram shown in Figure 1b. The domain is primarily composed of seven α-helices (α1 H326-S341, α2 I344-E353, α3 H361-E373, α4 S378-K392, α5 I398-I408, α6 S422-A436 and α7 K441-L446) linked by turns and loops. The six N-terminal helices are arranged as three pairs of antiparallel helix turn helix hairpins, which stack upon each other, with each pair of helices offset by approximately 45° from the previous layer. The C-terminal helix, α7, is located centrally above helices α5 and α6. One notable feature of the protein backbone is the loop between α5 and α6. In contrast to the short well-defined turns between α1 and α2 (two residues) and between α3 and α4 (four residues), the long loop between α5 and α6 (13 residues) appears to be relatively mobile and exist in multiple conformations. In part, the poor definition of this loop reflects the absence of backbone amide resonance assignments for several residues in this region, resulting in fewer NMR structural constraints; however, the broad lines observed for many signals are strongly indicative of the presence of conformational heterogeneity. Chemical shift and NOE data indicate that part of the loop (Y409–L417) has a propensity to adopt a helical conformation. This region of Pdcd4 MA-3$_C$ forms part of the interaction site with eIF4A (discussed below) and binding of eIF4A may stabilize the helical conformation, either extending the C-terminal end of α5 by up to three turns or forming a separate shorter helix. The amino acid sequence in this region is fairly well conserved across the different MA-3 domains (Supplementary Figure S1), with two aspartate residues (D414 and D418) being conserved in over 65% of MA-3 domains, and may reflect some functional importance of this region.

The stabilizing interactions between the helical hairpins and between their constituent helices rely almost entirely on favorable van der Waals contacts. A hydrogen bond between the carbonyl of L390 and the amide proton of I396, which are located in the C-terminal end of α4 and the following linker to α5, helps to stabilize this region of the domain. A second hydrogen bond was also identified between the side-chain hydroxyl group of Y409 and the carbonyl of L443, which contributes to a stable interaction between α5 and α7.

During the review of this paper the crystal structure of Pdcd4 MA-3$_C$ was reported on line (Laronde-Leblanc *et al.*, 2007). The solution and crystal structures contain nearly identical secondary and tertiary structure, which is reflected by an RMSD of 1.0 Å for the best fit of the backbone atoms for residues H326, V328-E353, E355-S374, G376-I412, N416 and H421-L446.

MA-3 domain structural homology

The structure of the C-terminal region of eIF4GI (residues 1234–1566) has recently been solved (Bellsolell *et al.*, 2006). This region contains a single MA-3 domain with approximately 36% homology to Pdcd4 MA-3$_C$. Comparison of the structures of the Pdcd4 MA-3$_C$ and eIF4G MA-3 domains clearly shows that they adopt similar secondary and tertiary structures, as highlighted by the ribbon representations shown in Figure 3a and by a backbone atom RMSD of 2.7 Å for the superposition of the two domains (residues Q321-V324, N325-S374, G376-I439, S440-P448 of Pdcd4 MA-3$_C$ and residues S1234-A1237, E1241-R1290, S1291-V1354, L1359-L1367 of eIF4GI MA-3). One major difference between the two MA-3 domain structures is the presence of an additional short helix (α5b) between helices five and six (α5 and α6) of eIF4G MA-3 (Figure 3b). As mentioned previously, the region between the fifth and


a
9.5
9.0
8.5
8.0
7.5
7.0
6.5
105
110
115
120
125
130
G376
T375
D414
S374
H361
A379
F1 15N (ppm)
F2 1N (ppm)
b
α1
α2
α3
α4
α5
α6
α7
0.100
0.080
0.060
0.040
0.020
0.000
Minimal Shift (ppm)
319
329
339
349
359
369
379
389
399
409
419
429
439
Residue Number
c
α1
α2
α3
α4
α5
α6
α7
0.100
0.080
0.060
0.040
0.020
0.000
Minimal Shift (ppm)
319
329
339
349
359
369
379
389
399
409
419
429
439
Residue Number

sixth helices of Pdcd4 MA-3$_C$ forms an ill-defined loop, part of which may adopt a helical conformation on eIF4A binding, as seen in eIF4G MA-3. The other notable difference between the two domains is the orientation of the seventh helix. The MA-3 domain of eIF4G forms part of a larger eIF4A-binding region, which comprises five antiparallel helical hairpins or HEAT repeats (Bellsolell *et al*., 2006). In contrast to the final C-terminal helix of Pdcd4 MA-3$_C$, which caps the previous layer of helices, α7 of eIF4G MA-3 forms part of a fourth helical hairpin, which is offset from the previous layer. Full-length Pdcd4 contains an additional 20 residues beyond the C terminus of Pdcd4 MA-3$_C$, however, this region is rich in glycine and polar residues, and is not predicted to adopt a helical conformation (Cuff *et al*., 1998).

Canonical HEAT repeats consist of tandem repeats of pairs of antiparallel α helices, which form right-handed, solenoid-type structures, with extended quasi-cylindrical shapes (Groves and Barford, 1999). The repeats contain a highly conserved proline residue within the first helix of each pair, which results in a marked bend in the helix (Andrade and Bork, 1995). Despite a lack of the conserved proline residue, the middle and C-terminal regions of eIF4G, including the MA-3 domain, are considered to be atypical members of the HEAT

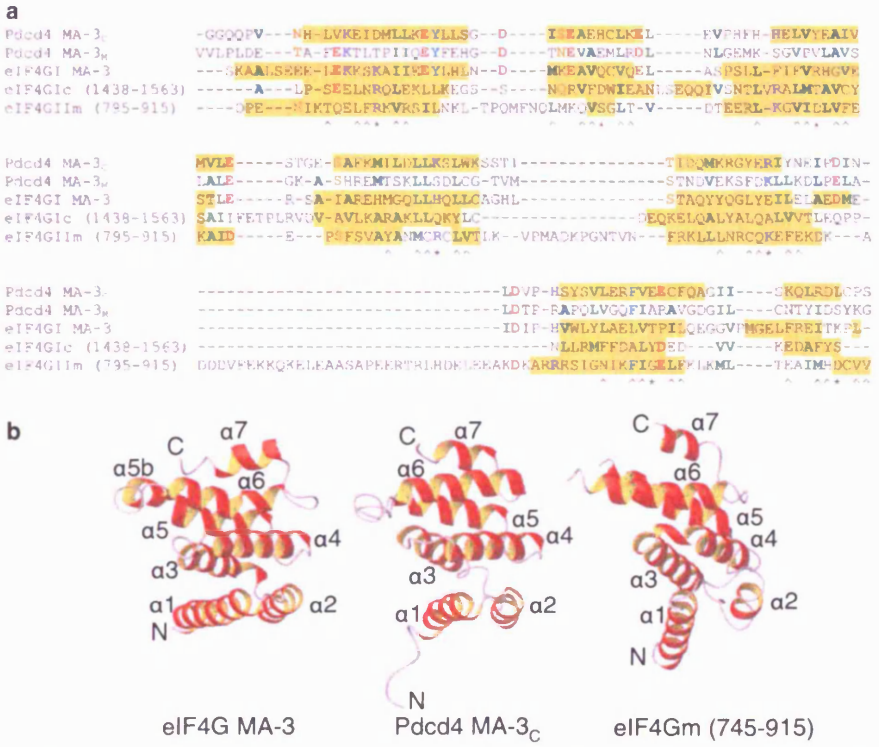


Figure 3 Structural homology between MA-3 domains and atypical HEAT repeats. (**a**) Shows a structure-based sequence alignment for Pdcd4 MA-3$_C$, eIF4GI MA-3 (residues 1234–1367), eIF4GI$_C$ C-terminal HEAT domain (residues 1438–1563) and eIF4GII$_M$ HEAT domain (residues 745–915), which was obtained using DALI (Holm and Sander, 1993). Pdcd4 MA-3$_M$ was aligned to Pdcd4 MA-3$_C$ on the basis of sequence conservation. Regions of helical secondary structure have been highlighted in yellow. Residues with conserved sequence similarity in three or more of the domains shown are highlighted as follows: hydrophobic (AVILM) in green, aromatic (FYW) in cyan, positively charged (KRH) in blue, negatively charged (DE) in red and polar (STNQ) in orange. The position of the conserved hydrophobic (^) and polar (*) residues found in the atypical HEAT repeats are highlighted below the alignment. (**b**) Shows equivalent views of ribbon representations of the backbone structure of the homologous regions of eIF4GI MA-3, Pdcd4 MA-3$_C$ and eIF4GII$_M$. Residues 851–880 of eIF4GII$_M$ are absent from the ribbon representation owing to the lack of structural data.

Figure 2 Mapping of the eIF4A-binding site on Pdcd4 MA-3$_C$. (**a**) Shows an overlay of two $^{15}N/^{1}H$ HSQC spectra of uniformly ^{15}N-labelled Pdcd4 MA-3$_C$ acquired in the presence and absence of full-length unlabelled eIF4A. The spectrum of the free Pdcd4 MA-3$_C$ domain (100 μM) is shown in black and for Pdcd4 MA-3$_C$ (100 μM) bound to full-length eIF4A (150 μM) in red. Peaks arising from a number of residues that were significantly shifted on formation of the Pdcd4 MA-3$_C$–eIF4A complex have been highlighted. (**b** and **c**) Show histograms of the minimal chemical shift changes observed for backbone amide groups of Pdcd4 MA-3$_C$ on binding of full length eIF4A (**b**) and the N-terminal domain (**c**). The positions of the helices in Pdcd4 MA-3$_C$ are highlighted above the histograms. In (**c**) minimal shift data for Cα/Hα is shown for a few residues where no information was obtained for the backbone amide groups (white bars). The bar for residue I396 has been truncated (minimal shift 0.113 p.p.m.).

repeat protein superfamily (Marcotrigiano *et al.*, 2001; Bellsolell *et al.*, 2006). HEAT repeats are generally involved in protein–protein interactions and several atypical HEAT repeats have been shown to play an essential role in the assembly of the protein synthesis machinery (Marcotrigiano *et al.*, 2001; Mazza *et al.*, 2001; Yamamoto *et al.*, 2005; Bellsolell *et al.*, 2006). The amino acid sequences of the helices in Pdcd4 MA-3$_C$ fit the proposed consensus sequence for atypical HEAT repeats (Bellsolell *et al.*, 2006). The consensus sequence features three clusters of conserved hydrophobic residues in each helix (Figure 3a), which allow favourable contacts within the HEAT repeat and with adjacent HEAT repeats. There is also a conserved solvent exposed polar residue, which lies between the second and third hydrophobic clusters.

Comparison of the backbone topology of Pdcd4 MA-3$_C$ with other known folds in the Protein Data Bank using Dali identified almost 600 structural neighbours (Holm and Sander, 1993). Four of the five closest structural homologues were of the type found in eIF4G and often referred to as mIF4G domains, which have been found in a number of eIFs and the nuclear cap-binding protein CBP80 (Ponting, 2000). This confirms previous suggestions that despite little sequence homology, MA-3 domains might be distantly related examples of mIF4G domains (Ponting, 2000).

Among the mIF4G domains of known structure, the closest structural homologue to Pdcd4 MA-3$_C$ is the central region of eIF4GII (eIF4G$_M$, residues 745–915); (Marcotrigiano *et al.*, 2001), which is reflected in a backbone RMSD of 3.55 Å (residues 322–326, 327–341, 343, 344–352, 354–374, 378–394, 397–413, 417–420, 421–449 of Pdcd4 MA-3$_C$ and residues 745–749, 751–765, 768, 773–781, 782–802, 804–820, 834–849, 882–885, 886–915 of eIF4GII$_M$) and is clearly apparent from the structural views shown in Figure 3b. eIF4G$_M$ is involved in binding to eIF3, RNA and eIF4A (Marintchev and Wagner, 2004). The interaction between eIF4G$_M$ and the C-terminal domain of eIF4A is believed to stabilize the closed, catalytically active orientation of the two domains of eIF4A (Oberer *et al.*, 2005). The most striking difference between Pdcd4 MA-3$_C$ and eIF4G$_M$ (745–915) is the loop between α5 and α6, which is approximately thirty residues longer than the corresponding loop in Pdcd4 MA-3$_C$. There is no conformational data available for this extended loop in eIF4G$_M$, as the region is not visible in the electron density maps used to determine the crystal structure (Marcotrigiano *et al.*, 2001). This region may serve a similar function to the flexible, ill defined loop of Pdcd4 MA-3$_C$, which is involved in binding partner proteins (see below). Pdcd4 MA-3$_C$ also shows structural homology to the C-terminal HEAT domain of eIF4GI$_C$, which is involved in binding to Mnk1 (Bellsolell *et al.*, 2006). This is reflected in a backbone RMSD of 3.5 Å (residues 324–354, 355–360, 361–373, 374–392, 397–416, 421–434, 438–446 of Pdcd4 MA-3$_C$ and residues 1438–1468, 1474–1479, 1481–1493, 1499–1517, 1521–1540, 1541–1554, 1555–1563 of eIF4GI). A structure-based alignment of the sequences from Pdcd4 MA-3$_C$ and the atypical HEAT domains of eIF4G is shown in Figure 3a.

Relative positioning of the MA-3 domains in Pdcd4

The two MA-3 domains of Pdcd4 share 18% sequence identity and 22% sequence similarity, which suggests they will adopt similar overall structures. The domains are linked by a stretch of 33 residues, which is predicted to contain a 16-residue long α-helix. This helix, together with the expected seventh C-terminal helix of the middle MA-3 domain of Pdcd4 (Pdcd4 MA-3$_M$), may form a fourth HEAT repeat and thereby a somewhat extended domain. The remaining 15 residues in the linker region are predicted to form an extended or irregular structure and the presence of six glycine residues implies significant flexibility. This region may therefore allow some flexibility in the positioning and orientation of the two MA-3 domains, perhaps facilitating interactions with both domains of eIF4A or other partner proteins. Comparison of the linewidths for backbone amide signals in $^{15}N/^{1}H$ HSQC spectra obtained from the isolated Pdcd4 MA-3$_C$ domain and a protein corresponding to the region encompassing both MA-3 domains, strongly suggests that there is a significant degree of independent motion of the two domains consistent with a flexible linker. Analysis of HNCO spectra for isolated Pdcd4 MA-3$_C$ and the dual MA-3 domain protein identified small, but significant shifts in signals from Pdcd4 MA-3$_C$ residues located on the exposed surface of α1 and α2 (data not shown), which suggests some transient interaction between the two MA-3 domains that may be stabilized on binding of eIF4A.

Characterization of eIF4A binding to Pdcd4

The chemical shift changes induced by eIF4A binding have clearly highlighted a number of residues on Pdcd4 MA-3$_C$ that are located at the interface with eIF4A, most notably E373, S374, T375 and G376, which are positioned at the C-terminal end of α3 and the following turn, and N410, I412, P413, D414, I415, L417 and D418, which are found at the C-terminal end of α5 and the following ill-defined loop (Figure 2 and Supplementary Figure S1). These residues are arranged in a broad strip on the surface of Pdcd4 MA-3$_C$, which runs over a quarter of the way around the molecule, as illustrated in Figure 4a and b. A few of the residues with shifted signals (Y405, I408 and S422) are hidden behind the α5–α6 loop and so observed changes here may reflect localized conformational changes in Pdcd4 MA-3$_C$ induced by eIF4A binding. As discussed previously, the N-terminal half of the ill-defined loop between α5 and α6 may adopt a helical conformation upon eIF4A binding. This conformational change could expose the hidden residues and form a groove in the surface of Pdcd4 MA-3$_C$ between residues E373-G376 and D414-H421. A small group of residues at the N-terminal end of α3 were also affected by eIF4A binding (Figures 2 and 4, and Supplementary Figure S1). These residues are somewhat separated from the main interaction site

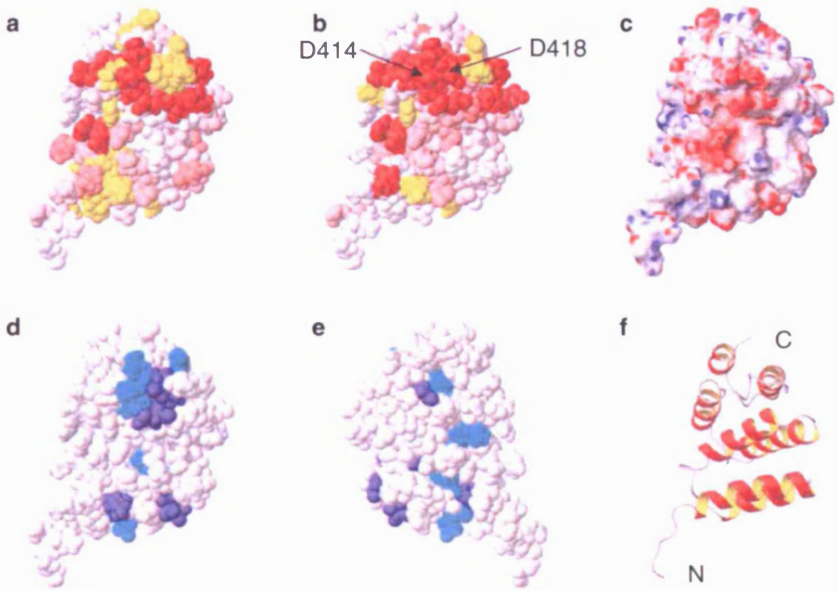


Figure 4 Location, features and properties of the eIF4A-binding site on Pdcd4 MA-3$_C$. (**a**) Shows a space-filled view of Pdcd4 MA-3$_C$, in which residues are coloured according to the perturbation of their backbone amide signals induced by eIF4A binding. Residues that showed a minimal shift change of less than 0.010 p.p.m. are shown in white, over 0.045 p.p.m. in red and between 0.010 and 0.045 p.p.m. are coloured according to the magnitude of the shift on a linear gradient between white and red. Residues for which no minimal shift data was obtained are shown in yellow. In (**b**) the backbone amide chemical shift changes induced by binding of the isolated N-terminal domain are highlighted using the same colour scheme as described above. Where possible, residues for which no backbone amide minimal shift data were available were coloured according to changes of their Cα/Hα signals (V328, V356, P357, K392, T395, I396, T397, M401, P413, D418, P420, S422 and P448) with the same colour scheme used. The position of D414 and D418, which have previously been reported to be involved in the interaction between Pdcd4 MA-3$_C$ and eIF4A, are also highlighted. (**c**) Shows a surface view of Pdcd4 MA-3$_C$, which is coloured according to electrostatic potential, with areas of significant negative charge shown in red, significant positive charge in blue and neutral in white. (**d** and **e**) Show space-filled views of Pdcd4 MA-3$_C$, in which residues are highlighted on the basis of sequence conservation, with equivalent residues identical in the MA-3 domains of Pdcd4 and eIF4G shown in blue and those conserved to a close homologue shown in cyan. (**e**) Rotated by 180° about the Y axis from the view shown in (**d**). (**f**) Shows a ribbon representation of the backbone topology of Pdcd4 MA-3$_C$, in the same orientation as (**a**–**d**).

and it seems probable that the changes arise from a slight change in orientation of α3 induced by the interaction of eIF4A with residues at the C-terminus of α3.

Very similar patterns of chemical shift changes were observed for Pdcd4 MA-3$_C$ in the presence of both full-length eIF4A and the N-terminal domain alone, which clearly demonstrates that only the N-terminal domain is required for binding to Pdcd4 MA-3$_C$. In addition, no changes were observed for Pdcd4 MA-3$_C$ in the presence of the isolated C-terminal domain. Previous studies have shown that both the N- and C-terminal domains of eIF4A, and the two MA-3 domains of Pdcd4 are required for the formation of a very high affinity complex (Yang *et al*., 2004; Zakowicz *et al*., 2005). As Pdcd4 MA-3$_C$ clearly interacts with the N-terminal domain of eIF4A, this strongly suggests that the C-terminal domain will bind to Pdcd4 MA-3$_M$, giving a head to tail arrangement to the two proteins in the complex.

Calculation of the electrostatic surface of Pdcd4 MA-3$_C$ revealed that the eIF4A interaction site is predominantly negatively charged, as shown in Figure 4c, which is consistent with the high number of acidic residues in this region. Interestingly, several of these acidic residues are conserved in the other MA-3 domains known to bind eIF4A. The structure of the N-terminal domain of human eIF4A, which is identical in sequence to mouse eIF4A, has recently been deposited (PDB code 2G9N). This reveals that the surface of the N-terminal domain of eIF4A is characterized by a number of large basic patches, located in and around the regions involved in ATP and RNA binding, as shown in Supplementary Figure S2 (Pause and Sonenberg, 1992; Pause *et al*., 1993; Svitkin *et al*., 2001; Tanner *et al*., 2003). It therefore seems very likely that Pdcd4 MA-3$_C$ will bind to one or more of these basic regions on eIF4A, which is supported by recent mutagenesis data (Zakowicz *et al*., 2005).

The degree of amino acid sequence conservation for surface-exposed residues in the MA-3 domains of Pdcd4 and eIF4GII, is illustrated by the space-filled views of Pdcd4 MA-3$_C$ shown in Figure 4d and e. There is clearly significantly more conservation in the region shown here to form the interface with eIF4A, which suggests that all three MA-3 domains share a common eIF4A-binding site. Interestingly, one of the strictly conserved residues, D414, is substituted by a lysine in the MA-3 domain of

the eIF4G homologue DAP-5/NAT1/p97, which is unable to bind eIF4A (Imataka and Sonenberg, 1997; Yang *et al*., 2004).

Recent mutagenesis studies of Pdcd4 have shown that the substitution of D414 with lysine, or D418 by alanine, results in a dramatic fall in eIF4A binding (Yang *et al*., 2004). These two aspartate residues are located adjacent to each other on the surface of Pdcd4 MA-3$_C$ and lie at the heart of the eIF4A-binding site identified by the NMR studies reported here, as highlighted in Figure 4b. The equivalent mutations in Pdcd4 MA-3$_M$ and eIF4G MA-3 also result in dramatic falls in eIF4A binding, which strongly supports the suggestion of a conserved eIF4A-binding site (Yang *et al*., 2004). The Pdcd4 D418A variant was shown to retain full eIF4G$_M$-binding activity, which clearly suggests that Pdcd4 interacts with eIF4G$_M$ through a site distinct from that involved in eIF4A binding (Yang *et al*., 2003a).

Similar site-directed mutagenesis studies of the interaction of eIF4GI (residues 1080–1600, eIF4G$_C$) with eIF4A, suggested that a cluster formed by R1281 and four surrounding glutamate residues on the MA-3 domain formed part of the probable-binding site for eIF4A (Bellsolell *et al*., 2006). We were not able to obtain chemical shift perturbation data for the equivalent residue of Pdcd4 MA-3$_C$ (Y365), however, it is located within the N-terminal half of α3. This region showed significant shifts in signals on eIF4A binding, which were attributed to localized conformational changes, rather than direct contact with eIF4A. As discussed previously, the eIF4A-binding site identified on Pdcd4 by NMR is largely conserved on eIF4G (see Figure 4d), which strongly implies a common binding site for eIF4A on both MA-3 domains. In contrast, neither R1281 nor the surrounding glutamate residues are conserved in Pdcd4 MA-3$_C$. The study of eIF4G$_C$ also suggested that the loop between helices five and six of eIF4G MA-3 was not required for eIF4A binding (Bellsolell *et al*., 2006). The work reported here clearly shows that the equivalent loop in Pdcd4 MA-3$_C$ forms a major part of the interface with eIF4A, which is supported by previous mutagenesis studies of Pdcd4 and eIF4G (Yang *et al*., 2003a, 2004). Given the structural similarity of the two MA-3 domains, the level of sequence conservation in this region and the earlier mutational data, it seems very likely that the α5–6 loop is an important element of the eIF4A-binding site in both Pdcd4 MA-3$_C$ and eIF4G MA-3.

Conclusions

The work reported here unambiguously identifies the binding site for eIF4A on the C-terminal MA-3 domain of Pdcd4, which is strongly conserved in other MA-3 domains involved in eIF4G binding. In addition, the N-terminal domain of eIF4A is shown to be sufficient for binding to Pdcd4 MA-3$_C$, which strongly implies that in complexes formed by the intact proteins the C-terminal domain of eIF4A will bind to the middle MA-3 domain of Pdcd4, resulting in an inhibitory complex with a head to tail arrangement of the proteins.

Materials and methods

Protein expression and purification

The unlabelled and uniformly ^{15}N- ^{13}C- and ^{15}N/^{13}C- labelled samples of Pdcd4 MA-3$_C$ and Pdcd4 (157–449) containing both MA-3 domains were prepared from pGex6P2-based *Escherichia coli* expression vectors as described previously (Waters *et al*., 2006b). The N-terminal histidine-tagged full-length mouse eIF4A, eIF4A N-terminal domain (residues 1–242) and eIF4A C-terminal domain (residues 237–406) were prepared from pET28a-based *E. coli* expression vectors. The proteins were initially affinity-purified on a Ni-NTA column under nondenaturing conditions, followed by gel filtration chromatography on a Superdex 75 16/60 prepacked column (Amersham Biosciences, GE Healthcare, Buckinghamshire, UK), essentially as described previously (Renshaw *et al*., 2002; Waters *et al*., 2006a).

NMR spectroscopy

NMR spectra required to determine the Pdcd4 MA-3$_C$ structure were recorded as described previously (Waters *et al*., 2006b). ^{15}N/^{1}H and ^{13}C/^{1}H HSQC spectra of Pdcd4 MA-3$_C$ were acquired in the presence and absence of eIF4A to identify the changes in the positions of signals induced by eIF4A binding. In these experiments, unlabelled full-length eIF4A or the N- or C-terminal domains alone, were added to 100 μM ^{15}N- or ^{13}C-labelled Pdcd4 MA-3$_C$ in a 25 mM sodium phosphate, 125 mM sodium chloride, 5 mM dithiothreitol, 10 μM ethylenediaminetetraacetic acid and 0.02% (w/v) sodium azide buffer at pH 7. The full-length eIF4A was titrated to a concentration of 150 μM (limiting solubility of eIF4A) and the isolated N- and C-terminal domains to between 400 μM and 1 mM. ^{15}N/^{1}H HSQC and HNCO spectra of 300 μM Pdcd4 (157–449) and 200 μM Pdcd4 MA-3$_C$ were acquired to identify shifts in the position of Pdcd4 MA-3$_C$ signals in the dual MA-3 domain protein. Typical acquisition times for the HNCO were 75 ms in F_3 (^{1}H), 14.5 ms in F_2 (^{15}N) and 8.6 ms in F_1 (^{13}C), with the spectra collected over approximately 40 h. The 2D and 3D NMR data were processed and analysed as described previously (Waters *et al*., 2006b).

Structural calculations

The family of converged Pdcd4-MA3$_C$ structures was determined in a two-stage process using the program CYANA, as described previously (Renshaw *et al*., 2005; Waters *et al*., 2006a). Hydrogen bond constraints, involving 17 residues with slowly exchanging backbone amide signals (residues 337, 338, 366–370, 372, 381, 384–388, 429–431) and where the hydrogen bond acceptor was unambiguous in preliminary structures, were added to the final round of calculations. Backbone torsion angle constraints derived from the protein backbone dihedral angle prediction program TALOS were included in both stages of the calculation (Cornilescu *et al*., 1999). Analysis of the final family of structures obtained was carried out using the programs CYANA and MOLMOL (Koradi *et al*., 1996; Guntert *et al*., 1997).

Mapping of binding sites

The minimal shift approach was used to determine the changes in the positions of Pdcd4 MA-3$_C$ NMR signals resulting from

eIF4A binding (Williamson *et al*., 1997; Muskett *et al*., 1998). The minimum change in the position for all backbone amide peaks between the free and eIF4A-bound Pdcd4 MA-3$_C$ was obtained by using Microsoft Excel to calculate the combined chemical shift difference in ^{15}N and ^{1}H for each assigned peak in the ^{15}N/^{1}H HSQC spectrum of the free protein compared to all peaks observed in the HSQC spectra of the complexes formed with eIF4A. The combined amide proton and nitrogen chemical shift difference ($\Delta\delta$) was defined according to the following equation: $\Delta\delta = \sqrt{(\Delta\delta_{HN})^2 + (\Delta\delta_N \cdot \alpha_N)^2}$ where $\Delta\delta_{HN}$ and $\Delta\delta_N$ correspond to the differences in ^{1}H and ^{15}N chemical shifts between pairs of compared HSQC peaks and α_N is a scaling factor of 0.2 required to account for differences in the range of amide proton and nitrogen chemical shifts. For each individual HSQC peak, the minimal shift induced by ligand binding was taken as the lowest possible combined shift value ($\Delta\delta$). The minimum change in the position of peaks arising from the well-resolved Cα/Hα and methyl signals in ^{13}C/^{1}H HSQC spectra was similarly determined, but using a scaling factor of 0.1 for the Cα shifts and 0.12 for the methyl carbon shifts. The minimal shift approach was also used to determine any significant changes in the positions of Pdcd4 MA-3$_C$ signals in HNCO spectra of the dual MA-3 domain protein, with a scaling factor of 0.35 used for carbonyl signals.

Acknowledgements

This work was supported by grants from the Wellcome Trust, Wilhelm-Sander-Stiftung and the Deutsche Krebshilfe. Lorna Waters was supported by a BBSRC PhD studentship.

References

Afonja O, Juste D, Das S, Matsuhashi S, Samuels H. (2004). Induction of PDCD4 tumor suppressor gene expression by RAR agonists, antiestrogen and HER-2/neu antagonist in breast cancer cells. Evidence for a role in apoptosis. *Oncogene* **23**: 8135–8145.

Andrade M, Bork P. (1995). HEAT repeats in the Huntingtons-disease protein. *Nat Genet* **11**: 115–116.

Bellsolell L, Cho-Park PF, Poulin F, Sonenberg N, Burley SK. (2006). Two structurally atypical HEAT domains in the C-terminal portion of human eIF4G support binding to eIF4A and Mnk1. *Structure* **14**: 913–923.

Bitomsky N, Bohm M, Klempnauer K. (2004). Transformation suppressor protein Pdcd4 interferes with JNK-mediated phosphorylation of c-Jun and recruitment of the coactivator p300 by c-Jun. *Oncogene* **23**: 7484–7493.

Bohm M, Sawicka K, Siebrasse J, Brehmer-Fastnacht A, Peters R, Klempnauer K. (2003). The transformation suppressor protein Pdcd4 shuttles between nucleus and cytoplasm and binds RNA. *Oncogene* **22**: 4905–4910.

Chen Y, Knosel T, Kristiansen G, Pietas A, Garber M, Matsuhashi S *et al*. (2003). Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis. *J Pathol* **200**: 640–646.

Cmarik J, Min H, Hegamyer G, Zhan S, Kulesz-Martin M, Yoshinaga H *et al*. (1999). Differentially expressed protein Pdcd4 inhibits tumor promoter-induced neoplastic transformation. *Proc Natl Acad Sci USA* **96**: 14037–14042.

Cornilescu G, Delaglio F, Bax A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**: 289–302.

Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**: 892–893.

Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T *et al*. (2006). Pfam: clans, web tools and services. *Nucleic Acids Res* **34**: D247–51.

Groves MR, Barford D. (1999). Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* **9**: 383–389.

Guntert P, Mumenthaler C, Wuthrich K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**: 283–298.

Herrmann T, Guntert P, Wuthrich K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**: 209–227.

Hershey JWB, Merrick WC. (2000). Pathway and mechanism of initiation of protein synthesis. In: Sonenberg N, Hershey JWB, and Mathews MB (eds). *Translational Control of Gene expression: Pathway and Mechanism of Initiation of Protein Synthesis*. Cold Spring Harbor Laboratory Press: New York, pp 33–88.

Holm L, Sander C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**: 123–138.

Imataka H, Sonenberg N. (1997). Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol Cell Biol* **17**: 6940–6947.

Jansen A, Camalier C, Stark C, Colburn N. (2004). Characterization of programmed cell death 4 in multiple human cancers reveals a novel enhancer of drug sensitivity. *Mol Cancer Ther* **3**: 103–110.

Koradi R, Billeter M, Wuthrich K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**: 51–55.

Laronde-Leblanc N, Santhanam AN, Baker AR, Wlodawer A, Colburn NH. (2007). Structural basis for inhibition of translation by the tumor suppressor Pdcd4. *Mol Cell Biol* **27**: 147–156.

Marcotrigiano J, Lomakin I, Sonenberg N, Pestova T, Hellen C, Burley S. (2001). A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell* **7**: 193–203.

Marintchev A, Wagner G. (2004). Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys* **37**: 197–284.

Mazza C, Ohno M, Segref A, Mattaj IW, Cusack S. (2001). Crystal structure of the human nuclear cap binding complex. *Mol Cell* **8**: 383–396.

Muskett F, Frenkiel T, Feeney J, Freedman R, Carr M, Williamson R. (1998). High resolution structure of the N-terminal domain of tissue inhibitor of metalloproteinases-2 and characterization of its interaction site with matrix metalloproteinase-3. *J Biol Chem* **273**: 21736–21743.

Oberer M, Marintchev A, Wagner G. (2005). Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev* **19**: 2212–2223.

Palamarchuk A, Efanov A, Maximov V, Aqeilan R, Croce C, Pekarsky Y. (2005). Akt phosphorylates and regulates Pdcd4 tumor suppressor protein. *Cancer Res* **65**: 11282–11286.

Pause A, Methot N, Sonenberg N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor-4A is required for RNA-binding and ATP hydrolysis. *Mol Cell Biol* **13**: 6789–6798.

Pause A, Sonenberg N. (1992). Mutational analysis of a DEAD box RNA helicase- The mammalian translation initiation-factor eIF-4A. *EMBO J* **11**: 2643–2654.

Ponting CP. (2000). Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem Sci* **25**: 423–426.

Renshaw P, Lightbody K, Veverka V, Muskett F, Kelly G, Frenkiel T *et al.* (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J* **24**: 2491–2498.

Renshaw PS, Panagiotidou P, Whelan A, Gordon SV, Hewinson RG, Williamson RA *et al.* (2002). Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6.CFP-10 complex. Implications for pathogenesis and virulence. *J Biol Chem* **277**: 21598–21603.

Rozen F, Edery I, Meerovitch K, Dever T, Merrick W, Sonenberg N. (1990). Bidirectional RNA helicase activity of eukaryotic translation initiation factor-4A and factor-4F. *Mol Cell Biol* **10**: 1134–1144.

Shibahara K, Asano M, Ishida Y, Aoki T, Koike T, Honjo T. (1995). Isolation of a novel mouse gene MA-3 that is induced upon programmed cell death. *GENE* **166**: 297–301.

Svitkin Y, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham G *et al.* (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5′ secondary structure. *RNA* **7**: 382–394.

Tanner N, Cordin O, Banroques J, Doere M, Linder P. (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* **11**: 127–138.

Waters L, Yue B, Veverka V, Renshaw P, Bramham J, Matsuda S *et al.* (2006a). Structural diversity in p160/CREB-binding protein coactivator complexes. *J Biol Chem* **281**: 14787–14795.

Waters LC, Bohm M, Veverka V, Muskett FW, Frenkiel TA, Kelly GP *et al.* (2006b). NMR assignment and secondary structure determination of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4. *J Biomol NMR* **36**: S5:18.

Williamson R, Carr M, Frenkiel T, Feeney J, Freedman R. (1997). Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry* **36**: 13882–13889.

Yamamoto Y, Singh CR, Marintchev A, Hall NS, Hannig EM, Wagner G *et al.* (2005). The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G. *Proc Natl Acad Sci USA* **102**: 16164–16169.

Yang H, Cho M, Zakowicz H, Hegamyer G, Sonenberg N, Colburn N. (2004). A novel function of the MA-3 domains in transformation and translation suppressor Pdcd4 is essential for its binding to eukaryotic translation initiation factor 4A. *Mol Cell Biol* **24**: 3894–3906.

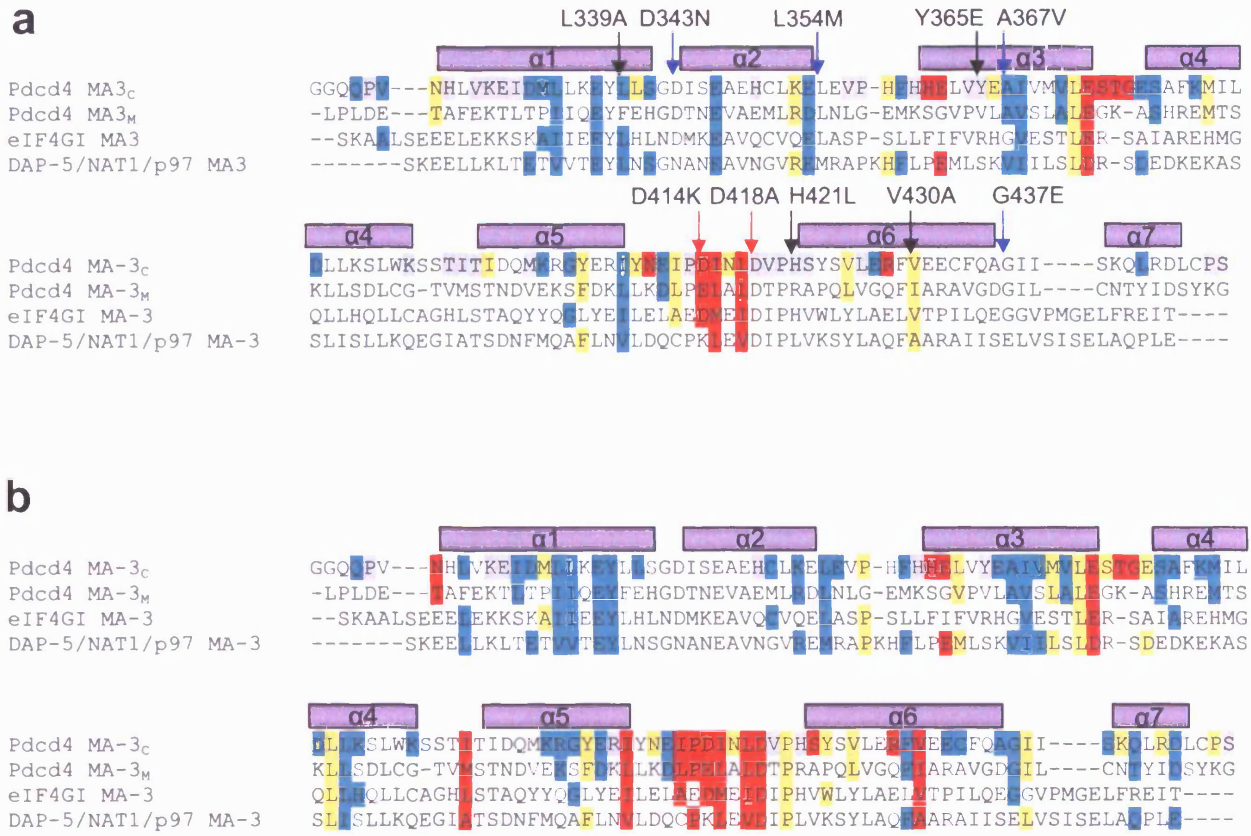
Yang H, Jansen A, Komar A, Zheng X, Merrick W, Costes S *et al.* (2003a). The transformation suppressor Pdcd4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation. *Mol Cell Biol* **23**: 26–37.

Yang H, Jansen A, Nair R, Shibahara K, Verma A, Cmarik J *et al.* (2001). A novel transformation suppressor, Pdcd4, inhibits AP-1 transactivation but not NF-kappa B or ODC transactivation. *Oncogene* **20**: 669–676.

Yang H, Knies J, Stark C, Colburn N. (2003b). Pdcd4 suppresses tumor phenotype in JB6 cells by inhibiting AP-1 transactivation. *Oncogene* **22**: 3712–3720.

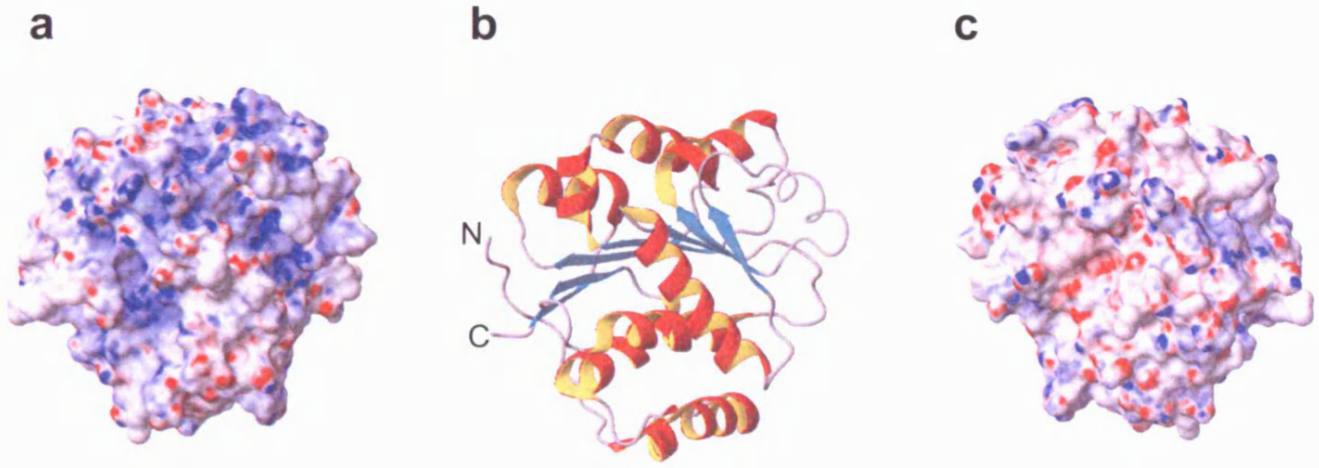
Zakowicz H, Yang H, Stark C, Wlodawer A, Laronde-Leblanc N, Colburn N. (2005). Mutational analysis of the DEAD-box RNA helicase eIF4AII characterizes its interaction with transformation suppressor Pdcd4 and eIF4GI. *RNA* **11**: 261–274.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc).



Supplementary figure S1 - Waters

Supplementary figure 1: A summary of the Pdcd4 MA-3_C domain residues implicated in eIF4A binding. Panel a shows a sequence alignment of the MA-3 domains from Pdcd4, eIF4GI and DAP-5/NAT1/p97. The MA-3 domains of Pdcd4 and DAP-5/NAT1/p97 are aligned on the basis of amino acid sequence similarity, whilst eIF4G MA-3 is aligned with Pdcd4 MA-3_C according to the structural based sequence alignment produced by DALI (Holm & Sander, 1993). The positions of the helices in Pdcd4 MA-3_C are shown above the alignment. Residues in Pdcd4 MA-3_C are colour coded to indicate the minimal shifts in backbone amide signals induced by eIF4A binding, with changes of between 0.015 and 0.025 ppm shown in blue, between 0.025 and 0.045 ppm in yellow and by over 0.045 ppm in red. Equivalent conserved residues are also highlighted in the other MA-3 domains. Residues for which no chemical shift data was obtained are shown in grey. The position of a number of site-directed mutants of Pdcd4 characterised by Yang and co-workers are also indicated. L339A, Y365E, H421L and V430A were found to have no effect on eIF4A binding. D343N, L354M, A367V and G437E reduced binding by 40-60%. D414K and D418A effectively blocked eIF4A binding. In panel b, residues of Pdcd4 MA-3_C with backbone amide signals that were significantly perturbed by binding of the N-terminal domain of eIF4A are colour coded as described for Panel a. For a few residues where no chemical shift perturbation could be obtained for backbone amide signals (V328, V356, P357, K392, T395, I396, T397, M401, P413, D418, P420, S422 and P448), the colour coding refers to minimal shifts for Cα/Hα signals.



Supplementary figure S2 - Waters

Supplementary figure 2: Structure and properties of the isolated N-terminal domain of human eIF4A.

Panels a and c show a surface view of the N-terminal domain of human eIF4A (PDB code 2G9N), which is coloured according to electrostatic potential, with areas of significant negative charge in red, positive charge in blue and neutral in white. The structure in panel c is rotated by 180° about the Y axis from the view shown in panel a. Panel b shows a ribbon representation of the backbone topology of the N-terminal domain in the same orientation as panel a.

NMR Constraints and Structural Statistics for Pdcd4 MA-3_C		
a) No. of Constraints used in Final Structural Calculation		
Intraresidue NOEs	614	
Sequential NOEs (*i*, *i*+1)	562	
Medium-range NOEs (*i*, *i*≤4)	599	
Long-range NOEs (*i*, *i* ≥5)	771	
Torsion angles	174	(87Φ and 87Ψ)
Hydrogen bonding	68	
b) Maximum and Total Constraint Violations in 49 Converged Pdcd4 MA-3_C Structures		
Upper distance limits (Å)	0.32 ± 0.02	2.9 ± 0.7
Lower distance limits (Å)	0.00 ± 0.01	0.0 ± 0.0
van der Waals contacts (Å)	0.27 ± 0.02	5.6 ± 0.8
Torsion angle ranges (°)	3.03 ± 0.67	16.2 ± 3.4
Average CYANA target function ($Å^2$)	1.46 ± 0.25	
c) Structural Statistics for the Family of Converged Pdcd4 MA-3_C Structures		
Residues within allowed region of Ramachandran plot (%)		99.7%
Backbone atom RMSD for structured region (residues 327-356, 359-412,421-448 of Pdcd4 MA-3_C)		0.49 ± 0.07 Å
Heavy atom RMSD for structured region (residues 327-356, 359-412,421-448 of Pdcd4 MA-3_C)		0.92 ± 0.07 Å

Supplementary Table S1