Structural Study of the Chromodomain Superfamily.

Nathan Philip Cowieson

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh, 2001
Declaration

I declare that this thesis was composed by myself and the research presented is my own except where otherwise stated.

Nathan Philip Cowieson
2001
Acknowledgments

First and foremost I would like to thank my supervisors Paul McLaughlin and Janet Partridge who both went above and beyond the call of duty to help me through my PhD. Their respective styles of supervision complemented each other very well and I feel lucky to have had them as supervisors. My funding was from the Medical Research Council who handled it very efficiently.

I’d also like to thank various people whom I have come into contact with during the course of this project and have helped me along, either by scientific advice and discussion or just witty banter. They’d be, in no particular order, Martin Weir, Iain McNae, Joel Tyndall, Lyndsay Sawyer, Brian Smith, Mary Russell, Jaqueline Dornan, Violet Anderson, Nichola Pichen, Sandra Bruce, Paul Taylor, Malcolm Walkinshaw, Marjorie Harding and Robin Allshire.
For Donna Cowieson (seeynmrn'n)
Abstract

Chromodomains are a superfamily of protein domains that are implicated in the modulation of chromatin structures. The chromodomain superfamily can be subdivided into two subfamilies, chromodomains and chromo shadow domains. Chromodomains function to localise proteins to specific sites on chromatin. An NMR structure of a chromodomain has been previously solved. Chromo shadow domains are protein-protein interaction domains. These recruit other chromatin associated proteins to their sites of action. The two domains have sequence similarities and are likely to have similar structures. The basis for their functional divergence is not known. In the present study, an X-ray structure of the chromo shadow domain of fission yeast Swi6 protein was solved to 1.9Å. The structure reveals that the chromo shadow domain is a dimer with monomers closely resembling the chromodomain structure. Dimerisation of the chromo shadow domains creates a cleft that is commensurate with peptide binding. Binding studies were carried out to further investigate chromo shadow domain function. Recently, proteins of the *Drosophila* dosage compensation complex were found to contain divergent chromodomain-like domains. These domains have been noted to have RNA binding properties. Preliminary structural studies were carried out to determine if these domains shared a common fold with chromodomains and chromo shadow domains.
Contents

1. Introduction 6

1.1 Chromatin Structure ..................................6
1.2 Chromatin function ....................................10
1.3 Chromodomains modulate chromatin structures ...............12
  1.3.1 Polycomb and Trithorax group mediated transcription regulation. ....................................15
1.4 Chromodomain proteins and heterochromatin ..................22
  1.4.1 Identification of heterochromatin associated proteins ....22
  1.4.2 Fission yeast centromeres. ..........................23
  1.4.3 Dynamics of HP1 in relation to heterochromatin ........26
1.5 HP1 chromodomain and chromo shadow domain function .....27
1.6 Chromodomains and dosage compensation. ....................34
1.7 Summary and questions about the
  chromodomain superfamily. ................................37

2. Materials and methods 40

2.1 Molecular biology ......................................40
  2.1.1 Expression Vectors ..................................40
  2.1.2 PCR primers .........................................40
2.1.3 PCR protocol ........................................... 42
2.1.4 Electrophoresis and purification of DNA ............ 42
2.1.5 Digestion of PCR product and vector ................ 43
2.1.6 Ligation ................................................. 43
2.1.7 Preprepared plasmids .................................. 43
2.1.8 Transformation of _E.coli_ ............................ 44

2.2 Preparation of protein .................................... 46
  2.2.1 Solubilisation and purification of Rad21 and Clr4 ... 47
  2.2.2 Purification of the Swi6 CSD .......................... 48
  2.2.3 Purification of the T323D Swi6 CSD ................... 49
  2.2.4 Purification of the chromodomain-like region of Alp13 ... 49
  2.2.5 Preparation of DNA polymerase α and CAF1 p150 fragment 50

2.3 Crystallisation of Swi6 CSD ................................ 50

2.4 GST pulldowns ............................................. 53
  2.4.1 GST-CSD interaction with CSD ......................... 53
  2.4.2 GST pulldowns with soluble proteins ................... 54
  2.4.3 GST pulldowns with insoluble proteins ................ 55

2.5 SELDI mass spectrometry .................................. 55

2.6 NMR ....................................................... 56

2.7 Alp13 circular dichroism .................................. 56

2.8 Alp13 analytical size exclusion .......................... 56

3. Crystal Structure Of a Chromo Shadow Domain .......... 57
  3.1 Introduction ............................................. 57
3.2 Expression of recombinant Swi6 CSD ........................................ 58
3.3 Purification of Recombinant Swi6 CSD ........................................ 59
3.4 Crystallisation of Swi6 CSD ...................................................... 61
3.5 X-ray diffraction by Swi6 CSD crystals ....................................... 62
3.6 The Matthews Coefficient .......................................................... 65
3.7 The Self Rotation Function .......................................................... 66
3.8 Phasing by Multiple Isomorphous Replacement ............................... 68
3.9 The CSD structure ...................................................................... 72
3.10 Comparison of Swi6 CSD with mouse HP1β CD ............................... 77
3.11 CSD molecules were organised in continuous helices in the crystal. . 77
3.12 Polyethylene glycol can be seen in the electron density .................. 80
3.13 Conclusions .............................................................................. 82

4. Dimerisation of Chromo Shadow Domains. ................................. 84
4.1 Introduction ................................................................................. 84
4.2 SELDI mass spectrometry ............................................................ 85
  4.2.1 Principle of SELDI mass spectrometry .................................... 85
  4.2.2 Details of SELDI experiment ................................................. 86
4.3 GST pulldown assay ................................................................... 88
  4.3.1 Principles of GST pulldown assay ......................................... 88
  4.3.2 Details of GST pulldown assay ............................................. 89
4.4 Gel exclusion chromatography .................................................... 92
4.5 Phenotypic characterisation of L315D mutation in S.pombe .......... 92
4.6 Dimerisation is likely to be a general property of CSDs................. 94
4.7 Conclusions ............................................................................... 94
5. CSD dimer interactions

5.1 Introduction ........................................ 98
5.2 Dimerisation of Swi6 CSD creates a potential binding site. ........ 100
5.3 Mutation of cleft residue Phe 324 in \textit{S.pombe}. .............. 102
5.4 Putative role of phosphorylation at the binding cleft. .......... 104
5.5 GST pulldowns with recombinant CSD interacting proteins. .... 106
  5.5.1 Chromatin Assembly Factor, p150 subunit. .................. 106
  5.5.2 \textit{S.pombe} DNA polymerase \(\alpha\) .................. 107
  5.5.3 \textit{S.pombe} Rad21 protein .......................... 109
  5.5.4 \textit{S.pombe} Swi2 protein ......................... 110
  5.5.5 \textit{S.pombe} Clr4 protein .......................... 111
5.6 Conclusions .......................................... 112

6. Analysis of a chromodomain-like domain. ........................ 116

6.1 Alignment of CLDs with CDs and CSDs ........................ 117
6.2 Analysis of interfamily relationship .......................... 119
6.3 Analysis of a CLD. ....................................... 119
6.4 Analytical size exclusion of CLD and CSD ........................ 121
6.5 One dimensional proton NMR of CLD .......................... 123
6.6 Circular dichroism of the CLD ................................ 124
6.7 Conclusions ............................................. 127

7. Conclusions ............................................. 129

7.1 Summary of Results ...................................... 129
7.2 Discussion ............................................. 130
Chapter 1

Introduction

1.1 Chromatin Structure

In eukaryotic cells, genomic DNA is wrapped around histone proteins into complexes called nucleosomes. The nucleosome is the basic unit of chromatin. A crystal structure of a nucleosome core particle has been solved (Luger et al 1997). Eight histone proteins, two pairs each of histones H2A, H2B, H3 and H4 form the core of the nucleosome structure. The overall topology of the complex is a flattened disk (figure 1.1). Two loops of DNA are wrapped around the narrow edge of the disk.

The histones H2A, H2B, H3 and H4 have a similar fold consisting of a major helix that runs down the long axis of the protein. At each end of this helix are small globular regions comprised of short helices and unstructured loops. The fold is approximately dumbbell shaped. The core folds of these proteins pack together into the nucleosome core particle. Two molecules each of histones H3 and H4 assemble into a tetramer that is capable of binding DNA in a histone-like manner (Camerini et al 1976 and Hayes et al 1991). Following this, two histone H2A/H2B heterodimers are assembled into the nucleosome core particle (Wolffe 1995). Histones H2A and H2B are less tightly associated with the DNA than H3 and H4 and are dissociated first by increasing salt concentration (Olenbusch
Figure 1.1: Crystal structure of a nucleosome from *Gallus gallus*. Two views are shown, the right hand view is rotated 90° around a vertical axis with respect to the left hand view. The two histone H2A subunits are coloured blue, two histone H2B subunits are green, two histone H3 subunits are red, two histone H4 subunits are yellow and the DNA is white. Alpha helices in the proteins are represented as cylinders. The figure is a Molscript rendering of structure 1EQZ from the protein databank.
et al 1967). It is thought that H2A and H2B may be dissociate in vivo during transcription as these histones are found to be depleted in transcriptionally active chromatin (reviewed by Hansen and Ausio 1992). The histones interact with DNA via arginine and lysine residues scattered throughout the core fold (Luger et al 1997).

At the N terminus of these histones is a long (~40 amino acids) unstructured ‘tail’ region. The tails protrude into solution from the surface of the nucleosome. These tails are lysine and serine rich; covalent modification of these residues by specific kinases (Hsu et al 2000), acetyl transferases (Bone et al 1994, Kruhlak et al 2001) and methyl transferases (Rea et al 2000) is a key mechanism by which a cell controls chromatin structure and gene transcription (reviewed by Grunstein 1997, Strahl and Allis 2000).

Nucleosomes associate with each other to form ‘higher order’ chromatin structures. Another histone protein (H1) is known as the linker histone and mediates these interactions. The linker histone is not part of the core nucleosome complex and was not crystallised with the other histones in figure 1.1. Several higher order conformations of chromatin structures can be visualised by electron microscopy. In some images the nucleosomes are spaced out along the DNA like ‘beads on a string’. The spacing of nucleosomes on DNA is the primary structure of chromatin. In preparations that include the linker histone H1 a 30nm wide fiber can be seen (reviewed by Felsenfeld and McGhee 1986).

Fibers are formed by associations between nucleosomes into secondary chromatin structures. There are a number of models describing how nucleosomes might interact together to form the 30nm fiber. Results obtained by fragmentation of the fiber by radiation (Rydberg et al 1998) and by stretching fibres with laser molecular traps (Katritch et al 2000) support a zig-zag model of nucleosome association. In this model each nucleosome associates with another spaced two along the chromatin fibre. Two columns of histones are formed with the DNA chain zig-zagging back and forth between them. The linker histone spans horizontally across the fibre on the zig zags of DNA. Electron micrographs of the ‘beads on a string’ and the 30nm fiber with a cartoon of the models is shown in figure 1.2.
Figure 1.2: a) Cartoon of chromatin primary and secondary structure. Nucleosomes are represented as red spheres with two turns of DNA wrapped around each shown in yellow. Open 'beads on a string' chromatin is shown on the left and the 'zig-zag' model for 30nm fiber assembly is illustrated on the right. b) Electron micrograph of open chromatin. c) Electron micrograph of the 30nm fiber. Both electron micrographs were taken from Rattner et al. 1979.
It is thought that the 30nm fiber folds into tertiary and even quaternary structures to form an array of higher order chromatin conformations. The classic ‘X’ shaped chromosomes visible by light microscopy during mitosis is an example of a structure formed by higher-order chromatin compaction. The way in which the 30nm fiber folds into structures of this magnitude is not understood. There is likely to be more than one way in which these structures assemble as chromatin is a heterogeneous entity.

Bulk chromatin is subdivided into euchromatin and heterochromatin. Euchromatin condenses into discrete chromosomes during mitosis and becomes diffuse during the rest of the cell cycle. Heterochromatin remains condensed throughout the cell cycle. During interphase heterochromatin is visible as densely staining patches at the nuclear periphery in yeast. Centromeres are heterochromatic and are visible in higher eukaryotic cells during mitosis as the primary constriction in the chromosomes.

1.2 Chromatin function

A simplistic textbook explanation of chromatin function is that it serves to compact DNA for packaging into the nucleus. In reality, the function of chromatin is much more interesting.

The structure of chromatin directly affects gene transcription. Almost every cell in the human body has an identical copy of the full genome. Despite this, there is a wide variation in cell morphology. Differentiation is achieved by altering the patterns of gene expression within a given cell type. A major way in which cells mediate alteration of expression patterns is through changes in chromatin structure. Condensing chromatin into a heterochromatin-like state prevents the transcriptional apparatus from reaching gene promoters and represses transcription. Conversely, opening up chromatin structures presents genes for efficient transcription.
In many species including yeast, Drosophila and mammals, specific acetylation of lysine residues on the tails of histones is an important mechanism for regulating chromatin structure (Turner et al 1992, Braunstein et al 1996). In general, underacetylation of histones is associated with heterochromatin and hyperacetylation with gene rich, transcriptionally active loci (Jeppesen et al 1992, Ekwall et al 1997). Though all four histone types of the nucleosome core particle are acetylated, histones H3 and H4 acetylation seems to be most important in terms of chromatin structure and transcriptional regulation (reviewed by Grunstein 1997).

Histones are specifically acetylated by histone acetyl transferases (HAT). There are two types of HAT, B-type HATs acetylate newly synthesised histones H3 and H4 in the cytoplasm. These are escorted to the nucleus and loaded into nascent chromatin by the chromatin assembly complex. Whilst being loaded into chromatin the histones are deacetylated in a regulated manner (reviewed by Roth and Allis 1996). A-type HATs acetylate histones in the nucleus. The integral role of Gcn5p, a yeast transcription factor, in an A-type HAT complex (Brownell et al 1996) has led to the idea that these HATs are targetted to specific chromatin regions by transcription factors. Indeed, several other transcription factors have been found to be important for recruiting A-type HATs (Yang et al 1996, Ogryzko et al 1996, Dunphy et al 2000). The HAT enzymes are antagonised by histone deacetylases (HDAC). Protein complexes associated with transcriptionally silent loci often have HDAC activity (van der Vlag and Otte 1999, Nielsen et al 1999).

How acetylation brings about changes in gene transcription and chromatin structure is not fully understood. Two models have been proposed. The first model suggests that neutralisation of the positive charge on histone tails by acetylation, 'loosens' the association of negatively charged DNA with the nucleosome. This makes the DNA more accessible to transcription factors. The second model suggests that the acetylated tails are recognised by specific binding proteins that in turn bring about changes in transcription. It is entirely possible that both of these models are correct (Roth and Allis 1996).

In addition to mediating transcription patterns, different chromatin structures also play important roles in chromosome mechanics. For example, heterochro-
matic structures at the centromere are likely to be essential for assembling the
kinetochore complex. The kinetochore binds microtubules and mediates chromo-
some segregation during cell division. Specialised chromatin structures are also
important at the telomeres. Here they protect the ends of the DNA molecule from
being degraded or from being recognised as a double stranded break by the cells
DNA repair machinery.

1.3 Chromodomains modulate chromatin structures

The dynamic and complex changes in chromatin structure are mediated by a
number of associated proteins. Chromodomains are a common feature of many of
these proteins.

The chromodomain (chromatin organiser modifier) was first described as a
region of homology between Drosophila proteins HP1 (Heterochromatin-associated
Protein 1) and Polycomb (Paro and Hogness 1991). HP1 and Polycomb both
repress gene expression in a dosage dependent manner. HP1 was known to be
a structural component of chromatin. This led to the idea that Polycomb acted
through chromatin structure modification and that chromodomains may play a
role in this.

A chromodomain (CD) is a ~70 amino acid domain, thought to mediate
protein-protein interactions. No DNA binding ability has been shown. To date,
the family of CD containing proteins has swollen to over 50. Initially defined
in Drosophila, CD proteins have now been found in yeast, nematodes, plants,
vertebrates and humans. The study of chromodomain proteins has led to the
identification of a second related domain. This homologous, but discrete, domain
has so far only been found in tandem with a classical CD in a subset of chromod-
omain proteins. It has been called the chromo shadow domain (CSD) (Aasland
and Stewart 1995). A multiple sequence alignment of chromodomains and chromo
shadow domains, made by Aasland and Stewart (1995) is shown in figure 1.3. A
schematic figure of some chromodomain proteins showing their various domains and their relationship to each other is also shown.

An NMR structure of the chromodomain of murine HP1β has been solved. The structure reveals a monomeric protein with a simple structure. A single α helix packs against a three stranded β-sheet (Ball et al 1997). This structure is shown in figure 1.4.

Sequence homology between chromodomains and chromo shadow domains is low (~20% identity between the chromodomain of murine HP1β and the chromo shadow domain of fission yeast Swi6). This makes alignment difficult but it is predicted that chromo shadow domains have a similar structure to chromodomains.

More recently, a set of sequences with low homology to chromodomains have been identified in proteins of the dosage compensation complex (Koonin et al 1995). This complex is found associated with the single X chromosome of male Drosophila. The complex doubles the gene transcription of certain regions of this chromosome to levels comparable to transcription from the two X chromosomes of females. This doubling is thought to be achieved by chromatin structure modulation. Though the homology of these regions to chromodomains is below statistical significance, the function of these domains in chromatin modulation lends weight to the argument that they are true chromodomains. Domains with high homology to these divergent chromodomains have been found in a number of other proteins. In this thesis these domains will be counted as a distinct domain family called the chromodomain-like domains (CLD).

Together, chromodomains, chromo shadow domains and the chromodomain-like domains make up the chromodomain superfamily. The chromodomain superfamily has a number of diverse functions that are fundamental to chromatin regulation. The functions of this superfamily will be investigated by reviewing several areas of cell biology. Firstly the Polycomb protein will be discussed. This chromodomain protein acts in concert with numerous other proteins to maintain transcription patterns through chromatin structure modulation. Secondly, the role of HP1 at the centromere will be discussed. Heterochromatin structure at the cen-
Figure 1.3: a) Multiple sequence alignment of chromodomains (top section) and chromo shadow domains (bottom section). Proteins and organism are noted in the left hand column followed by the amino acid numbering of the start and end of the domain. Amino acids are colour boxed according to chemical type. Structure prediction is shown along the bottom H = α-helix, E = β-sheet. The alignment is copied from Aasland and Stewart (1995). b) A modular schematic of selected chromodomain proteins. A ruler is shown marked with amino acid numbers. Chromodomains are shown in green, chromo shadow domains in red, SET domains in purple, a domain shared with SNF2 and other transcriptional control and DNA repair proteins is blue and a helicase domain in yellow.
tromere is important for stable chromosome segregation. Finally, the *Drosophila* dosage compensation complex will be discussed. This complex, containing chromodomain proteins, hyperactivates transcription from the male X chromosome through chromatin structure modulation.

Several interesting properties of chromodomains arise from these literature reviews. Firstly chromodomains function to target proteins to specific chromatin sites. Chromo shadow domains facilitate the recruitment of other proteins to these sites and mediate the assembly of large multi-protein complexes. In addition chromodomain proteins can spread along chromatin from initiation sites and are regulated by post-translational modification.

### 1.3.1 Polycomb and Trithorax group mediated transcription regulation.

During *Drosophila* development, cells differentiate to form the various tissues that will make up the embryo, the larva and eventually the adult fly. Differentiation is largely brought about by altering the expression patterns of the homeotic genes. The homeotic genes are found in two clusters in the fly genome, the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C). In early development concentration gradients of transcription factors, both self and maternally derived,
are set up across the embryo. The gap and pair-rule gene products play a key role in this process. Gap genes repress transcription of certain homeotic genes whilst pair-rule gene products activate them (reviewed by McGinnis and Krumlauf 1992). This complex interplay of factors is transient: they set up transcription patterns early in development but are then themselves degraded. Maintenance of these transcription patterns, after the degradation of the early transcription factors, is the role of the polycomb and trithorax group proteins. Polycomb group proteins repress, and Trithorax group proteins activate gene transcription.

Proof of the role of Polycomb in repressing homeotic gene expression was demonstrated by fusing a β-Galactosidase reporter gene with regulatory elements from the Abdominal-B gene. Abdominal-B is a homeotic gene from the bithorax cluster. In Drosophila embryos, expression of this β-Galactosidase construct is seen in tightly regulated segments similar to the expression pattern of the Abdominal-B gene. Embryos that lack Polycomb set up the distinctive β-Galactosidase expression patterns in early development but later the pattern is lost and the reporter gene becomes diffusely expressed throughout the embryo (Busturia and Bienz 1993). Hox genes, also regulated by the gap and pair-rule transcription factors, were also found to be dependent on polycomb group (PcG) proteins for maintenance of transcription patterns (Beuchle et al 2001). The activating role of Trithorax proteins was demonstrated using similar techniques (Breen et al 1995).

Approximately 13 protein members of the PcG have been identified but it has been estimated that there may be as many as thirty more undefined members of this group (Mahmoudi and Verrijzer 2001). These proteins are thought to act together as a large complex (Shao et al 1999). This has been demonstrated by immunofluorescent staining showing that PcG proteins co-localise on chromosomes (Buchenau et al 1998, Franke et al 1995). In addition, proteins of the PcG co-fractionate on sucrose gradients (Gunster et al 1997), are immunoprecipitated with Pc antibodies and show interactions through yeast two hybrid and in vitro screens (Kyba and Brock 1998). More recently, PcG complexes have been purified from Drosophila embryos. These experiments have revealed that the components of these complexes vary from complex to complex (Ng et al 2000, Saurin et al...
Fewer members of the trithorax group (T\(x\)G) proteins are known but these are also likely to function as a complex (Shao et al 1999, Papoulas et al 1998).

The polycomb protein was the first member of the PcG to be identified and gave its name to the group. Polycomb (Pc) was found to contain a chromodomain, also found in HP1 (Paro and Hogness 1991). As HP1 was known to promote heterochromatin (Eissenberg et al 1992), this observation led to the idea that Polycomb group proteins may silence genes through modulation of chromatin structure. Some Polycomb group proteins are modifiers of position effect variegation (Grigliatti 1991) and the PcG protein EED has been shown to have histone deacetylase activity (van der Vlag et al 1999). These results also link the Polycomb group complex with regulation of chromatin structure.

In 1930, HJ. Muller described mosaic expression of the white\(^+\) gene in the eyes of Drosophila melanogaster resulting in patches of red and white colour (reviewed by Wakimoto 1998). The phenomenon is known as position effect variegation (PEV) and is an important concept in the study of chromatin associated proteins. Genes that are adjacent to heterochromatic loci often have variable expression levels from cell to cell within one organism. This mosaic variegation of expression is thought to be due to changes in the extent of heterochromatin. A good example of PEV is in the mosaic expression of the white\(^+\) gene in the eyes of a certain Drosophila lineage. The product of the white\(^+\) gene confers a red colouring to the eye. This phenotype was found to be due to a chromosomal rearrangement event in which the white\(^+\) gene had been translocated close to the pericentric heterochromatin. In some cells the heterochromatin covers and silences the gene (white eyes): in others the gene is clear and is transcribed (red eyes). These differences in gene expression are heritable and give rise to clonal patches of red and white in the eyes of these flies (reviewed by Reuter and Spierer 1992).

This chromosomal rearrangement has been useful in studying chromatin modulation. Mutations in proteins that promote heterochromatin formation cause the pericentric heterochromatin to recede and yield a completely red eye colour in these flies. These proteins are called Suppressors of variegation (Su(var)) and include Su(var)3-9, a histone methyl transferase and a homologue of fission yeast Clr4 and
Su(var)205, the HP1 protein. Similarly mutations in proteins that promote open chromatin conformations cause the heterochromatin to advance and yield a white eye colour. These genes are called the Enhancers of variegation (E(var)). This screen and the Su(var) and E(var) proteins are reviewed by Reuter and Spierer (1992) and by Walrath (1998).

PcG proteins act as large complexes in maintaining repressed transcription states of homeotic genes after early development. There is a strong indication that this repression may be mediated through changes to chromatin structure. Trithorax group proteins also act as complexes and maintain active gene transcription. What is the role of the chromodomain in Polycomb function? In discussing this, the site of action of the Polycomb group proteins and their movements through the cell cycle must be examined more closely.

The location of Polycomb proteins on chromosomes can be mapped directly by using polytene chromosomes. These chromosomes are found in cells of the salivary glands of Drosophila. The cells undergo multiple rounds of DNA replication without division. Homologous chromosomes in the resulting polyploid cells align side by side to form single, giant, polytene chromosomes. Because of their size, structures like centromeres and other sites of heterochromatin can be seen by light microscopy.

Indirect fluorescent labelling shows that Pc is localised to multiple euchromatic loci in these chromosomes (Zinc and Paro 1989). Interestingly, Trithorax, which also localises to euchromatic sites, colocalises with Pc at many of these sites. This result suggests a model in which PcG and TxG proteins antagonise each other at the same locus to achieve a correct balance of transcription (Strutt et al 1997; Chinwalla et al 1995).

Messmer et al (1992) noted that an N-terminal fragment of Pc that includes the chromodomain is sufficient for targeting a β-galactosidase fusion construct to Pc localisation sites. Further, mutations in the chromodomain disrupt Pc localisation on polytene chromosomes and alleviate gene silencing (Eissenberg et al 1990, Lehming et al 1998).
An interesting experiment was performed in which the chromodomain of HP1 was replaced with the chromodomain of Pc. HP1 also has an N-terminal chromodomain but is localised to heterochromatic loci in contrast to Pc's euchromatic localisation. This chimeric protein localises to euchromatic Pc sites (Platero et al 1995). This experiment shows that the chromodomain of Pc is sufficient for targeting a protein to Pc localisation sites. An unexpected result of this experiment was that the chimera was also targeted to heterochromatic HP1 sites. However, the chromo shadow domain at the C-terminus of HP1 can localise to heterochromatin independently of the chromodomain (Powers et al 1993). This observation may explain this result.

The Pc localisation sites are characteristic DNA sequence elements. Fusion of certain non-coding DNA sequences from the ANT-C and BX-C gene clusters with the reporter gene LacZ, results in LacZ expression in Drosophila embryos that is under control of the PcG and TxG proteins. These experiments led to the idea that there were specific DNA sequences to which these protein complexes associated. These are known as polycomb resonsive elements (PRE) (Simon et al 1993). It is important to note that both PcG and TxG proteins associate with PREs. PREs stretch over hundreds or even thousands of base pairs and lack easily identifiable consensus sequences. A consensus sequence has been described common to several PREs but the relevance of this has not been tested (Mihaly et al 1998). Localisation of PcG and TxG proteins to PREs can be studied at higher resolution by a technique called chromatin immunoprecipitation (CHIP).

In the CHIP protocol, cells are fixed with paraformaldehyde to cross link proteins to DNA. Chromatin is then prepared and sonicated. Sonication breaks the chromatin into small fragments. DNA fragments associated with a protein of interest are immunoprecipitated with specific antibodies. After de-cross-linking, the immuno-precipitate can be probed for the presence of specific DNA sequences by Southern analysis using specific radioactive probes or by the polymerase chain reaction.

CHIP analysis of Drosophila embryos in one study shows that maternally derived Pc associates with PREs early in development. Initially, Pc is narrowly
localised at the PREs and later in development spreads out from these sites. Trx colocalises with Pc at many sites in early development. Later the two proteins resolve to more mutually exclusive sites (Orlando et al 1998). This result suggests a model in which transcription states are established early in development by antagonistic interplay of PcG and TxG proteins. Later in development the protein groups become set in their respective positions and there is less interaction.

PcG proteins can spread outwards along chromatin from a nucleation site. As the dosage of PcG proteins is increased, their localisation on polytene chromosomes becomes more ubiquitous until the entire chromosome appears to be covered (Sharp et al 1997). This dosage dependence is also seen by CHIP. Polycomb concentration is found to be highest at the PRE and levels decay gradually with distance on either side. The gradual decay implies that Pc localisation is not tight and specific, but a concentration dependent equilibrium. (Orlando et al 1998, Strutt et al 1997).

PcG mediated silencing patterns are stably maintained through mitosis and even meiosis. Fab7 is a DNA element within the BX-C homeotic-gene cluster that has a regulatory function. It contains several PREs and is known to localise Pc. A transposable P element was made with the Fab7 element and the Drosophila white+ gene flanking a gal4 upstream activation sequence in control of a LacZ gene. Transcription from the lacZ gene can be induced by heat shock in this construct. The product of the white+ gene confers red eye colouring.

When this construct was transfected into flies the white+ gene was silenced in a Pc dependent manner. Strong activation of the LacZ gene by heat shock was able to overcome the repressive state, presumably by driving PcG proteins from the site. Interestingly, if the heat shock is performed late in development when transcription patterns have largely been defined the repressive state is quickly reestablished after the heat shock event. If embryos are heat shocked early in development the repressive state is not reestablished. Furthermore the derepressed state is inherited through mitosis and to some extent through meiosis (Cavalli and Paro 1999).
These results demonstrate how stable levels of transcription can be maintained by PcG and TxG proteins. In addition it points to the existence of a mechanism by which distinct phenotypes can be inherited independently of genotype. This phenomenon is known as epigenetic inheritance. It is not clear how PcG and TxG proteins reassociate with regions of the genome after DNA replication. It is likely that the chromatin is marked in some way, for example by methylation or acetylation (Brannan et al 1999).

The above results suggests that PcG silencing is very stable, however immunofluorescence staining suggests that Pc localisation changes throughout the cell cycle. Staining is distributed throughout the nucleus in granules and is at its most intense during interphase. The staining decreases dramatically in late S phase and remains low throughout mitosis (Alkema et al 1997, Gunster et al 1997). Changes in PcG localisation may be mediated by post-translational modification. It has also been demonstrated that Bmi-1, a human PcG protein, is hypophosphorylated in G1 and S phase and hyperphosphorylated through M phase (Voncken et al 1999). TxG proteins have also been shown to have cell cycle specific localisation patterns (Muchardt et al 1996).

In summary, the chromodomain of Pc is sufficient for targetting to PRE sites in euchromatin. Though capable of binding alone, the chromodomain functions as part of a multi-protein complex. The PRE sites are not well defined sequence motifs and the chromodomain does not seem to interact directly with the DNA. Pc binds to PREs in a dosage dependant manner and spreads out from them. There is dynamic competition for these sites between PcG and TxG proteins. At some stage in development a molecular ‘decision’ is made and either PcG or TxG proteins become predominant at a given site. This ‘decision’ is inherited epigenetically through cell division. PcG and TxG proteins disappear from their chromatin locations during mitosis and reappear during the rest of the cell cycle. This suggests that the chromatin may be marked in some way to allow faithful reassociation of these proteins.

Many of the features of the chromodomain containing PcG complex are also observed in the complex of proteins assembled on the pericentric heterochromatin.
The chromodomain protein HP1 plays a key role in this complex. More is understood about the molecular mechanisms of HP1 chromodomain function. Moreover, HP1 also contains a chromo shadow domain, consideration of which brings wider insight into the functions of the chromodomain superfamily.

1.4 Chromodomain proteins and heterochromatin.

PcG proteins maintain gene repression at euchromatic loci. In contrast, HP1 acts at heterochromatic loci. Although gene transcription is repressed in heterochromatin this is unlikely to be the main function of the chromatin structure in these regions as the underlying DNA is gene poor. It is thought that chromatin structures in these regions prevent recombination (reviewed Weiler and Wakimoto 1995). Chromatin structure at the centromere also seems to be important for faithful chromosome segregation (Allshire et al 1995) and at the mating type loci for mating type switching (Lorentz et al 1992). This review will focus mainly on the role of HP1 proteins and more specifically on the molecular activities of chromodomains within these proteins.

1.4.1 Identification of heterochromatin associated proteins

Many heterochromatin associated proteins, such as HP1, were originally identified in Drosophila as a suppressors of position effect variegation (PEV). PEV is discussed in the previous section. The chromodomain protein Su(var)3-9 was also discovered in this way. Other genetic screens have been used to find heterochromatin associated proteins in fission yeast. One of these uses mating type switching and a second the alleviation of silencing of reporter genes in heterochromatic loci.

Fission yeast have two mating types, plus (P) and minus (M). When nutrients are limiting a cell can fuse with another of the opposite mating type to form a single diploid cell. Meiosis occurs within the diploid followed by sporulation. An ascus is generated with four haploid spores. The spores are starch rich and yeast that have sporulated can be easily visualised by staining with iodine.
Fission yeast mating type is switched via a translocation event at the mating type locus. The mating type locus is heterochromatic. There are three regions within the locus, mat1, mat2 and mat3. mat2 contains the P allele and the mat3 locus the M allele. mat2 and mat3 are heterochromatically silenced. Mating type switching is brought about by a translocation event that places a copy of either the P or M allele into the transcriptionally active mat1 locus.

Mutants defective in mating type switching do not produce spores and so are not darkly stained with iodine. One mating type switching mutant swi6 is the fission yeast homologue of HP1. This indicates that heterochromatin structure at the mating type loci is critical to the switching process. This screen also identified a fission yeast homologue of su(var)3-9 called clr4+.

A further strategy that has been successfully used to identify heterochromatin associated proteins is the insertion of a marker gene into heterochromatic loci. Fission yeast with mutations in the ura4+ gene cannot produce their own uracil and therefore cannot grow on minimal media without uracil supplementation. A ura4+ gene placed within the centromere is silenced by the heterochromatin and does not help the cells grow on minimal media. Mutations in heterochromatin associated proteins often alleviate silencing at the centromere and allow these cells to grow on minimal media (Allshire et al 1995, Ekwall et al 1996). This assay confirms the role of Swi6 and Clr4 in centromeric silencing.

A further two fission yeast chromodomain proteins were identified from the S.pombe genome project. These are Chp2, another fission yeast HP1 homologue, and Chp1 with an N-terminal chromodomain and a unique C-terminus (Thon and Verhein-Hansen 2000).

1.4.2 Fission yeast centromeres.

Fission yeast have 'complex' centromeres that are more similar to the centromeres of higher eukaryotes than to those of budding yeast. The centromeres extend between 40 and 120 Kb on each of the three fission yeast chromosomes. The three centromeres are quite distinct in sequence though there are many conserved
regions. Generally, there is a central core region surrounded on either side by inverted repeating sequences (reviewed Allshire 1996)

Numerous fission yeast proteins that play a role at the centromere have been identified. Among these there are four chromodomain proteins, these are Swi6, Clr4, Chp1 and Chp2. Swi6 and Chp2 are homologues of Drosophila HP1 protein having a chromodomain and a chromo shadow domain. Clr4 is a homologue of Drosophila Su(var)3-9 with a chromodomain and a SET domain. Other than its chromodomain no other domain has been recognised in Chp1. Non-chromodomain proteins Mis6 and Riki also play important roles at the fission yeast centromeres.

Deletion of the Swi6+ gene results in chromosome segregation defects and loss of centromeric silencing. Indirect immunofluorescent labeling of Swi6 reveals punctate nuclear staining at centromeres, silent mating type loci and telomeres (Ekwall et al 1995). The chromodomain of Swi6 is sufficient to target a GFP fusion protein to these sites (Wang et al 2000). Localisation of Swi6 within the centromere was assayed at higher resolution by CHIP (CHIP was explained in the previous section). Swi6 was found to be associated with the outer repeats of centromere 1 but not with the inner core (Partridge et al 2000). Interestingly, there is a tRNA gene at the boundary between the outer repeats and the inner core. Reporter genes inserted into this region are less repressed than in other regions of the centromere.

Using the same methods, it was found that chromodomain protein Chp1 was also associated with the outer repeats. In contrast, the Mis6 protein is localised to the inner core of the centromere and excluded from the outer repeats (Partridge et al 2000). Mis6 is found exclusively at centromeres, while most of the other proteins also affect silencing at telomeres and the mating type loci. (Saitoh et al 1997, Thon and Verhein-Hansen 2000).

The function of Clr4 and Riki at the centromeres has been harder to assess as it has not been possible to raise antibodies against these proteins that allow localisation to be assessed. However, knockout of clr4+ or riki1+ genes results in Swi6 delocalisation indicating that these may have roles in recruiting Swi6 to heterochromatin (Ekwall et al 1996).
However it is recruited, Swi6 can spread outwards from these sites along chromatin. A 3Kb fragment of exogenous DNA including the ura4+ gene was inserted into the outer repeats of centromere 1. It was found by CHIP that Swi6 was able to spread from the endogenous centromeric sequences to cover this insert (Partridge et al 2000). Chp1 is unable to spread over the same insert. This result reveals both a spreading property of Swi6 and also that DNA sequence is not vital for heterochromatisation.

The mechanisms by which centromeres are initiated and maintained are not fully understood. In humans, functional neocentromeres have occasionally been observed to assemble at other points on the chromosome when the centromere is disrupted (Wandall et al 1998). Conversely, centromeric DNA elements inserted into a minichromosome confer mitotic stability indicating that a functional kinetochore is assembled (Heller et al 1996, Masumoto et al 1998). It may be that the extended repetitive DNA sequence at the centromere is involved in initiating heterochromatin and a kinetochore structure. But once this is formed it can be propagated epigenetically. In favour of this hypothesis is the observation that centromeric chromatin is marked by different histone acetylation states compared to the rest of the genome (Ekwall et al 1997).

Histone acetylation is an important factor in the centromere of fission yeast. When cells are treated with trichostatin A, a specific inhibitor of histone deacetylases, histones become hyperacetylated. Trichostatin A induced hyperacetylation of histones results in increased expression of a marker gene inserted at the centromere and chromosome segregation defects (Ekwall et al 1997). Remarkably the changes in acetylation pattern at the centromere, induced by trichostatin A, are maintained through multiple rounds of mitosis and even through meiosis (Nakayama et al 2000). Acetylation is also linked to pericentric heterochromatin in mammalian cells (Taddei et al 2001).
1.4.3 Dynamics of HP1 in relation to heterochromatin

It is clear that the HP1 homologue Swi6 plays an important role at the fission yeast centromere. However, HP1 proteins are not restricted to centromeres but act widely on heterochromatin. The dynamics of HP1 have been studied extensively in both *Drosophila* and mammals.

*Drosophila* and humans have at least three homologues of HP1 each (Singh *et al* 1991, Smothers and Henikoff 2001). The different homologues have different localisation patterns within the nucleus: these patterns also change throughout the cell cycle. Complexity is added to this situation as HP1 localisation differs between mouse and human cells.

Humans and mice also have three HP1 homologues named HP1α, HP1β and HP1γ. The proteins are almost identical between these two organisms but cannot be equated by sequence homology to the three *Drosophila* HP1 homologues. During interphase, mouse HP1α and HP1β are found associated with heterochromatin whilst HP1γ is scattered in small foci throughout the nucleus. In mitosis HP1α is concentrated at the centromeres. HP1γ associates diffusely with the chromosomes and mitotic spindle. HP1β diffuses throughout the cytoplasm only reconverging at the nucleus by telophase. In mitotic chromosomal spreads HP1α and HP1β can be seen localised at centromeres. HP1γ is localised to multiple sites on the chromosomal arms (Minc *et al* 1999, Nicol and Jeppsen 1994).

HP1 proteins have complicated localisation patterns that change throughout the cell cycle. These changes may be mediated by phosphorylation. Human HP1γ has been shown to be phosphorylated by the kinase Pim-1 (Koike *et al* 2000). Pim-1 levels are up-regulated during G1/S phase of the cell cycle. The enzyme is a ubiquitously expressed kinase (Liang *et al* 1996). It has been suggested that HP1γ is phosphorylated at a cluster of serine residues in the linker region between the chromodomain and the chromo shadow domain and that this down-regulates its silencing function (Koike *et al* 2000). However, the effect of HP1 phosphorylation is not well understood.
There are two slightly contradictory studies of HP1 phosphorylation in *Drosophila* (Huang *et al* 1998). In the first, underphosphorylation of HP1 is associated with its incorporation into large protein complexes. *Drosophila* embryos do not begin transcribing their own genes until after cell division has begun. The early embryos rely on a pool of maternally derived proteins for their development. These are found in large complexes. HP1 was found in three complexes of differing size. These were purified from each other and HP1 was found to be multiply phosphorylated in all. However, the level of phosphorylation of HP1 decreased with increasing complex size. The HP1 found in the largest complex was found to phosphorylated to a lesser degree than in smaller complexes and to be tightly associated with subunits of the origin recognition complex (ORC) (Huang *et al* 1998).

In a second study, HP1 from *Drosophila* nuclear extracts was also found to be multiply phosphorylated. Recombinant HP1 was similarly phosphorylated when incubated with the extracts. Inhibitors of casein kinase II prevented this. Phosphorylation sites on the recombinant HP1 were mapped to a serine residue slightly N-terminal of the chromodomain and two serine residues at the C-terminal end of the chromo shadow domain. The C-terminal sites were within a casein kinase consensus sequence. Mutation of these residues to glutamic acid had no effect on the normal localisation of HP1 to heterochromatin. Mutation of these residues to alanine was lethal. These results imply that phosphorylation of HP1 is required for its assembly into heterochromatic complexes (Zhao and Eissenberg 1999).

1.5 **HP1 chromodomain and chromo shadow domain function**

HP1 contains a chromodomain and a chromo shadow domain joined by a flexible linker region. A variety of studies have attempted to dissect the functions of these domains. Recently there have been significant advances in understanding their functions.
The chromodomains of murine HP1α, HP1β and HP1γ have been shown to bind histone H3 N-terminal tail peptides only when these peptides are specifically methylated on lysine 9 (Lachner et al 2001, Bannister et al 2001). In contrast, Pc and Su(var)3-9 chromodomains bind to unmodified histone H3 tails as well as tails methylated on Lys 9 or phosphorylated on Ser 10 or to histone H4 tails (Lachner et al 2001).

This result led on from the observation that the SET domain of Drosophila Su(var)3-9 methylates lysine 9 of histone H3 (Rea et al 2000). This activity is conserved in the homologous S.pombe C1r4 protein (Rea et al 2000). C1r4 is required for the correct localisation of the HP1 homologue Swi6 (Ekwall et al 1996).

Serine 10 of histone H3 is phosphorylated by Ipl1/Aurora kinase (Hsu et al 2000) and is linked to chromatin condensation at mitosis (Wei et al 1999). Methylation of lysine 9 inhibits phosphorylation of serine 10 on histone H3 tails. Conversely, phosphorylation of serine 10 inhibits lysine 9 methylation in vitro (Hsu et al 2000). Interestingly, the localisation on chromatin of Su(var)3-9, a histone H3 methyl transferase, is known to be cell cycle regulated (Aagaard et al 2000). Histone H3 tails in silent regions are underacetylated (Jeppesen et al 1992, Ekwall et al 1997). It is known that acetylation of lysine 9 prevents subsequent methylation in vitro (Rea et al 2000). Recently it has been discovered that a histone deacetylase (Clr3) is required for methylation of H3 by C1r4 and association of Swi6 (Nakayama et al 2001b). It is likely that covalent modifications of histone tails are at least partly responsible for governing changes in chromatin condensation during the cell cycle, epigenetically inherited transcription patterns and the extent of heterochromatic domain boundaries.

These findings led to a hypothesis that there was a 'histone code', an interlinked network of histone modification events that are recognised by specific binding proteins. These modifications mark the chromatin and allow precise modulation of local higher order structures. The marks are maintained through cell replication and mediate reassociation of proteins after division. This allows inheritance of chromatin states. As chromodomain proteins are known to specifically
recognise modified histone tails and are conserved in numerous proteins, they may be a major component of this system (Strahl and Allis 2000).

This is an attractive hypothesis. However there are some results that would seem to be in conflict. Several studies suggest different modes of nucleosome binding by HP1 proteins. It has been demonstrated that Drosophila HP1 can bind reconstituted nucleosomes in vitro. These nucleosomes do not have post translational modifications. Further, binding can take place independently of the histone tails. In this study HP1 was also found to have some affinity for DNA (Zhao et al 2000).

In another study, the chromodomains of the three murine HP1 homologues were shown to bind to the core folds of histone H3 and H1 both in vivo and in vitro (Nielsen et al 2001). It has also been shown that the linker region between the chromodomain and the chromo shadow domain plays an important part in the localisation of HP1 to heterochromatin (Smothers and Henikoff 2001). Finally, there is evidence that HP1 may rely on interaction with specific DNA binding proteins such as DNA polymerase α (Nakayama et al 2001a) or the origin recognition complex (Pak et al 1997, Shareef et al 2001) for targetting to chromatin.

A structure of the chromodomain of murine HP1β has been solved by NMR. The domain fold is a three stranded, anti-parallel β-sheet with an α-helix packing along one face. Approximately 50 amino acid residues are involved in the core structure. Amino acids important for the fold are conserved in other chromodomains and so it is likely that all adopt this structure (Ball et al 1997). An NMR structure of the chromodomain of the fission yeast Clr4 protein confirms this (Horita et al 2001). It was suggested that a hydrophobic groove on the surface of the molecule was likely to function as a protein-protein interaction site. Mutations close to this groove disrupt the function of the HP1 protein (Ball et al 1997, Platero et al 1995, Messmer et al 1992). This groove is also seen on the surface of a recent NMR structure of the chromodomain of fission yeast Clr4 protein (Horita et al 2001).

Although a great deal of work has been done to characterise the interactions
of the chromo shadow domain, its contribution to heterochromatin is less well understood than the chromodomain. This reflects the complexity of the CSD interactions.

Mouse homologues of HP1 were shown to self associate and bind TIF1α and TIF1β in a yeast two hybrid assay (Le Douarin et al 1996). However, it was not until some years later that these interactions, together with binding of the lamin B receptor, were shown to be mediated through the CSD (Ye et al 1997). Subsequently, the yeast two hybrid assay has been used to demonstrate CSD interactions with the p150 subunit of the mouse chromatin assembly factor 1 (CAF1) (Murzina et al 1999), the human DNA repair protein dimer Ku70 (Song et al 2001), a human kinase involved in cell cycle regulation (Koike et al 2000) and human SP100, a protein found in nuclear bodies (Seeler et al 1998). Many of these interactions have also been demonstrated by coimmunoprecipitation and in vitro by GST pulldown. In addition, CSD interactions with DNA polymerase α (Nakayama et al 2001a) and subunits of the origin recognition complex (Pak et al 1997) have been shown by GST pulldown and coimmunoprecipitation respectively. Unpublished results also indicate interaction between Swi6 and the mitotic cohesin Rad21. These interactions are summarised in table 1.1. It should be noted that in all of these in vitro CSD interaction experiments, at least one of the binding partners was either in vitro translated or the binding done in cell lysates. Therefore, these experiments do not rule out the involvement of post translational modifications or intervening molecular bridges.

These results suggest a model in which the CSD functions as a promiscuous protein-protein interaction domain. It is worth speculating on the possible physiological relevance of some of these interactions.

During S phase the Chromatin assembly factor (CAF) complex is found at the site of DNA synthesis (Shibahara and Stillman 1999). The complex is involved in incorporating histones into the chromatin of the newly formed DNA (Kaufman et al 1995). The interaction with HP1 may only be important for recruiting HP1 to chromatin. Alternatively, HP1, bound to CAF through the CSD, binds methylated histone H3 through its chromodomain. This may allow methylated
<table>
<thead>
<tr>
<th>H1 interacting protein</th>
<th>Function</th>
<th>H1 protein/region</th>
<th>Experimental method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. pombe</em> DNA polymerase α</td>
<td>DNA replication</td>
<td><em>S. pombe</em> Swi6/ unmapped</td>
<td>in vivo: mutations in pola disrupt Swi6 localisation in vitro: GST pulldown</td>
<td>Nakayama 2001a</td>
</tr>
<tr>
<td><em>S. pombe</em> Rad21</td>
<td>cohesin</td>
<td><em>S. pombe</em> Swi6/ unmapped</td>
<td>genetic interaction</td>
<td>unpublished</td>
</tr>
<tr>
<td>Mouse TIF1α TIF1β</td>
<td>mediate nuclear receptor functions</td>
<td>mouse H1α and H1β/ CSD</td>
<td>yeast two hybrid immuno-precipitation GST pulldown</td>
<td>Douris 1996 Nielsen 1999</td>
</tr>
<tr>
<td>human lamin B receptor</td>
<td>nuclear membrane bound receptor</td>
<td>human H1α/ CSD</td>
<td>in vivo: yeast two hybrid in vitro: GST pulldown</td>
<td>Ye 1997</td>
</tr>
<tr>
<td>human Sp100</td>
<td>component of nuclear bodies</td>
<td>human H1α, H1β/ CSD</td>
<td>in vivo: yeast two hybrid</td>
<td>Seeler 1998</td>
</tr>
<tr>
<td>Drosophila ORC components</td>
<td>DNA replication origin recognition complex</td>
<td>Xenopus Drosophila H1/ both CD and CSD</td>
<td>immuno-precipitation</td>
<td>Pak 1997</td>
</tr>
<tr>
<td>methylated histone H3 tail peptides</td>
<td>core nucleosome component</td>
<td>mouse H1α, H1β/ CD</td>
<td>in vitro: pulldown</td>
<td>Lachner 2001 Bannister 2001</td>
</tr>
<tr>
<td>human Pim-1</td>
<td>kinase</td>
<td>human H1α/ CSD</td>
<td>yeast two hybrid</td>
<td>Koike 2000</td>
</tr>
<tr>
<td>mouse core fold of histone H3</td>
<td>core nucleosome component</td>
<td>mouse H1α, H1β/ CD</td>
<td>immuno-precipitation GST pulldown</td>
<td>Nielsen 2001</td>
</tr>
<tr>
<td>human Ku70</td>
<td>DNA repair telomere maintenance</td>
<td>human H1α/ CSD</td>
<td>in vivo: yeast two hybrid in vitro: GST pulldown</td>
<td>Song 2001</td>
</tr>
<tr>
<td>p150 subunit of mouse CAF1</td>
<td>nucleosome assembly into chromatin</td>
<td>mouse H1α, H1β/ CSD</td>
<td>in vivo: yeast two hybrid in vitro: GST pulldown</td>
<td>Murtina 1999</td>
</tr>
<tr>
<td><em>S. pombe</em> Swi2</td>
<td>silencing at mating type loci</td>
<td><em>S. pombe</em> Swi6</td>
<td>yeast two hybrid</td>
<td>unpublished</td>
</tr>
<tr>
<td>human INCERB</td>
<td>inner centromere binding protein</td>
<td>human H1α hinge</td>
<td>yeast two hybrid GST pulldown</td>
<td>Ainsztein 1998</td>
</tr>
</tbody>
</table>

Table 1.1: Proteins interacting with HP1 homologues.
histones, displaced during DNA synthesis, to be held in the same location and placed back into the two new DNA strands. This mechanism would allow localised histone modifications to be conserved through cell division.

DNA polymerase α is involved in the initiation of DNA replication and priming of the Okazaki fragments on the lagging strand (Burgers 1998). The polymerase interacts directly with the chromo shadow domain of fission yeast Swi6. The interaction is required for Swi6 localisation (Nakayama et al 2001a). There may be a general mechanism whereby HP1 homologues are delivered to chromatin sites through CSD interactions with proteins such as CAF subunits and DNA polymerase α.

The CSD of human HP1α has been shown to bind to the lamin B receptor. This receptor spans the inner nuclear membrane (Ye et al 1997). Heterochromatic regions are often found around the nuclear periphery. It is conceivable that interactions between the CSD and the lamin B receptor mediate this localisation. However, fission yeast do not seem to have proteins homologous to the lamin B receptor and so the relevance of this interaction in other species is not clear.

CSD interacting proteins TIF1α, TIF1β and CAF p150 were all found to contain a pentapeptide motif with the consensus motif P-x-V-x-L (where x is any amino acid) (Murzina et al 1999, Anders et al 1999). The region of TIF1α and TIF1β sufficient for binding HP1 in vitro was mapped to a 25 amino acid oligopeptide containing this pentapeptide sequence (Anders et al 1999). Mutations in the pentapeptide region of CAF p150 prevent it being co-immunoprecipitated with HP1 antibodies (Murzina et al 1999).

A random phage display experiment was used to refine the CSD binding consensus sequence (Smothers and Henikoff 2000). This refined consensus sequence is shown in figure 1.5. Some, but not all, of the CSD interacting proteins shown in table 1.1 contain a consensus binding pentapeptide.

In summary, HP1 proteins localise to chromatin by direct binding of nucleosomes via the chromodomain. Covalent modification of histones may play a role
Figure 1.5: Pentapeptide consensus sequence derived from phage display and sequence analysis of CSD interacting proteins. The sequence is shown in LOGO format with the frequency of an amino acid occurring at a certain position represented by letter size. The figure is reproduced from Smothers and Henikoff (2000).

In regulating this binding, marking of histones has been suggested to be the mechanism for epigenetic inheritance.

The binding of chromo shadow domains to other proteins may be important for delivering HP1 proteins to chromatin or for recruiting other proteins to chromatin or both.

A number of features of HP1 action is held in common with Pc. Both proteins act in the context of large protein assemblies. Both complexes can spread out over chromatin. Epigenetically inherited marking of chromatin is the main factor contributing to assembly of the complexes though broad patterns of DNA sequence are also associated with binding. Localisation of these proteins may be regulated by post-translational modification throughout the cell cycle.
1.6 Chromodomains and dosage compensation.

Sequences with similarity to chromodomains are found in the *Drosophila* proteins Mof and Msl3 (Koonin *et al* 1995). The sequence similarity falls below statistical significance. However, these proteins are part of the dosage compensation complex. This complex has functional similarity with the chromodomain containing PcG and HP1-associated protein complexes in that they are all thought to alter gene transcription rates by acting upon chromatin. The functional homology lends weight to the argument that the Mof and Msl3 sequences are true chromodomains.

The *Drosophila* dosage compensation complex doubles transcription from certain genes on the single X chromosome of male flies to equal that from the two X chromosomes of females. In *Drosophila*, sex is determined by the ratio of X chromosomes to autosomes.

Different karyotypes are found between the sexes of a number of organisms. In humans, females have two X chromosomes and males one X and one Y. Female is the default sex and maleness is determined by the presence of a Y chromosome. Flies have a similar karyotype but sex is determined by the ratio of X chromosomes to autosomes. In birds males are homozygous and females are heterozygous.

In many of these species some form of compensation mechanism keeps the transcription levels of certain genes on the sex chromosomes the same in both genders. In humans, one of the two female X chromosomes is inactivated. In flies, transcription from the single male X chromosome is doubled (reviewed Marin *et al* 2000).

The *Drosophila* dosage compensation complex consists of at least five proteins, maleless (mle), the male specific lethal genes (msl-1, msl-2 and msl-3) and males absent on the first (mof). In addition there are two non-coding RNA molecules integral to the complex, roX1 and roX2.

In females, the action of a protein called sex-lethal prevents translation of the msl-2 protein. In the absence of msl-2 the dosage compensation complex does not
form. Males do not express the sex-lethal gene and therefore the complex is sex specific.

The dosage compensation complex protein msl-3 contains a region with sequence similarity to chromodomains. A region of msl-3 was found to have highly significant sequence similarity with some uncharacterised human ESTs (P <10^{-3}). These in turn were found to have some similarity to the chromodomain of Pc (P ~0.08). The sequence comparison algorithms used in the study could not detect any direct similarity between msl-3 and the Pc chromodomain. Also the indirect similarity is below the widely accepted significance cutoff of P=0.05. The comparison was noted because of the functional homology between the dosage compensation complex and that of known chromodomain proteins (Koonin et al 1995).

A number of features of the dosage compensation complex make it reminiscent of the chromodomain protein complexes discussed in previous sections.

The dosage compensation complex is likely to upregulate gene transcription by altering chromatin structures. In support of this hypothesis the X chromosome in males shows a distinct expanded morphology when compared to female X chromosomes suggesting a diffuse chromatin structure (Kageyama et al 2001). The dosage compensation complex acetylates lysine 16 of histone H4 specifically on the male X chromosome (Bone et al 1994, Smith et al 2000). Acetylation of histone H4 is linked to gene activation via chromatin structure modulation (Grunstein 1997). Interestingly, the mof protein has homology to acetyl transferases (Hilfiker et al 1997) and has been shown to acetylate lysine 16 on histone H4 (Akhtar and Becker 2000).

Proteins of the dosage compensation complex act together as a complex. This is inferred from the observations that the proteins colocalise on the X chromosome (Bone et al 1994, Kelley et al 1995), that all are required for correct functioning of the complex, that loss of one effects the stability of the others and that proteins of the dosage compensation complex can be immunoprecipitated with anti msl-2 antibodies from nuclear extracts from male Drosophila (Smith et al 2000).
The dosage compensation complex binds to initiation sites and is then able to spread outwards in cis along the chromosome. Msl-1 and msl-2 can localise to sites on the X-chromosome in the absence of the other proteins of the complex (Palmer et al. 1994). The minimal complex is not capable of hyperactivation and only localises to around 35 sites compared to the hundreds occupied by the full complex. Two of the ~35 'minimal' sites were mapped and found to encode roX RNAs (Kelley et al. 1999).

When one of these roX encoding DNA regions is transferred to an autosomal chromosome the dosage compensation complex is assembled at the new locus. Autosomal localisation is independent of transcription from the transgene. The minimal requirement for binding by the dosage compensation complex has been mapped to a 217bp region of roXi (Kageyama et al. 2001). Interestingly, other genes from the X-chromosome have been moved to autosomes and were found to lose hyperactivation (Bhadra et al. 1999). This implies that the initiation sites for the dosage compensation complex are encoded by DNA sequence and not by epigenetic, chromatin markings.

Immunofluorescence staining of the dosage compensation complex at the initiation sites with msl-1 antibodies reveals that the complex spreads outwards in cis to surrounding chromosomal regions. (Kelley et al. 1995, Kageyama et al. 2001).

There is no evidence that members of the dosage compensation complex are subject to post-translational modification. However, Jil-1, a kinase, has been shown to physically associate with the complex (Jin et al. 2000). Jil-1 has been previously shown to phosphorylate histone H3 on Ser 10 (Jin et al. 1999). It has been demonstrated that phosphorylation of histone H3 protects it from subsequent methylation by SET domains. Methylation of H3 on Lys 9 is associated with heterochromatin (section 1.4.3)(Noma et al. 2001). Therefore, phosphorylation of histone H3 may also be an important factor in the modulation of chromatin structure by the dosage compensation complex.

Recently, it was shown that the chromodomain-like domains of msl-3 and mof were able to bind RNA in vitro (Akhtar et al. 2000). If the chromodomain-like
domain is truly a sub-group of the chromodomain superfamily then this represents another diverse function.

In summary, the dosage compensation complex modifies chromatin structure via post translational modification of histone tails. The complex is localised to specific sites on chromatin from which it spreads onto surrounding regions. In these aspects the complex is similar to the PcG and HP1 complexes discussed in previous sections. However, there are also several important differences. The PcG and HP1 complexes are targetted to chromatin, in part, by epigenetic mechanisms: it seems likely that the dosage compensation complex is not. Secondly, post translational modifications are likely to have a role in regulating the PcG and HP1 complexes: it is not apparent that the dosage compensation complex is regulated in a similar way.

1.7 Summary and questions about the chromodomain superfamily.

Chromodomains are a 70 amino acid conserved domain found in proteins that modulate chromatin structures. Generally, chromodomain proteins function as part of large protein complexes that accumulate at specific chromatin sites. These complexes are likely to be non-stoichometric, implied from the dosage dependant localisation of Pc and HP1 (Sharp et al 1997, Nakayama et al 2000), and are dynamic through the cell cycle altering in size, position and content. Post-translational modification events, both to the chromodomain proteins themselves and to their chromatin substrate, are likely to play an important role in regulating chromodomain proteins. The structural changes in chromatin, brought about by chromodomain proteins are stable and can be inherited epigenetically.

The chromodomain superfamily is divided into two sub-classes. The division is based both upon sequence similarities and on functional distinctions. The first class, called chromodomains includes the first chromodomains identified and gives its name to the superfamily. A structure of a chromodomain has been solved
by NMR. Chromodomains function in localising proteins to sites on chromatin. Chromatin binding properties of chromodomains have been observed and these are dependent on post-translational modification of histones. The second chromodomain subclass are chromo shadow domains.

The second sub-class is termed the chromo shadow domain. These occur in tandem with chromodomains in HP1 homologues. There are no structures of chromo shadow domains. Sequence identity is low between the chromodomains and chromo shadow domains (the average sequence identity between the chromodomains and chromo shadow domains shown in the alignment in figure 3.11 is 15.2% and the minimum identity is 4%). Though these subclasses are predicted to be related by experts in the field (Aasland and Stewart 1995) this relationship is not obvious from the sequence identity. Chromo shadow domains appear to function as promiscuous protein-protein interaction domains. Having been localised to its site of action by the chromodomain, chromo shadow domains may recruit various other proteins to the site. The chromo shadow domain can be phosphorylated in vitro and this may alter its binding affinities with other proteins.

It has been suggested that there are divergent chromodomains in the Drosophila dosage compensation complex proteins Ms13 and Mof. These domains have low sequence identities with both the chromodomain subfamily and the chromo shadow domain subfamily (The average sequence identity between the divergent chromodomains and the classical chromodomains in the alignment shown in figure 6.1 is 13.5%, the lowest sequence identity between these families is 4%. The average sequence identity between the divergent chromodomains and the chromo shadow domains shown in this figure is 8.6%, the lowest sequence identity is 2%). Despite the low sequence identity, the relationship of these chromodomain-like domains with chromodomains has been widely accepted and incorporated into the literature (Prakash et al 1999, Akhtar et al 2000).

These divergent chromodomains are found, among other places, in the dosage compensation complex. The chromodomain-like domains of mof and msl3 have RNA binding properties and may bind roX RNAs that are a functional part of the dosage compensation complex. The complex has features that are homologous.
to other protein complexes in which chromodomain proteins are found. There is no structural information about this class and it is not clear that they share a common fold with the other chromodomain sub-classes.

Several questions arise from this literature review. Firstly, do chromo shadow domains and the chromodomain-like domains of msl3 and mof have a similar structure to chromodomains? The answer to this question will determine if it is consistent to consider these three domain families as sub-classes of a single superfamily. Secondly, assuming that the three chromodomain sub-classes do share a common fold, how has this fold evolved to perform three distinct functions? Thirdly, are there key residues conserved within each of the three classes that can be used to assign new sequences to a particular family and predict its function?

In an attempt to answer these questions, this study will describe the solution of a chromo shadow domain structure by X-ray crystallography. The structure will be the first crystal structure of any chromodomain. The structure will be compared with the NMR structure of the murine HP1β chromodomain (Ball et al 1997). The function and behaviour of the chromo shadow domain will be assayed in vitro and in vivo. Finally, preliminary structural studies of a divergent chromodomain-like domain will be described.
Chapter 2

Materials and methods

2.1 Molecular biology

2.1.1 Expression Vectors

pMW172 is based on the commercially available expression vector pBR322 (Bolivar et al 1977). A tetracycline resistance gene in pBR322 is replaced by a T7 insert in pMW172. The insert contains a $\phi$10 promoter, a start codon at the beginning of a multiple cloning site and stop codons in each frame after the multiple cloning site (Way et al 1990). The vector allows isopropyl-$\beta$-D-thiogalactopyranoside (IPTG) inducible expression of recombinant proteins in \textit{E.coli}. The vector confers ampicillin resistance to host \textit{E.coli}.

pGEX-KG allows IPTG inducible expression of glutathione-S transferase (GST) fusion proteins in \textit{E.coli}. Transcription apparatus transcribes continuously through a GST open reading frame, a thrombin cleavage site, a poly-glycine linker and into a multiple cloning site where a target gene can be cloned (Gaun and Dixon 1991).

Both pMW172 and pGEX-KG are based on the pET expression system (Studier and Moffatt 1986).

2.1.2 PCR primers

The PCR primers used to make the expression vectors described in this thesis are shown below. All primers are shown from 5’ to 3’.
Table 2.1: Primers for cloning the CSD of fission yeast Swi6 protein, amino acids 261 to 328, into *E.coli* expression vectors pMW172 and pGEX are shown below.

Table 2.2: Primers for cloning the T323D mutant Swi6 CSD into pMW172 and pGEX. The same coding primer was used for cloning the T323D mutant as was used to clone the native protein.

Table 2.3: Primers for cloning the F324A mutant Swi6 CSD into pMW172 and pGEX. The same coding primer was used for cloning the F324A mutant as was used to clone the native protein.

Table 2.4: Primers for cloning the thumb domain of *S.pombe* DNA polymerase α (amino acids 1057 to 1225) into pMW172.

Table 2.5: Primers for cloning a fragment of the *S.pombe* Swi2 protein (amino acids 102 to 149 ) into pGEX.
Table 2.6: Primers for cloning the chromodomain-like region of the *S.pombe* Alp13 protein (amino acids 6 to 76) into pMW172.

### 2.1.3 PCR protocol

Primers were used at a final concentration of 5 ng/μl. Template DNA for all Swi6 subdomains was pAL2 (Lorentz *et al* 1994). The Swi2 fragment, the thumb domain of DNA polymerase α and the chromodomain-like region of Alp13 were amplified from a fission yeast genomic library. The PCR reactions were done with recombinant *Thermus aquaticus* (Taq) DNA Polymerase (Amersham Pharmacia Biotech) at a final concentration of 125 units per ml of PCR reaction mix. The polymerase enzyme was supplied with a 10X reaction buffer containing 100mM Tris-HCl (pH8.6), 500mM KCl, 1% Triton X-100. MgCl₂ concentrations in the PCR reaction were titrated from 0.5 to 4 mM. All the PCR reactions were most efficient at MgCl₂ concentrations of 2.5mM. The four deoxyribonucleotides ATP, TTP, CTP and GTP were each used at a final concentration of 200 μM.

PCR reaction followed a touchdown program. One cycle of this program involved a 30 second melting step at 94°C followed by a 30 second reannealing step at variable temperature followed by a 60 second extension step at 72 °C. The cycle was repeated 30 times with the temperature of the annealing step gradually reducing from 65°C to 45°C with each subsequent step.

### 2.1.4 Electrophoresis and purification of DNA

The 100 μl PCR reaction mixes were run out on 1-1.5% (w/v) agarose gels made up with 89 mM Tris base, 89 mM Boric acid, 2 mM EDTA buffer (TBE) and 0.3μg/ml ethidium bromide. The gels were run at 100V for between 40 and 60
mins. DNA was visualised under UV light. PCR product bands were excised from
the gels and purified into 50 µl dH₂O using Quiagen gel extraction kits.

The expression vectors pGEX or pMW172 were prepared from 5 ml overnight
cultures of *E.coli* strain XL1 Blue (Stratagene), transformed with one or the other
plasmid, using Quiagen plasmid miniprep kits.

### 2.1.5 Digestion of PCR product and vector

10 µl of vector or 5 µl of PCR product were digested with the restriction enzymes
EcoR1 and BamH1 (Roche). The 50 µ double digests contained 10 units each of
the two enzymes with 10 mM Tris-HCl pH 8.0, 5 mM MgCl₂, 100 mM NaCl, 1
mM 2-Mercaptoethanol (SuRE/Cut Buffer B). The DNA was digested for 1 hour
at 37°C.

After digestion the cut vector and PCR product were purified away from en-
zymes, small fragments of DNA and buffer B with Quiagen PCR purification kits
and exchanged into 50 µl of dH₂O.

### 2.1.6 Ligation

10 µl ligation reactions were set up containing 1 µl of the digested vector, a titration
of 1-3 µl of digested PCR insert, 1 µl T4 DNA ligase (1 unit/µl)(Roche), 1 µl of
10X ligation buffer (supplied with the enzyme), 660 mM Tris-HCl, 50 mM MgCl₂
, 10 mM dithio-erythritol, 10 mM ATP, pH 7.5. These ligation reactions were left
overnight at room temperature.

### 2.1.7 Preprepared plasmids

A pET16 vector for expressing amino acids 204 to 268 of mouse CAF-1 p150 pro-
tein fused to a ten histidine tag in *E.coli*, as described by Murzina *et al* (1999),
was a gift from Alain Verreault, Imperial Cancer Research Fund, Clare Hall Lab-
oratories Blanche Lane, South Mimms, UK.
The pGEX vector for expressing the full length *S. pombe* Rad21 protein fused to a GST tag was constructed by Birkenbihl and Subramani (1995).

A vector for expressing in *E. coli* amino acids 127 to 490 of the *S. pombe* Clr4 protein fused to a maltose binding domain was a gift from Janet Partridge, MRC Human Genetics Unit, Edinburgh.

2.1.8 Transformation of *E. coli*

Two *E. coli* strains were used, XL1 Blue (Stratagene) for amplification of expression vectors and BL21 (DE3) (Doherty et al 1995) (Novagen) for expressing protein.

To transform the *E. coli* strains with the plasmid they were made calcium competent. The bacteria were grown to mid-logarithmic phase of their growth curve (optical density of 0.6 measured at 600 nm) in 100 mls Luria Bertani (LB) broth at 37°C in a 500 ml conical flask with shaking. The cells were transferred to two 50 ml Falcon tubes before being chilled on ice for 15 mins. Cells were pelleted by 10 mins centrifugation at 1500xg. The growth media was aspirated and the pellet resuspended in 10 mls ice cold 100 mM CaCl₂. This was left on ice for 30 mins before the cells were pelleted by a similar centrifugation and resuspended in 2 mls ice cold CaCl₂. The calcium competent cells can be stored at -80°C by making them upto 15% with glycerol and flash freezing in aliquots in liquid nitrogen.

100 μl of calcium competent cells were pipetted onto 1-2 μl of plasmid miniprep DNA or the whole 10 μl ligation reaction mix. The cells and the DNA were mixed by inverting and incubated on ice for 30 mins. They were then heat shocked for 1 min in a waterbath at 42°C and placed back on ice for two mins. One ml of LB broth with no antibiotic was pipetted onto the cells. Transformations involving ligation reactions were incubated with shaking at 37°C for 30-60 mins, transformations involving plasmid minipreps were not. Cells were pelleted by centrifugation at 3000 rpm in a microcentrifuge and 800 μl of the supernatant removed. Cells were resuspended in the remaining 300 μl of supernatant and spread onto LB broth agar plates containing 50 μg/ml ampicillin with a glass spreader flamed in ethanol.
LB broth agar plates were prepared by melting LB broth agar in a microwave on a low power setting until liquid. The media was cooled until hand hot (~50°C). A 1000X stock of ampicillin was prepared by dissolving ampicillin in 70% ethanol to a concentration of 50 mg/ml, this was stored at -20°C. This stock was added to the molten agar to a final concentration of 50 µg/ml. The agar was poured into Petri dishes and allowed to cool (approximately 25 mls of agar to one Petri dish). The plates were allowed to dry slightly be inverting them in a 37°C oven with the lids slightly off.

After spreading the transformed cells on the plate they were incubated overnight at 37°C. Colonies from the plates were picked into 5 ml LB broth overnight cultures with 50 µg/ml ampicillin. These were used to inoculate larger scale cultures. Inoculation was typically done at a dilution of 1 part overnight culture to 50 parts fresh media with 50 µg/ml ampicillin.

For preparation of protein, BL21 (DE3) cells containing the expression vector of interest were grown to logarithmic phase of their growth curve before induction with IPTG. IPTG stock solutions were at 1M and dissolved in water. The stock was kept in 1ml aliquots at -20°C. These were added to the cultures at a 1000X dilution. Cells were incubated at 37°C with shaking for 3 hours after induction.

Cells were pelleted by centrifugation, large scale cultures (1-6 L) were centrifuged in 1L centrifuge pots at 2000xg for 20 mins, small scale cultures (10-100 mls) were centrifuged in 15 or 50 ml Falcon tubes at 3000xg for ten mins. Pellets from large scale cultures were resuspended in the large centrifuge pots in 50mls of phosphate buffered saline (PBS) 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4·7H2O, 1.4 mM KH2PO4 and transferred to Falcon tubes. The cells were repelleted in the Falcon tubes by centrifugation, the PBS removed and the pellets were flash frozen in liquid nitrogen before storage at -80°C.
2.2 Preparation of protein

As described in the previous section cell pellets from 11 cultures of *E. coli* that had overexpressed recombinant proteins were washed in PBS and frozen in Falcon tubes. The recombinant proteins described in this thesis include various constructs from the Swi6 protein: wild-type untagged CSD, GST fused CSD, L315D CSD, GST fused L315D, T323D CSD and GST fused T323D CSD. In addition, a fragment of mouse CAF p150 subunit with a 10His fusion tag, the thumb domain of fission yeast DNA polymerase α, GST tagged fission yeast Rad21 protein, GST tagged fragment of the fission yeast Swi2 protein, MBP tagged fragment of the fission yeast Clr4 protein and a chromodomain-like region from the fission yeast Alp13 protein. These proteins were extracted from the *E. coli* cell pellets and resuspended in the following ways.

The pellets containing all the GST fusion constructs, with the exception of the Rad21 fusion protein, were resuspended by vortexing and aspiration into 10 mls of cold PBS + 200 mM KCl, 1mM ethylenediaminetetraacetic acid (EDTA), 1mM ethylene glycol bis(beta-aminoethyl ether)N,N',N,N'-tetraacetic acid (EGTA), 1mM phenylmethylsulfonylflouride (PMSF), 1mM dithio-threitol (DTT) (buffer PEEPD). A small spatula-tip of solid lysozyme (Sigma) (~10 mg) was added to the cell suspension. This was left on ice for 10 mins. The suspension was sonicated on ice with 4-6 30 second bursts of sonication with 30 secs rests between bursts with medium to low power until the suspension assumed a slightly glossy, transparent appearance. Insoluble material was spun out of the cell lysate by centrifugation in a 50 ml centrifuge tube at 25,000 rpm for 20 mins at 4°C. The supernatant was carefully poured into a 15 ml Falcon tube and kept on ice.

1 ml of a 50% (w/v) slurry of glutathione sepharose beads in 20% (v/v) ethanol (Amersham Pharmacia Biotech) was equilibrated in PBS by several cycles of pelleting, aspirating supernatant and resuspension in fresh PBS. This was done in an Eppendorf tube and pelleting was done in a microcentrifuge at 4000 rpm. The
beads were added to the lysate in the 15 ml Falcon tube and incubated on a wheel at 4°C for 30 mins.

After incubation the beads were pelleted by centrifugation and the lysate removed by aspiration. Unbound protein was washed from the beads by 6 cycles of pelleting the beads, aspirating the supernatant and resuspending the beads in 15mls of fresh PEEPD buffer. After washing the supernatant was removed and the beads transferred to a 1.5ml Eppendorf tube. 250 µl of elution buffer PEEPD + 10 mM glutathione was pipetted onto the beads. The beads were resuspended by flicking and incubated on a wheel at 4°C for 15 mins. The beads were pelleted and the elution buffer removed to a fresh Eppendorf tube on ice. This was repeated twice more so that the final volume of elution buffer was 750 µl. 40 mls of cold PEEPD buffer + 15% glycerol was poured into a Petri dish. A 47mm, 0.2 µm filter disk (Millipore) was carefully floated on top of the buffer. The 750 µl elution buffer containing the GST fusion protein was pipetted onto the disk. This was left to dialyse at 4°C for 15 mins. After this the liquid on the filter containing the GST protein in PEEPD + 15% glycerol was removed to a fresh Eppendorf tube before splitting into 100 µl aliquots, flash freezing in liquid nitrogen and storing at -80°C.

2.2.1 Solubilisation and purification of Rad21 and Clr4

The Rad21 GST fusion protein and the Clr4 MBP fusion protein containing pellets were resuspended in 10 mls PEEPD buffer - the 1 mM DTT + 1.5% N-lauryl sarcosyl + lysozyme. The use of N-lauryl sarcosyl in the resolubilisation of fusion proteins is discussed by Frangioni and Neel (1993). The cell suspension was sonicated and the insoluble material spun out as described above. The soluble lysate was transferred to a 15 ml Falcon tube, and Triton X-100 added to a concentration of 2%, DTT was added to 1 mM. Glutathione sepharose beads were added to the Rad21 lysate and amylose resin (New England Biolabs) to the Clr4 lysate. Binding of the proteins to these beads and washing was done in a similar way to that described above for the other GST fusion proteins except that all washing
and binding steps used the basic PEEPD buffer + 1.5% N-lauryl sarcosyl and 2% Triton X-100. These proteins were not eluted but left bound to the beads for the purpose of GST pulldown experiments.

2.2.2 Purification of the Swi6 CSD

The wildtype, untagged Swi6 CSD containing pellet was resuspended in 10 mls 50 mM NaCl, 25 mM Tris pH 7.0, 1 mM EDTA on ice. The cells were resuspended by vortexing and aspiration. Cells were lysed by a combination of lysozyme and sonication and the insoluble material spun out as described above for GST fusion proteins. The lysate was loaded onto a 50 ml diethyl aminoethyl (DEAE) fast flow Sepharose (Amersham Pharmacia Biotech) anion exchange column pre-equilibrated in buffer A (50 mM NaCl, 25 mM Tris pH 7.0, 1 mM EDTA). The column was run at 0.5 mls/min in buffer A until unbound proteins were eluted and the absorbance at 280nm of the eluent returned to baseline (approximately 2-3 column volumes). A gradient was run from 100% buffer A to 100% buffer B (500 mM NaCl, 25 mM Tris pH 7.0, 1 mM EDTA) over 6 column volumes. Finally, the column was run at 100% buffer B for 2 column volumes. 10 ml fractions were collected throughout the column run and the absorbance of these at 280 nm was measured. Samples from peak fractions were analysed by running them on a 15% SDS PAGE gel. A protein with migration on these gels consistent with CSD eluted from the column in the first third of the gradient. Fractions containing this protein were pooled and concentrated to 5 mls using a Vivaspin concentrator with a 5KDa molecular weight cutoff (Vivascience). The concentrated fractions were dialysed overnight at 4°C against 1 l of 25 mM NaCl, 25 mM Tris pH 6.0, 1 mM EDTA.

After dialysis the sample was loaded onto a 50 ml carboxy methyl (CM) fast flow Sepharose (Amersham Pharmacia Biotech) cation exchange column pre-equilibrated in this columns buffer A (25 mM NaCl, 25 mM Tris pH 6.0, 1 mM EDTA). The column was run at 0.5 mls/min for 3 column volumes collecting 5 ml fractions. Two distinct peaks of absorbance at 280nm were measured. The CSD
was detected by SDS PAGE gel in both of these peaks but was relatively pure in the second. Fractions in this peak were pooled and concentrated to 1ml using a Vivaspin concentrator with a 5KDa molecular weight cutoff (Vivascience).

2.2.3 Purification of the T323D Swi6 CSD

The *E. coli* cell pellet containing the mutant CSD was resuspended in 10 mls 25 mM BisTris pH 6.5, 5mM EDTA, 50mM NaCl with some lysozyme. Cell lysis and preparation of a soluble lysate are as described above for the wild type CSD. The soluble lysate was loaded onto a 50 ml DEAE fast flow Sepharose column pre-equilibrated in buffer A (25 mM BisTris pH 6.5, 5mM EDTA, 50mM NaCl). The column was run for 3 column volumes at 100% buffer A and then a gradient from 100% buffer A to 100% buffer B (25 mM BisTris pH 6.5, 5mM EDTA, 500mM NaCl) was run over 6 column volumes. Finally the column was run for 2 column volumes at 100% B. The mutant CSD was eluted around the middle of the gradient. Fractions containing the CSD protein were pooled and concentrated to 2 mls.

A 120 ml high resolution 16/60 Sephacryl™S-200 size exclusion column (Amersham Pharmacia Biotech) was pre-equilibrated in 25 mM BisTris pH 6.5, 5mM EDTA, 300mM NaCl. The concentrated T323D CSD containing column fractions from the previous purification step were loaded onto the column. The mutant CSD eluted with a retention volume of about 2 column volumes. The column was run at 0.3 ml/min and 20 ml fractions were collected throughout. Fractions containing the mutant CSD were pooled and concentrated to about 1 ml.

2.2.4 Purification of the chromodomain-like region of Alp13

The *E. coli* cell pellet containing the CLR from Alp13 was resuspended in 10 mls 25 mM BisTris pH 6.5, 50 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT with some lysozyme. Cell lysis and preparation of a soluble lysate are as described above for the wild type CSD. The soluble lysate was loaded onto a 50 ml DEAE
fast flow Sepharose column pre-equilibrated in buffer A (25 mM BisTris pH 6.5, 50 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT). The column was run for 4 column volumes at 100% buffer A collecting 5 ml fractions. Two distinct peaks of absorbance at 280nm were measured. The CLR was detected by SDS PAGE gel in both of these peaks but was relatively pure in the second. Fractions in this peak were pooled and concentrated to 1ml using a Vivaspin concentrator with a 5KDa molecular weight cutoff (Vivascience).

2.2.5 Preparation of DNA polymerase α and CAF1 p150 fragment

The thumb domain of DNA polymerase α was expressed in *E.coli* BL21 (DE3) using the standard protein expression protocol described above. The 10xHis tagged fragment of mouse CAF1 p150 subunit was expressed in *E.coli* strain JM109 (DE3). After transformation a colony of these cells were picked fresh from the selective nutrient agar plates to 1l liquid cultures with ampicillin. These were grown to logarithmic phase before IPTG induction. For the purpose of pulldown experiments neither protein was purified. Soluble lysates from these cultures were prepared as described above and frozen with 15% (v/v) glycerol.

2.3 Crystallisation of Swi6 CSD

Crystallisation was achieved by hanging drop vapour diffusion. 1 ml of a mother liquor (mother liquor is the solution that the protein crystals are grown in) is pipetted into one well of a 24 well Limbro plate (Molecular Dimensions). 1-10 μl of the well solution is removed to a siliconised glass coverslip, an equal volume of protein solution is added to the drop on the coverslip. Vacuum grease is applied around the rim of the well and the coverslip is pressed into the grease, sealing the well with the drop hanging down inside the well. Over time the solution in the drop equilibrates with the solution in the well through the vapour phase. The slow equilibration is amenable to crystal formation.
As described above, the wild-type CSD of Swi6 was purified from a 11 culture of recombinant _E. coli_ and concentrated to around 1 ml. The concentration of protein in this sample was measured using the BCA assay (Pierce) with a bovine serum albumin standard (Pierce) and found to vary between 10 and 20 mg/ml from preparation to preparation. This solution was diluted or concentrated to a standard concentration of 15 mg/ml for crystallisation.

To find crystallisation conditions a two dimensional ammonium sulphate screen was set up varying pH from 4.9 to 9.1 against ammonium sulphate concentrations from 0.5-3.0M, a polyethylene glycol (PEG) screen varying PEG 5000 concentrations from 7% to 28% and pH from 4.9 to 9.1 and Structure Screen 1 (Molecular Dimensions), a sparse matrix screen (Jancarik and Kim 1991).

Initial crystallisation conditions were 30% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5 discovered in Structure Screen 1. These were refined to 20% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5.

To protect the crystals for cooling in liquid nitrogen the coverslip from the well over which the crystals had been grown was transferred to a fresh well containing 1 ml of 30% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5. The coverslip was sealed over the new well with vacuum grease and left to reequilibrate with the higher PEG concentration for one week at 4°C. The coverslip was removed from the well and the crystals removed from the drop, using a mounted cryoloop (Hampton Research), to a drop of cryoprotectant 30% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5 4% (v/v) glycerol.

To prepare heavy atom derivatives for multiple isomorphous replacement, compounds ethyl mercury thiosalicylate (EMTS), _K_2PtCl_4_ and Au(CN)_2_ were dissolved in the 30% (w/v) PEG mother liquor to a concentration of 10 mM. These stock solutions were pipetted into a hanging drop after formation of CSD crystals and reequilibration to 30% PEG and the coverslip resealed over the well of mother liquor. The three soaks were 1mM _K_2PtCl_4_ for 24 hrs, 1mM EMTS for 26 hrs and 1mM EMTS + 0.5 mM Au(CN)_2_ for 26 hrs. All the soaks were done at 4°C.

A cryoloop is a fibre of nylon that has been folded in half and the two ends
twisted together to make a loop. The twisted ends of the loop are inserted into
the bore of a hollow metal post such as a blunt hypodermic needle. The post
is attached to the lid of a cryovial (Hampton Research). The fibre, post and lid
are held together with araldyte. The lid can be screwed onto the body of the
vial with the post and the loop inside. A metal disk on the outside of the lid
allows the mounted cryoloop to be attached to the magnetic mounting platform
of a goniometer head.

For cooling in liquid nitrogen the crystals were removed from the cryoprotectant
with a mounted cryoloop. The wide end of a 1 ml pipette tip (Gilson) fits securely
onto the outside of the cryovial lid, this can be used as a handle to dip the cryoloop,
with mounted crystal, into a flask of liquid nitrogen. The body of the cryovial is
held in a clothes peg and submerged under the liquid nitrogen in the flask. The lid
can be screwed onto the body of the vial beneath the surface of the nitrogen and
the pipette tip removed. The vial, capped and full of nitrogen can be attached to
a storage cane (Hampton Research) for long term storage of the crystal in a dewar
of liquid nitrogen.

To collect diffraction data the lid of the cryovial with the mounted crystal is
attached to a goniometer head. A cryostream of dry nitrogen gas at 100 K flows
over the crystal to keep it cool and prevent build up of ice. To mount the crystal
onto a goniometer head the vial is removed from the cane with a precooled clothes
peg. A 5 ml pipette tip is trimmed to fit securely onto the bottom of the vial as
a handle, which is also precooled. The lid of the vial is loosened slightly before
removing the peg. The magnetic head of the goniometer is tilted to about 45°
down so that when the vial is attached to it the nitrogen does not spill out of the
vial and aligned so that when the crystal is mounted it will be in the center of the
cryostream. After the top of the lid of the cryovial is attached to the magnetic
goniometer head the body of the vial is unscrewed and removed in one, quick
movement.

The native CSD crystal diffraction dataset was collected using a CCD (Mar)
scanner at SRS Daresbury beamline PX9.6. The crystal to detector distance was
190 mm, the oscillation range 1.5° and the exposure time of 30 seconds. The
derivative datasets were collected on a FR591 rotating anode X-ray generator (Bruker Nonius) at 80 V, 40 mA with a Mar image plate detector. The crystal to detector distance was 200 mm and the exposure time 1800 seconds. The vertical and horizontal slit widths on the collimator were 0.3 mm.

Diffraction data was indexed with the program MOSFLM (Leslie 1992). An initial electron density map was produced from the native data set and the three derivative data sets by the program SOLVE. The quality of the map was improved by density modification and solvent flipping with the program DM (CCP4) (Cowtan 1994). A model of the CSD was built into the density using the baton-build, auto-mainchain and auto-sidechain algorithms in the modelling program O (Jones et al 1991). The model was refined using the program CNS (Brunger et al 1998).

2.4 GST pulldowns

2.4.1 GST-CSD interaction with CSD

GST-CSD and unfused GST were prepared and purified as described in section 2.2. 100 μl of the GST-CSD (approximately 1-2 mg/ml) or 20 μl of GST (approximately 10 mg/ml) was added to 400 μl of E.coli cell lysate either without recombinant protein or containing recombinantly expressed untagged CSD. 40 mls each of 6M urea, 4M urea, 2M urea, 1M urea, 0.5M urea + PBS + 300 mM KCl and 0M urea + PBS + 300 mM KCl were made up and decanted into 6 Petri dishes. 47mm, 0.2 μm filter disks (Millipore) were carefully floated on the surface of the 6M urea solution and the mixtures of purified protein and cell lysate were pipetted onto the disks. The disks were left on the surface of the solution for 15 mins for the protein mixtures to equilibrate with the solution in the dish. Using two pairs of tweezers the disks were moved to the 4M urea solution for 15 mins and so on from 4M to 0M urea. After the urea was dialysed away in this stepwise way the mixtures were transferred to Eppendorf tubes. Glutathione Sepharose beads were preequilibrated in PBS + 300 mM KCl and 70 μl of beads were added to each
tube. The tubes were incubated on a wheel at 4°C for 30 mins before the beads were pelleted by centrifugation in a micro-centrifuge at 4000 rpm for 1 min. The supernatant was removed and the beads resuspended in 1.5 mls PBS + 300 mM KCl, this was repeated six times to wash unbound proteins from the beads before they were finally pelleted, the wash buffer removed and 25 μl of 2X SDS PAGE gel loading buffer (125 mM Tris pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, 2.5 mg/ml bromophenol blue, 5 mM DTT) added onto bead. The samples were boiled for 3 mins during which time the beads settle to the bottom of the tube. The dye was removed from top of the beads with a glass syringe (Hamilton) to a 15% SDS PAGE gel.

2.4.2 GST pulldowns with soluble proteins

The GST pulldown protocol for the following pairs of proteins are similar to the protocol described above for GST-CSD with CSD except that the urea denaturation step was missed. Protein mixtures were incubated together for 30 mins before glutathione agarose beads were added.

As described above the thumb domain of fission yeast DNA polymerase α and the fragment from mouse CAF1 p150 subunit were expressed in E.coli and and soluble lysates from these cultures were prepared. 100 μl of GST-CSD or GST-T323D (approximately 15 mg/ml) or 20 μl of unfused GST were added to 400 μl of the lysates. The GST pulldown protocol is similar to that described for the pulldown of CSD with GST-CSD but the urea denaturation step was missed.

The GST-Swi2-fragment pulldowns were purified as described above. The concentration of the purified protein was approximately 7 mg/ml. 20 μl Swi2 or 4 μl GST (approximately 35 mg/ml) were added to 400 μl binding buffer (25 mM Tris pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT) with either 15 μl purified CSD (approximately 10 mg/ml) or 10 μl T323D CSD (approximately 15 mg/ml). The proteins were incubated together for 30 mins before the beads were pelleted and the supernatant removed. The beads were washed 6 times in 1.5 mls binding buffer.25 μl 2x SDS gel loading buffer was added to the beads and they were boiled
for 3 mins before the dye was removed from the beads with a syringe (Hamilton) and loaded onto a 15% (w/v) acrylamide, SDS PAGE gel.

2.4.3 GST pulldowns with insoluble proteins

The GST-Rad21 and MBP-Clr4-fragment proteins were solubilised with 1.5% (w/v) N-lauryl sarcosyl, as described above. During purification these proteins were bound to glutathione beads or amylose beads and washed but not eluted from the beads. 100 μl of beads, loaded with the fusion protein, was added to 400 μl of binding buffer (0.5% (w/v) N-lauryl sarcosyl, 1% (v/v) Triton X-100, 25 mM Tris pH 7.5, 200 mM KCl, 1 mM EDTA) with 15 μl purified CSD (approximately 10 mg/ml) or 10 μl purified T323D CSD (approximately 15 mg/ml). The mixture was incubated at 4°C for 30 minutes. After incubation the beads were pelleted by centrifugation, the supernatant removed and the beads resuspended in 1.5 mls of binding buffer. This was repeated six times to remove all unbound protein. 25 μl 2x SDS gel loading buffer was added to the beads and they were boiled for 3 mins before the dye was removed from the beads with a syringe (Hamilton) and loaded onto a 15% (w/v) acrylamide SDS PAGE gel.

2.5 SELDI mass spectrometry

GST and GST-CSD were purified by the protocol detailed above. 1 μl of a 1 mg/ml solution of either GST-CSD or GST in PBS was pipetted onto a PS2 amine coupling ProteinChip®. Protein can be detected bound to the biochip from solutions with protein concentrations as low as 500 fM to 1 pM. The chip was incubated with the protein solution for 1 hr in a humid environment. The chip was then blocked with 3 μl 0.5 M Tris pH 8.0 to prevent further protein binding and washed 3 times with PBS + 1% (v/v) Triton X-100 + 0.5 M NaCl and a final wash of PBS + 0.05% (v/v) Triton X-100. 4 μl of 2 pM CSD in bacterial extract was spotted on the biochip. This was incubated for 90 mins at room temperature,
washed 3 times in PBS + 0.05% (v/v) Triton X-100, 1 time in PBS and 1 time in H$_2$O. The biochip was analysed by SELDI mass spectrometry.

2.6 NMR

Approximately 10 mM solutions of the Swi6 CSD and Alp13 CLD were made in 90% H$_2$O, 10% D$_2$O. 1 dimensional NMR spectra were collected with a sweep width of 7200 Hz using a VARIAN ANOVA 600 MHz NMR spectrometer.

2.7 Alp13 circular dichroism

Circular dichroism spectra were collected from the chromodomain-like region of Alp13 (amino acids 6 to 76) and from the chromo shadow domain of Swi6 (amino acids 261 to 328). The purified proteins were diluted in PBS, the concentration of the Alp13 chromodomain-like region was estimated to be 0.39 mg/ml and the Swi6 CSD to be 0.47 mg/ml by BCA assay. The spectra were collected on a JASCO J-600 spectropolarimeter from 190 to 260 nm.

2.8 Alp13 analytical size exclusion

A Superdex 75 PC 10/30 column (Amersham Pharmacia Biotech) was equilibrated in column buffer (25 mM Tris pH 7.5, 300 mM NaCl, 1 mM EDTA). 100 µl of 10 mg/ml solution of Swi6 CSD or Alp13 chromodomain-like region, dialysed into the column buffer, were loaded onto the column. The column was run at room temperature at 5 mls/min. The column was calibrated with a low molecular weight calibration kit (Amersham Pharmacia Biotech).
Chapter 3

Crystal Structure Of a Chromo Shadow Domain

3.1 Introduction

Swi6 is a fission yeast homologue of HP1. The chromo shadow domain of Swi6 is similar in sequence to those of HP1 homologues in higher organisms. Thus, it is likely that conclusions drawn from the structure of the Swi6 CSD will be applicable to other CSDs. In addition, the advantage of studying this protein in fission yeast are that hypotheses resulting from the structure can be tested in vivo in this genetically tractable organism.

Specifically, the aims of this structural study are to test the hypothesis that the CSD has a similar structure to the CD. This is predicted to be the case despite low sequence similarity. Also to determine how the CSD may interact with other proteins and, by comparison with the CD structure, define why these structurally homologous protein domains have such different functions.

In this chapter the steps towards solution of the Swi6 CSD structure will be presented. Since several of the stages of the structure determination were of special interest from a crystallographic perspective, these will be discussed. The CSD model will be described and an initial comparison made with the CD NMR structure.
3.2 Expression of recombinant Swi6 CSD

Swi6 is a protein of 328 amino acids containing an amino terminal CD and a carboxy terminal CSD. The CD is preceded by 80 amino acids and is linked to the CSD by a 120 amino acid region. These regions are predicted to be without secondary structure. The CSD extends right to the C terminus. The boundaries of the CSD were determined by similarity to the CD structure. Oligonucleotide primers were designed to amplify this region of the Swi6+ gene by means of the polymerase chain reaction (PCR). The PCR product was ligated into the plasmid vector pMW172. This vector is based on the commercially available pBR322 (Bo- livar et al 1977) but has a modified polylinker to facilitate subcloning of fragments excised from other plasmids. pMW172 and pBR322 are pET based vectors that allow expression of protein from cloned genes in *E. coli*.

pET vectors maintain a cloned gene under the control of a T7 RNA polymerase promoter. The promoter is not recognised by *E. coli* RNA polymerase and so is normally silent in the bacteria. A host strain transfected with λ prophage DE3, carries a T7 RNA polymerase gene integrated in its genomic DNA. The gene is under control of the lac operon allowing lactose inducible expression of protein from the pET vector. A non-hydrolysable analogue of lactose, isopropyl-beta-D-thiogalactopyranoside (IPTG), is used for practical reasons. The inducibility of expression from pET vectors is useful since overexpression of recombinant proteins can be toxic to the host cells giving selective advantage to cells that do not express the protein. Thus cells are usually grown to mid-logarithmic phase of their growth curve before IPTG is added to induce expression of the recombinant protein (Studier 1986).

The *E.coli* strain BL21 (DE3) was transfected with the pMW172 vector containing the cloned CSD construct. When induced with IPTG in mid-logarithmic phase of their growth, the bacteria overexpressed the CSD yielding roughly 20-30mg of pure protein per litre of cell culture after two purification steps.
3.3 Purification of Recombinant Swi6 CSD

From the primary amino acid sequence of the CSD a theoretical pI of 5.25 was calculated (Bjellqvist 1993). The CSD bound to DEAE Sepharose anion exchange resin in a buffer containing 25mM Tris pH 7.0, 50mM NaCl and 1mM EDTA. Protein with an SDS PAGE migration consistent with CSD was eluted by approximately 150 mM NaCl during a salt gradient from 50mM to 500mM NaCl (figure 3.1).

Most *E. coli* proteins have slightly acidic pI's, similar to that of the Swi6 CSD. This makes separation by the ion exchange protocol described above difficult. The CSD was eluted from a carboxymethyl (CM) Sepharose cation exchange column at pH 6.5, before application of a salt gradient but retarded in relation to the initial breakthrough peak of unbound proteins. This suggests that the pI of the CSD is close to pH 6.5. However, the pI was calculated to be 5.25, more than one pH unit below that at which the column was run. This error may be due to the small size of the CSD. In small proteins a single charged residue has a greater effect on the overall charge than in a larger protein. This means that a salt bridge, which neutralises two charges, can have a large effect on the pI (figure 3.1).

Separation of CSD from contaminating proteins on the CM column was enhanced by various methods. Firstly, by use of a long, narrow column, the retardation effect is amplified along its length. Secondly, the load protein was concentrated to 2mls, which made the elution peaks sharper. Thirdly, the protein was loaded at low ionic strength (25mM NaCl). This was found experimentally to further increase CSD retardation. In this way an ion exchange column can be used in a way that is more analogous to gel filtration chromatography than conventional ion exchange.

Purified Swi6 CSD was concentrated to approximately 15mg/ml. A 1D proton NMR spectrum was taken of the protein sample on a 600MHz instrument to check that the CSD was correctly folded. The NMR spectrum is shown in figure 3.2 (NMR data was collected by B. Smith).
Figure 3.1: a) Column traces from the two ion exchange chromatography steps used in the purification of Swi6 CSD. The absorbance at 280 nm is shown as a blue line with values on the left-hand Y axis. Salt concentration is shown as a red line with values plotted in mM on the righthand Y axis. Fractions were collected and the fraction number is plotted on the X axis. Fractions containing the CSD, pooled for the next stage are boxed in green. b) Lanes from a 15% (w/v) SDS PAGE gel showing samples at different stages of the preparation. Molecular weight markers are shown (MRK) with relevant sizes indicated in KDa. Total cell protein before (-IPTG) and after (+IPTG) induction are shown. The pooled CSD containing fractions from column 1 (step 1) and column 2 (step 2) are also shown.
Figure 3.2: 1D proton NMR spectrum of Swi6 CSD. Chemical shift in parts per million is noted along the X axis. Water quenching is seen at 4.8 ppm.

The sharp peak shape throughout the spectrum is due to a rapid tumbling time that is characteristic of a globular protein. The methyl $H_1$ resonances from unfolded proteins tend to cluster in a single peak around 1 parts per million (ppm). Methyl peaks dispersed between 1 and -1 ppm in the CSD spectra are indicative of a folded protein. $Ca$ protons from $\beta$ sheet structures typically lie between 5 and 6.5 ppm. A peak from the proton of the sidechain nitrogen of a tryptophan residue is seen as a low, broad peak at 11 ppm.

### 3.4 Crystallisation of Swi6 CSD

When the concentration of organic precipitants such as polyethylene glycol (PEG) or inorganic precipitants such as ammonium sulphate (AS) are raised, proteins come out of solution. If the change is gradual the protein solution can become supersaturated. In the supersaturated phase, and with other variables such as
temperature, pH or other chemical additives at the correct values, proteins can crystallise.

Hanging drop vapor diffusion was used to achieve a gradual increase in precipitant concentration. Typically, one ml of a crystallisation solution is pipetted into the bottom of a well. A small volume (1-10µl) is removed to a siliconised coverslip (the silicone presents a hydrophobic surface to the drop preventing it from spreading out over the surface of the glass). This drop is mixed with an equal volume of the protein solution and the coverslip sealed over the well, drop downward. Via the vapor phase, the drop will equilibrate slowly with the well solution. Protein crystals can form in the equilibrating drop.

To find conditions under which CSD would crystallise, screens were set up varying concentrations of AS and PEG with pH. In addition, a sparse matrix screen (structure screen 1 from Hampton Research), which covers a wide range of pHs, precipitants and additives in 50 wells, was set up. These screens were duplicated at 4°C and 17°C. The sparse matrix screen yielded crystals after two days at 17°C and after 5-10 days at 4°C. These grew in 30% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5. These initial conditions were optimised and the final conditions for crystal growth were 20% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5 at 4°C; the protein concentration was approximately 15 mg/ml (figure 3.3).

### 3.5 X-ray diffraction by Swi6 CSD crystals

X-ray diffraction data from protein crystals can be collected at room temperature but damage by free radicals in the crystal means that more than one crystal is often needed to collect a full dataset. Alternatively the crystals can be transferred to a cryo-protectant such as glycerol before rapid cooling down to liquid nitrogen temperatures. Diffraction data is collected while the crystal is in a stream of dry nitrogen at 100K.
Figure 3.3: protein crystals of the Swi6 chromo shadow domain grown in 20\% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris.HCl pH 8.5. Largest crystals are approximately 0.75 mm long.

CSD crystals, after growth in 20\% (w/v) PEG 4000, were equilibrated slowly by vapor diffusion to 30\% (w/v) PEG 4000. To achieve this, the coverslips from wells containing crystals were removed to new wells containing the mother liquor but with 30\% (w/v) PEG 4000. One week at 4\( ^\circ \)C was allowed for equilibration to this new PEG concentration. The crystals were then transferred to a solution for freezing containing 30\% (w/v) PEG 4000, 0.2M Sodium Acetate, 0.1M Tris HCl pH 8.5 plus 4\% (v/v) glycerol. Diffraction data to a maximum resolution of 1.9\( \text{Å} \) were collected at the SRS Daresbury beamline PX9.6, with a wavelength of 0.87\( \text{Å} \), using a CCD (Mar) scanner. Data were indexed with the trigonal spacegroup P3\(_2\)21 using the CCP4 program MOSFLM (Collaborative Computational Project, Number 4. 1994).

Indexing and refinement statistics are summarised in table 3.1.

The unit cell has axes a=b=59.3\( \text{Å} \), c=90.6\( \text{Å} \). All trigonal systems have axes a=b without constraint on c and two threefold axes parallel the [001] direction related by a twofold axis along the [110] direction.
Data collection, phase determination and refinement.

Space group P3221 a=59.30 A c=90.63 A
Two molecules in asymmetric unit (0.61 solvent)

<table>
<thead>
<tr>
<th>Data set</th>
<th>Native</th>
<th>EMTS*</th>
<th>EMTS+AUCN2†</th>
<th>PT‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.87</td>
<td>Cu Kα</td>
<td>Cu Kα</td>
<td>Cu Kα</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>1.9</td>
<td>3.0</td>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>14648</td>
<td>3838</td>
<td>6473</td>
<td>3933</td>
</tr>
<tr>
<td>Overall completeness (%)</td>
<td>99.8</td>
<td>98.6</td>
<td>96.4</td>
<td>99.4</td>
</tr>
<tr>
<td>Redundancy</td>
<td>14.0</td>
<td>11.9</td>
<td>7.3</td>
<td>6.4</td>
</tr>
<tr>
<td>R mean I$^#$ (outer shell)</td>
<td>0.061 (0.24)</td>
<td>0.135 (0.26)</td>
<td>0.074 (0.22)</td>
<td>0.066 (0.162)</td>
</tr>
<tr>
<td>Number of heavy atom sites</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Phasing power* (centric/acentric)</td>
<td>0.81/0.98</td>
<td>0.77/0.91</td>
<td>0.11/0.12</td>
<td></td>
</tr>
<tr>
<td>Figure of merit (outer shell)</td>
<td>0.64/0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>22.3-1.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflection used (R$_{free}$ set)</td>
<td>14578 (733)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R$<em>{free}$/R$</em>{free}$$^#$</td>
<td>0.226 (0.259)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms (solvent atoms)</td>
<td>990 (129)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyethylene glycol (subunits)</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. bond lengths/Angles (Å/degrees)</td>
<td>0.009/1.23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*EMTS, ethyl mercuirthiosalicylate, incubation at 1mM for 26h. †EMTS+AUCN2, ethyl mercuirthiosalicylate at 1mM and 0.5 mM Kau(CN)$_2$ for 26 h. ‡PT, 1 mM K$_2$PdCl$_4$ for 24 h. $^\#$R$_{free}$ = $\Sigma$[FI - BF]/$\Sigma$FI, in which FI and BF are the observed and calculated structure factor amplitudes respectively. R$_{free}$ is calculated from a set of randomly chosen reflections omitted from the refinement, and R$_{free}$ is calculated for the remaining reflections included in the refinement.

Table 3.1: Table of statistics relating to native and heavy atom derivative crystal diffraction data processing, structure solution and refinement.
<table>
<thead>
<tr>
<th>No. CSD molecules/unit cell</th>
<th>% solvent in crystal</th>
<th>Vm (Å³/Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 molecule</td>
<td>81.86</td>
<td>6.84</td>
</tr>
<tr>
<td>2 molecules</td>
<td>63.73</td>
<td>3.42</td>
</tr>
<tr>
<td>3 molecules</td>
<td>45.59</td>
<td>2.28</td>
</tr>
<tr>
<td>4 molecules</td>
<td>27.46</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Table 3.2: Showing the percentage solvent and the Matthews number (Vm) for 1, 2, 3 or 4 molecules per unit cell.

3.6 The Matthews Coefficient

The number of protein molecules in the unit cell of an indexed crystal can be predicted using the Matthews coefficient (Matthews 1968). The coefficient is proportional to the volume of a unit cell excluded from solvent by a given number of protein molecules. The volume of a protein can be calculated from its molecular weight given that the average partial specific volume of protein is 0.75 ml/g. Matthews coefficients are shown in table 3.2 for one to four CSD molecules per asymmetric unit.

The solvent content of most protein crystals falls in the range 30-70% and the average Matthews number is 2.8Å³/Da (Matthews 1968). The presence of one or four molecules of CSD in the unit cell seems unlikely. Two or three molecules per cell give values within the standard range but the coefficient cannot resolve between these two possibilities. One method that may distinguish between these is a self rotation function.
3.7 The Self Rotation Function

A self rotation function can be plotted using phaseless $|F|^2$ values. Fourier transform of this phaseless data gives peaks that represent the vectors between atoms in the crystal rather than the atom positions themselves. A Patterson map plots out these vector peaks. A self rotation function can be used to detect symmetry within the Patterson map.

In practice, a sphere is defined enclosing a subset of the Patterson peaks. The sphere is rotated on top of itself in a manner that can be described by two angles, $\omega$, which is the angle of rotation from a polar axis, and $\phi$, which is the angle of rotation around the equator. $\omega$ and $\phi$ rotation will bring a vertical vector to a new position within the sphere. Finally, the sphere is rotated around this new vector (an angle defined as $\kappa$) (figure 3.4). When the $\kappa$ rotation axis lies along a symmetry axis in the data this rotation superimposes symmetry related peaks. A peak on the self rotation function indicates that lots of Patterson peaks are superimposed by that $\kappa$ rotation. When the $\kappa$ rotation axis lies along a twofold symmetry axis there is a high peak at $\kappa=180^\circ$; similarly a threefold symmetry axis gives a peak at $\kappa=120^\circ$ and so on.

Molecules in the unit cell of the CSD crystal are related by the crystallographic symmetry elements of the spacegroup $P3_2121$. In addition, molecules can associate with each other in multimers with internal, non-crystallographic symmetry. Two molecules associated in a dimer may or may not be related by twofold, non-crystallographic symmetry. Similarly, three molecules associated in a trimer may have threefold, noncrystallographic symmetry. However, a dimer does not have threefold, non-crystallographic symmetry and a trimer does not have twofold, non-crystallographic symmetry. The absence of a non-crystallographic twofold or threefold axis peak in the self rotation function does not rule out the possibility of dimers or trimers. A self rotation function peak corresponding to a twofold, non-crystallographic symmetry symmetry axis will support the hypothesis that there
A self rotation function was performed on native data from the CSD crystal. As expected, the crystallographic symmetry operators of the spacegroup P321 resulted in peaks of maximum density (100σ) at φ angles that were multiples of 60°. As well as twofold axes (κ=180°) peaks of maximum density at κ=120° result from rotation around the two threefold axes of the spacegroup. The mean density of the rotation function was -0.024σ with a rmsd of 6.78σ.

After the crystallographic symmetry peaks, the next most dense peaks occur at 38σ, this is significantly above the mean level. These peaks occur at an ω rotation of 90° and κ rotations of 180°. The φ rotation of the peak was at 46.6° and every 60° after this related by a crystallographic threefold axis. This peak may represent a non-crystallographic twofold axis between molecules in the unit cell. The results of the self rotation function support the hypothesis that there are two molecules in the asymmetric unit.
Table 3.3: Equivalent positions for spacegroup \( P3_21 \) in relation to an atom with coordinates \( x, y, z \). Positions are from the International Tables For Crystallography Volume A: Space-group Symmetry.

3.8 Phasing by Multiple Isomorphous Replacement

The structure factors of the CSD crystal diffraction data were phased by multiple isomorphous replacement (MIR). In MIR a heavy atom is introduced into a crystal by soaking it in a heavy metal solution. Because of their large electron density relative to protein atoms, vectors between heavy atoms can be resolved above other atomic vectors on a Patterson map. It is often possible to calculate the atomic coordinates of a heavy atom site from the vectors on a Patterson map and from a knowledge of the symmetry operations of a spacegroup.

The symmetry of spacegroup \( P3_21 \) is described by two threefold screw axes parallel to the \( Z \) axis of the unit cell. These are related by a twofold axis along \( X=Y \ Z=0 \). A heavy atom at position \( x, y, z \) has an equivalent atom across the twofold. Symmetry related atoms are generated from these equivalents at points rotated \( 120^\circ \) and \( 240^\circ \) around the screw axes and translated one third and two thirds along \( Z \) respectively. Thus there are six symmetry related equivalent heavy atoms per unit cell. This is illustrated in figure 3.5.

The equivalent positions shown in table 3.3 can be subtracted from each other to give the vectors between them. For example, the vector between atom \( c (y, x - y, z + \frac{2}{3}) \) and atom \( e (x - y, \bar{y}, \bar{z} + \frac{1}{3}) \) is simply \( \bar{x}, \bar{x}, 2z + \frac{2}{3} \). Matching these vectors to the peaks on a Patterson map allows calculation of the heavy atom position.
Figure 3.5: The edges of a cell with spacegroup $P3_21$ are shown. The view looks down the Z axis. Two planes within the unit cell are represented by broken lines. The threefold screw axes are shown as filled triangles with radiating lines. Equivalent positions are shown as open circles. The fractional position along the Z axis compared to atom 'a' at coordinate Z is indicated beside each atom. Each atom is labelled with a letter a-f, the atomic positions of each can be expressed as a function of the coordinates of the original atom $(x,y,z)$ these are coordinates are given for each in table 3.3.

$(x,y,z)$. Approximate phases for the structure factors can be calculated from this position.

The two screw axes of $P3_21$ translate atoms $\frac{1}{3}$ along the z axis. Vectors from one atom to the next along the two screw axes are found on planes $\frac{1}{3}$ and $\frac{2}{3}$ along the z axis. These are known as Harker planes. Each Harker plane should contain six vector peaks relating each heavy atom with the next along the two screw axes.

Many heavy atom compounds were soaked into CSD crystals at different temperatures, concentrations and for different times. Most of these soaks resulted in crystal cracking, high mosaicity, non-isomorphism with the native dataset or failure of the heavy atom to bind to protein sites in the crystal. However, three useful derivative datasets were collected and together these allowed the CSD structure to be solved.

Crystals of CSD were soaked for 26 hours in a solution of 1mM ethyl mercu-
Figure 3.6: Harker plane \( z = \frac{1}{3} \) from a difference Patterson of data from EMTS soaked crystals and a native dataset. The resolution range of the data is 15-4.0Å. Peaks are contoured at 15, 20, 25, 30, 40, 50, 55 and 60 sigma. The scale is 2mm/Å. The \( x \) axis is horizontal and the \( y \) axis is at 59.3°.

rithiosalicylate (EMTS) at 4°C. A Patterson map was calculated and the peaks at the Harker plane \( z = \frac{1}{3} \) are shown in figure 3.6.

The heavy atom vectors on the Harker plane of the EMTS derivative were very strong because it is a good derivative with high occupancy. It is also evident that there are twelve peaks on the Harker plane instead of the expected six. This doubling of peaks comes about because the \( Z \) coordinate of the heavy atom is close to \( \pm \frac{1}{6} \). Because the two screw axes are related across a twofold at \( Z = 0 \), the distance along \( Z \) from one atom to its equivalent is \( \frac{1}{3} \). This means that at the Harker plane there are the expected vectors relating one molecule to another within a screw axis but also vectors relating molecules between the two screw axes.

In the lower left hand corner of the Patterson map shown in figure 3.6 there is a constellation of six vector peaks. Three of these peaks are vectors between molecules related by one of the threefold screw axes and the other three are from the other. When the \( x \) and \( y \) coordinates of the heavy atom position lie on the [1\( \bar{1} \)0] plane (the plane is defined in figure 3.5), the six peaks merge into three. In
Figure 3.7: Vector peaks calculated from the heavy atom positions. Blue shows peaks predicted if an atom lay directly on the [110] plane. The gold coloured peaks show vectors for the gold or mercury position. The grey peaks correspond to the platinum vectors. The view is down the Z axis of the Harker plane at Z=\frac{1}{3}. The green arrows show how moving an atom from the [110] plane to the gold or mercury position and then to the platinum position resolves the peaks.

In this special case, the vectors between molecules in one screw axis are the same as the vectors between molecules in the other screw axis. As the coordinates diverge from this plane, the vectors in this six peak constellation move further apart. This is illustrated in figure 3.7. It can be seen in figure 3.6 that the peaks in this constellation are close to merging into three and so it is predicted that the x and y coordinates of the heavy atom will be close to the [110] plane. Thus by superficial examination of the Patterson map, a rough idea of where the heavy atom must be can be deduced.

The position of the mercury atom in this crystal was calculated to be 0.616 0.544 0.835, in fractional coordinates. The calculation was performed by SOLVE (Terwilliger and Berendzen 1999). From the heavy atom position, phased structure factors can be calculated for the contribution of the heavy atom to the X-ray scattering. From the heavy atom structure factors it is possible to estimate phase angles for the protein structure factors. However, a single heavy atom derivative leaves ambiguity in the phase angles. A second, different derivative is required to break this ambiguity.
A CSD crystal was soaked for 26 hours in 0.5mM KAu(CN)2 at 4°C. The soaking resulted in a second isomorphous derivative. However, the gold heavy atom occupied a similar site to that of the mercury in the previously described derivative. The result of this was that the gold derivative was not sufficient to break the phase ambiguity. A 26 hour double soak in 0.5mM KAu(CN)2 + 1mM EMTS gave a similar result.

A third heavy atom derivative was made by soaking CSD crystals in K2PtCl4. The platinum soaking treatment was found to induce non-isomorphism and disorder in the crystals. However, a weak platinum derivative was produced after a 24 hour soak with 1mM K2PtCl4 at 4°C. The platinum position in this derivative was moved relative to the gold and mercury positions. All three heavy atoms were bonded to Cys 310 of the CSD. The movement of the Pt was due to this Cys residue adopting a different rotamer conformation in this derivative (figure 3.8).

A native data set to 1.9Å and the gold, mercury and platinum derivative datasets were entered into the phasing program SOLVE (Terwilliger and Berendzen 1999). This gave rise to an interpretable map after solvent flattening and density modification using the program DM (Cowtan 1994). A model of the CSD was built into this electron density map.

3.9 The CSD structure

The N-terminus of the Swi6 CSD construct starts at amino acid 261 of the full length protein. Thereafter, residues 272 to 302 form a three stranded anti-parallel $\beta$-sheet. A short helical section, which will be referred to as helix 1 extends from 303 to 310. The longest helix, helix 2, packs against the $\beta$-sheet, running perpendicular to it from residues 311 to 322 (figure 3.9a).

Flexibility at the N and C termini of the molecules prevented these from being built. The CSD construct crystallised encompassed residues 261 to 328. Chain A was built from 262 to 322 and B from 266 to 324.
Figure 3.8: Heavy atom positions calculated for the three derivatives used in the structure solution of the CSD. The backbone of the CSD protein is shown as a blue ribbon and some selected sidechains are shown in ball and stick representation. The Cys rotamer bonded to the gold and mercury atoms is shown in ball and stick representation and the bonds of the alternate rotamer bonded to the platinum as thin green lines. The heavy atoms are not drawn to scale with each other or the protein atoms.
Figure 3.9: A) Structure of the CSD, the three β-strands and two α-helices are labelled. B) 3D alignment of CSD (dark grey) and CD (light grey) structures by a least squares superposition.
There are two molecules of Swi6 CSD in the asymmetric unit related by a non-crystallographic twofold axis with spherical polar angles \( \omega \, 92.6^\circ \) and \( \phi \, 347.9^\circ \) (figure 3.10a). The largest non-crystallographic twofold peaks in the self rotation function were at \( \omega \, 90^\circ \) and \( \phi \, 46.6 + n60^\circ \) where \( n \) is an integer (section 3.7). The peaks are close to the non-crystallographic twofold axis seen in the structure.

The CSD molecules were built independently of each other and initial refinement was done without non-crystallographic symmetry restraints. Despite being built independently, the two molecules are nearly identical with a few minor differences that can all be attributed to constraints made by crystal contacts.

The two molecules contact each other through the helical region. Helix 2 is closest to helix 2 of the other molecule at their N-terminal end where they cross each other, at an angle of 35°. Residues at the N-terminus of helix 1 make intermolecular interactions with the residues at the N-terminus of helix 2. The interface centres around hydrophobic associations between Leu 315, Met 314 and the side of Tyr 318 on neighbouring molecules. A structural water molecule is coordinated by the aromatic hydroxyl of Tyr 318 and the side chain amino group of Asn 307 from one molecule. The carboxyl group of Glu 319 from the other molecule occupies the third coordination site (figure 3.10d). These interactions are all reciprocated around the twofold axis. Ser 293 and Arg 281 make asymmetrical bonds through water molecules, between the two CSD monomers (figure 3.10b).

Association between the two asymmetric molecules of CSD in the unit cell, excludes water from a total protein surface area of 1300Å\(^2\). This is a larger area than is usual for nonspecific protein-protein interaction in a crystal. Crystal contact areas fall within the range 200-1200Å\(^2\) (Janin 1995). Protein-protein interactions have an average size of 1600Å\(^2\) (±400Å\(^2\)) (Lo Conte et al 1999). Thus, the contact area between the two CSD molecules is more typical of a true dimer than of a non-specific crystal contact. The interface is largely hydrophobic in nature, but has polar groups around the periphery (figure 3.10c).
Figure 3.10: a) stereo diagram of the CSD dimer viewed along the twofold axis. Secondary structure features are labelled in the left hand view of the stereo pair. b) Detail of the interacting residues. For clarity only those residues involved in interactions and a backbone trace of the helices are shown. A structural water molecule coordinated by two sidechains from one monomer and one from the other is labelled W. c) A surface plot of a CSD monomer shown in the same orientation as the cyan molecule in 'a'. The surface that is excluded from water upon dimerisation is shown coloured according to its chemical nature. Carbon is green, Oxygen is red, Nitrogen is blue. d) Electron density around the dimer interface, Tyr 318 and Asn 307 from one monomer occupy two coordination sites of a structural water molecule, Glu 319 of the other monomer occupies the third. The map coefficients are 2Fo-Fc and the electron density is contoured at 1.5σ.
3.10 Comparison of Swi6 CSD with mouse HP1\(\beta\) CD

Overall the CSD resembles the chromodomain (CD) structure with a root mean square deviation (r.m.s.d) of 1.8Å from the start of the sheet to the end of the helix. Major changes are the extent of helix 1, two complete turns in the CSD structure compared to a single turn with helical nature in the CD structure. Alignment of the \(\beta\)-sheets of each result in the turns of helix 2 being out of phase (figure 3.9b).

The sequence identity between chromodomains and chromo shadow domains is around 20%. At this level it is difficult to accurately align sequences or to put confidence in the alignments. Alignment algorithms such as ClustalW (Thompson \textit{et al} 1994) align chromodomains or chromo shadow domains but do not manage accurate alignments between the families, as judged by structural equivalents. The two structures were superimposed allowing an accurate between-family alignment (figure 3.11). The alignment of CDs with CSDs reveals the residues that are conserved between the families.

Eight amino acid positions were identified that are conserved in both CDs and CSDs (boxed in grey in figure 3.11). The eight residues were all hydrophobic in nature and distributed throughout the sequence. The sidechains of the conserved residues are buried in the interior of the CSD structure. A hydrophobic core is created by the tight packing of these residues (figure 3.12).

3.11 CSD molecules were organised in continuous helices in the crystal.

Electron density in the phased map formed continuous spirals along the Z direction of the crystal (figure 3.13). Two superhelices can be seen in the unit cell, each around a threefold axis. The seventh molecule in each helix is in an equivalent ori-
Figure 3.11: Multiple alignment of CDs and CSDs. Protein names are as standard and species are abbreviated as sp, S. pombe, hs, Homo sapiens, mm, Mus musculus, xl, Xenopus laevis, ce, Caenorhabditis elegans, gg, Gallus gallus. Residues conserved between families are boxed in grey, residues conserved only in the CD family are boxed in red and residues conserved only in the CSD family are boxed in blue. Alignment within the families was carried out by ClustalW and between the families using a least squares algorithm.

Figure 3.12: a) sidechains of residues that are conserved between CD and CSD families are shown in ball and stick representation on the backbone trace of a Swi6 CSD molecule. The two molecules are a stereo pair.
entation to the first but translated one unit cell length along the $Z$ axis. Contacts between the superhelices are between the second and third molecule of the seven molecule repeating unit. Large solvent cavities exist between the non-contacting regions of the helices.

Figure 3.13: Electron density is shown in blue. The outline of the unit cell is shown in red.
3.12 Polyethylene glycol can be seen in the electron density

After building the CSD structure, a continuous feature of unexplained electron density was observed. The density resembled a gently spiraling tube and was still visible at 3σ. The feature was in a hydrophobic region of the structure. The hydrophobic environment and the continuous nature of the density was not consistent with a network of water molecules. The position of the density relative to the N and C terminii of the CSD structure ruled out the possibility that it was an unbuilt part of the protein structure. The density is likely to represent part of a molecule of polyethylene glycol (PEG). PEG is a polyether with repeating structure \((\text{CH}_2\text{CH}_2\text{O})_n\). A commonly used organic precipitant, PEG was used in the crystallisation of the CSD. Although PEG is widely used for crystallisation it has only rarely been described in a crystal structure.

Structures of *E. coli* metalloenzyme peptide deformylase have 10 repeating units of a PEG molecule ordered in the active site. PEG has been shown to be a competitive inhibitor of this enzyme (Becker *et al* 1998). Six repeating units of PEG were also described in the crystal structure of the receiver domain of the *S. meliloti* enzyme FixJ. FixJ is part of the bacteria's nitrogen fixing apparatus. The PEG molecule in this structure is not in an active site. Oxygen atoms of the PEG make hydrogen bonds with the sidechain nitrogen of a lysine residue as well as several neighbouring waters (Gouet *et al* 1999).

The CSD was crystallised in the presence of 20% (w/v) PEG 4000. Such a solution of PEG 4000 is a distribution of sizes with a mean molecular weight of 4000 Da or 90 repeating units. The density in the CSD structure accommodates 11 repeating units of PEG. At either end, the density enters solvent filled cavities in the crystal. Thus there is room to accommodate additional PEG subunits but these, making no contact with the protein, may be too disordered to detect in the electron density map.
The tube-shaped electron density is unbranched, like a molecule of PEG, but the lack of recognisable features raises a problem in correctly defining the register of the PEG model. Two features help in establishing the correct register. A water molecule, hydrogen bonded to Asn 295, is also hydrogen bonded to the PEG molecule at a position that must correspond to an oxygen atom. The PEG molecule crosses a crystallographic twofold axis. PEG, unlike peptides, have internal twofold symmetry around a central carbon to carbon bond. Positioning a C-C bond across the twofold axis brings a PEG oxygen atom within hydrogen bonding distance of the water molecule (2.6Å). The PEG molecules lie along a cleft formed by Trp 269, Trp 293, Ala 297, and Asn 295. As well as forming hydrogen bonds via water to Asn 295, the PEG molecule makes hydrophobic interactions with the other residues described in this cleft (figure 3.14).

The PEG molecule extended along a cleft at the crystal contact site between neighbouring molecules of the superhelix described in section 3.11. The three β-strands of one CSD molecule would form a continuous sheet with the neighbouring molecule if PEG could be considered as an intervening β strand.
Electrospray mass spectrometry was attempted in order to prove the existence of PEG in the crystals. However, the size and viscosity of PEG 4000 prevented it entering vapor phase when a solution of PEG 4000 was introduced into the flight tube of an electron ionisation mass spectrometer by micro-injection.

3.13 Conclusions

The two crystallographically unrelated molecules of CSD in the asymmetric unit associate to form a dimer in the CSD crystals. These dimers associate via the edge of their $\beta$-sheets with other dimers to form large superhelices that run through the crystal. A molecule of PEG at the crystal contacts occupies a position between the neighbouring $\beta$-sheets. Occasionally the packing of a protein in a crystal mimics associations of the protein in vivo. It was hypothesised that the interesting packing observed in the CSD crystals could be mimicking a biological property of the CSD.

Swi6 is known to spread along chromatin and it was tempting to speculate that the spirals observed in the crystal were reproducing the mechanism by which Swi6 organises itself along chromatin in vivo. In this model the PEG molecule could be mimicking an extended peptide such as a histone tail.

If this model were correct, the crystal contact sites making up the helix would be expected to be conserved between HP1 homologues. However, there is no obvious conservation at the contact sites.

CDs and CSDs have limited sequence homology and divergent functions. Despite obvious differences the domains were predicted to have the same fold. The crystal structure of the CSD of *S. pombe* Swi6 protein, reported here to a resolution of 1.9Å, confirms that CSDs have a fold similar to CDs.

The structure of Swi6 CSD as a representative of the CSD family allows sequence alignment between CSDs and CDs. Alignment between the families identifies residues that are conserved throughout the chromodomain superfamily. It is interesting that these residues are all in the hydrophobic core of the domain.
Interactions with a protein occur at the surface and are mediated by surface amino acids. The lack of conservation of surface amino acids between CDs and CSDs reflects their distinct functions. Thus the chromodomain fold, set up by the core conserved residues, can be seen as a scaffolding that correctly orientates the surface residues that take part in protein interactions.

The large surface area of the CSD interaction in the crystals suggests that the Swi6 CSD may be a dimer in solution as well as in the crystal. Dimerisation of CSDs would represent a major difference when compared to monomeric CDs. The structure of the proposed dimer interface could be used to test dimerisation in vitro and in vivo.
Chapter 4

Dimerisation of Chromo Shadow Domains.

4.1 Introduction

In the Swi6 CSD crystal there are two asymmetric CSD molecules per unit cell. These have a contact area of 1300 Å², approximately twice that of the other protein-protein contacts in this crystal. The contact is hydrophobic in the center and polar around the periphery. This is a common feature of protein-protein interfaces (Lo Conte et al 1999). These features of the interaction suggest that the intimate contact between the CSD molecules in the crystal may represent a true mode of dimerisation in solution.

Several experiments were designed to test the dimerisation of the Swi6 CSD both in vitro and in vivo and to confirm that the interface seen in the crystal is the true dimerisation interaction.
4.2 SELDI mass spectrometry

SELDI analysis was performed by Kathryn Vardy of Ciphergen Biosystems, Inc.

4.2.1 Principle of SELDI mass spectrometry

Surface-Enhanced Laser Desorption/Ionization (SELDI) is an adaption of matrix-assisted laser desorption (MALDI) mass spectrometry. SELDI is useful for identifying species that interact with a target molecule.

In this procedure, a target molecule is covalently cross-linked to a biochip. This is commercially called a ProteinChip®. Various ProteinChips® are available with a variety of surface chemistries that can be chosen to optimise the binding of a bait molecule. When bound, the molecule is covalently cross-linked to the biochip. A chemical modification blocks the surface of the chip to prevent any further binding. The biochip is then incubated with a target. The target can be a single molecule or a complex mixture such as a cell lysate. Subsequently, the biochip is washed to remove all unbound material (figure 4.1).

Molecules associated with the target, bound to the ProteinChip®, can be analysed by SELDI mass spectrometry.

In final preparation, the ProteinChip® is coated in energy absorbing molecules (EAM). The prepared ProteinChip® is transferred to a SELDI time of flight (TOF) mass spectrometer. A weak pulsing laser excites the EAM causing embedded molecules to be ionised and desorbed. Only molecules that are not covalently bound to the biochip enter the flight tube of the mass spectrometer. In this way the masses of molecules that interact with a target can be analysed.

---

1 SELDI technology is patented by Ciphergen Biosystems, Inc. and ProteinChip® is a registered trademark of this corporation. http://www.ciphergen.com/.
4.2.2 Details of SELDI experiment

The CSD of Swi6 was expressed in *E.coli* with an N-terminal glutathione S-transferase (GST) affinity tag. The fusion protein (GST-CSD) was purified in a single chromatography step using glutathione agarose as an affinity resin. Another advantage of using this fusion protein is that it has a large mass difference when compared to CSD alone. GST-CSD is ~35KDa compared to ~8KDa for CSD alone. The mass difference allows easy resolution of the two proteins by mass spectrometry or denaturing gel electrophoresis.

In this SELDI experiment GST-CSD is the bait protein. GST without a CSD fusion was also expressed in *E.coli*, purified and used as a negative control bait protein.

CSD was over-expressed in *E.coli* and a soluble lysate prepared from these cells. This lysate, including *E.coli* proteins and recombinant CSD, is the target. A mutant form of CSD was produced in which leucine 315, a key dimer interface residue, was mutated to aspartate. This protein was found to be completely insoluble (the implications of this will be discussed later). The mutant CSD was spun down by centrifugation and was not found in the soluble fraction. The sol-
Figure 4.2: Peaks from the output of a SELDI mass spectrometry experiment. Peak height is represented as greyscale. The top three lanes are the SELDI output from the bait/target combinations indicated in the left margin. The lower two lanes are the total protein from the CSD lysate (CSD+) and the negative lysate (CSD-). Molecular weights in Da are marked along the bottom. CSD has a predicted molecular weight of 8060 Da. A peak assumed to be CSD is labelled.

uble fraction from this *E. coli* lysate did not contain CSD but had been treated in exactly the same way as the lysate containing the native CSD. This lysate was used as a negative control target in this assay.

The bait proteins, GST-CSD and GST, were covalently bound to a ProteinChip®. The chip was incubated with either the target lysate containing CSD or the lysate depleted of the mutant CSD. Unbound protein was washed from the biochips. Bound proteins were analysed by SELDI mass spectrometry (figure 4.2).

A peak with the same molecular weight as Swi6 CSD was seen bound to GST-CSD but not to GST alone. This peak was absent when the biochip-bound GST-CSD was incubated with lysate deficient in CSD.

The SELDI results are consistent with a hypothesis that Swi6 CSD associates with GST-CSD with higher affinity than GST alone. The result suggests the the CSD can self-associate. The technique does not give any indication of stoichiometry or that association is via the interface seen in the crystal structure. The SELDI experiment was a ‘one off’ as the machine was only available for a limited time. These experiments could therefore not be repeated varying stringency or injecting
purified CSD into the flight chamber of the instrument as a control. However, the results of this experiment are consistent with the other results detailed in this chapter.

A GST pulldown assay was optimised to test for self-association. Although this alternative technique still does not test stoichiometry, it is more stringent. A positive result by GST pulldown will indicate that the self-association is relatively tight.

4.3 GST pulldown assay

4.3.1 Principles of GST pulldown assay

The dimerisation assay was optimised and carried out with the help of Dr. Janet Partridge.

Glutathione S-transferase (GST) is a 26 KDa dimeric enzyme from the trematode worm *Schistosoma japonicum*. Recombinant GST expressed in *E.coli* maintains its enzyme activity and dimerises. The *E.coli* expression vector pGEX allows expression of an open reading frame with an N-terminal GST fusion. GST fusion proteins bind with high affinity to glutathione Sepharose beads allowing affinity purification (Guan and Dixon 1991).

A bait GST fusion protein is incubated with an untagged target protein. After incubation, glutathione Sepharose beads are added to the protein mixture. The fusion protein binds to the beads through the GST moiety. Target proteins that associate with the fusion protein will also be immobilised on the beads. Unbound proteins are removed by repeated centrifugation and resuspension of the beads in a wash buffer. Proteins that are retained can be detected by gel electrophoresis. The stringency of this assay can be controlled by varying the salt or detergent concentration of the wash buffer.
A target protein can bind to the beads directly, through the GST protein or to the protein of interest that is fused to GST. In order to establish that binding is to the protein of interest, GST alone is used as a negative control.

### 4.3.2 Details of GST pulldown assay

A GST pulldown assay was designed to determine if the CSD of Swi6 could self-associate.

The bait proteins used in this assay were GST-CSD, GST-L315D and, as a negative control, GST. The L315D mutant form of the CSD has been discussed in section 4.2.2. This mutant form of the domain was found to be insoluble. However, a GST fusion construct of the L315D mutant was found to be soluble. These were incubated with one of two target lysates, an *E.coli* lysate containing untagged CSD or a lysate from which the insoluble mutant CSD L315D had been depleted by centrifugation.

It is known that GST can dimerise and it is suspected that the CSD can also dimerise. Upon mixing of the GST-CSD and the CSD two species were predicted, GST-CSD dimers and CSD dimers. During incubation it was expected that exchange would take place such that three species would be present at the end of the assay, GST-CSD dimers, CSD dimers and a new species GST-CSD/CSD heterodimers. This is illustrated in figure 4.3.

The formation of a GST-CSD/CSD heterodimer would be demonstrated by the presence of untagged CSD bound to glutathione beads. However, untagged CSD was not detected by SDS PAGE and Coomassie staining in the bead bound fraction. Two hypotheses are suggested to explain this: first the CSD does not self-associate or, second, the CSD does dimerise but exchange does not take place. The GST pulldown assay was refined to test between these possible hypotheses.

GST-CSD, GST-L315D or GST was mixed with *E.coli* lysates with or without untagged CSD, as before. The proteins in the mixtures were then denatured in 6M urea. This process denatures the proteins and separates any pre-formed
Figure 4.3: *Schematic showing some theoretical species formed by dimer exchange between GST-CSD homodimers and CSD homodimers upon incubation.*

dimers. The urea was dialysed away in a stepwise fashion allowing the proteins to refold and dimers to reform. Glutathione Sepharose beads were incubated with the refolded mixtures as before and washed.

Figure 4.4 shows the results of this experiment. A band of the expected CSD size is seen in the bead bound fraction of the GST-CSD + CSD lane. The band is not detected in the GST negative control lane indicating that is specifically interacts with the CSD moiety. The band is also not detected in the GST-L315D + CSD lysate lane. This is evidence that the CSD dimerisation is through the interface seen in the crystal structure. The band is also absent in the CSD negative lysate indicating that the band is CSD, this was confirmed by Western blotting with Swi6 antibodies (data not shown).

Urea mediated unfolding of a mixture of GST-CSD and CSD followed by refolding allows the formation of GST-CSD/CSD multimers. GST/CSD multimers were not detected in the negative controls. The results of these experiments indicate that Swi6 CSDs are able to self-associate tightly enough to be detected in this stringent assay. The inability of the GST-L315D mutant CSD fusion protein
Figure 4.4: SDS PAGE gel resulting from a pulldown assay. Molecular weights are indicated on the left in KDa. Identity of the bands are indicated on the right. Lanes contain, from left to right, molecular weight markers, cell lysate without CSD, cell lysate with CSD, bead immobilised proteins from GST plus CSD lysate mixture, bead immobilised proteins from GST-CSD plus negative lysate mixture, bead immobilised proteins from GST-CSD plus CSD lysate mixture, bead immobilised proteins from GST-L315D plus CSD lysate mixture. Below the coomassie stained gel is a western blot, stained with affinity purified, anti-Swi6 antibodies.
to interact with wildtype CSD suggests that interaction may be through the same interface as that seen in the crystal structure.

4.4 Gel exclusion chromatography

A sample of purified, untagged CSD was analysed by size exclusion chromatography. This experiment is detailed in section 6.4. The mass of the CSD monomer is calculated to be 8.06 KDa and the dimer twice this. From the calibration curve of the column, monomeric CSD would be expected to elute with a $K_{AV}$ of 0.47 (retention volume = 74.8 mls) and dimeric CSD to elute with a $K_{AV}$ of 0.356 (retention volume = 64.4 mls). In fact, the CSD was eluted from the column with a $K_{AV}$ of 0.31 (retention volume = 61 mls) and an extrapolated molecular weight of 21.4 KDa. Within the error margins of the column this weight suggests that the CSD is dimeric in solution.

4.5 Phenotypic characterisation of L315D mutation in *S. pombe*

*All work in fission yeast was carried out in collaboration with Dr. Janet Partridge.*

In the crystal structure, the CSD dimer interface centers around a leucine residue, number 315 (see figure 3.10b). This residue was changed to an aspartate by site directed mutagenesis in the genomic *Swi6* gene of *S. pombe*. The effect of this mutation on *Swi6* localisation and centromeric silencing was assayed. Starting from the assumptions that the CSD dimerises *in vivo* and that the L315D mutation disrupts dimerisation, the phenotype of this mutation may give clues about the function of CSD dimerisation.

*Swi6* can be visualised in distinct nuclear foci when wild type *S. pombe* are stained with anti-*Swi6* antibodies (α-*Swi6*). These foci do not show up above background staining in *Swi6* L315D mutant cells (figure 4.5).
Deletion of genomic Swi6 from *S. pombe* causes a reduction in gene silencing in the outer repeat regions of the centromere (Ekwall *et al* 1999). Centromeric silencing is monitored by inserting an ade6 gene into the outer repeats of the centromere of chromosome 1. In wild type cells the centromere is heterochromatinic and the gene is silent. Silencing of the ade6 gene causes a build up of a red coloured metabolite that turns the *S. pombe* red when grown on limiting adenine media. In Swi6 null cells silencing is alleviated at the centromere allowing expression of the ade6 gene. These cells are white.

Mutation of Swi6 L315D in *S. pombe* allows expression of a centromeric ade6 gene. This phenotype is similar to Swi6 knockout (figure 4.5).
Mutation of a dimer interface residue *in vivo* leads to loss of Swi6 localisation and loss of centromeric silencing. The results imply that dimerisation may be an important function of the CSD *in vivo*.

### 4.6 Dimerisation is likely to be a general property of CSDs.

The CSD of Swi6 has been shown to dimerise *in vitro* in the previous experiments. Mutagenesis of a dimer interface residue suggests that the dimerisation may be important for Swi6 function in fission yeast. The CSDs of human HP1α and HP1γ have been shown to self interact by yeast two hybrid assay and *in vitro* pulldown (Ye *et al* 1997). It is likely that dimerisation is a common property of CSDs.

Those residues that are conserved throughout CSDs but not in CDs as seen in a multiple alignment (blue boxed residues in figure 3.11) are shown highlighted in the structure of the Swi6 CSD (figure 4.6). These conserved residues cluster around the site of the dimer interface. This clustering of conserved residues supports the evidence presented in this chapter that the CSD of Swi6 dimerises in solution by the same dimer interface as seen in the crystal structure. Moreover, the conservation of interface residues between CSDs suggests that the interface characterised in the Swi6 CSD structure is relevant to all CSDs.

### 4.7 Conclusions

Two non-crystallographically related molecules of the Swi6 CSD were observed to closely associate in the crystal structure (section 3.9). The extent and hydrophobic nature was more typical of a true dimer interface than of a crystal contact. The experiments outlined in this chapter were designed to test this.

The CSD was shown to associate in solution by two alternative techniques, SELDI mass spectrometry and GST pulldown. Direct evidence for CSD dimeri-
Figure 4.6: *Residues conserved within the CSD family, but not between the CSD and CD families, were coloured blue on a molecular surface of a CSD monomer. The other monomer is shown in magenta as a backbone trace. The dimer is oriented in a similar way to the dimer shown in figure 3.10a.*

sation was shown by size exclusion chromatography. A mutation in the dimer interface disrupted Swi6 function in *S. pombe*. This indicates both that dimerisation may occur *in vivo* and that the interface seen in the crystal may be the true mode of dimerisation.

The residues that make contact between the monomers in the crystal structure are conserved throughout the CSD family. This implies that the monomer interactions of the Swi6 CSD may be common to all CSDs.

After this work was completed an NMR structure of the CSD of mouse HP1β was published (Brasher *et al* 2000). The CSD in this solution structure was also a dimer and the interface was very similar to that seen in the crystal structure.

Several point mutations of surface residues of the Swi6 CSD have been made during the course of this study. Mutations include, Phe 324 to Ala, Thr 323 to Asp and Leu 315 to Glu. Of these, only the Leu 315 dimer interface mutation was insoluble. It is possible that dimerisation is essential for the solubility and correct folding of the CSD. GST proteins of the alpha/mu/pi and *S. japonicum*
Figure 4.7: The dimer interface of the Swi6 CSD is compared to the interface of the *S. japonicum* GST dimer interface. The Phe residue of the GST lock and key mechanism and Leu 315 of the CSD are shown with the bonds represented as sticks. The Ca backbone of the top molecule of each dimer is shown as a trace and the lower molecule of each as a solvent accessilbe surface. The surfaces are coloured by electrostatic potential.

classes similarly rely on dimerisation for correct folding and solubility (Kaplan et al 1997). There are features of the GST dimer interface that are similar to the CSD dimer interface.

GST molecules dimerise via a hydrophobic interface. A central feature of this interface is a phenylalanine residue that rests in a hydrophobic pocket on the surface of the opposing monomer. The phenylalanine residue essentially makes up part of the hydrophobic core of the other monomer. This means that the monomers are stable only as dimers. The Phe residue and the hydrophobic pocket are known as the lock and key (Sayed et al 1999).

The dimer interface of the CSDs is also hydrophobic. In the center of the interface leucine 315 makes contacts with hydrophobic core residues of the opposing monomer. It is possible that the Leu residue of the CSD may be analagous to the Phe residue in the lock and key mechanism of the GST dimer. The interfaces of the CSD and *S. japonicum* GST dimers are shown in figure 4.7.

It was found that when GST-CSD dimers were mixed with CSD dimers ex-
change did not take place without urea denaturation. This result suggests that the CSD dimers are very tight. However, the inability of GST-CSD dimers to exchange as studied in section 4.3.1 will be affected by the dimerisation of GST. Two molecules of GST-CSD are likely to associate with each other through both the CSD and the GST moiety. Transient dissociation of either of these domains would not result in dissociation of the two molecules. The dissociated dimer is likely to reassociate with its own partner as this is held in a local high concentration by the physical association at the other end of the molecule. This effect makes the dimerisation of CSD to CSD in GST fusion proteins appear higher than it actually is. In retrospect, it may have been better to use a monomeric fusion tag (such as a hexahistidine tag) for these experiments.

Dimerisation may be an important property of CSDs in vivo. The insolubility of the bacterially expressed CSD monomers makes the importance of dimerisation difficult to assess. However, CSDs do not simply function as dimerisation domains. They have been shown to associate with numerous other proteins as shown in table 5.1. The protein-protein interaction function of the CSD will be investigated in the following chapter.
Chapter 5

CSD dimer interactions

5.1 Introduction

HP1 homologues have been shown to bind a number of proteins. A literature review of these interactors is shown in table 5.1, section 1.4.3. The CD of HP1 has been implicated in binding histone H3 (Bannister et al 2001, Lachner et al 2001, Nielsen et al 2001). In contrast to the CD, the CSD has been reported to bind numerous partners (table 5.1).

A random phage display experiment identified peptides that bound to immobilised mouse HP1α CSD (Smothers and Henikoff 2000). Many of the peptides contained the five amino acid consensus sequence shown in figure 1.5. The pentapeptide largely had a proline in position one, a valine in position three and hydrophobic residues in positions four and five (Smothers and Henikoff 2000). The consensus pentamer has been called the MOD1 (mouse HP1β) interacting region (MIR). Many, but not all, of the CSD interacting proteins shown in table 5.1 contain a sequence similar to the consensus.

The experiments described in this chapter, were designed to further characterise the binding of the CSD to other proteins. The surface topology of the Swi6 CSD was explored for a site commensurate with binding an unstructured pentapeptide. A deep cleft was found between the ends of helices 2 where the two CSDs meet in the dimer. A residue situated at the edge of the cleft was mutated in S.pombe and the phenotype observed. GST pulldown experiments were attempted
<table>
<thead>
<tr>
<th>HP1 interacting protein</th>
<th>Function</th>
<th>HP1 protein/region</th>
<th>Experimental method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. pombe DNA polymerase a</td>
<td>DNA replication</td>
<td>S. pombe Swi6/ unmapped</td>
<td>in vivo; mutations in pola disrupt Swi6 localization in vitro; GST pulldown</td>
<td>Nakayama 2001a</td>
</tr>
<tr>
<td>S. pombe Rad21</td>
<td>cohesin</td>
<td>S. pombe Swi6/ unmapped</td>
<td>genetic interaction</td>
<td>unpublished</td>
</tr>
<tr>
<td>Mouse TIF1α TIF1β</td>
<td>mediate nuclear receptor functions</td>
<td>mouse HP1α and HP1β/ CSD</td>
<td>yeast two hybrid immunoprecipitation GST pulldown</td>
<td>Dourin 1996 Nielsen 1999</td>
</tr>
<tr>
<td>human lamin B receptor</td>
<td>nuclear membrane bound receptor</td>
<td>human HP1α/ CSD</td>
<td>in vivo; yeast two hybrid in vitro; GST pulldown</td>
<td>Ye 1997</td>
</tr>
<tr>
<td>human S100</td>
<td>component of nuclear bodies</td>
<td>human HP1α HP1β, HP1γ/ CSD</td>
<td>in vivo; yeast two hybrid in vitro; GST pulldown</td>
<td>Seeler 1998</td>
</tr>
<tr>
<td>Drosophila ORC components</td>
<td>DNA replication origin recognition complex</td>
<td>Xenopus Drosophila both CD and CSD</td>
<td>immunoprecipitation</td>
<td>Pak 1997</td>
</tr>
<tr>
<td>methylated histone H3 tail peptides</td>
<td>core nucleosome component</td>
<td>mouse HP1α,β,γ/ CDS</td>
<td>in vitro; pulldown</td>
<td>Laechner 2001 Bannister 2001</td>
</tr>
<tr>
<td>human Pim-1</td>
<td>kinase</td>
<td>human HP1α/ CSD</td>
<td>yeast two hybrid</td>
<td>Koike 2000</td>
</tr>
<tr>
<td>mouse core fold of histone H3</td>
<td>core nucleosome component</td>
<td>mouse HP1α,β,γ/ CDS</td>
<td>immunoprecipitation GST pulldown</td>
<td>Nielsen 2001</td>
</tr>
<tr>
<td>human Ku70</td>
<td>DNA repair telomere maintenance</td>
<td>human HP1α/ CSD</td>
<td>in vivo; yeast two hybrid in vitro; GST pulldown</td>
<td>Song 2001</td>
</tr>
<tr>
<td>p150 subunit of mouse CAF1</td>
<td>nucleosome assembly into chromatin</td>
<td>mouse HP1βγ/ CSD</td>
<td>in vivo; yeast two hybrid in vitro; GST pulldown</td>
<td>Martínez 1999</td>
</tr>
<tr>
<td>S. pombe Swi2</td>
<td>silencing at mating type loci</td>
<td>S. pombe Swi6</td>
<td>yeast two hybrid</td>
<td>unpublished</td>
</tr>
<tr>
<td>human INCENP</td>
<td>inner centromere binding protein</td>
<td>human HP1α hinge</td>
<td>yeast two hybrid GST pulldown</td>
<td>Ainsztein 1998</td>
</tr>
</tbody>
</table>

Table 5.1: Literature review of proteins interacting with HP1 homologues.
to study *in vitro* binding between CSD and a number of the proteins shown in table 5.1.

5.2 Dimerisation of Swi6 CSD creates a potential binding site.

As discussed in section 3.9, Swi6 CSD dimerises via interactions between helices 1 and 2. The major helices (helix 2) from associated monomers are closest to each other at their N terminal end and diverge towards the C terminus. This divergence creates a cleft at their C terminal end. The cleft is approximately 15Å long, 5Å wide and runs across the dimer axis. The bottom of the cleft is formed by the edges of Tyr 318 and Phe 317 from each monomer; the edges of the cleft are formed by residues 321-324, His, Leu, Thr and Phe respectively; the ends of the cleft are delimited by Ile 279 (figures 5.1 and 5.3).

A non-crystallographic twofold symmetry axis runs through the centre of the cleft. As a peptide is a non-symmetrical molecule it may be that peptide binding in this cleft induces asymmetry in the CSD dimer. However, some traces of cleft asymmetry can be seen in the structure. Phe 324 from one molecule is stabilised by interaction with the hydrophobic portion of Arg 281 from the other. This interaction is not reciprocated across the non-crystallographic twofold and therefore cleft residues Phe 324 and Thr 323 can only be seen in one half of the dimer. This asymmetry may be due to crystallographic packing constraints (though none is obvious) or may be a true feature of the dimer.

The role of the cleft in CSD interactions was confirmed by an NMR structure of the CSD of mouse HP1β (Brasher *et al* 2000). A detailed comparison of the NMR structure with the crystal structure of the Swi6 CSD will be made in the discussion chapter. After peak assignment and modelling, the HP1β CSD was mixed with the CSD binding region of chromatin assembly factor subunit p150 that has been shown to bind to the CSD of HP1β (Murzina *et al* 1999). The shifts corresponding to residues around the cleft were strongly perturbed indicating that these residues
Figure 5.1: *Swi6 CSD dimer has been rotated 90° around an X axis compared to the dimer view in figure 3.10a.* a) A molecular surface has been drawn over the dimer. The cleft can be seen running horizontally over the surface. A pentameric peptide has been fitted on the structure, the fit serves to highlight the cleft and give a sense of scale, it is not a formal model. b) A backbone trace illustrating the orientation of the molecule in 'a'. Cleft residues are shown in a ball and stick representation and labelled.
are implicated in binding the CAF peptide. Gel filtration experiments show the HP1\(\beta\) CSD bound the peptide with a stoichiometry of 2:1. This is predicted if the peptide binds to a site across the two monomers (Brasher et al 2000).

5.3 Mutation of cleft residue Phe 324 in \textit{S.pombe}.

\textit{All work in fission yeast was carried out in collaboration with Dr. Janet Partridge.}

It was hypothesised that mutations in cleft edge residues of Swi6 CSD would not affect Swi6 localisation but would prevent recruitment of heterochromatin proteins and alleviate centromeric silencing. Residue Phe 324 was mutated to an Ala in \textit{S.pombe} genomic \textit{Swi6}\(^+\). To achieve this the genomic Swi6 gene was replaced with a \textit{ura4}\(^+\) gene by homologous recombination. This was in turn replaced by a mutated \textit{swi6} gene. The Phe 324 residue is a few amino acids from the C terminus of the Swi6 protein and site directed mutagenesis was carried out with a mutagenic non-coding PCR primer. The localisation of the mutant Swi6 was investigated by indirect, fluorescent staining with Swi6 antibody. Centromeric silencing was assessed as described in section 4.5.

Localisation of Swi6 to heterochromatin is mediated by the chromodomain and therefore mutation of residues around the CSD cleft would not be hypothesised to disrupt localisation. Indeed, F324A mutation in Swi6 does not affect its localisation in fission yeast nuclei (figure 5.2a and b).

An interesting result was noted in the centromeric silencing assay (figure 5.2c and d). Swi6 null cells lose centromeric silencing and express an \textit{ade6}\(^+\) gene inserted there. These cells are white. In contrast, Swi6 wild type cells have strong centromeric silencing and are completely red. \textit{S.pombe} with an F324A mutation in Swi6 have variegating colour. Roughly 50% of the colonies appeared red and 50% appear white. Some of the colonies retain centromeric silencing and some lose it. If a red colony is streaked out on fresh media the progeny also variegate red and white. The implications of this variegation will be discussed at the end of this chapter.
Figure 5.2: Mutation of F324A in fission yeast genomic Swi6+. a) Immunofluorescence of Swi6 in wild type S. pombe. DNA is stained in blue with DAPI. Swi6 is stained green with α-Swi6. b) Immunofluorescence of Swi6 F324A in S. pombe. Staining is similar to ‘a’. c) Swi6 null cells grown on nutrient agar with limiting adenine. Wild type S. pombe are streaked in the top right of the plate. d) Swi6 F324A mutant S. pombe on ade+ nutrient agar. Wild type cells streaked as in ‘c’.
It had been hoped that the loss of silencing phenotype of the cleft mutation would facilitate a genetic, over expression screen looking for cleft interacting proteins. The rational for this screen is that proteins whose interaction with the CSD has been attenuated by the F324A mutation can be forced to reassociate by overexpression. In this screen, cells of the Swi6 mutant background are transformed with a library of vectors that drive over expression of different cDNAs. Colonies are selected in which the white, loss of centromeric silencing phenotype reverts to the red wild type phenotype. Sequencing of the cDNA vectors from these colonies identifies proteins that may interact with the CSD cleft. However, such a screen is not possible in this case because the cells revert from the white loss of centromeric silencing phenotype to the red wild type phenotype and back again without overexpression from any cDNA vector.

5.4  Putative role of phosphorylation at the binding cleft.

Orientation of the cleft residues of the Swi6 CSD crystal structure and the NMR structure of HP1β CSD (Brasher et al 2000) are quite different. Residue F324 in the Swi6 structure points up and out of the cleft into solution whilst its equivalent in the HP1β structure (W170) sits inside the cleft making hydrophobic contacts with A125. In Swi6, residue T323, situated in the cleft wall, is angled away from the twofold axis leaving the Swi6 cleft open. In the HP1β structure the equivalent residue, T169, is angled towards the twofold axis and covers the centre portion of the cleft (figure 5.3).

The deep cleft seen in the Swi6 CSD structure is completely closed in the HP1β CSD structure. This could simply be a difference between the two proteins but it was hypothesised that both domains could adopt an open or a closed conformation. The open cleft of the Swi6 structure may be stabilised by crystal packing as opposed to the closed cleft solution structure. Opening and closing the cleft may
Figure 5.3: The two figures show the backbone traces of Swi6 CSD crystal structure (left hand image) and mouse HP1β (right hand image). Both structures are oriented in the same way. Important cleft forming residues are shown as ball and stick representations.

be an important regulatory mechanism of CSDs. There are several observations to support this hypothesis.

Alignment of HP1 sequences show that when the position equivalent to Swi6 F324 is a Trp, the position equivalent to I279 is an Ala. When the F324 position is a smaller residue such as a Phe, the I279 position is a larger residue such as an Ile or a Val (figure 3.11). The smaller residues such as Ala are needed to make room for the bulky Trp residue within the cleft. It is possible that the larger I279 position residues are required to make close contact with the smaller F324 position residues implying that these spend at least part of their time within the cleft.

Swi6 T323 is conserved throughout CSDs as a serine or a threonine and is equivalent to serine 199 Drosophila HP1. It has been reported that serine 199 of Drosophila HP1 is phosphorylated in vivo. Mutation of this serine to alanine reduced the ability of a β-galactosidase fused HP1 to localise to heterochromatin (Zhao and Eissenberg 1999). Phosphorylation of this residue in the closed cleft state would place two negatively charged residues close to each other and tend to push the cleft towards the open conformation (figure 5.3).
Finally, large perturbations of W170 shifts were observed in the NMR spectra of mouse HP1β when CAF p150 peptide was added (Brasher et al 2000). This suggests that there is a large change in the environment of W170 upon peptide binding as expected for a shift of this residue out of the cleft.

To attempt to mimic phosphorylation, threonine 323 of Swi6 CSD was mutated to an aspartic acid. The mutant protein was expressed in *E.coli* and a strategy for purification was developed. The crude bacterial lysate containing the T323D protein was loaded onto a DEAE Sepharose column in 25 mM BisTris pH 6.5, 5 mM EDTA, 50 mM NaCl). After elution of unbound proteins the CSD was eluted at approximately 300 mM NaCl during a salt gradient from 25 to 500 mM NaCl. The CSD containing fractions were concentrated to 1 ml and loaded onto a 120 ml high resolution 16/60 Sephacryl™S-200 size exclusion column (Amersham Pharmacia Biotech) in 25 mM BisTris pH 6.5, 5 mM EDTA, 300 mM NaCl. The mutant CSD was ~95% pure after this purification step and was used along with the wild type CSD in the GST pulldown assays.

5.5 GST pulldowns with recombinant CSD interacting proteins.

To study the interaction between CSD and other proteins GST pulldown experiments were set up. Several proteins from table 5.1 were cloned for recombinant expression in *E.coli*. These proteins included, the large p150 subunit of mouse chromatin assembly factor, *S.pombe* DNA polymerase α, *S.pombe* cohesin Rad21, *S.pombe* Swi2 protein and *S.pombe* Clr4 protein.

5.5.1 Chromatin Assembly Factor, p150 subunit.

Chromatin assembly factor (CAF) is a protein complex comprised of three polypeptide subunits. CAF deposits free histones into forming chromatin fibres and is conserved from yeast to man (Kaufman et al 1995). The large subunit of the
mouse CAF complex (p150) has been shown to bind to mouse HP1β and HP1γ \textit{in vitro} and \textit{in vivo}. An N-terminal sequence of CAF p150, containing a consensus motif like that discussed above, binds the HP1 proteins (Murzina \textit{et al} 1999). \textit{S.pombe} have a CAF complex with a subunit similar to mouse p150 but this does not contain the N-terminal sequence with the consensus pentapeptide.

Amino acids 204 to 268 of mouse CAF p150, containing the HP1 interacting region, was overexpressed in \textit{E.coli} fused to a ten histidine tag for nickel affinity purification. The construct used to express this CAF protein was made by N. Murzina (1999) and obtained from A. Verreault (ICRF, London). The recombinant p150 polypeptide did not bind to nickel resin except when denatured in 6M guanidine. This implies that the histidine tag interacts with the protein preventing nickel binding. Renaturation of the polypeptide was problematic and so unpurified CAF p150 peptide was used in the pulldown assay.

GST, GST-CSD or GST-T323D (mutant CSD mimicking phosphorylation of T323) were mixed in equivalent amounts with an \textit{E.coli} cell lysate containing the recombinant p150 polypeptide. After incubation, glutathione agarose bead were added and unbound proteins washed from the beads. The beads were boiled in SDS PAGE loading buffer and run on a 15\% (w/v) acrylamide denaturing gel (figure 5.4 lanes 1-4). There was no enhancement of CAF p150 peptide binding to either GST-CSD or GST-T323D when compared to the GST negative control.

Mouse CAF p150 residues 204 to 268 do not bind Swi6 CSD in the conditions studied. It is possible that while the CAF polypeptide binds to mouse HP1β and HP1γ it does not bind to Swi6 CSD. Misfolding of the CAF polypeptide may also be a factor in the failure of this assay in detecting binding.

5.5.2 \textit{S.pombe} DNA polymerase α

Mutations in \textit{S.pombe} DNA polymerase α (polα) have been shown to affect Swi6 localisation and heterochromatic silencing \textit{in vivo} (Nakayama \textit{at al} 2001a). GST tagged Swi6 interacts with \textit{in vitro} translated polα \textit{in vitro} (Nakayama \textit{at al}
Figure 5.4: SDS PAGE gels from in vitro binding assays with various proteins recombinantly expressed in E. coli. Target proteins are incubated with fusion tagged bait proteins. The bait proteins are bound to affinity resins. Unbound proteins are washed away. Lanes contain: 1) *E. coli* lysate containing amino acids 204-268 of mouse CAF p150; 2) Target: lysate in lane 1 Bait: GST-CSD; 3) Target: lysate in lane 1 Bait GST-T323D; 4) Target: lysate in lane 1 Bait GST; 5) *E. coli* lysate containing amino acids 1057-1225 of polα; 6) Target: lysate in lane 5 Bait GST-CSD; 7) Target: lysate in lane 5 Bait GST; 8) Target: CSD Bait: GST-CSD; 9) Target: CSD Bait: GST; 10) Target: T323D Bait: GST; 11) Target: CSD Bait: GST-Rad21; 12) Target: T323D Bait: GST; 13) Molecular weight markers with sizes indicated.
2001a). The role of Swi6 polα binding in vivo is not clear. It has been speculated that polα may recruit and maintain Swi6 at heterochromatic loci.

Polα from S.pombe contains a MIR like peptide near the C-terminus of the protein. There is no evidence to suggest that Swi6 CSD binds to this region of polα but it is tempting to speculate that it does. A crystal structure of a DNA polymerase from Archaeon pyrococcus has been solved (Hashimoto et al 2001) that is highly homologous to polα. By mapping the polα MIR region onto this structure it was noted that the pentapeptide was in an α-helical conformation with the conserved position three valine pointing inward to the proteins core.

Polα is a 1405 amino acid protein. Because of its size it was decided to use a subdomain of the protein in the pulldown assay. DNA polymerases have three domains, which have been called the thumb, palm and fingers domains. The enzyme is shaped like an open hand and closes around DNA like a clenched fist. The thumb domain of polα (residues 1057 to 1225) were expressed in E.coli. This fragment contains the putative MIR peptide sequence (residues 1129 to 1133). Unpurified polα thumb was used in a GST pulldown assay with GST and GST-CSD (figure 5.4 lanes 5-7).

The thumb domain of polα was not found to bind Swi6 CSD under these conditions. It is possible that the MIR-like region of polα is not the region that binds the CSD or that post translational modification of either Swi6 or polα or both is required for interaction to occur.

5.5.3 S.pombe Rad21 protein

Rad21 is a 628 amino acid protein that was initially identified as being important for DNA double stranded break repair (Birkenbihl and Subramani 1995). Rad21 has since been identified as part of the mitotic cohesin complex, which holds sister chromatids together prior to the onset of anaphase. Unpublished results from the laboratory of JP. Javerzat suggest that there is an interaction between Rad21 and Swi6. This interaction is very interesting as it provides a physical link between the cohesin complex and centromeres. Rad21 contains a MIR-like sequence.
Full length Rad21 with a GST tag was overexpressed in *E. coli* from a plasmid constructed by Birkenbihl and Subramani (1999). The 95 KDa fusion protein was found to be insoluble but could be resolubilised with 1% (w/v) sarcosyl (sodium laurel sarcosinate). However, sarcosyl prevented the solubilised protein binding to glutathione agarose beads. This effect could be overcome by incorporation of 2% triton X-100 in the binding buffer. Removal of sarcosyl from the buffer prevents elution of the protein from the beads. This is likely to be due to hydrophobic ‘sticking’.

Glutathione agarose beads were incubated in an *E. coli* lysate containing GST-Rad21 or GST, sarcosyl and triton X-100. Unbound proteins were washed away and the beads incubated with purified, untagged Swi6 CSD or mutant T323D CSD. After a further wash, proteins remaining bound to the beads were analysed by SDS PAGE. Neither CSD or T323D were found to be bound to either the GST negative control or the GST-Rad21 loaded beads (figure 5.4 lanes 9-12). A western blot using αSwi6 failed to resolve any bound CSD (data not shown).

Failure to detect CSD binding by Rad21 could reflect that this interaction is not direct. The failure to produce soluble Rad21 in *E. coli* indicates that this protein may be misfolded. The CSD interacting region of recombinant Rad21 may be aggregated or not in its correct context.

5.5.4 *S. pombe* Swi2 protein

*S. pombe* deficient in Swi2 are unable to switch their mating type. Mating type is determined by the translocation of one of two mating type genes into a heterochromatically silent loci (Klar *et al* 1998). Swi6 and Swi7 (polα) were also identified in a screen for mating type switching defects (Gutz and Schmidt 1985). No homology has been found between Swi2 and any protein of known function. Unpublished results from the laboratory of R. Allshire show that Swi2 directly interacts with Swi6 in yeast two hybrid assays. Swi2 also contains a MIR like peptide.
Swi2 was cloned from an *S. pombe* cDNA library into vector pMW172, suitable for recombinant expression in *E. coli*. The vector was transformed into a number of *E. coli* expression strains but did not over-express.

The full length protein is large (722 amino acids). A small, 48 amino acid fragment of Swi2 (amino acids 102-149), including the MIR pentamer (amino acids 135-139) was cloned and expressed in *E. coli* as a GST fusion protein. Purified GST-Swi2 polypeptide or GST alone was incubated with purified CSD or T323D mutant CSD. Proteins that remained bound to glutathione agarose beads after washing were analysed by SDS PAGE (figure 5.5 lanes 14-17). CSD could not be detected bound to the Swi2 peptide despite the assay being repeated at several pHs and salt concentrations.

Swi2 is known to bind directly to Swi6. This interaction was not detected by GST pulldown in this assay. It is possible that the region of Swi2 cloned is not the CSD interacting region or that the peptide was not correctly presented for binding.

### 5.5.5 *S. pombe* Clr4 protein

Though there is no direct evidence that *S. pombe* Clr4 is a heterochromatin localised protein, it is likely that it is as knockout causes loss of centromeric silencing and disruption of Swi6 localisation (Ekwall et al 1996). Clr4 contains an N terminal chromodomain and a C terminal SET domain (Ivanova et al 1998). The SET domain of Clr4 has been shown to methylate lysine 9 on histone H3 (Rea et al 2000). Swi6 can bind methylated histone H3 (Bannister et al 2001) and is dependent on Clr4 for correct localisation (Ekwall et al 1996). The chromodomain of Clr4 cannot bind methylated histone H3 and the mechanism of Clr4 localisation to heterochromatic loci is unclear (unpublished results). Clr4 contains a MIR-like peptide and so it was speculated that it may bind to Swi6 CSD. In support of this hypothesis, human and mouse homologues of Clr4 have been shown to co-immunoprecipitate with HP1 (Aagaard et al 1999). This interaction was later mapped to the first 44 amino acids of the human and mouse proteins (Melcher et
However, this region is not conserved between the human and mouse proteins and Clr4.

Clr4 is insoluble when expressed in *E. coli*. A portion of Clr4 (amino acids 127 to 490) was expressed in *E. coli* fused to a maltose binding domain (MBP). The MBP binds with high affinity to amylose resin allowing affinity purification with amylose resin. In addition the MBP is an extremely soluble protein and can often help solubilise fused proteins. The MBPClr4 construct (made by J. Partridge) was insoluble and was resolubilised using 1% sarcosyl and 2% Triton X-100. MBPClr4 was bound to amylose beads and proteins from the *E. coli* lysate were washed away. The loaded beads were incubated with either CSD or T323D CSD, proteins remaining bound to the beads after washing were analysed by SDS PAGE (figure 5.5 lanes 18-19).

Clr4 did not bind CSD in the conditions screened. This may indicate that this region of Clr4 does not bind CSD or that the resolubilised recombinant Clr4 is not capable of carrying out its normal interactions.

### 5.6 Conclusions

Examination of the surface topology of the dimeric Swi6 CSD structure revealed a cleft that ran along the dimer interface. The interior of the cleft is hydrophobic and is of a size that would accomodate a pentapeptide such as that identified by Murzina *et al* (1999) and Smothers and Henikoff (2000). Cleft edge residues are conserved between Swi6 and other HP1 homologues suggesting that the cleft may be a conserved feature. However, the cleft is not seen in an NMR structure of the CSD of mouse HP1β (Brasher *et al* 2000). It is speculated in this chapter that there is an equilibrium between an open cleft conformation (seen in the Swi6 CSD structure and a closed cleft conformation (seen in the mouse HP1β structure). The open conformation may be stabilised by phosphorylation.

In order to test the hypothesis that the cleft is important for the function of Swi6 in *vivo* a mutation was made in a cleft edge residue (F324A). The mutation
Figure 5.5: Continuation of figure 5.4. Lanes contain, 1) Target: CSD Bait: GST-Swi2 amino acids 102-149; 2) Target: T323D Bait: GST-Swi2 amino acids 102-149; 3) Target: CSD Bait: GST; 4) Target: T323D Bait: GST; 5) Target: T323D Bait: MBPClr4 amino acids 127-490; 6) Target: CSD Bait: MBPClr4 amino acids 127-490; 7) 1/10 T323D CSD load; 8) 1/10 CSD load; 9) Molecular weight markers with sizes indicated.
Figure 5.6: Schematic figure showing one model of the molecular basis for position effect variegation.

did not disrupt Swi6 localisation at heterochromatic loci but caused defective silencing of an ade6+ reporter gene in the outer repeats of centromere 1. In wild type *S.pombe* this reporter gene is invariably silent leading to a red cell colouring. Cells with defective centromeric silencing fail to repress this reporter gene and have a white colouration. Interestingly, the defective silencing phenotype of the F324A mutant variegated red and white from colony to colony. This variegating phenotype is reminiscent of *Drosophila* position effect variegation and it is worth noting that HP1 (a *Drosophila* homologue of Swi6) is a suppressor of position effect variegation (Eissenberg *et al* 1990). Phenotypic variegation has been observed before in fission yeast centromeres (Allshire *et al* 1994). However, there are several features of the F324A variegation that are worth elaboration.

Variegation is thought to result because of natural differences in the extent of heterochromatin from cell to cell. This is illustrated in figure 5.6. In the top part of this figure the ade6+ gene is clear of the centromeric heterochromatin (represented by grey circles). Transcription of this non-repressed gene causes a white colouration in the cell. The lower part of the figure shows another case in which the heterochromatin covers the gene causing transcriptional repression and a red cell colour.

Several types of variegation can be observed in fission yeast brought on by different mutations. Repression or activation of a gene can switch randomly from parent to daughter cell. This results in a mixture of red and white cells and pink colonies. Alternatively, repression can be inherited stably over several generations.
before switching to an active state or vise versa. This form of variegation results in colonies with red and white sectors. The variegation seen in the F324A mutant does not follow either of these forms. Repressive states appear to be stable within a colony and only switch between repressed and active when the colony restreaked on fresh media. It is difficult to propose a model to explain this data. In a mature colony of fission yeast nutrients are likely to be limiting and the cells fairly senescent. When the colony is restreaked onto fresh media nutrients become abundant and cell growth approaches logarithmic. It is possible that during this phase of their growth cycle fission yeast are more likely to variegate but this is highly speculative. What is certain, is that in the F324A cleft mutant there are deficiencies in centromeric silencing and a reduced fidelity of epigenetic inheritance.

To define the mechanism of interaction between CSD binding proteins and the cleft, \textit{in vitro} GST pulldown experiments were attempted. These studies used recombinant protein expressed in \textit{E.coli}. Despite screening eight recombinant proteins in a variety of conditions, no interactions were detected. There may be several reasons for this lack of interaction. Firstly, post-translational modification is known to play an important role in HP1 mediated modulation of chromatin structure. It is possible that such modifications are necessary for binding. Binding studies with HP1 homologues have largely made use of \textit{in vitro} translated protein or the yeast two hybrid assay, which allow the possibility of post-translational modification. Secondly, in the present study a number of these proteins seem to have misfolding and solubility problems when expressed in bacteria. One possible reason for this is that their natural context in the cell is within large complexes and not free in solution. It is possible that these proteins are folded with the assistance of chaperones and escorted to their target site in chromatin. Expressing such proteins in bacteria may be fundamentally problematic. Thirdly, the stability of heterochromatin associated protein complexes may be in multiple weak interactions rather than few strong ones. This would allow the complexes to be dynamic. It may also explain why identifying interactors with HP1 homologues has been difficult \textit{in vitro}. 

115
Chapter 6

Analysis of a chromodomain-like domain.

Recently chromodomains from *Drosophila* dosage compensation complex proteins msl3 and mof have been shown to bind RNA (Akhtar *et al* 2000). Analysis of these sequences suggests that it may be more accurate to classify them as a third subfamily of the chromodomain superfamily distinct to both chromodomains and chromo shadow domains. However, as shown later, expression of A1p13, a member of this family, and analysis by circular dichroism, suggests that these domains may not have the characteristic chromodomain superfamily fold.

A group of chromodomains with divergent sequence was first proposed by Koonin *et al* (1995). A region with highly significant similarity (P<10^-3) to human retinoblastoma-binding protein 1 (RBP-1) was observed in the *Drosophila* dosage compensation protein msl-3. Low similarity (~0.08%) was observed between this region and the chromodomain of *Drosophila* Polycomb protein. Though the sequence similarity is not statistically significant enough to conclude that msl-3 has a chromodomain, the chromatin modifying role of msl-3 lent force to this hypothesis (Koonin *et al* 1995). The chromodomain-like domain of msl-3, and sequences with highly significant similarity to it, will be called chromodomain-like domains (CLDs) in this chapter.

The dosage compensation complex (DCC) contains five proteins, including msl-3 and mof (mof contains a domain similar to the msl-3 CLD) and two non-coding RNA molecules (Lucchesi 1998). It is found localised to hundreds of sites on the X chromosome of male *Drosophila* (Copps *et al* 1998). Male *Drosophila*
have only a single copy of the X chromosome compared to two in females. Despite the difference in copy number, X gene transcripts are found at the same level in males and females. This is achieved by doubling the rate of X gene transcription in males, a process mediated by the DCC (Lucchesi 1998).

Transcriptional up-regulation of the X chromosome is likely to be brought about through modification of chromatin structure. Nucleosomes of the male X chromosome are found to be hyperacetylated on lysine 16 of histone H4 (Gu et al 2000). Histone acetylation states are closely linked to chromatin structure. MLE (one of the protein components of the DCC) has homology to a family of DNA helicases known to modulate chromatin structure (Sanjuan and Marin 2001). Moreover, the male X chromosome has a characteristic expanded morphology (Gorman et al 1995).

If the CLD of msl-3 is a true chromodomain it is divergent in sequence. It has been reported that the CLDs of msl-3 and mof are able to bind either of the two RNA molecules of the DCC (Akhtar et al 2000). The RNA binding function of these CLDs is divergent from the known functions of classical chromodomains and CSDs. It is not clear if these divergent domains share the chromodomain superfamily fold with CDs and CSDs.

In this chapter, alignment of homologues of the msl-3 chromodomain-like domain with CDs and CSDs is described. Comparison is made between the groups. Expression in E.coli of a chromodomain-like domain from S.pombe Alp13 protein is described along with preliminary structural characterisation.

6.1 Alignment of CLDs with CDs and CSDs

An iterative PSI-BLAST (Altschul et al 1997) search of the non-redundant sequence database with residues 14-99 of Drosophila msl-3 protein identified a number of similar domains. The PSI-BLAST output contained human RBP-1, S.pombe Alp13, Drosophila polycomb and yeast esal-associated factor 3 (eaf-3). Eaf-3 is a component of the NuA4 acetyl transferase complex (Eisen et al 2001). All of
Figure 6.1: Multiple sequence alignment of CD, CSD and CLD sequences. Sequences within the families were aligned by ClustalW. CD and CSD were aligned by superposition of representative structures. The CLDs were aligned with the other two families by eye. Residues conserved in chemical type throughout are boxed in grey. Residues conserved only within the CSD family are boxed in blue, CD family in red and CLDs in green (marked variant). Protein name and species annotations are similar to those in figure 3.11 with the addition of sc: Saccharomyces cerevisiae.

The residues conserved in both CDs and CSDs that make up the hydrophobic core of the chromodomain fold are largely conserved in the sheet region of the CLD family. The helical region of the CLDs is extremely divergent in sequence and there is no detectable conservation of residues in this region between CLD sequences. Conservation within the CLD family is predominantly in the sheet
portion. In this respect they are more similar to CDs than CSDs. The key CSD dimerisation residues are not found in CLDs and so they are unlikely to dimerise in the same way as CSDs.

6.2 Analysis of interfamily relationship

The relationship between the sequences in the alignment (figure 6.1) can be displayed as a dendrogram. Sequences are connected by lines whose lengths are proportional to their relatedness (figure 6.2).

The three chromodomain sub-families radiate from a single branchpoint in the centre of the dendrogram. By this analysis, the CLDs are as divergent from the CD group as they are from the CSD group. The three groups of domains form three distinct classes of protein. The CLDs should be considered as a separate class and not grouped with either of the existing chromodomain sub-families.

Another interesting feature of the dendrogram is the clustering of the CDs. CDs from HP1 homologues, Polycomb homologues and from Su(var)3-9 homologues cluster together within the CD family cluster. CDs from helicase proteins are spread around within the CD cluster. It was discussed in chapter 1 that the chromodomains of HP1 homologues bound methylated histone H3 tails but the chromodomain of Pc and Su(var)3-9 homologues have no dependance on methylation. The sub-clustering of the chromodomains may be a reflection of sequence divergence leading to different binding specificities.

6.3 Analysis of a CLD.

Cloning, purification and analysis of the Alp13 CLD was done with the assistance of R. Errington.

The CLD of *S.pombe* protein Alp13 was chosen as an archetypal CLD for structural studies. Mutations in Alp13 disrupt the structure of the mitotic spindle
Figure 6.2: Dendrogram created by ClustalW from the multiple alignment shown in figure 6.1 and rendered using Phylip (Felsenstein 1985). The proteins are represented by numbers which appear next to the protein names at the left hand side of figure 6.1. Red symbols are CDs, blue are CSD and green are CLDs. The CD subfamily is further split into the type of protein which the CD came from, this is represented by the symbol shapes as shown in the key.
Alp13 may be a chromatin associated protein, inferred from the presence of a CLD, and may also be linked to the spindle, inferred from the mutant phenotype. The possibility that the Alp13 links chromatin to the spindle is the reason for interest in this protein. The CLD of *S. pombe* Alp13 protein was cloned into vector pMW172 for expression in *E. coli*. The plasmid was sequenced to check fidelity. A soluble polypeptide of approximately 8 KDa could be seen on an SDS PAGE gel after transformed *E. coli* BL21 (DE3) were induced with IPTG. Mass spectrometry (mass spectrometry was carried out by A. Cronshaw) of the band confirmed that its molecular weight was consistent with that predicted.

The recombinant CLD was easily purified from a crude *E. coli* lysate in two chromatography steps. Lysate was loaded onto a diethylaminoethyl weak anion exchange column at a pH close to the calculated pI of the CLD (6.5). The CLD eluted from the column after the initial breakthrough of unbound proteins but before the salt gradient in a manner similar to the elution of Swi6 CSD from the carboxymethyl column (section 3.3). This strategy was extremely effective and no major protein contamination was detected by SDS PAGE. Pooled fractions containing the CLD were concentrated to 1 ml and loaded onto an S200 gel filtration column. The elution trace showed a single major peak, which corresponded to CLD.

### 6.4 Analytical size exclusion of CLD and CSD

A high resolution Pharmacia Superdex™-75 column was used to determine the mass of CLD and CSD. The mass will indicate if the CLD is mono or multimeric in solution.

Superdex™-75 is a resin of highly cross-linked agarose beads that can resolve proteins in a mass range between 3 and 70 KDa. The calculated chromodomain masses are, Alp13 CLD (9.5 KDa) and Swi6 CSD (8.2 KDa). The chromodomains fall outside the range of the mass standards used to calibrate the column. More suitable standards could not be obtained due to the small size. Despite this,
Figure 6.3: Calibration curve from Superdex\textsuperscript{TM} 75 column. \( \log_{10} \) molecular weight (KDa) is plotted on the Y axis. The Y axis plots \( K_{AV} \). 
\[ K_{AV} = \frac{V_C - V_O}{V_T - V_O} \]
where \( V_C \) is the retention volume of a protein, \( V_O \) is the void volume of the column and \( V_T \) is the total column volume. A best fit line is drawn through four data points corresponding to the elution of four protein calibration standards, ribonuclease (13.7 KDa), Chymotrypsinogen A (25 KDa), Ovalbumin (43) and Albumin (67 KDa). An equation for a best fit line is shown at the top right of the figure with an \( R^2 \) value.

The relatively large mass difference between monomers and dimers should allow them to be resolved on this column (monomeric CLD is calculated to elute with a retention volume of 72.3 mls, dimeric CLD at 62.6 mls).

The column was calibrated with the size standards. The retention volumes were plotted against \( \log_{10} \) molecular weight (KDa) (figure 6.3).

Both the CSD of Swi6 and the CLD of Alp13 were purified and 100 \( \mu l \) of a 1 mg/ml solution of each were loaded onto the size exclusion column. The absorbance at 280 nm of the eluate from the column was measured. Each protein was eluted from the column in a single peak. The CSD was eluted from the column with a \( K_{AV} \) of 0.31, this corresponds to a molecular weight of 21.4 KDa, 2.6 times its calculated mass. This is consistent with the CSD being a dimer in solution.
Figure 6.4: 1D proton NMR spectrum of Alp13 CLD. Chemical shift in parts per million is noted along the X axis. Water quenching is seen at 4.8 ppm.

The CLD was eluted from the column with a $K_{AV}$ of 0.41, the calculated molecular weight is 11.67 KDa which is 1.2 times the calculated mass. These results suggest that the CLD may be a monomer in solution.

6.5 One dimensional proton NMR of CLD

*NMR spectra were gathered by B. Smith.*

A one dimensional proton NMR experiment was carried out to determine if the CLD of Alp13 was a folded domain. The CLD was concentrated to around 8 mg/ml (~1mM). The 1D spectrum was collected on a 600MHz instrument. The NMR spectrum is shown (figure 6.4).

The spread of sharp peaks on the spectrum and especially the methyl peaks dispersed between 1 and -1 ppm are indicative of a folded domain.
Protons associated with the protein backbone Ca atoms are typically spread on either side of the suppressed water signal at 4.8 ppm. Secondary structure elements can affect these shifts. The cluster of peaks to the left of the water signal are indicative of β sheet. It is not possible to infer a helical content from this spectrum. We may conclude that CLD is a folded domain with some β sheet content.

6.6 Circular dichroism of the CLD

*Circular dichroism analysis was carried out by Dr. Sharon Kelly at the Scottish Circular Dichroism Facility, Glasgow.*

Circular dichroism (CiD) measures the difference in absorbance of right and left handed circularly polarised light by an optically active sample. The peptide bonds of a protein absorb radiation in the far-UV region of the spectrum (190-250nm). Regular secondary structure elements induce asymmetry to the absorption. Information about a protein's secondary structure can be extracted from CiD spectra.

The wave function of circularly polarised light has a fixed magnitude and a revolving vector angle. The wave describes a screw along the direction of propagation. Two circularly polarised waves that are in phase but have opposite screw directions will combine to form a simple linear wave. When a simple linear wave is passed through an optically active sample that absorbs one of the two component circular waves to a greater degree than the other the emerging wave will be elliptically skewed from the linear. The ellipticity of a wave in degrees is corrected for the pathlength of the cell and the molecular weight and concentration of the sample. The primary units of circular dichroism are degrees cm²/dmol.

Circular dichroism analysis was done on the CLD of Alp13 and on the CSD of Swi6. If the CLD is related to the chromodomain superfamily it would be expected to have a similar fold when compared to the CSD. Similarly folded proteins will
give similar Cid spectra because they have the same percentages of secondary structure. Cid spectra collected for the two domains are shown in figure 6.5.

Visual examination of the data would strongly suggest that the two domains have different folds as reflected by the large differences between the two spectra.

The percentages of α-helix, β-sheet and random coil of a protein can be calculated from a Cid spectrum. This process involves fitting the Cid spectra curves from proteins of known secondary structure onto the curve of the unknown protein. To get accurate results the protein concentration of the sample must be known very precisely. The results also vary according to which Cid spectra are included in the comparison datasets.

Concentrations of CSD and CLD were estimated by their absorbance at 280 nm and by a BCA assay. Absorbance at 280 nm is related to protein concentration by an extinction coefficient that can be calculated from the sequence. The coefficient does not take into account the effect of sidechain interactions in a folded protein and so this is not always accurate. The BCA assays measures the reduction of Cu++ to Cu+ by cysteine, tyrosine and tryptophan residues of a protein in
an alkaline solution. The BCA reagent (bicinchoninic acid) reacts with Cu⁺ to produce an intense purple colour at 562 nm. Protein concentrations are measured by comparing the absorbance at 562 nm of an unknown protein sample with a serial dilution of a protein standard of known concentration. The accuracy of this assay relies on the unknown protein reducing Cu²⁺ with similar efficiency to the standard. A₂₈₀ estimates CLD at 0.41 mg/ml and the CSD at 0.72 mg/ml. BCA estimates CLD at 0.39 mg/ml and CSD at 0.47 mg/ml.

Two programs were used to estimate the secondary structure content of the two chromodomains. The structure of CSD is known and so this can be used to gauge the performance of the fitting programs. The two programs are CONTIN and SELCON.

CONTIN (Provencher and Glöckner 1981) contains a database of spectra from a large number of proteins of known conformation. These spectra are fitted to the curve of the unknown protein by a least squares algorithm. Spectra that do not fit well are weighted so that they have a low contribution to the final result. The results of the CONTIN program are seen in table 6.1.

SELCON (Sreerama and Woody 1994) works in a similar way to CONTIN, fitting spectra from a database to the unknown spectrum. Spectra that do not fit well are deleted from the dataset to speed up the program. The 5D spectrum from the unknown protein is included in the dataset with an estimate of the secondary structure. The program is run once to derive a better estimate of secondary structure and this new value is used to re-run the program. SELCON is run again and again in an iterative way until the secondary structure estimate does not change. The results of the SELCON program are shown in table 6.1.

The crystal structure of Swi6 CSD has 30.6% α-helix and 26.1% β-sheet (calculated by the sub-program YASSPA within the structure manipulation program O (Jones et al 1991)). Neither 5D analysis program came up with secondary structure composition close to that seen in the crystal structure. This disparity could be due to inaccuracy in estimates of protein concentration. The chromodomains are very small compared to most of the proteins used in the databases.
### Table 6.1: Secondary structure percentages for Swi6 CSD and Alp13 CLD calculated from their CiD spectra by the programs CONTIN and SELCON. Protein concentrations were CLD 0.39 mg/ml and CSD 0.47 mg/ml as estimated by a BCA assay.

<table>
<thead>
<tr>
<th>Domain</th>
<th>α-helix</th>
<th>β-sheet</th>
<th>β-turn</th>
<th>remainder</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swi6 CSD</td>
<td>15%</td>
<td>55%</td>
<td>11%</td>
<td>19%</td>
</tr>
<tr>
<td>Alp13 CLD</td>
<td>20%</td>
<td>40%</td>
<td>19%</td>
<td>21%</td>
</tr>
<tr>
<td>SELCON</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- no solutions for Swi6 CSD -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alp13 CLD</td>
<td>16.7%</td>
<td>23%</td>
<td>22.3%</td>
<td>38%</td>
</tr>
</tbody>
</table>

6.7 Conclusions

There are a number of proteins in species from budding yeast to man that have regions with highly significant similarity with the chromodomain-like domains of *Drosophila* proteins mof and msl3. These proteins have previously been described in the literature (Koonin *et al* 1995). In this chapter these domains were aligned with chromodomains and chromo shadow domains. In this alignment it is noted that not all of the core hydrophobic residues, conserved in both CDs and CSDs are conserved in the variant sequences. The lack of conservation of core residues casts doubt on the structural similarity between these chromodomain-like domains and the chromodomain superfamily. However, the extremely low sequence similarity between these sequences and the other chromodomains provide few anchor points for confident alignment. It is possible that these core residues are conserved but are misaligned.
A dendrogram generated from this alignment shows clearly the relationship between the three sequence subsets. The CLDs, CDs and CSDs are all equally divergent from each other. Each subfamily clusters discretely from the others. The level of sequence divergence between the CDs and CSDs has led to the two families having very distinct functions and biochemical properties. Equally, the sequence divergence between the CLDs and the other two families may have led to similar functional differences. Though the CLDs may share functions with the other two families this cannot be implied without experimental data.

Structural studies were carried out on a CLD from the fission yeast protein Alp13. These studies determined that this region was a folded domain and that is was likely to be monomeric in solution. Interestingly, a circular dichroism spectrum of the CLD was found to be very different to a spectrum of the CSD of Swi6. If these domains shared a similar structure, the spectra would be expected to be similar. This preliminary evidence suggests that the CLDs may not be structurally related to the chromodomain superfamily.

Since this work was completed a paper has been published describing the conservation of CLD domains in 23 different species. The paper proposes that the CLDs be considered as a third subclass of the chromodomain superfamily (Bertram and Pereira-Smith 2001). However, unlike the present work, no experimental evidence was reported to support that these domains have structural homology with the chromodomain superfamily.
Chapter 7

Conclusions

7.1 Summary of Results

In this thesis the solution of a crystal structure of the CSD of the fission yeast Swi6 protein is described. This is the first CSD structure and the first crystal structure of any chromodomain. The domain is similar in structure to an NMR structure of the chromodomain of murine HP1β. The major structural difference between the structures is the insertion of a helical loop between the β-sheet and α-helix of the CSD compared to the CD.

The CSD dimerises. This can be seen in the crystal structure and has been demonstrated by several complementary techniques in solution. The helical loop, found in the CSD structure, but not in the CD structure, is important for dimerisation. Residues surrounding the dimer interface are conserved in chemical nature in all CSDs implying that this mode of dimerisation may be a common property of CSDs. In support of this an NMR structure of a CSD, published after completion of this work, has a similar mode of dimerisation (Brasher et al 2000).

A cleft was observed on the surface of the Swi6 CSD dimer. The cleft is of a size that would accommodate an extended pentapeptide such as the CSD-binding consensus sequence PxVxL. The cleft is not observed in the HP1β CSD NMR structure. It was speculated that these proteins may switch between an open cleft conformation such as that seen in the Swi6 CSD structure and a closed cleft conformation seen in the HP1β CSD structure. This switch was proposed to be mediated by CSD phosphorylation.
In an attempt to study the interaction of proteins with the cleft a number of in vitro GST pulldown experiments were attempted. Despite prolonged effort no CSD binding by these proteins could be detected. CSD binding by the proteins screened may be dependent on other cellular processes such as post-translational modification, chaperones or other cooperatively binding proteins.

Finally, a domain from the fission yeast protein Alp13, with some sequence similarity to chromodomains was expressed in *E. coli* and characterised. This domain is part of a family of homologous domains that are proposed to be chromodomains. Sequence analysis revealed that these domains are equally divergent from CDs and CSDs. Biochemical analysis suggests that the Alp13 chromodomain-like domain is a folded domain and is monomeric. Structural analysis by circular dichroism however, strongly suggests that the domain does not share structural similarity with the chromodomain superfamily.

7.2 Discussion

This discussion comprises three sections. First, a molecular basis for the divergent function of CDs and CSDs will be proposed. Key surface residues will be identified that are involved in domain interactions. Secondly, the divergence of the chromodomain superfamily to perform different functions, while preserving a common structure, will be discussed. Some general points will be made about the evolution of protein domains and their categorisation into families and superfamilies. Finally, a model of the function of HP1 homologues will be discussed in light of the results of this thesis.
7.3 Key residues for the functions of CDs and CSDs

Sequence alignment of the CD and CSD families, in light of the representative structures, reveals residues conserved throughout the two families and residues conserved only within one or the other of the two families (section 3.10). It can be seen that the residues conserved between the families are involved in forming the hydrophobic core of the structure (figure 3.12). Selective pressure conserving these residues over large evolutionary distance, illustrates their importance.

It would be useful to use these core residues to generate a search pattern for identifying chromodomains in sequence databases. However, the small size of the chromodomain, the variable lengths of the sections between the core hydrophobic residues and the amino acid variation at each position make such search patterns rather non-specific. The core residues are more useful diagnostically in sequence analysis. That is, when a potential new chromodomain sequence is added to a chromodomain alignment, it is useful to check that the core residues are conserved.

The CD and CSD subfamilies have distinct functions (section 1.4.3). The residues that are conserved within, but not between, the families are likely to be involved in their respective functions. Understanding the role of these conserved residues helps in grouping the chromodomains into functionally similar families.

The residues that are conserved only in the CSD subfamily (figure 3.11) mainly cluster around the dimer interface (figure 4.6). Key CSD dimerisation residues are a tyrosine residue directly followed by a glutamate in the helix H2 (Swi6 residues Y318 and E319) and a Asn (Swi6 N307) residue in helix H1. An insertion in helix H1 of CSDs compared to CDs is important for the correct positioning of the Asn residue. The positions of the residues at the dimer interface are very similar in the Swi6 CSD structure and the murine HP1β CSD structure (Brasher et al 2000). The occurrence of these residues in a chromodomain sequence is a strong
indication that the domain is of the CSD class and will dimerise in a similar way to the Swi6 CSD.

Upon dimerisation a cleft is seen running along the dimer interface of the Swi6 CSD structure. This cleft is hypothesised to be the site of protein-protein interaction with the CSD. Residues His 321, Thr 323 and Phe 324 in Swi6 form the edge of this cleft. These residues are conserved in type in most CSDs indicating that the mode of protein-protein interaction is common.

The pattern \([\text{NH}]\)-\([\text{X(3)}]\)-\([\text{PR}]\)-\([\text{X(2)}]\)-\([\text{VLIM}]\)-\([\text{ILD}]\)-\([\text{X(2)}]\)-\([\text{YG}]\)-\([\text{E-X(4)}]\)-\([\text{WFKLIP}]\) was generated from the dimerisation and cleft residues of an alignment of CSDs. In the pattern amino acids are represented as one letter code. Letters inside the square braces represent a single position in which the amino acid can be any of the letters shown. ‘X’ represents any amino acid and the following numbers signify that the preceding letter is repeated that number of times.

When the Swissprot and Trembl protein databases were searched with this pattern 1, all of the CSDs listed in the Pfam database 2 were found. Pfam uses a Hidden Markov Model to assemble multiple alignments of protein domains without structural input. Though the pattern recovered all of the CSDs in this database it was not particularly specific and recovered many spurious sequences. It did not recognise any CDs or CLDs.

Some chromodomains have been shown to bind directly to the methylated tails of histone H3. It is likely that there are other mechanisms by which chromodomains bind to nucleosomes (section 1.4.3). The surface residues that are conserved in chromodomains but not in CSDs are likely to mediate this function.

It was suggested that a hydrophobic cleft observed on the surface of the murine HP1\(\beta\) CD structure may be responsible for protein-protein interaction (Ball et al

---

1Swissprot and Trembl are anotated databases of protein sequences maintained at the Swiss Institute of Bioinformatics. http://www.expasy.ch/

2Pfam is a database of domain families maintained at the Sanger Center. http://www.sanger.ac.uk/Software/Pfam/
Figure 7.1: Residues conserved within the CD family are coloured according to their chemical type on a molecular surface representation of the CD of mouse HP1β. Carbon atoms are coloured green, nitrogen blue and oxygen red. Non-conserved residues are coloured yellow. The orientation of the left hand molecule is indicated with a thumbnail picture of a backbone trace. The right hand image is related to the left by a 180° rotation around a horizontal axis. The figure is taken from Cowieson et al. (2000).

1997). The equivalent cleft in the Swi6 CSD structure is covered by the N-terminus of the molecule (Cowieson et al. 2000). This suggested that perhaps the cleft was artificially produced by N-terminal truncation in the NMR structure. However, the same cleft can be seen on the NMR structure of the Clr4 chromodomain (Horita et al. 2001).

When the residues conserved only in the CD family and not in the CSD family (figure 3.11) are mapped onto the surface of the CD structure they form a continuous ‘sash’ that circumscribes the surface of the molecule (figure 7.1).

Conservation of surface residues in proteins implies that they may be important for function. Residues of the ‘sash’ may be important for the interaction of CDs with histone proteins. The ‘sash’ stretches right around the CD molecule implying that interaction with the histone may involve more than one face. A possible model for this interaction is that the methylated tail of histone H3 interacts with ‘sash’ residues on one side of the CD while residues on the other side of the CD are free.
Figure 7.2: *A model of a chromodomain bound simultaneously to the tail and the core fold of histone H3 in a nucleosome. The histones in the nucleosome are coloured similarly to those in figure 1.1 and the DNA is shown in white. The model is roughly to scale in terms of protein sizes and histone tail length.*

to bind to the main body of the histone protein. A diagram of this is shown in figure 7.2. This model supports both the experimental results that suggest CDs bind methylated histone tails and the results that show tail independent histone binding.

7.4 The chromodomain superfamily fold

Considerable effort has been made in recent years to group related proteins into hierarchical phylogenies (Levitt and Chothia 1976). The purpose of categorising proteins is so that the function and structure of newly discovered proteins can be guessed by similarity to known proteins.

The SCOP database ³ groups proteins into classes, folds, superfamilies and families. At the top of the hierarchy, proteins are grouped into five broad classes based on the secondary structure content. Within these classes proteins with common folds are sub-grouped. These have similar secondary structure elements

³http://scop.mrc-lmb.cam.ac.uk/scop/
occurring in the same order in the peptide chain but are not necessarily related in sequence or function. Within each fold there are numerous superfamilies. Proteins in the same superfamily have the same structure and are often functionally related. Each superfamily is subdivided into families of proteins with greater than 30% sequence identity, similar structures and similar functions. Two proteins are considered to be related in evolutionary terms if they are homologous in structure, sequence and function, or certainly in two of these (Murzin et al 1995).

In the SCOP database chromodomains are in the class ‘a+b’. This groups proteins that have α-helix and β-sheet elements that do not alternate on the chain. There are 184 folds within this class and chromodomains are in the fold interleukin 8-like (IL8-like). Within this fold there are two superfamilies, interleukin 8-like chemokines and chromo domain-like. There is one family in the IL8-like chemokine superfamily containing all these structurally related chemokines. The chromo domain superfamily contains two families, chromo domains and "Histone-like" proteins from archaea.

In figure 7.3 some example structures from the IL8-like and chromo domain superfamilies are shown. These have all been orientated in a similar way for easy comparison. The chromo shadow domain of Swi6 and the chromodomain of murine HP1β are in the chromo domain family of the chromo domain-like superfamily. Sac7d is in the ‘histone like protein from archaea’ family. These proteins are thought to be involved in the packaging of archaeal DNA in the absence of histones. Sac7d binds to DNA in a non-sequence specific manner via positively charged residues on the back of the β-sheet (Robinson et al 1998, Su et al 2000). Neutrophil activating peptide-2 (NAP2) and interleukin 8 (IL8) are both α-chemokines. These proteins both homodimerise making contacts on the edge of the beta sheet such that the helices from opposing monomers align side by side (Malkowski et al 1995, Baldwin et al 1991). Rantes is a β-chemokine. It also dimerises but by a different mechanism to NAP2 and IL8. The N-termini of two rantes monomers line up antiparallel to each other in an extended conformation. The helices of the two monomers make no contacts and are on opposite sides of the monomers from the dimer interface (Wilken et al 1999). Stromal derived factor 1
alpha (SDF1a) shows sequence divergence from both \( \alpha \) and \( \beta \) chemokines, but in action and mode of dimerisation is more similar to IL8 and NAP2 (Dealwis et al 1998).

The different dimerisation modes of these proteins are interesting. The chromodomain and sac7d are both monomeric. The chromo shadow domain dimerises through the helical portion of the structure. IL8, NAP2 and SDF1a dimerise through the edge of the beta sheet. Rantes dimerises through associations between the N-terminal unstructured part of the protein. It is interesting that these proteins have numerous modes of dimerisation despite sharing a common fold.

It can be clearly seen in the overlay in figure 7.3 that if the beta-sheet portions of these structures are aligned, the helices do not overlay. This reflects that in these IL8-like structures, the position of the helix relative to the beta-sheet is relatively mobile through evolution. It is tempting to speculate that that the movability of the helix relative to the sheet may be one reason why this fold has been reused by evolution to achieve numerous different functions.

Statistics from a comparison of the different proteins with the chromodomain of murine HP1\( \beta \) are shown in table 7.1. It can be seen that in terms of sequence identity none of the proteins is very similar to the chromodomain. Proteins of similar structure are placed in the same family in the SCOP database if they have greater that 30% sequence identity. Despite being in the same family as the chromodomain, the CSD falls well below this cutoff. In fact, a different criterion for grouping these proteins is being used. This criterion is their context.

CSDs are considered to be related to CDs despite having sequence identity of less than 30% and having different functions. Relatedness by evolution is inferred because they are found in tandem with CDs on the same protein sequence. This suggests that they may have arisen from a gene duplication event. It is likely that the two domains are related by evolution as they have similar structures and core hydrophobic residues are conserved between the two. In the case of CDs and CSDs relatedness cannot be inferred from sequence or function but from the fact that they are found in a similar context, in tandem on the same protein.
Figure 7.3: A number of proteins with structural homology to the chromodomain superfamily is shown. The chromodomain of mouse HP1β is shown in white in the left of centre. Clockwise from top left, the chromo shadow domain of fission yeast Swi6, the archaean Sac7d protein, the human chemokine interleukin 8, an overlay of the backbone Ca traces of all the proteins, the human chemokine neutrophil activating protein 2, the human chemokine stromal derived factor 1 alpha and finally the human chemokine RANTES.
Table 7.1: Comparison of the sequence and structure of various proteins with the chromodomain of murine HP1β. Sequences were aligned individually with the chromodomain sequence using CLUSTALW and percentage identity derived from this alignment. This is shown in column % ID vs CD. Structural alignment was done using the CCP4 program lsqkab over the extent of the structured region. Root mean square deviations for all atoms in the structures are output from lsqkab in Å and are shown in the column ‘rmsd vs CD’ (Å).

Another example of the use of context in assigning protein relationships is that the chromodomain-like domains of proteins such as Msl3, Mof and Alp13 are considered to be related to the other chromodomains despite having sequence identity well below 30%, different functions and completely unknown structures. This relationship is inferred because these domains function in chromatin structure modulation similar to the other chromodomains. As a result of the assumption that the chromodomain-like domains were related to chromodomains it was proposed that the RNA binding function of the chromodomain-like domains may be a common property of all chromodomains (Akhtar et al 2000).

The present structural study of the chromodomain-like domains suggests that they may have been wrongly categorised, certainly in terms of family, but also possibly in the wrong class. This miscategorisation resulted from the use of context to judge relatedness. Though including context in categorisation criteria can be useful, such as in the case of grouping CDs and CSDs, it reduces the stringency of the procedure and therefore may lead to erroneous categorisation.
7.5 A model of Swi6 function at fission yeast centromeres

It is known that Swi6 binds nucleosomes via the chromodomain. Methylation of histone H3 on lysine 9 by the Su(var)3-9 homologue Clr4 is the initiation event for chromodomain binding. It is not clear how Clr4 methylation is controlled so that HP1 homologues are correctly localised. Upon binding to an initiation site Swi6 can spread out along chromatin. Fission yeast chromodomain protein Chpl cannot spread and so the spreading of Swi6 may be a function of the CSD. Knockout of the fission yeast Swi6 gene causes alleviation of centromeric gene silencing and defects in chromosome segregation. Knockout is not lethal and chromosome segregation and centromeric silencing, though defective, still take place. This implies that although Swi6 is an important part of centromeric heterochromatin and the kinetochore these can form in its absence, albeit with impaired function.

In this thesis it is reported that the CSD can dimerise. Upon dimerisation a cleft is formed that may be important for interactions. Mutations around this cleft causes a variegating, reduced centromeric silencing phenotype. Conserved surface residues on the CD form a ‘sash’ that circumscribes the domain. This implies that interactions occur on both sides of the domain. These results add some details to previously suggested models of the fission yeast centromere.

Allshire (1996) suggests a model of the fission yeast centromere in which the chromatin fibre folds back on itself at the central core bringing the outer repeats into close proximity. The protein complement of the central core region is distinct to that of the outer repeats. Using this model as a foundation, several modes of binding of Swi6 can be suggested.

Four possible modes of Swi6 binding are shown in figure 7.4. These are all theoretically possible and not mutually exclusive.

First, one chromodomain of a Swi6 dimer associates with a methylated histone H3 tail on one nucleosome. The second chromodomain of the dimer is then
Figure 7.4: A model of a fission yeast centromere. The chromatin fibre doubles back on itself to form a loop. The central core region of the centromere is at the top of the loop in the middle of the fibre. Four theoretically possible modes of Swi6 binding to the centromere are shown, numbered 1 to 4. The figure is not to scale with respect to the length of DNA in a *S. pombe* centromere or the size of proteins involved.
constrained in its movement to a sphere with a radius equal to the length of two Swi6 flexible linker lengths (one linker is approximately 400 Å in length assuming a peptide bond length of 3.8 Å and extended conformation). This chromodomain binds a methylated histone H3 tail of the next nucleosome along the chromatin chain (The inter nucleosome distance in a chromatin fiber is not known but one nucleosome is approximately 100 Å wide and 50 Å thick). In this way the Swi6 dimer cross links neighbouring nucleosomes and perhaps fixes the chromatin chain into distinct higher order structures. Zhao et al (2000) reports a cooperative binding of HP1 to nucleosomes. The binding of a chromodomain from one Swi6 dimer to one histone H3 monomer of a nucleosome may induce the binding of a second Swi6 dimer to the second histone H3 of the same nucleosome. This would leave the second chromodomain of this newly bound Swi6 dimer free to bind to the next nucleosome in the chain. This process could be responsible for Swi6 spreading.

Second, the two chromodomains of a Swi6 dimer could bind directly to the two histone H3 molecules of a single nucleosome. In favour of this mode of binding, the two histone H3 tails protrude from the same side of the nucleosome core structure (figure 1.1). This mode of binding would not cross link nucleosomes. The binding of both halves of the Swi6 dimer may exert conformational change on the nucleosome and this in turn may lead to higher order chromatin association or silencing.

Third, if the centromere structure does loop back on itself and the two outer repeat sequences are brought close together spatially then the Swi6 dimer may be able to cross link the two halves of the centromere. Cross linking loops of chromatin may not be a specifically centromeric function of Swi6 but this could occur at other regions of heterochromatin.

The first three models of binding all leave the CSDs free for interaction with other proteins. This recruitment property may be an important part of the Swi6 function.

The fourth model is that primary recruitment of Swi6 to the centromere is through the CSDs binding other centromerically localised proteins. For example it has been demonstrated that the CSD can bind DNA polymerase-α. The poly-
merase may process along the DNA with an associated Swi6 dimer. When the protein complex reaches a nucleosome that has been methylated the chromodomains bind and the Swi6 dimer dissociates from the polymerase.

7.6 Future work

In this thesis a study of the binding of a CSD to several other proteins in vitro was attempted. This binding study was unsuccessful and it was hypothesised that this may be due to solubility problems with the recombinant proteins and the importance of post-translational modifications for binding. To solve the problems, GST-pulldown binding experiments with GST fused Swi6 CSD could be repeated with target proteins in the presence of protein extracts from S.pombe. Alternatively GST pulldowns could be performed in S.pombe protein extracts. Bound proteins would be resolved by 2D gel electrophoresis and identified by mass spectrometry.

Part of the aim of studying Swi6 CSD binding was to define a region of an interacting protein sufficient for binding. This region would be synthesised as a peptide and co-crystallised with the CSD. A co-crystal structure would better define the residues important for binding.

Testing models one to four shown in figure 7.4 would be a longer term aim. The easiest way to start studying interaction of Swi6 with chromatin would be to disrupt dimerisation and study the effect of this on Swi6 localisation and spreading. However, the results of this thesis suggest that the CSD may be unstable as a monomer. This may be a fundamental problem limiting future study of HP1 function.

Other possibilities include studying Swi6 interaction with in vitro reconstituted nucleosomes, methylated on lysine 9 of histone H3. The sizes of the complexes formed by mixing Swi6 with these nucleosomes could be studied by ultracentrifugation or size exclusion chromatography. This approach may determine if it is
physically possible for Swi6 to crosslink nucleosomes. But it may be hard to prove the relevance of these results \textit{in vivo}.

7.7 Concluding remarks

The simplicity and predictability of genetic inheritance has allowed huge advances in the understanding and application of molecular biology. However, the literature review and experimental results outlined in this thesis are concerned with epigenetic inheritance. Recent advances in this field are revealing the molecular mechanisms of a type of inheritance that is independent of DNA sequence. This is primarily mediated through changes in higher order chromatin structure. If the simplicity of genetic inheritance was a gift to modern science then it is certainly true in this case that we've had as much fun playing with the packaging!
Bibliography


[66] J. Hsu, Z. Sun, X. Li, M. Reuben, K. Tatchell, DK. Bishop, JM. Grushcow, CJ. Brame, JA. Caldwell, DF. Hunt, R. Lin, MM. Smith, and CD. Allis. Mitotic phosphorylation of histone h3 is governed by ipl1/aurora kinase and


[74] TA. Jones, JY. Zou, and SW. Cowan. Improved methods for building protein models in electron density maps and the location of errors in these models.


158


[143] MM. Shareef, C. King, M. Damaj, R. Badagu, DW. Huang, and R. Kellum. Drosophila heterochromatin protein 1 (hp1)/origin recognition complex (orc)


