THE UNIVERSITY
of EDINBURGH

Insight into Stc1-interactions bridging RNAi and chromatin modification in *Schizosaccharomyces pombe*

Sreerekha Sreedharan Pillai

Thesis presented for the Degree of Doctor of Philosophy

University of Edinburgh

September 2016

Declaration

I hereby declare that this thesis was composed by me, and that the research presented is my own except where otherwise stated. This work has not been submitted for any other degree or professional qualification.

Sreerekha Sreedharan Pillai

2016

"*I survived because the fire inside me was brighter than the fire around me*"

Joshua Graham

TABLE OF CONTENTS

Acknowledgement

I would like to express my sincere thanks to my supervisor Dr. Elizabeth Bayne for providing me the opportunity to undertake my PhD while working as a research staff. I extend my gratitude towards her for the support provided throughout my study and my appreciation for the fast track correction of this thesis.

I wish to express my gratitude to the University of Edinburgh, especially School of Biological Sciences for providing the financial support for my studies. A sincere thanks also goes to MRC funding body for supporting my research job and for the extensions. I would like to take this opportunity to express my thanks to the College of Life Sciences, University of Dundee where it all began and to all my friends and past colleagues. I also wish to thank my PhD committee members: Dr Irina Stancheva, Dr Jeyaprakash Arulanandam (JP) and Dr Ken Sawin for their valuable comments.

I greatly appreciate and thank my dear friend and lab-companion, Dr Francesca Taglini, for the sweet and memorable lab life, for the continuous support and love and for the uncountable episodes of 'going crazy with scientific talks'! Without you my dear Franci, my lab life wouldn't have been this colourful. I also express my sincere thanks to my dear friend and previous lab-mate, Dr Laure Verrier. I thank you for the support you have given me, for inspiring me throughout and for the precious 'life outside lab' talks. I extremely enjoyed being one amongst the '*pombe-puff girls'*!

I wish to express my sincere thanks and appreciation to my present lab-mates Elliott Chapman and Dr Jo Strachan who have been so much fun to work with. Your support and friendliness are greatly appreciated and cherished. I also wish to thank my dear friends and colleagues from '4th floor of Darwin building' for providing such a great sense of community and for all the support and love. I thank Dr Benura Azeroglu, Dr Heidrun Interthal and Dr Meriem El-Karoui for the lovely chats. I also express my thanks to my dear friend Dr Keerthi Chathoth for her continuous support, love, inspiration and for listening to my 'venting-it-out' episodes!

None of this would have been possible without the tremendous love and rock-solid support of my loving family. I am extremely grateful to my ever-loving parents, my sweetheart sister and my grandmother for their unconditional love and support. Thank

you Amma (mother), for believing in me and supporting my decision to pursue higher studies in the UK. Thank you Acha (father), for sacrificing your comfort to provide me with the educational path I wished to follow. Thank you Ammae and Acha for providing me with the 'epigenetics' for endurance, strong will and determination! Thank you my darling chechi (big sister), for the inspiring person you are, and for the precious support and love you provide me. Thank you Ammommae (grandmother), for believing in me and for the priceless love and support. I also wish to express my sincere thanks to my ammavan (uncle) for his support and love and to all my extended family members close to my heart. I immensely thank my dear mother-in-law for the tremendous love and support provided during the write-up of this thesis. Thank you for taking care of my baby boy for me and for the delicious food and the tea!

A special thanks goes to my own product of meiosis, my cute little Aadi for being such a well behaved baby and letting me write up this thesis. Last, but not at all the least, I express my sincere thanks to my soul-mate, my darling husband Sanal, who have been the 'Iron-man alloyed with titanium' in my life. Your unconditional love and abilities of 'Professor X' is greatly appreciated and beyond saying thanks. Thank you my love for being my strength, especially throughout the journey of my PhD and writing up of this thesis, and for putting up with my emotional turmoil. I dedicate this thesis to each of you, my family, as a token of appreciation.

"*Family is not an important thing, it's everything!"*

Michael J. Fox

Lay Summary

Cell division is a vital event required for the sustainability of all living organisms. To ensure that next generation cells inherit the correct complement of DNA - the blueprint of life, cells require a way to transfer this information with precision. Within the cell, DNA is packaged with specialised proteins called histones into chromatin, and chromatin is further organised into structures called chromosomes. During cell division, the information within the chromatin must be passed on equally to the next-generation cells through the duplication and segregation of chromosomes. Accurate segregation depends on the proper functioning of specialised regions of the chromosomes termed centromeres. The proper functioning of centromeres requires them to be in a specialised compact structure called heterochromatin, where the chromatin is tightly packed through its modification by the addition/removal of chemical groups. Defects in heterochromatin formation can lead to improper segregation of chromosomes resulting in diseases including cancer. Hence it is important to understand how heterochromatin is properly formed and maintained at centromeres as this information can lead to better understanding of underlying causes of such diseases and potentially help scientists find a cure.

To gain better understanding of heterochromatin formation, fission yeast has been extensively used as a model system since it has centromere architecture similar to higher organisms including humans. Interestingly, various cellular machineries come together to promote the formation of heterochromatin. In particular, chromatin modification at fission yeast centromeres involves small-RNAs (siRNAs). siRNAs generated from centromeres target the siRNA-machinery (termed 'RNAi') to

centromeres helping in the recruitment of the chromatin modification machinery (termed 'CLRC') to chromatin. The CLRC in turn modifies chromatin by adding chemical groups required for the compaction of chromatin, thereby ensuring proper heterochromatin assembly.

A protein named Stc1 was identified as a key component that recruits CLRC to chromatin in an RNAi-dependent manner. Understanding the interactions of Stc1 is important for understanding how the connection between RNAi to chromatin modification is made. My work provides important insights into direct interactors of Stc1, and provides data on how Stc1 mediates the interaction between the RITS and CLRC complexes. My study reveals that association with the RNAi machinery and CLRC is mediated largely by distinct regions within Stc1. Additionally, I found that Stc1 has the potential to act as a 'sociable' protein by virtue of its association with multiple proteins, opening up future avenues of research into the roles of other Stc1-like 'sociable' proteins. Thus my study helps in understanding how two different molecular machineries can be utilised to bring about a single outcome by the use of 'sociable' proteins like Stc1.

Abstract

Compact heterochromatin is essential for genome stability and hence cell survival. Studies in many organisms including humans underline the importance of pericentromeric heterochromatin in centromere function. Fission yeast centromeres share a common structural organisation with those of their metazoan counterparts. The fission yeast model has been pivotal in understanding many key events in the pathway leading to the assembly of pericentromeric heterochromatin. In particular, studies in this system have revealed that the RNA interference (RNAi) pathway connects with the chromatin modification machinery to impart proper heterochromatin formation.

Transcription of the pericentromeres by RNA polymerase II (Pol II) produces double stranded RNA (ds RNA) which is processed by Dicer(Dcr1) into small interfering RNAs (siRNAs). These siRNAs are loaded onto the Argonaute protein Ago1, and target the Ago1-containing RITS (RNA-Induced Transcriptional Silencing) complex to the pericentromeres via complementary base-pairing of the siRNA to the nascent centromeric transcript. RITS then recruits the sole Histone H3-K9-methyl transferase, Clr4, as part of the Clr4-complex, CLRC. The resulting H3K9-methyl marks further result in the recruitment of downstream chromatin binding proteins including the HP1-homolgue Swi6 which plays a key role in cohesin retention. Additionally, the H3K9-methyl marks are required for stabilising the association of CLRC and RITS, thereby promoting a reinforcing loop within the RNAi-mediated heterochromatin pathway. Thus crosstalk between RITS and CLRC is important in establishing and maintaining silent chromatin at the pericentromeres.

Stc1 has been proposed to act as a critical link that connects the RITS and CLRC complexes. Stc1 is required for heterochromatin establishment and maintenance at the pericentromere and association of RITS with CLRC is lost in the absence of Stc1. Moreover, Stc1 directly interacts with Ago1 and is essential for siRNA production. These and other previous observations (Bayne et al. 2010) highlight the key role played by Stc1 in the RNAi-mediated heterochromatin pathway. To understand how Stc1 mediates the specific cross-talk between RNAi and chromatin modification, I have investigated the nature of Stc1 interactions with the RNAi and chromatin modification machineries.

Using *in-vitro* binding assays, I found that Stc1 directly interacts with the CLRC subunits Dos2 and Clr4. I also identified the RITS subunit Tas3 as a potential interactor of Stc1, in addition to Ago1. A collaborating research group elucidated the structure of Stc1 using NMR (He et al. 2013) and my study provides evidence for interactions via the distinct domains of Stc1. Stc1 utilises its disordered C-terminus to bind to Dos2 while the N-terminus, which contains a tandem zinc finger domain, acts as a multi-protein interaction interface binding the CLRC subunit Clr4 and RITS subunits Ago1 and Tas3, opening up possibilities for Stc1-containing distinct-complexes. My work provides new insights into the role of Stc1 and opens up future avenues of research key to understanding how heterochromatin domains are defined and maintained.

Abbreviations

ARC	Argonaute siRNA chaperone complex
ATP	Adenosine triphosphate
CD	Chromo-domain
CENP-A	centromere protein A
CENP-B	centromere protein B
CerFP	cerulean fluorescent protein
CLRC	Clr4-containing complex
CoIP	co-immunoprecipitation
CSD	chromoshadow domain
CTD	carboxy terminal domain
Cul4	Cullin 4
DCAF	DDB1-Cul4 associated factors
DDB1	DNA damage binding protein 1
DNA	Deoxyribonucleic acid
dsRNA	Double stranded RNA
GST	glutathione S-transferase
H3K9	Histone H3 Lysine 9
H3K9me2	Histone H3 Lysine 9 dimethyl
HDAC	histone deacetylase
HP1	heterochromatin protein 1
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
imr	inner most repeats
IP	immunoprecipitation
kDa	kilo Dalton
Mda	mega Dalton
miRNA	micro RNA
MoRF	molecular recognition features
mRNA	messenger RNA
NLS	nuclear localisation signal

otr	outer repeats
PAGE	polyacrylamide gel electrophoresis
PCD	Post-chromo-domain
PDB	protein databank
PEV	position effect variegation
piRNA	PIWI-interacting RNA
priRNA	primal RNA
PTGS	post transcriptional gene silencing
PTM	post translational modification
rDNA	ribosomal DNA
RDRC	RNA dependent RNA polymerase complex
RdRp	RNA dependent RNA polymerase
RFTS	Replication foci targeting signal
RISC	RNA-induced silencing complex
RITS	RNA induced transcriptional silencing
RNA	Ribonucleic acid
RNA Pol II	RNA Polymerase II
RNAi	RNA interference
SDS	sodium dodecyl phosphate
SEC	size exclusion chromatography
SHREC	Snf2/histone deacetylase repressor complex
siRNA	small interfering RNA
sRNA	small RNA
ss-siRNA	single stranded siRNA
Stc1	siRNA to chromatin 1
Su(var)3-9	suppressor of variegation 3-9
TGS	transcriptional gene silencing
WT	wild type
YFP	yellow fluorescent protein
ZF	zinc finger

List of figures

List of tables

CHAPTER 1

Introduction

1.1 Chromatin Organisation

Genome packaging involves association of DNA with specialised proteins, primarily histones. This DNA-protein complex is referred to as chromatin. The basic unit of chromatin is called the nucleosome. A canonical nucleosome harbours two copies each of core histones H2A, H2B, H3 and H4, wrapped by 146 base pairs (bp) of DNA (Richmond et al. 1997). The core nucleosomes are connected via linker DNA that is bound by linker histone H1, giving rise to the first level of DNA folding in the form of the 10nm "beads-on-a-string" chromatin fibre (Kornberg 1977). Further packaging results in the formation of higher order chromatin fibres.

The 'textbook model' of chromatin structure, depicting a 10nm chromatin fibre further folded into a 30nm fibre, is still debated (Belmont et al. 1999). The classical 'solenoid model' proposes a regular packaging of nucleosomes in which the 10nm chromatin fibre is coiled around a central axis (Fussner et al. 2011). Although the 30nm chromatin fibre (whose structure still remains elusive) has been reconstituted *in-vitro*, the existence of this structure *in-vivo* is controversial (Li & Zhu 2015). It has been proposed that eukaryotic genome organisation could be explained by the folding and packaging of the 10nm chromatin fibre via frequent bending and kinking to achieve a linear compaction similar to a 30nm fibre (Fussner et al. 2011). Regardless of the debate on how the 10nm chromatin fibre is further packaged, higher order packaging of nucleosomes is evident *in vivo* and is a dynamic process (Fussner et al. 2011). Indeed, chromosomal processes like DNA replication and repair require dynamic chromatin assembly and reassembly. Major molecular machineries involved in modulating chromatin structure include chromatin remodelers, histone chaperones and

histone modification machineries. Chromatin remodelers act by either ejecting, moving, destabilising or restructuring nucleosomes utilising energy released from ATP hydrolysis (Clapier & Cairns 2009). Histone chaperones on the other hand are key proteins that regulate nucleosome assembly (Burgess & Zhang 2013). Histone-modifying enzymes are responsible for depositing a variety of chemical moieties like acetyl and methyl groups on lysine residues in the N-terminal tail of histones. Histone tail modification can influence chromatin structure in a number of ways; for example, acetylation of histones neutralises the positive charge of histones and this results in weakened histone-DNA interactions. The relaxed state of DNA allows access to DNA-binding proteins and promotes activities including gene transcription. In contrast, methylation of histones does not alter the electric charge of histones, but can regulate the binding of chromatin-associated proteins (Zhang et al. 2016).

Interplay between chromatin remodelers and histone modifications is yet another powerful mechanism for altering the structure and function of chromatin. For example, the CHD (Chromo-domain helicase DNA-binding) family of chromatin remodelers (conserved from yeast to humans) harbour tandem chromo-domains that mediate binding to methylated lysine residues. This cross-talk has been shown to be crucial for localisation and/or activity of CHD family chromatin remodelers in a variety of organisms including humans (Swygert & Peterson 2014). In addition to the concerted actions of proteins and protein-modifications, specific features of DNA are also thought to play important roles in functional genome organisation. For example, emerging evidence points towards functional genome organisation in space (Filion 2015). For example, clustering of invasive transposon elements within the genome in various organisms is not only thought to facilitate silencing of these invasive elements,

but also to provide a co-opted mechanism for controlling other genome functions through higher order genome organisation (Mizuguchi et al. 2015). Clustering of gypsy retrotransposon by 'insulator' proteins in *Drosophila* has been implicated in the establishment of higher order chromatin structure (Capelson & Corces 2005; Gerasimova et al. 2000). Similarly CENP-B protein (belonging to transposase family) mediated transposon clustering into Tf ('Transposon of fission yeast') bodies in fission yeast recruits chromatin modifying enzymes to assemble 'closed' chromatin and could also facilitate genome organisation (Mizuguchi et al. 2015). Thus genome organisation seems to be the outcome of actions-in-concert involving chromatin associated factors with a role for features embedded in the underlying DNA.

1.2 Heterochromatin

Early cytological studies led to the classification of chromatin structures into two types: euchromatin and heterochromatin (Heitz 1928). Based on the classical view, euchromatin represents the decondensed form of chromatin which is usually associated with gene-rich regions of the genome, while heterochromatin embodies the condensed state of chromatin typically found in gene-poor genomic regions. Heterochromatin is generally associated with transcriptional repression. Pioneering studies on position effect variegation (PEV) demonstrated that the juxtaposition of a gene with sites of heterochromatin as a result of chromosomal rearrangement, rendered a euchromatic gene, inactive (Grewal & Elgin 2002). As the term implies, the variegated phenotype is a result of the change in the position of the gene on the chromosome, rather than a change to the underlying DNA sequence. This suggested that the chromatin context in which a gene resides can potentially determine its

activity. The phenomenon of PEV is thought to occur due to spreading of heterochromatin onto the juxtaposed gene, rendering it transcriptionally silent. In comparison to euchromatin, heterochromatin displays reduced sensitivity to nucleases, and the nucleosomes are more tightly arranged in an orderly fashion. This highly ordered chromatin could thus limit or inhibit access to a variety of molecular machines involved in transcription, replication etc. (Elgin & Reuter 2013).

Classical views on what defines heterochromatin, and the mechanisms underlying its formation, have begun to change, thanks to extensive research carried out on various model organisms ranging from yeast to mammals. A proposal based on the 'histone code hypothesis' put forward the idea of classifying chromatin as active or inactive based on the combination of histone modifications present and the local concentration of these modifications (Jenuwein & Allis 2001). For example, the phosphorylated-acetylated state of histone H3 (H3S10-phosphorylation, and H3K9 and/or H3K14 acetylation) has been linked to the transcriptionally active state. Conversely, transcriptional repression is associated with methylation of H3K9, which requires deacetylation of H3K14 and antagonises H3S10-phosphorylation. Studies so far indicate that various enzymatic machineries including histone deacetylases (HDACs) and histone methyltransferases (HMTs) act in concert for the formation of heterochromatin (Wang et al. 2016). In many organisms, repressive chromatin is additionally marked by DNA methylation.

In the past, heterochromatin was thought to represent a static state of chromatin. However, pioneering work spanning over a decade, has led to the discovery that heterochromatic regions are in fact dynamic in nature. Heterochromatic regions are often associated with structural proteins like HP1 (Heterochromatin Protein 1) and

other non-histone proteins bearing chromatin-binding domains. Although HP1 was initially considered a stably associated structural protein (Grewal & Elgin 2002), FRAP (Fluorescence Recovery After Photobleaching) studies from yeasts to mammals have subsequently shown HP1 to be highly dynamic in nature (Cheutin et al. 2003; Cheutin et al. 2004; Festenstein et al. 2003). Thus constitutive heterochromatin assembly at regions such as centromeres and telomeres has gained attention since studying this could help better understand the dynamic mechanisms involved in preserving genome integrity.

1.3 Centromeres

Most eukaryotes possess monocentric chromosomes in which a centromere is assembled at a single localised region (McKinley & Cheeseman 2015). The centromere in such chromosomes, represents the primary constriction cytologically observed in a condensed metaphase chromosome. It is the site of kinetochore assembly and is essential for the segregation of sister chromatids. A key feature of active centromeres in eukaryotes is the presence of the centromere specific histone H3 variant CENP-A (centromere protein A), which is considered a structural and functional foundation for kinetochore assembly (Allshire & Karpen 2008). The prevailing classification of centromeres in monocentric chromosomes, based on underlying DNA, defines point centromeres and regional centromeres (**Fig 1.1**) (McKinley & Cheeseman 2015). A typical example of a point centromere is that of *Saccharomyces cerevisiae*, where a single nucleosome (approximately 125bp DNA) acts as a functional centromere. In this case, the short stretch of specific DNA sequence is essential and sufficient to assemble a functional centromere. Regional centromeres are

widely seen across various taxa including *S. pombe*. The underlying DNA in a regional centromere extends from a few kilobases to megabases, although DNA sequence alone is not sufficient to define centromere identity. Unlike point centromeres, the modular structure of a regional centromere typically consists of a non-repetitive central core flanked by pericentromeric repetitive elements that are assembled in heterochromatin (McKinley & Cheeseman 2015). Considering the context of this thesis, I will focus on regional centromeres in the following sections.

1.4 Centromere identity

Centromeres function in a highly conserved process in which chromosomes attach to microtubules and segregate by movement along the microtubules to the spindle poles. This ensures that following replication, the daughter cells inherit equal genetic material. However, centromere identity does not appear to be conserved, based on the underlying DNA sequence (Karpen & Allshire 1997). Throughout evolution, centromeric and pericentromeric DNA seems to rapidly evolve, and therefore it has been proposed that factors beyond the primary DNA sequence could play a role in centromere identity (Henikoff et al. 2001) . Hence, the specification of centromeres is considered to be epigenetic. The term 'epigenetics' is widely used in different contexts. In the context of faithful propagation of centromere identity, epigenetic inheritance can be defined as "heritable changes in gene or centromere activity without a corresponding change in the primary DNA sequence" (Karpen & Allshire 1997).

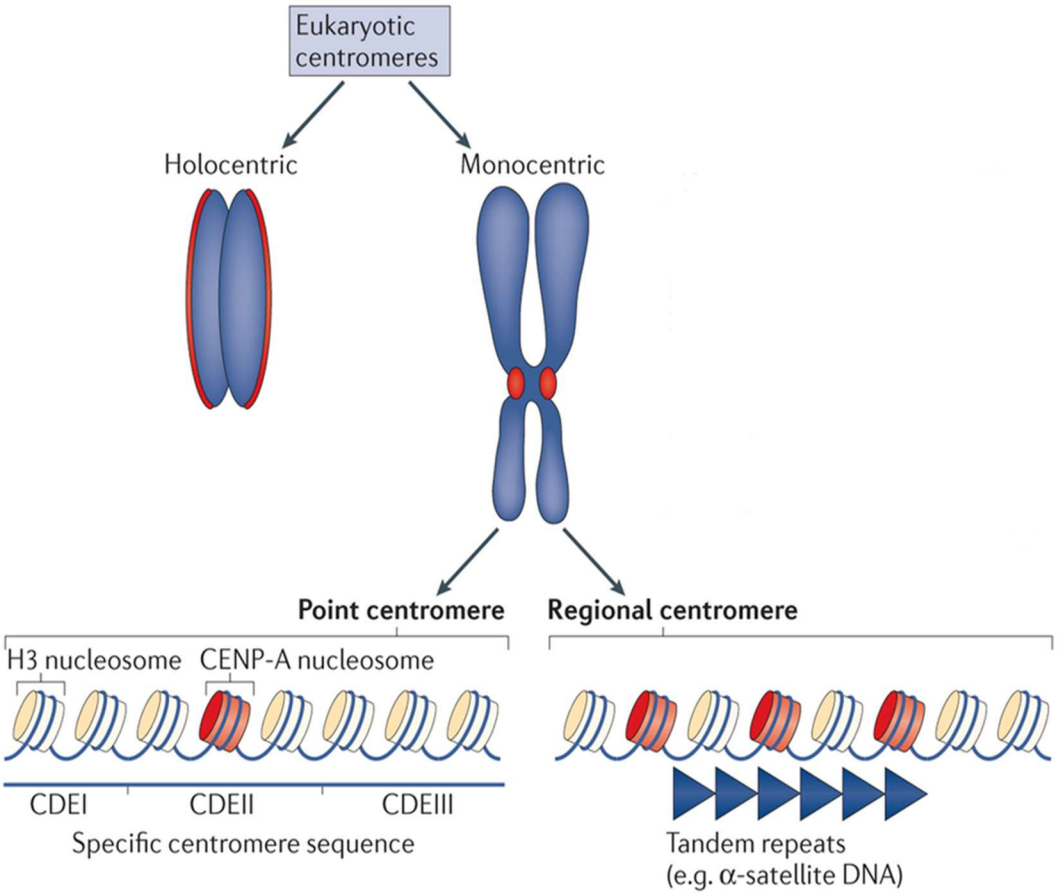


Fig 1.1 Types of centromeres

Eukaryotes possess either holocentric or monocentric chromosomes. In holocentric chromosomes, centromeres are diffused across the whole chromosome while monocentric chromosomes have a single, localised centromere. Two types of centromeres found in monocentric chromosomes are point and regional centromeres. Point centromeres are defined by short DNA sequence motifs as observed in *S. cerevisiae* (illustrated), consisting of three specific centromere DNA elements (CDEI-III) essential for a fully functional centromere. These elements assemble a single centromere specific CENP-A (centromere protein A) nucleosome which act as a centromere. In contrast to point centromeres, regional centromeres typically consist of large arrays of repetitive elements (e.g. alpha-satellite DNA in humans) and assembles numerous CENP-A nucleosomes. Figure adapted from McKinley and Cheeseman (2015).

1.4.1 Sequence identity of centromeric DNA

Several studies ranging from yeast to humans have demonstrated that the underlying DNA sequence is insufficient to nucleate a functional kinetochore and that additional features define the centromere (Karpen & Allshire 1997). Current models for centromere identity include both genetic and epigenetic features (McKinley & Cheeseman 2015). Centromeric DNA often contains repetitive sequences, AT-rich segments and/or remnants of transposable elements. However, mere presence of these elements alone does not result in the formation of a functional centromere. Despite conservation of their repetitive nature, centromeres lack evidence of sequence conservation throughout evolution, except for the presence of short stretches of DNA called CENP-B boxes in the centromeres of rodents and great apes (McKinley & Cheeseman 2015). CENP-B boxes are found in the highly repetitive alpha-satellite (alphoid) DNA, and are binding sites for the human centromere protein CENP-B (Masumoto et al. 1989; Muro et al. 1992). There is some evidence that CENP-B, and hence the CENP-B box sequence elements, are dispensable for centromere function, including experiments on CENP-B null mice and the existence of human neocentromeres and Y-chromosome lacking the CENP-B box (McKinley & Cheeseman 2015). However, a recent study by Fachinetti et al. suggests that CENP-B has a redundant role in stabilising key centromere proteins and is required for faithful chromosome segregation (Fachinetti et al. 2015).

Fission yeast consists of three putative CENP-B homologues Abp1, Cbh1 and Cbh2 (Murakami et al. 1996; Lee et al. 1997; Irelan et al. 2001). All three proteins have been shown to localise at fission yeast pericentromeres and might be involved in redundant pathways for recruiting the HP1 homolog Swi6 (Nakagawa et al. 2002). Abp1 has been

additionally shown to bind to AT-rich sequence within the centromeric central core *in vitro*; however, information on a consensus motif is lacking due to the AT-rich nature of the sequence (Halverson et al. 1997). Cbh1, on the other hand, binds to a consensus motif within a specific pericentromeric repeat region (Lee et al. 1997) that has been shown to be essential and sufficient for establishing a functional centromere in *S. pombe* along with the central core (Baum et al. 1994). Thus the underlying DNA sequence can contribute to proper centromere function.

1.4.2 Epigenetic identity of centromeres

Centromere identity is thought to be governed largely epigenetically. The first evidence for the epigenetic nature of centromere identity came from work by Earnshaw and Migeon, who demonstrated that a dicentric chromosome from a clinical sample had one of its two centromeres inactivated, even though the underlying sequence of both centromeres remained the same (Earnshaw & Migeon 1985). Subsequent work has indicated histone H3 variant CENP-A as the crucial epigenetic mark required for the faithful inheritance of the centromere. CENP-A is actively retained at centromeres during replication and is inherited by the newly synthesised DNA (Chen & Mellone 2016). This ensures the inheritance of the specific and consistent centromere location in future generations. Tethering of CENP-A or the histone chaperone HJURP onto non-centromeric DNA has been demonstrated to be sufficient for the formation of a functional centromere that can be propagated faithfully over several cell divisions (Mendiburo et al. 2011; Barnhart et al. 2011). These and other studies have provided strong evidence that epigenetic mechanisms play a crucial role in specifying the centromere. Intriguingly, comparison of CENP-A^{Cid} in closely related species of *Drosophila* indicates rapid evolution of this histone variant (Malik & Henikoff 2001).

Sequence variation observed within the CENP-A^{Cid} amino terminal tails and H3-DNA contact regions (shown to be critical for localisation to centromere), parallels patterns of variation within centromeric satellite repeats, suggesting possible co-evolution of centromeric DNA and histone variants (Henikoff et al. 2001).

1.5 Centromere Domain Organisation

In organisms including plants and animals, domains of CENP-A nucleosomes can be found interspersed with stretches of H3 nucleosomes within the centromeric chromatin (Chen & Mellone 2016). For example, in human and flies, CENP-A nucleosomes are interspersed with H3K4me2 nucleosomes, while in chicken, H3K9me3 nucleosomes (and low levels of H3K4me2 nucleosomes) are observed along with CENP-A nucleosomes (Stimpson & Sullivan 2010). This indicates that centromeric chromatin organisation can vary in different organisms. Indeed, studies using naturally occurring neocentromeres indicate the underlying plasticity associated with *de-novo* assembly of CENP-A-rich regions and hence centromeres (Stimpson & Sullivan 2010). For example, a lack of shared chromatin organisation within the CENP-A domains among different neocentromeres has been observed. While one particular neocentromere revealed interspersed CENP-A and H3 domains, another neocentromere displayed a much larger CENP-A-rich domain. Additionally, these neocentromeres displayed variability in associated heterochromatin (some neocentromeres completely lacked a heterochromatic environment). Despite the observed plasticity, evidence for domain organisation in eukaryotic centromeres exists (Stimpson & Sullivan 2010). In fission yeast, a clear boundary exists between the central core of centromere and the outer heterochromatic region. It has been shown that the presence of a tRNA gene in this

boundary region and, importantly, the transcription of this tRNA gene, is essential to maintain the distinct centromere domains (Scott et al. 2006). Selective recruitment of specific proteins to the distinct domains of the centromere was also demonstrated in fission yeast (Partridge et al. 2000). In human and flies, there is also evidence of distinct domains within the centromeric region corresponding to core centromere flanked by heterochromatic domains, although boundary elements are thought to span a larger region and display lower sequence-specificity in these species (Blower et al. 2002; Lam et al. 2006). Thus a common organisational theme for centromeric chromatin seems to emerge, consisting of a core centromeric sequence flanked by heterochromatic domains rich in repetitive DNA sequences (**Fig 1.2**).

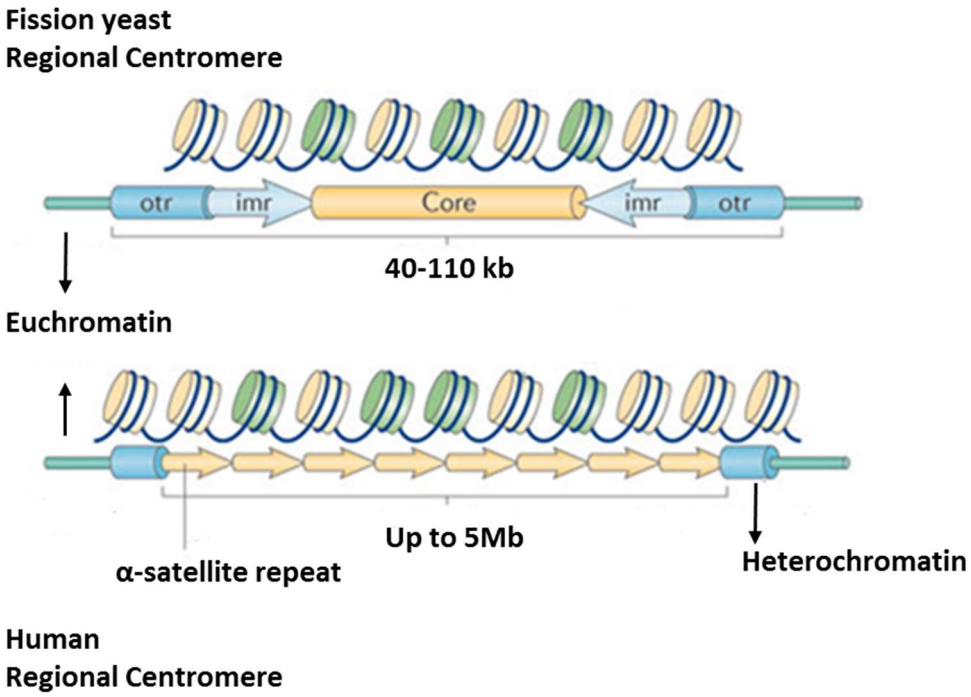


Fig 1.2 A common theme for centromeric chromatin organisation

Regional centromeres in fission yeast and human, feature a central core region consisting of interspersed CENP-A and H3 nucleosomes (represented in green and yellow cans respectively). The central core domain varies in size in different organisms and consists of repeat elements: fission yeast central core domain comprises of inner most repeats (*imr*) and unique central core DNA while human centromeres harbour large areas of alpha-satellite

repeats. The CENP-A associated central region is further flanked by heterochromatin regions (*imr* and *otr* (outer repeats) in fission yeast) marked by repressive histone modifications including H3K9-methylation marks. Figure adapted from Verdaasdonk and Bloom (2011) (Verdaasdonk & Bloom 2011).

1.6 Emerging roles of non-coding RNA

The pathways involved in bringing about chromatin modifications are well studied in different model systems, and can involve non-coding RNAs (Malecová & Morris 2010). Classic examples of transcription-coupled silencing through long non-coding RNAs can be found during X-inactivation in female mammals and during programmed DNA elimination in ciliates like *Tetrahymena*. Further well-studied examples occur in *Drosophila*, where polycomb-trithorax proteins act antagonistically to regulate gene transcription through long non-coding RNAs (Malecová & Morris 2010).

Small RNA (sRNA) guided silencing referred to as RNA interference (RNAi) has emerged as a key regulatory mechanism of cellular processes in eukaryotes. Three major species of sRNAs are small interfering RNAs (siRNAs), micro RNAs (miRNAs) and Piwi-associated RNAs (piRNAs) (Verdel et al. 2009). Although the various sRNA-based RNAi mechanisms share a unifying theme (i.e. silencing utilising Argonaute proteins), key differences can be observed in sRNA biogenesis and mode of function. Proteins within the Argonaute family are principal players in all sRNA-mediated silencing pathways. The number of Argonautes present varies dramatically between species, from a single Argonaute in fission yeast (Ago1) to 27 in the worm *C. elegans* (Ketting 2011). Despite this variability, all Argonaute members share a common purpose – they associate with sRNAs to license the RNAi machinery for target identification through base-pairing (Ketting 2011). Argonautes possess a catalytically active RNase-H-like domain which is required for cleaving RNA, referred

to as ‘slicer’ activity (Huisinga & Elgin 2009). Argonautes are classified into three clades: the AGO clade, the PIWI clade (specific function in the metazoan germline) and WAGO clade (specific to worms) (van Wolfswinkel & Ketting 2010). Besides Argonaute proteins, key players in RNAi are members of the RNase III family of nucleases including Dicer (van Wolfswinkel & Ketting 2010). Dicer plays a crucial role in sRNA biogenesis through recognition and cleavage of long dsRNA or precursor RNAs with hairpin structures (a process referred to as ‘dicing’) to give rise to siRNAs or miRNAs (Kurzynska-Kokorniak et al. 2015).

siRNAs are approximately 21 nucleotides in length and are the products of Dicer cleavage (van Wolfswinkel & Ketting 2010). The substrate RNA for Dicer is long dsRNA, which can come from either exogenous (viral replication) or endogenous sources (for example, bidirectional transcription due to the presence of internal cryptic antisense promoters). In organisms including yeast, plant and worms, secondary siRNA generation involving an RNA-dependent RNA polymerase (RdRP) is observed. In comparison, piRNAs are longer than siRNAs (24-30 nucleotides) and are mainly involved in transposon silencing in the animal germline. piRNA biogenesis is Dicer-independent, and involves an Argonaute-based amplification loop referred to as the ping-pong cycle. It is thought to be triggered by single-stranded transcripts arising from piRNA-generating loci (**Fig 1.3**) (van Wolfswinkel & Ketting 2010).

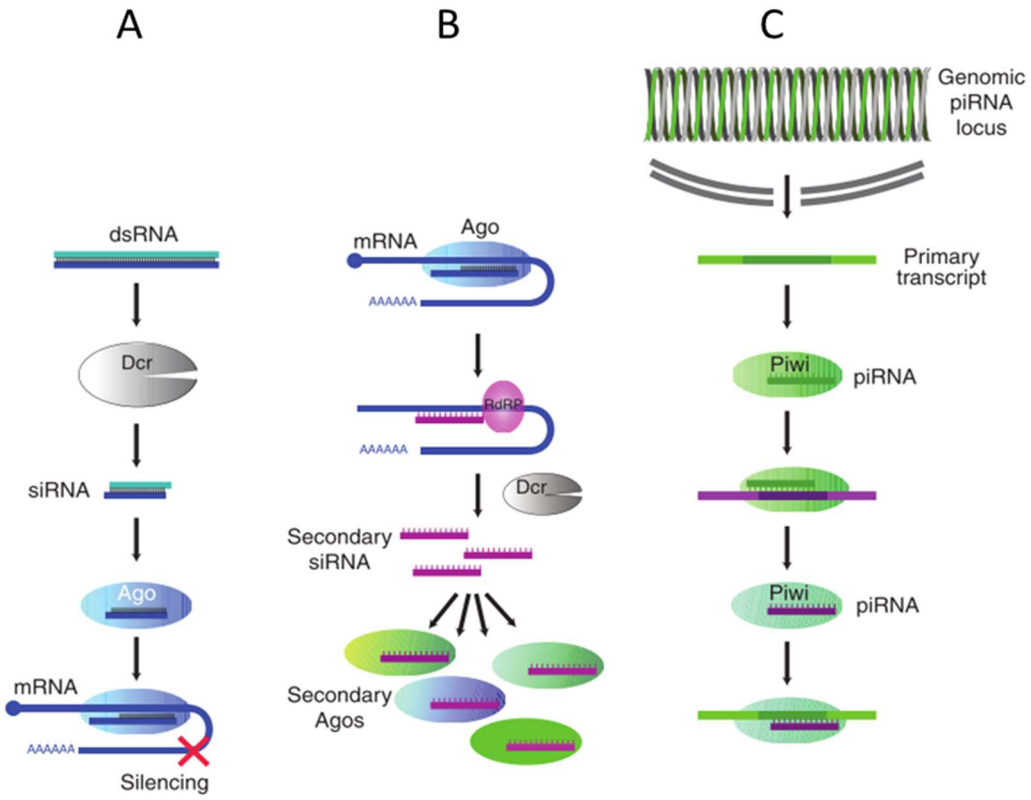


Fig 1.3 Characteristics of siRNA and piRNA mediated silencing

(A) siRNA generation from dsRNA template. Dicer cleaves dsRNA template to give rise to approximately 21 nucleotide siRNAs which are bound by Argonaute protein (Ago). Slicer activity of Ago results in the release of 'passenger strand'. The retention of single strand (called 'guide strand') licenses Ago for target recognition as part of the RISC complex via base pairing of guide strand and target mRNA, resulting in silencing of mRNA, commonly by degradation.

(B) secondary siRNA generation. In many organisms, targeting of RISC results in the recruitment of dsRNA-synthesising RdRP enzyme which produces more dsRNA template to be processed into siRNAs by Dicer. In organisms with more than one Ago, the subsequent loading of these secondary siRNAs into specific Ago proteins can determine the downstream silencing events.

(C) piRNA production. piRNA mediated silencing is present in the animal germline and is involved in transposon silencing. Major sources of piRNAs are piRNA clusters consisting of transposon elements and the initial trigger is thought to be single stranded piRNA transcript which is processed by Piwi-clade Ago proteins. piRNA mediated silencing additionally involves a piRNA-specific amplification 'ping-pong' cycle analogous to the RdRP-mediated secondary siRNA generation in organisms like yeast and plants. **Figure 1.3A-C** adapted from van Wolfswinkel and Ketting (2010).

In the best characterised mode of post transcriptional gene silencing (PTGS), siRNAs associate with the Argonaute-containing RNA-induced silencing complex (RISC) and direct specific transcript degradation (Malecová & Morris 2010). Besides PTGS, sRNAs can also mediate transcriptional gene silencing (TGS), which involves recruitment of chromatin modifying machineries for target silencing. TGS is seen for siRNAs in plants, *C. elegans* and *S. pombe,* and for piRNAs in developmental regulation in *Drosophila* and mice. In recent years *in vitro* studies have also implicated siRNAs in transcriptional gene silencing in human cells. Moreover, investigations on the mechanism of epigenetic silencing of tumour suppressor genes in various cancers also indicate a role of siRNAs in transcriptional gene silencing, leading to the view that defects in this pathway may underlie a variety of diseases including cancer (Grewal & Elgin 2007; Malecová & Morris 2010).

1.7 Fission yeast: A powerful model system to study RNAi mediated heterochromatin formation

The role of RNAi in heterochromatin assembly is well characterised in the fission yeast *Schizosaccharomyces pombe*. The fission yeast genome consists of three chromosomes with major constitutive heterochromatic loci found at centromeres, sub-telomeres, mating type locus (which determines the cell mating type) and ribosomal DNA (rDNA) (**Fig 1.4**) (Allshire & Ekwall 2015). Studies using this model system were pivotal in elucidating the connection between RNAi and chromatin modification. Most of the components involved in RNAi and chromatin modification mechanism in fission yeast are conserved in higher eukaryotes and the fission yeast genome harbours only a single copy of most of these components. (Allshire & Ekwall 2015).

Additionally, fission yeast centromeres resemble those of higher eukaryotes, including humans, both in organisation and in the conserved factors involved in epigenetic regulation of the loci (Pidoux & Allshire 2004). These features make it a valuable and simple model for studying the more complex heterochromatic centromere (Pidoux & Allshire 2004) .

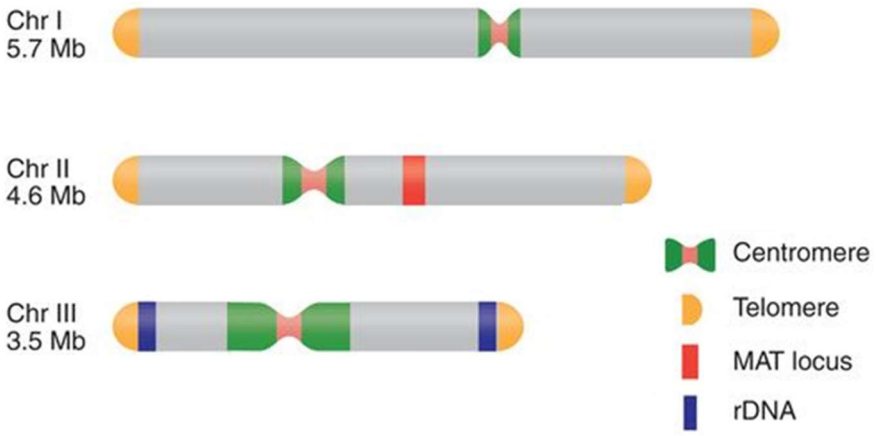


Fig 1.4 Major constitutive heterochromatic loci in *S. pombe*

A schematic representation of the three chromosomes in *S. pombe* highlighting regions of constitutive heterochromatin formation. Centromeric and telomeric heterochromatin plays a key role in maintaining genome stability. Heterochromatin at the mating type locus is required to determine the cell type identity. A function for heterochromatin at the rDNA locus is unclear. Figure adapted from Allshire and Ekwall (2015).

1.8 Heterochromatin in *S. pombe*

Fission yeast centromeres range from 40-110kb in size and efficient centromeric heterochromatin formation is crucial for proper chromosome segregation. Fission yeast centromeres share similar organisation with their metazoan counterparts, in that the centromeric central core DNA is flanked by heterochromatic repetitive elements. The central core of each of the fission yeast centromeres consists of unique sequence, and is flanked by inner-most repeats (*imr*), which are further flanked by outer-repeats

(*otr*) comprising tandem repeats of *dg* and *dh* elements (**Fig 1.5A**) (E Lejeune et al. 2010; Allshire & Ekwall 2015). At centromeres and sub-telomeric regions, heterochromatin imparts structural support required for proper chromosome segregation and telomere integrity respectively. On the contrary, at the mating type locus, heterochromatin is required for gene silencing that determines the cell type identity. Besides constitutive heterochromatin, three additional classes of heterochromatin have been reported in *S. pombe*: facultative heterochromatic islands found at a subset of meiotic genes (Zofall et al. 2012; Hiriart et al. 2012), HOODs (heterochromatin domains) at developmentally regulated genes and retrotransposons (Yamanaka et al. 2013) and transient heterochromatin formed at convergent genes (Gullerova & Proudfoot 2008). This thesis will focus on the role of RNAi in constitutive heterochromatin assembly.

1.9 Role of RNAi in constitutive heterochromatin formation

Extensive studies have been undertaken to understand the mechanism by which RNAi promotes constitutive heterochromatin formation. In fission yeast, the absence of RNAi components has been shown to impair heterochromatin formation at centromeres. Key components of the RNAi pathway are the Dicer ribonuclease Dcr1, the Argonaute protein Ago1, and Rdp1, an RNA-dependent RNA polymerase thought to be involved in dsRNA generation and/or secondary siRNA biogenesis. Two studies more than a decade ago demonstrated the requirement for Dcr1, Ago1 and Rdp1 for proper silencing at centromeres (Volpe et al. 2002; Volpe et al. 2003). While RNAi is essential for the establishment and maintenance of heterochromatin at centromeres, RNAi has been found to be dispensable for maintenance of heterochromatin at mating

type locus and telomeres (Volpe et al. 2002; Hall et al. 2002; Kanoh et al. 2005; Hansen et al. 2006). This apparent discrepancy is due to the existence of redundant pathways for heterochromatin formation at these two other loci. At the mating type locus, the ATF/CREB family transcription factors Atf1 and Pcr1 facilitate heterochromatin formation by directly recruiting the chromatin modifying machinery (Jia et al. 2004; Kim et al. 2004) . Similarly, at the sub-telomeric regions, the telomere-binding protein Taz1 functions in an alternative pathway for formation of heterochromatin (Kanoh et al. 2005; Hansen et al. 2006). Despite the existence of redundant pathways, it has been demonstrated that RNAi is indeed required for the efficient establishment of heterochromatin at these loci. Interestingly, within all the three major heterochromatic loci, there are regions of homology (96% identical) that have been shown to act as nucleation centres for heterochromatin assembly via RNAi (Wood et al. 2002; Hall et al. 2002; Sadaie et al. 2004). Thus RNAi-targeted heterochromatin assembly seems to be a general mechanism for the initial seeding of chromatin modification at all the three major constitutive heterochromatic loci.

1.10 RNAi mediated co-transcriptional gene silencing at fission yeast centromere

According to the generally accepted model of heterochromatin formation at fission yeast centromeres (**Fig 1.5B**), the centromeric outer repeats are transcribed bidirectionally during S-phase by RNA Polymerase II (Pol II), giving rise to non-coding centromeric transcripts that are bound by the RNAi machinery(Djupedal et al. 2005; Kato et al. 2005; Chen et al. 2008; Kloc et al. 2008; Zhang et al. 2008). The centromeric dsRNAs are processed by the RNAse III-like enzyme Dicer to produce

approximately 21-nucleotide long short interfering RNAs (siRNAs) which are loaded onto the RITS (RNA induced transcriptional silencing) complex via the Argonaute-siRNA chaperone (ARC) complex (Verdel et al. 2004; Reinhart et al. 2002; Buker et al. 2007). The siRNA-bound RITS complex targets the nascent complimentary transcript and recruits the H3K9-methyltransferase complex CLRC (Clr4-containing complex), promoting H3K9 methylation and hence heterochromatin formation. RITS additionally recruits the RDRC (RNA dependent RNA polymerase) complex which promotes dsRNA production, ensuring an efficient positive feedback mechanism (Motamedi et al. 2004; Colmenares et al. 2007).

H3K9 methylation further acts as a platform for the association of structural proteins including the chromo-domain proteins Swi6 and Chp2 (HP1 homologues). The ability of Swi6 to self-associate is thought to be important in the spreading of heterochromatin via higher order oligomerisation, facilitating bridging of nucleosomes (Cowieson et al. 2000; Canzio et al. 2011). Swi6 and Chp2 are thought to be involved in recruitment of histone deacetylase complexes including SHREC (Snf2/HDAC-containing repressor complex) which plays crucial role in transcriptional gene silencing of the major heterochromatic loci (Yamada et al. 2005; Sugiyama et al. 2007; Fischer et al. 2009). Swi6 is additionally involved in cross-talk with Epe1, a negative regulator of heterochromatin, implying a role for Swi6 in defining heterochromatic domains (Ayoub et al. 2003; Zofall & Grewal 2006). Indeed, a recent study has challenged the view that Swi6 is required for maintenance and spreading of heterochromatin and instead suggests a potential role for Swi6 in demarcating constitutive heterochromatin, preventing spreading into the adjacent euchromatin (Stunnenberg et al. 2015). Further roles for Swi6 in recruiting cohesin and promoting loading of the centromere-specific

histone variant CENP-A^{cnp1} underpin the importance of pericentromeric heterochromatin (Fischer et al. 2009). Thus a plethora of molecular machineries seem to act in concert at centromeres in fission yeast to ensure chromosomal integrity.

Evidence so far is suggestive of RNAi-dependent and independent mechanisms of silencing of centromeric transcripts. The reverse centromeric transcript is silenced via RNAi, indicated by the accumulation of reverse transcripts in RNAi mutants. This silencing appears to be independent of structural components of heterochromatin such as Swi6, and hence to occur post-transcriptionally. However, silencing of the forward strand depends on Swi6 but is resilient to defective RNAi, indicative of transcriptional silencing (Volpe et al. 2002). Although classical post-transcriptional gene silencing (PTGS) involves degradation of mRNA in the cytoplasm (*trans*-silencing: the site of origin of the RNAi trigger is separate to its site of action), the mechanism by which RNAi silences centromeric transcripts appears to be different in that the RNAi response acts in *cis* on chromatin. Key observations supporting the idea of non-canonical RNAi-mediated silencing are: (1) RNAi components localise to the chromatin and are involved in recruitment of the chromatin modifying machinery; (2) RNAi mediated silencing in *S. pombe* functions efficiently only in *cis* (ie, unlike most PTGS systems, the dsRNA that triggers the RNAi response originates from the silenced locus); (3) Efficient production of centromeric siRNAs appears to be dependent on the H3K9-methyltransferase Clr4 (Noma et al. 2004; Sugiyama et al. 2005; Hong et al. 2005). The prominent *cis* response of RNAi in fission yeast was demonstrated by a tethering assay in which the RITS component Tas3 was tethered to RNA transcript of a normally active *ura4*$^{+}$ gene (Bühler et al. 2006). siRNAs generated from the tethered site were capable of silencing

complementary RNA only in *cis* (i.e. they failed to silence a second *ura4*$^{+}$ gene placed in a different chromosomal location implying failure to silence in *trans*) (Bühler et al. 2006). This chromatin-associated RNAi response, however, does not seem to affect RNA Pol II occupancy, implying an unusual, co-transcriptional silencing mechanism (Bühler et al. 2006). In addition to RNAi, the TRAMP complex which degrades aberrant RNA via the exosome is also involved in the proper silencing of pericentromeric repeat transcripts, though it does not affect the main structural features of heterochromatin including H3K9 methylation (Bühler et al. 2007). Thus, at fission yeast pericentromeres, it appears that RNAi-dependent and independent mechanisms which act on chromatin in *cis* act in concert for proper centromeric transcript silencing. To reflect the nature of these combined mechanisms, silencing at pericentromeres is referred to as either co-TGS or *cis*-PTGS (Martienssen & Moazed 2015).

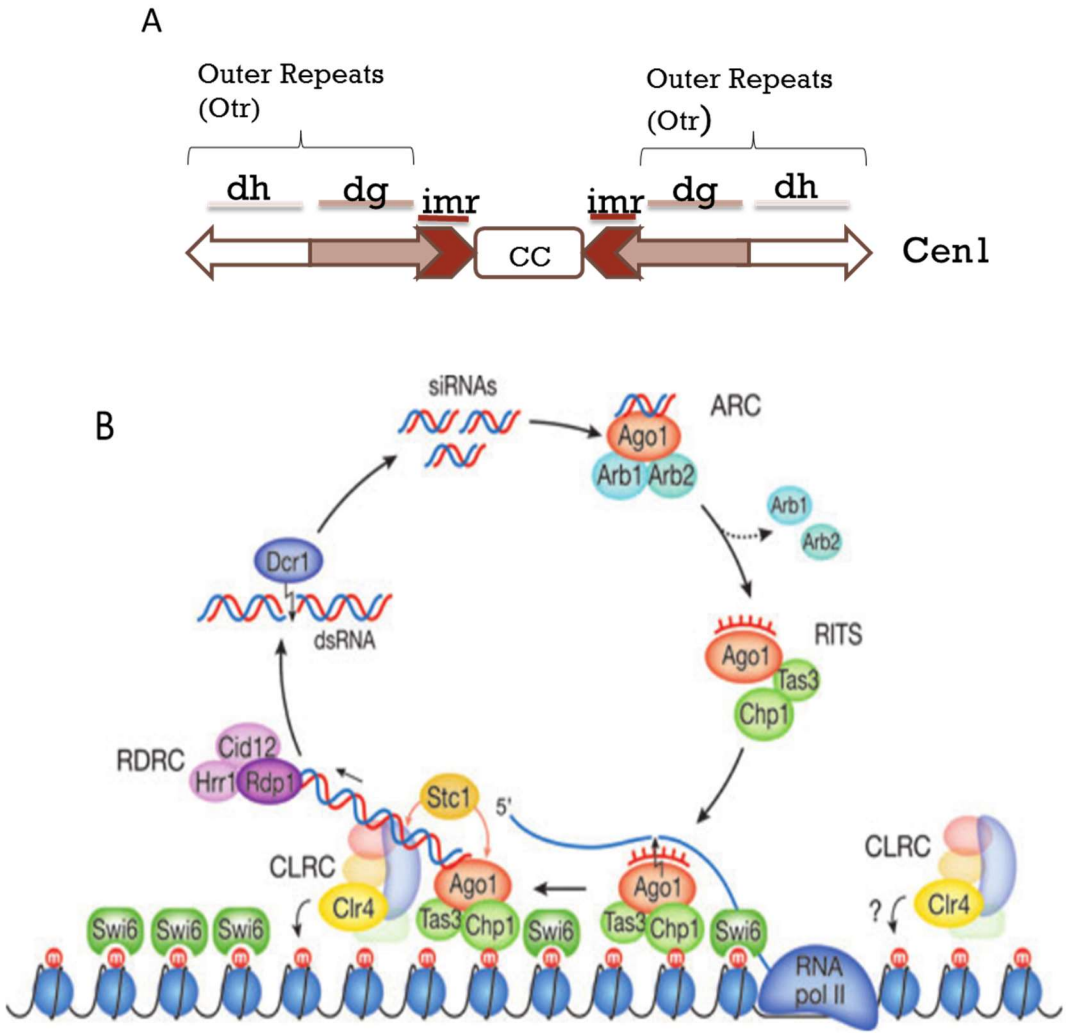


Fig 1.5 RNAi mediated heterochromatin formation at fission yeast centromere

(A) Schematic representation of centromere 1 (cen1). The central core domain consists of *cc* (central core) and *imr* (inner most repeats) regions upon which the kinetochore gets assembled and the outer repeat regions consists of the *dg* and *dh* repetitive elements that forms the heterochromatic pericentromere. The overall structure is similar to that of metazoans where the kinetochore is surrounded by heterochromatic region. **(B) RNAi mediated recruitment of chromatin modification machinery**. In fission yeast, the outer repeat sequences (otr) of the pericentomeric region flanking the central non-repetitive core are transcribed by RNA polymerase II (Pol II) during S phase, giving rise to double stranded RNA (dsRNA). These dsRNAs are processed by RNAse III-like enzyme Dicer to produce short interfering RNAs (siRNAs) which are loaded onto the RITS complex via the Argonaute-siRNA chaperone (ARC) complex. The siRNA-bound RITS complex targets the nascent complimentary transcript and also recruits the methyltransferase complex CLRC (Clr4 -containing complex), promoting H3K9 methylation and hence heterochromatin formation. RNAi-independent mechanisms are also additionally thought to act at the pericentromeres for CLRC recruitment. More details in the text. **Figure 1.5B** adapted from Goto and Nakayama (2012).

1.10.1 The RITS complex

The RNAi-effector RITS (RNA induced transcriptional silencing) complex consists of the Argonaute family member Ago1, a chromodomain protein Chp1 and the GW (Glycine-Tryptophan)-repeat protein Tas3 (Verdel et al. 2004; Partridge et al. 2007). Ago1, the sole Argonaute protein in fission yeast belongs to the AGO clade and consists of N-PAZ and MID-PIWI domains. The PIWI domain harbours the endonuclease activity (slicing) required for cleaving RNA targets, and additionally required for the release of the siRNA passenger strand; the latter licenses Ago1-associated RITS for target recognition. RITS localises to centromeres via Ago1 in an siRNA-dependent manner and the centromeric association is thought to be reinforced by chromatin association through Chp1, which bears a chromodomain that binds to H3K9-methyl marks (Verdel et al. 2004; Partridge et al. 2002; Noma et al. 2004). RITS localisation strongly correlates with peaks of siRNAs and H3K9-methylation across all three major heterochromatic loci in *S. pombe.* Moreover, RITS has been shown to delocalise when Clr4, the sole H3K9-methyltransferase, is removed (resulting in a lack of H3K9-methyl mark for RITS binding via Chp1) (Cam et al. 2005). Similar to the role of Chp1 in reinforcing RITS association with chromatin, it has been proposed that in the miRNA pathway, additional proteins might be involved in stabilising the interaction of miRNA-bound AGO proteins with mRNA, since the association of AGO proteins with their target mRNA is only indirect via sRNA-interaction and therefore might not be sufficiently stable (Meister 2013).

The Tas3 subunit of RITS has functional similarity to GW proteins, which bind AGO proteins through a conserved GW repeat-motif. It has been shown that Tas3 binds to Ago1 via the 'Ago-hook', which consists of GW repeats and shows sequence

similarity to the GW-protein TNRC6A (trinucleotide repeat-containing gene 6A) implicated in PTGS in humans (Partridge et al. 2007; Till et al. 2007; Meister 2013). Besides the Ago-hook motif in Tas3, the rest of the protein has no known functional domains. It has been suggested that Tas3 could act as a sensor for Ago1-loading, based on the observation that Ago1 defective in siRNA-association fails to interact with Tas3 (Holoch & Moazed 2015a). Structural studies on yeast and human AGO proteins indicate distinct proximal interaction interfaces for siRNA binding and GW-protein binding within the PIWI domain (Nakanishi et al. 2012; Schirle et al. 2012). These structural data further support the possibility that GW proteins like Tas3 could act as sensors for siRNA-loaded Ago1, thereby ensuring fidelity of RITS in target recognition.

1.10.2 The ARC complex

The ARC complex is thought to be involved in the process of RITS maturation (Buker et al. 2007). Besides Ago1, the ARC complex consists of Arb1 and Arb2; while Arb1 is conserved within fungi, Arb2 seems to be conserved from fission yeast to humans. However, none of the putative homologues of Arb2 are characterised. Arb1 and Arb2 have no known functional domains, except that the C-terminus of Arb1 resembles organellar maturases involved in self-splicing of introns (Buker et al. 2007). Assembly of Ago1 in the ARC complex appears to inhibit Ago1-slicing activity, since ARC is mostly associated with duplex centromeric siRNAs. The ARC complex is therefore proposed to be involved in the delivery of duplex siRNA to Ago1 (**Fig 1.5B**) (Buker et al. 2007). Recent evidence confirms that ARC complex acts upstream of RITS, based on the requirement for ARC subunit Arb1 in siRNA-loading of Ago1, and the

fact that Ago1 siRNA-loading is a prerequisite for RITS assembly (Holoch & Moazed 2015a).

1.10.3 The RDRC complex

Though primary siRNA generation is thought to occur as a result of bidirectional transcription of the centromeric repeats (Volpe et al. 2002), this alone is not sufficient to support the efficient and dynamic silencing processes acting on pericentromeres. Indeed, through the discovery of the ping-pong cycle of piRNA amplification in *Drosophila,* it is well understood that amplification loops for continual sRNA generation are critical for efficient RNAi-based silencing (Czech & Hannon 2016). The RNA-dependent RNA polymerase Rdp1 associate with the RNA helicase Hrr1 and the polyA polymerase family protein Cid12 to form the RDRC complex (**Fig 1.5B**), the crucial RNAi complex required for secondary siRNA generation in *S. pombe*. All three RDRC components are required for centromeric silencing and act in *cis* at centromeric chromatin (Volpe et al. 2002; Motamedi et al. 2004). Centromeric transcripts generated by RNA Pol-II form the template for RDRC for generation of more dsRNA. This dsRNA gets rapidly processed by Dcr1 to produce secondary siRNAs, thereby ensuring a continuous supply of siRNAs for licensing of Ago1 (RITS) (Volpe et al. 2002; Djupedal et al. 2005; Sugiyama et al. 2005). A subset of splicing factors have also been implicated in the RDRC-dependent siRNA generation process, although their function is unclear as heterochromatin formation does not require the splicing process or the spliceosome (Bayne et al. 2008; Bernard et al. 2010). RITS and RDRC associate physically and this association is dependent on Dcr1 and the H3K9-methyltransferase Clr4. Additionally, RITS and RDRC localisation to

centromeric chromatin is inter-dependent (Motamedi et al. 2004; Sugiyama et al. 2005), implicating a tight feedback loop mechanism in siRNA generation.

1.10.4 The CLRC complex

Efficient siRNA production from centromeric repeats requires Clr4, the sole H3K9-methlytransferase in *S. pombe* (Noma et al. 2004; Motamedi et al. 2004). Clr4 is observed to be in complex with Rik1, Dos2/Raf2, Dos1/Raf1, the RING-finger protein Pip1 and the ubiquitin ligase scaffold family protein cullin 4 (Pcu4/Cul4). This Clr4-containing complex, CLRC, is crucial in heterochromatin formation. Consistent with its critical role, deletion of CLRC subunits results in accumulation of centromeric transcripts, impaired siRNA production and complete loss of H3K9 methylation (Sadaie et al. 2004; Horn et al. 2005; Hong et al. 2005; Li et al. 2005; Jia et al. 2005; Thon et al. 2005). CLRC subunits colocalise to the major heterochromatic domains in *S. pombe;* however, some of the subunits, Rik1 and Cul4 in particular appear to have additional target sites based on genome-wide profiling studies, suggesting potential additional roles for these proteins independent of their function as part of CLRC (Zhang et al. 2008).

The subunit organisation of CLRC is thought to resemble the CRL4 complex in humans, which contains the cullin Cul4 that forms a scaffold for the E2-ubiquitin-conjugating enzyme, bringing E2 close to its substrate. Within the CRL4 complex, Cul4 interacts with a RING finger protein that associates with the E2-enzyme, and with DDB1(DNA damage binding protein 1), which acts as an adaptor and recruits WD-40-containing substrate receptor proteins known as DCAFs (DDB1-Cul4

associated factors) (Fischer et al. 2011). DCAFs subsequently recognise specific substrates to be presented to the E2-enzyme. A well characterised DCAF is DDB2, which binds to UV-damaged lesions in DNA as part of the $CRL4^{DDB2}$ complex's role in DNA damage sensing (Fischer et al. 2011). A similar organisation can be observed in CLRC, where Cul4 associates with the RING finger protein Pip1 and the WD-40 protein Rik1. The Dos1 subunit of CLRC is thought to act as the DCAF for the DDB1-like protein Rik1, and hence potentially present a yet-to-be identified substrate for modification (Kuscu et al. 2014; Buscaino et al. 2012). It is interesting to note that in the case of $CRL4^{DDB2}$ complex, the substrate bound by the DCAF DDB2 (i.e. DNA), is not the target for modification. Instead the association of $CRL4^{DDB2}$ with DNA results in histone ubiquitination at sites of DNA damage, indicating that the substrate associated with the DCAF does not have to be the target for ubiquitin ligase activity (Scrima et al. 2011).

In comparison with human $CRL4^{DDB2}$ subunit organisation, a marked difference in the fission yeast CLRC complex is the presence of a non-analogous subunit, Dos2. Dos2 interacts with all other CLRC subunits except Clr4, thereby acting as the hub of CLRC (Kuscu et al. 2014). Dos2 contains an N-terminal replication foci targeting signal (RFTS) similar to that found in the DNA methyltransferase DNMT1 and harbours a C2H2-type zinc finger at the C-terminus. Interestingly, a CLRC-independent role for Dos2 and Rik1 has been proposed in the regulation of centromeric transcription during S-phase, operating in association with Cdc20 (DNA Polymerase-ε) and the transcription factor Mms19 at replication forks (Li et al. 2011). Further molecular characterisation is required to reveal any potential role for the RFTS domain in targeting Dos2 and associated proteins to the replication fork. Notably, the CLRC

complex has been shown to exhibit E3 ubiquitin ligase activity on histone H2B *in-vitro* (Horn et al. 2005). However, the functional significance of the observed ligase activity of CLRC in heterochromatin formation *in-vivo* is not yet clear.

1.11 Establishment of constitutive heterochromatin

Elucidating the order of events in *de novo* formation of heterochromatin has proved difficult because RNAi and chromatin modification mechanisms are functionally tightly connected in *S. pombe.* This is evidenced by the observations that defective RNAi abrogates H3K9 methylation at heterochromatic regions, and conversely, that lack of Clr4 impairs siRNA production. Consequently, different studies have led to the proposal of two opposing models for the initiation of heterochromatin assembly: (1) H3K9 methylation and hence CLRC association as the initial trigger with a downstream role for RNAi; and (2) RNAi as the priming factor for subsequent recruitment of H3K9-methyltransferase activity (CLRC).

The first model for *de novo* heterochromatin formation argues that H3K9 methylation is the initial trigger essential for the downstream recruitment of RNAi. *De novo* assembly via RNAi is thought to be a highly stochastic and inefficient process and hence without prior chromatin stabilising entities like H3K9 methylation, RNAi-associated processes are proposed to be insufficient to act as the initial trigger for establishment of heterochromatin (Noma et al. 2004; Sugiyama et al. 2005; Allshire & Ekwall 2015). Experiments utilising a Tas3 mutant (that retains Chp1 association but not Ago1 interaction) demonstrated that heterochromatin establishment requires association of Ago1 with chromatin via H3K9-methyl association via Chp1 (Partridge et al. 2007). Several other lines of evidence, support the proposed ability of the H3K9-

methyl mark to establish heterochromatin in the absence of an RNAi-trigger, including the observations that (1) residual H3K9 methylation is observed at centromeres in RNAi mutant background (Partridge et al. 2007; Bayne et al. 2010), (2) a subset of RNAi-deficient cells can eventually form heterochromatin at mating type locus over a few generations (Hall et al. 2002) and, (3) centromeric siRNAs are lost in the absence of Clr4 (Noma et al. 2004; Nakayama et al. 2001).

On the contrary, recent evidence for the presence of low levels of centromeric siRNAs in CLRC-mutant backgrounds, and a potential role for Clr4 in siRNA generation independent of H3K9 methylation challenge the view of absolute requirement of H3K9 methylation for RNAi (Bayne et al. 2010; Gerace et al. 2010). The RNAi-dependent recruitment model supports the view that non-coding centromeric transcripts act as a scaffold for RITS recruitment and that RNAi acts as the initial targeting mechanism for localising CLRC to chromatin to initiate H3K9 methylation (Allshire & Ekwall 2015; Martienssen & Moazed 2015). The role of RNAi in recruiting CLRC was demonstrated by artificially tethering the RITS component Tas3 to a nascent transcript at a euchromatic locus. Tethering of Tas3 resulted in transcriptional silencing via heterochromatin formation that was dependent on Clr4, Swi6 and Sir2 (HDAC implicated in H3K9 and H4K16 hypoacetylation) (Bühler et al. 2006; Shankaranarayana et al. 2003). However, a key question remained in this model as to how the first, 'primary' siRNAs are generated. The observation that centromeric siRNAs can be detected in the absence of Rdp1 (RDRC component involved in the positive feedback mechanism) indicated the existence of primary siRNAs (Djupedal et al. 2009). Recently, advances in high-throughput sequencing enabled the discovery of a class of Ago1-associated Dcr1-independent sRNAs called primal RNAs

(priRNAs), which has been proposed to act as the initial targeting signal for RITS and hence as a trigger for *de novo* heterochromatin assembly (Halic & Moazed 2010). These priRNAs are proposed to be generated as a result of degradation of centromeric transcripts, and are thought to be subsequently loaded onto Ago1 in a non-specific fashion. However, further characterisation is required to validate the Dcr1-independent heterochromatin assembly pathway. An alternative and well received model is that of Dcr1-dependent primary siRNA production. Secondary structure probing experiments revealed that centromeric transcripts are partially double-stranded and that these dsRNA regions can be cleaved by Dcr1 *in-vitro* (Djupedal et al. 2009). It is therefore proposed that RNA Pol II transcripts derived from the centromeric repeat elements fold onto themselves to form hairpin-like structures *in-vivo* which serve as the template for Dcr1-mediated primary siRNA generation (Djupedal et al. 2009). The significant role of RNAi in establishing heterochromatin is also supported by the additional requirement of RNAi for the *de novo* assembly of heterochromatin at mating type locus and sub-telomeres. Although redundant pathways exist for the maintenance of heterochromatin at these loci (see section 1.9), RNAi recruited to nucleation sites that share homology to pericentromeric outer repeats has been shown to be essential for efficient *de novo* assembly of heterochromatin (Grewal & Jia 2007).

1.12 Stc1

A systematic genetic screen for components involved in centromeric silencing (Bayne et al. 2014) led to the discovery of Stc1, a protein later implicated in RNAi-mediated centromeric heterochromatin formation (Bayne et al. 2010). Stc1 is required for establishing heterochromatin at centromeres, mating type locus and telomeres. Similar to RNAi components, Stc1 is not required for maintenance of heterochromatin at mating type locus and telomeres, however, it is indispensable for the maintenance of centromeric heterochromatin. Thus, the *stc1* mutant has an RNAi-component-like phenotype. Several key lines of evidence suggested a potential function for Stc1 as a linker that connects RITS and CLRC (Bayne et al. 2010): (1) Stc1 associates with both RITS and CLRC *in vivo*; (2) Stc1 directly interacts with Ago1 *in-vitro;* (3) In the absence of Stc1, Ago1 fails to associate with Clr4 *in vivo*. Notably, when tethered to a euchromatic locus, Stc1 has been shown to recruit CLRC to chromatin, independently of RNAi. This suggested that (1) Stc1 is capable of recruiting CLRC to chromatin; (2) RNAi is required upstream of Stc1 localisation to chromatin; and (3) Association of Stc1 with chromatin-bound RITS could serve as the localisation determinant for CLRC. Thus, Stc1 is envisaged as the key factor mediating the RNAi-dependent recruitment of CLRC to chromatin (**Fig 1.5B**). In addition to this bridging role, it is thought that Stc1 may have a further function in the RNAi pathway, since unlike for CLRC components, deletion of Stc1 results in complete loss of siRNAs. However, intriguingly, an investigation of dependency relationships revealed that Stc1 association with RITS is CLRC dependent while Stc1-CLRC association is independent of RNAi (Bayne et al. 2010).

Stc1 is a 25kDa protein, well conserved in the fission yeast clade but with no known homologs in higher eukaryotes. While the Stc1-N terminus shows homology to other fungal proteins, no sequence homology has been found for the Stc1 C-terminus (Bayne et al. 2010). A collaborating group utilising NMR (Nuclear magnetic resonance), revealed that the N-terminus of Stc1 harbours a tandem zinc finger (ZF) domain similar to a LIM-domain. However, the Stc1 ZF domain differs from a LIM-domain in that the individual zinc fingers are linked by a flexible linker such that each of the zinc fingers has the potential to fold separately. The two fingers could therefore act as two distinct units (He et al. 2013). The Stc1-C terminus remained unstructured in solution and hence no structural information is available for this domain.

1.13 Aims of this study

Stc1 acts as the key factor which connects RNAi to chromatin modification in fission yeast. Understanding the interactions of Stc1 is vital in improving current knowledge of how histone modifying complexes and the RNAi pathway are integrated in order to ensure targeting of histone modifications to precise genomic regions. Insights gained from detailed studies on the interactions of Stc1 and Stc1-containing complexes in fission yeast should also help us better understand the missing factors involved in heterochromatin formation in other systems. In order to understand how Stc1 functions in the RNAi-mediated heterochromatin pathway, I focussed on two major aims:

Aim 1

How does Stc1 act as a critical link between RITS and CLRC?

Stc1 has been shown to physically associate with RITS via Ago1 (Bayne et al. 2010). However, the basis of this interaction was unknown, and, although Stc1 co-immunoprecipitates CLRC components, it was unclear if and how Stc1 directly associates with CLRC. Utilising *in-vitro* methods, I aimed to determine whether Stc1 interacts directly with any of the CLRC components, and identify the region(s) of Stc1 involved in specific interactions with RITS and CLRC components. I additionally aimed to study Stc1-RITS interaction and Stc1-specific complexes *in-vivo* with the ultimate aim of characterising the RITS-Stc1-CLRC bridging complex.

Aim 2

How does CLRC interact with Stc1?

The findings from **Aim 1** indicated Dos2 and Clr4 as the potential direct binding partners of Stc1 amongst the CLRC subunits. To obtain a better picture of the molecular basis and functional significance of Stc1-CLRC interactions, I aimed to investigate the regions/residues of Dos2 and Clr4 involved in Stc1-binding *in-vitro*.

CHAPTER 2

Materials and methods

2.1 Yeast culture and media

S. pombe cultures were grown at 32°C until log phase ($2x10^6$ cells/ml – $1x10^7$ cells/ml) as indicated for each experiment. Approximate generation times for *S. pombe* cells in commonly-used growth media are shown in **Table 2.1** and list of strains used are shown in **Table 2.11**.

Medium	**Temperature °C**	**Generation time**
YES	32	2 hours 10 minutes
4x YES	32	2 hours 30 minutes

Table 2.1 Generation times for *S. pombe* haploid strains

2.1.1 Growth media

All solutions were prepared according to Moreno *et al.,* Nurse, and Forsburg and Rhind ((Nurse 1975; Moreno et al. 1991; Forsburg & Rhind 2006)) by the University media service. Details of the media and supplements used are shown in **Table 2.2**.

Yeast extract was from DIFCO. Agar and malt extract were from OXOID while all amino acids were from Sigma-aldrich. Yeast strains carrying specific resistance markers are cultured in media containing the respective antibiotic resistance. Media for bacterial cultures were obtained from the University media service.

YES Agar (-ade) (1L):	Yeast extract	5.0g
	D-glucose anhydrous	30.0g
	Arginine	0.2g
	Lysine	0.2g
	Histidine	0.2g
	Uracil	0.2g

	Leucine	0.2g
	Agar	20.0g
YES Liquid (1L):	Yeast extract	5.0g
	D-glucose anhydrous	30.0g
	Arginine	0.2g
	Lysine	0.2g
	Histidine	0.2g
	Uracil	0.2g
	Leucine	0.2g
4x YES Liquid:	All reagents as above 4x	

ME plates (1L):	Malt extract	30.0g
	Adenine	250.0g
	Arginine	250.0g
	Histidine	250.0g
	Uracil	250.0g
	Leucine	250.0g
Vitamins 1000x (100ml)	Pantothenic acid	0.5g
	Nicotinic acid	1.0g
	Inositol	1.0g
	Biotin	1mg
	Filter sterilised	
Minerals 10,000X (100ml)	Boric acid	5.0g
	$MnSO_4$	4.0g
	$ZnSO_4$	4.0g
	$FeCl_2.6H_2O$	2.0g
	Molybdic acid	1.6g
	$CuSO_4.5H_2O$	0.4g
	Citric acid	10.0g
	Filter sterilised	

Salts 50X	Magnesium chloride	53.5g
	Calcium chloride	1.0g
	Potassium chloride	50.0g
	di-sodium sulphate	2.0g
Supplement stocks/L	Adenine 50X	5.0g
	Arginine 100X	10.0g
	Histidine 100X	10.0g
	Uracil 20X	2.0g
	Leucine 100X	10.0

Additional Supplements	Nourseothricin (cloNAT) (Werner BioAgents)	0.4mg/ml
	Geneticin (G418) (Gibco)	0.1mg/ml

Table 2.2 Composition of yeast growth media and supplements used

2.1.2 Cell culture

Cells cultured in YES media were harvested during log phase, between $6x10^6$ cells/ml and $1x10^7$ cells/ml, as indicated for each experiment. This was achieved by inoculating 50ml of media with a loop full of freshly growing yeast and growing in a shaking incubator overnight at 32˚C (36˚C for experiments involving temperature sensitive mutants). This starter culture was subsequently diluted to the required volume and allowed to reach log phase and harvested at the required concentration. For optimal growth conditions, all cultures were grown in flasks of atleast double the volume of culture required.

2.1.3 Cell counting

Cell counting was done using a haemocytometer. A haemocytometer is a calibrated microscopic slide with grids etched onto the counting chamber. The grid consists of 25 large squares which are further subdivided into 16 smaller squares. The grid area is $1mm^2$ and has a depth of 0.1mm once a cover slip is placed over the grid. This forms a known volume of $0.1mm^3$. 10µl of cell suspension is pipetted and taken up into the grid by capillary action. The number of cells/ml can then be calculated by counting the number of cells in 25 large squares, and multiplying it by $1x10^4$.

2.2 *S. pombe* Molecular Genetics

2.2.1 Mating and random spore analysis

Crosses were done on ME medium to induce sporulation by nitrogen starvation (Forsburg & Rhind 2006). Cells of opposite mating type (h+/h-) were mixed together on ME agar plates and left for 1-2 days at 25˚C. The cells were checked under light microscope for the presence of asci. Cells were resuspended in 300µl 1/100 diluted Glusulase (Perkin Elmer) and either incubated for 6 hours at 37˚C or overnight at 32˚C. This glusulase treatment leads to the digestion of vegetative cells and asci walls, leaving behind intact spores. After incubation in Glusulase, the spores were resuspended in 300µl dH_2O and dilutions ranging from 1/10 – 1/200 were plated on appropriate selective media and grown at 32˚C until colonies formed.

2.2.2 C-terminal tag integration

C-terminal tagging of Stc1 and Dos2 was achieved by homologous recombination using DNA fragments carrying the tag sequence flanked by 80-100 nucleotides homologous to the target site. DNA fragments for epitope tagging were generated using PCR amplification of Bahler cassettes from the *PFA6a* plasmid series (Bähler et

al. 1998). Proper insertion at the desired locus was always confirmed by PCR and/or sequencing.

2.2.3 Lithium Acetate Transformation

A 50ml culture was grown to log phase ($5x10^6$ – $8x10^6$ cells/ml) in YES medium and harvested by centrifugation at 3000rpm for 3 minutes. Cells were washed once with dH_2O and then resuspended in 10ml 0.1M lithium acetate in TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA) pH 4.95. Following incubation for 1 hour at 32˚C, cells were pelleted and resuspended in 0.1M lithium acetate in TE buffer pH 4.95 at a concentration of $1x10^9$/ml. 3µg of DNA was mixed with 150µl of cells followed by the addition of 370µl 50% PEG-3350 (Sigma) in TE buffer. Cells were then incubated for an hour at 32˚C, heat-shocked at 42˚C for 20 minutes and collected by centrifugation. The cell pellets were resuspended in 6ml of fresh YES media and left to grow for 6 hours or overnight at 32˚C for recovery before being plated onto YES plates containing appropriate antibiotics.

2.2.4 Centromeric silencing assay

Reporter genes inserted at sites of heterochromatin are known to be subject to silencing, and this has been utilised as a method to assay heterochromatin integrity. I used *ade6*$^+$ as the reporter gene for centromeric silencing assays, as this provides a convenient colour read out for the presence or absence of silencing. *ade6*$^+$ is inserted at the pericentromeric outer repeats, specifically on the right side of centromere 1 at the *SphI* site (otr1R(*SphI*):*ade6*$^+$; **Fig 2.1**) (Allshire et al. 1994; Ekwall et al. 1997). Cells carrying a mutation in the endogenous *ade6*$^+$ gene (*ade6-210*) were used as a background for these assays, thus ensuring that the centromeric reporter gene is the only source of functional Ade6. In WT cells, the *ade6*$^+$ gene inserted at the centromeres is silenced, and this renders the cells red in colour when grown under

limited adenine conditions. This is due to a block in the adenine biosynthesis pathway which causes accumulation of the red pigmented amino-imidazole ribonucleotide (Fisher 1969). In contrast, in mutants in which heterochromatin is disrupted, the *ade6+* gene is expressed and cells are white. A 10-fold serial dilution plating assay was employed to assess colony colour.

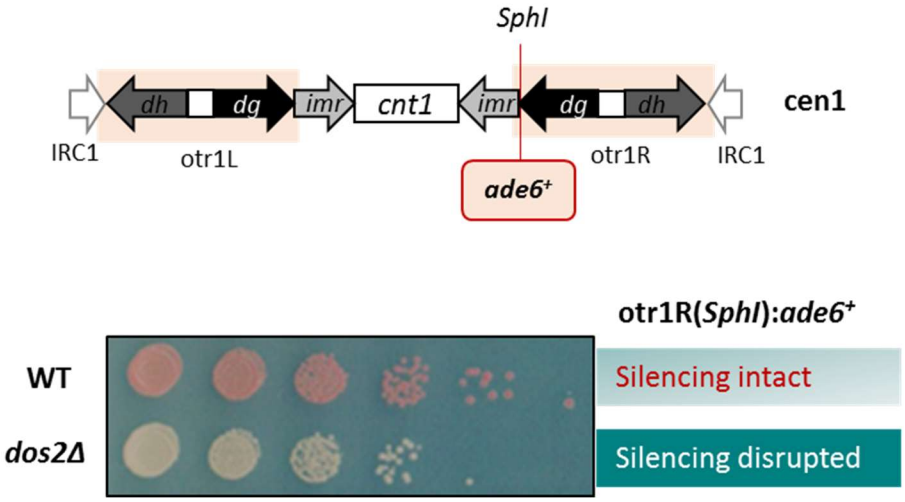


Fig 2.1: Centromeric silencing assay

The diagram is a schematic representation of the *ade6*$^{+}$ insertion at the *SphI* site on the right side of centromere 1 (cen1). *cnt1* is the central core; *imr* are inner most repeats; otr1R and otr1L represent the outer repeat regions on the right and left arms of cen1, consisting of tandem repeats of *dg* and *dh* elements. IRC1 are boundary elements required for demarcating the heterochromatic region (shaded region) from the nearby euchromatin. Below, the schematic, is an example of a centromeric silencing assay where wild type (WT) and mutant (*dos2Δ*) cells have been plated on low adenine medium.

2.3 Molecular Cloning

2.3.1 Plasmid construction

For generating Stc1-CFP-Nat and Dos2-YFP-Kan strains, I created *pFA6a-YFP-Kan* and *pFA6a-Cerulean-Nat* plasmids. CFP and YFP fragments were cloned (from *pCerulean-C1* and *pEYFP-C1* respectively) by replacing *GFP* in *PFA6a-GFP* plasmids bearing Nourseothricin and Kanamycin resistance markers respectively, using *PacI/AscI* restriction sites.

clr4-CD (chromo-domain) alanine-scanning mutations were generated in *pGBK-clr4* using QuickChange Lightning Site Directed Mutagenesis Kit (Stratagene) following manufacturer's instructions. *clr4* truncation plasmids were constructed in *pGBK* by cloning *clr4* fragments generated by PCR using *pGBK-clr4* as template. *Swi6* plasmids were created in pGAD vector by cloning DNA encoding either Swi6 full length or CD using *pREP41-swi6* as template. PCR fragments with overhanging restriction sites (*Nde/Xma*) were amplified in the case of all *clr4* and *swi6* plasmid constructions using Platinum *Pfx* Taq DNA polymerase (Invitrogen) according to manufacturer's instructions. Plasmids are listed in **Table 2.10**.

2.3.2 Generation of chemically competent bacterial cells

A single fresh DH5α bacterial colony was inoculated in 5ml LB and cultured overnight at 37ºC. The culture was diluted 1:200 in pre-warmed 100ml LB containing 20mM $MgSO_4$ and cultured until OD_{600} reached 0.4-0.6. Following incubation on ice for 10 minutes, cells were pelleted at 5000rpm at 4ºC for 5 minutes and gently resuspended in 40ml (per 100ml of culture) of ice-cold TFB1 buffer (300mM potassium acetate, 100mM RuCl, 10mM $CaCl_2$, 50mM $MnCl_2$, 15% glycerol, pH 5.8 with acetic acid) and incubated on ice for 5 minutes. Cells were harvested at 3000rpm at 4ºC for 10 minutes and gently resuspended in 4ml (per 100ml of culture) of ice-cold TFB2 buffer

(10mM MOPS, 10mM RuCl, 75mM $CaCl_2$, 15% glycerol, pH 6.5 with KOH) and incubated on ice for 15 minutes. 50µl of cells were aliquoted into pre-chilled Eppendorf 1.5ml tubes and stored at -80ºC.

2.3.3 Bacterial transformations

50µl of chemically competent DH5α were thawed on ice, and incubated with 10ng of plasmid on ice for 30 minutes. Cells were heat-shocked at 42ºC for 30s and incubated on ice for 2 minutes. 250µl pre-warmed SOC broth (Invitrogen) was added and cells grown at 37ºC for 1 hour before plating on LB agar supplemented with the appropriate antibiotic (50mg/ml carbenicillin or 50mg/ml kanamycin).

2.3.4 Plasmid miniprep

Single bacterial colonies were inoculated in 5ml of LB containing the appropriate antibiotic and cultured overnight at 37ºC. Plasmids were extracted from the bacterial cells using QIAprep Miniprep Kit (Qiagen) based on manufacturer's instructions.

2.4 DNA Protocols

2.4.1 Polymerase Chain Reaction (PCR)

2.4.1.1 Pfx PCR

For generation of DNA fragments with high fidelity required for molecular cloning, yeast transformation or genotyping, the proof reading Platinum *Pfx* Taq DNA polymerase was used according to manufacturer's protocol and the PCR program provided in **Table 2.3**. 1-5µl of PCR products were analysed by agarose gel electrophoresis and the rest were gel-purified using a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's protocol.

Denature		95ºC, 2min
Denature		95ºC, 30s
Anneal	35 cycles	50ºC, 45s
Extend		68ºC, 1min/Kb
Final extension		68ºC, 10min

Table 2.3: Pfx PCR program

2.4.1.2 Colony PCR

For the selection of desired clones following yeast genetic crosses or transformations, colonies were initially screened by PCR using genomic DNA template generated by simple yeast lysis. A tiny blob of yeast was resuspended in 10µl of 20mM NaOH and incubated at 95ºC for 10 minutes for cell lysis. 1µl of this lysis solution served as the source for template in a 20µl PCR reaction using Taq DNA Polymerase (Roche). Reactions were set up as shown in **Table 2.4.** All PCR reactions were performed using a T3000 Thermocycler (Biometra), using the conditions shown in **Table 2.5.** A list of DNA oligonucleotides used are provided in **Table 2.9**.

10X Buffer with $MgCl_2$	2µl
10mM dNTP mix	0.3µl
10µM forward primer	0.3µl
10µM reverse primer	0.3µl
Taq (5U/µl)	0.1µl
Water (Total 20µl volume)	16µl

Table 2.4: Colony PCR reaction set up

95ºC		2min
95ºC		30s
50ºC	35 cycles	30s
72ºC		1min
72ºC		10min

Table 2.5: Taq-colony-PCR program

2.4.2 Agarose gel electrophoresis

1% agarose (Melford) was prepared in 1X TBE buffer and ethidium bromide (Sigma) was added to a final concentration of 0.2µg/ml to visualise DNA. A safer alternative to ethidium bromide, SYBR Safe DNA Gel Stain (Thermofisher), was also used at 1:10,000 dilution.

2.4.3 Sanger sequencing

Sequencing reactions were set up using BigDye Terminator Cycle v3.1 Cycle Sequencing RR-100 (Applied Biosystems) as shown in **Table 2.6.** The reaction programme used is shown in **Table 2.7**. Sequencing reactions were sent to the Edinburgh Genomics sequencing facility for analysis.

5X sequencing buffer	2.0µl
10µM primer	3.2µl
Template DNA	x µl (250ng plasmid, 70ng PCR product)
BigDye	1.0
water	10.0µl

Table 2.6: Sanger sequencing reaction

95ºC		5min
95ºC		30s
50ºC	25 cycles	20s
60ºC		4min
60ºC		1min

Table 2.7: Sanger sequencing PCR program

2.5 Protein Protocols

2.5.1 Recombinant protein purification

PGEX-4T1-constructs (**Table 2.10**) were transformed into *E. coli* BL21 derivative Rosetta2(DE3)*pLysS* (Novagen) according to manufacturer's protocol. The

transformed cells were grown at 37°C in 2xYT medium until OD600 reached 1.0, cooled on ice, and then induced by adding 0.2mM isopropyl-β-D-thio-galactoside (IPTG) to the medium. After overnight incubation at 16°C, the bacteria were harvested by centrifugation Bacterial lysis was performed using Bugbuster protein extraction reagent (Novagen) according to the manufacturer's protocol. 10ml Bugbuster solution containing Benzonase nuclease (25units/ml Bugbuster), rLysozyme (1KU/ml Bugbuster) and Protease inhibitor cocktail (Sigma) was used to resuspend 1g of bacterial cells (wet weight). The cell suspensions were incubated at room temperature for 30 minutes followed by centrifugation at 16,000g for 20minutes at 4ºC and supernatant collected. GST-tagged proteins were purified from the soluble cell lysate on glutathione- sepharose-4B (GE-Healthcare). 250µl sepharose beads (prewashed with 1:10 diluted Bugbuster solution) were used per 10ml soluble lysate and incubated for 45 minutes at 4ºC to allow binding. GST-protein bound beads were then collected by centrifugation at 500g for 5 minutes at 4ºC. Beads were subjected to three rounds of washes using 1:10 diluted Bugbuster solution and then loaded onto Poly-Prep chromatography columns (Bio-rad). GST- proteins were eluted using 15mM reduced glutathione in 1M Tris HCl pH 8.0. Eluates were analysed by SDS-PAGE and coomassie staining. Elutions containing GST-proteins were pooled and subjected to further purification and buffer exchange using Vivaspin columns (10K cut off; Starsted).

2.5.2 Protein immunoprecipitation

S. pombe strains were cultured in 50ml YES to log phase ($5x10^6$ – $8x10^6$ cells/ml). Pelleted cells were washed once with distilled water and stored at -80°C until processed. Stored pellets were thawed on ice and resuspended in 500µl ice-cold lysis buffer (50mM Hepes-NaOH pH7.5, 150mM NaCl, 5mM EDTA, 0.1% NP-40)

containing freshly added 5mM DTT, 1X protease inhibitors (EDTA-free, Roche), 1X Proteinase Inhibitors Cocktail (Sigma), 1mM PMSF and 0.2mM benzamidine. Cells were lysed using 500µl acid-washed glass beads (425-600µm, Sigma) on a bead-beater (BioSpec Products) for 3 minutes. For lysate collection, the bottom of each tube was pierced with a 25G needle, inserted into another tube and the assembly centrifuged at 1000rpm for 1 minute at 4°C. The recovered lysate was subjected to two rounds of centrifugation at 13,000rpm for 10 minutes at 4°C to remove cell debris. The crude lysate was pre-cleared with 25µl Protein-G-Agarose (Roche) for 30 minutes at 4°C. Samples were spun at 500g for 2 minutes at 4°C and the supernatant transferred to a new tube. A 50µl aliquot of sample was stored as 'input' control for the experiment and the remaining sample mixed with 25µl of Protein-G-Agarose slurry and 1µl of anti-flag (Sigma) or anti-c-myc (Roche) antibody. Samples were incubated for approximately 2 hours at 4°C with rotation. The beads were then collected by centrifugation at 500g for 2 minutes at 4°C and washed twice with ice-cold lysis buffer. Proteins were eluted by resuspending beads in 20µl 2X SDS sample buffer (60mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.2% bromophenol blue, 100mM DTT) and incubating at 65°C for 20 minutes. Denatured proteins were analysed by SDS-PAGE (section **2.5.4**)

For protein co-immunoprecipitation (CoIP) experiments in Chapter 3 (**Fig 3.9**) and Chapter 4 (**Fig 4.1** and **Fig 4.3 A** and **C**), *S. pombe* strains grown to a density of $1x10^8$ cells/ml in 4xYES medium were harvested and drop-frozen in liquid Nitrogen. Cells were lysed mechanically using a mortar and pestle with cooling at regular intervals using liquid Nitrogen. The resulting cell powder was resuspended in ice-cold lysis buffer and incubated for 1 hour with rotation at 4°C. The soluble fraction was collected

by centrifugation (13,000 rpm for 10 minutes at 4°C), and then incubated with 8µl dynabeads (Invitrogen) coupled to 8µl anti-flag antibody per sample, at 4°C for 1 hour. Beads were collected, washed and processes as described above.

The CoIP experiments shown in **Fig 3.7** were performed according to the bead-beating method described above, with changes to the lysis buffer and the amount of source material. Approximately 1.5g of cells (wet weight) at log phase (5-6x10^6 cells/ml) were used per CoIP. 100µl lysis buffer (50mM HEPES-NaOH (pH7.5), 150mM NaCl, 2mM EGTA, 2mM EDTA, 0.25% NP-40) containing freshly added 3mM DTT, 1mM PMSF, 1X Protease Inhibitor (EDTA-free, Roche) was used to resuspend 0.1g cell pellet (wet weight).

2.5.3 GST pulldown assay

Soluble fractions of cell lysates were obtained using the same mechanical lysis method (mortar and pestle) mentioned in section **2.5.2**. Soluble lysates were pre-cleared by incubating with glutathione-sepharose-4B beads for 1 hour at 4°C. The pre-cleared lysates were incubated with approximately 100µg of GST-fusion protein or equimolar concentration of GST-alone protein as a control, along with glutathione-sepharose-4B beads for 3 hours at 4°C. The beads were then recovered and washed four times in lysis buffer as described in section **2.5.2**. Samples were resuspended in 2x SDS sample buffer and were analysed by SDS-PAGE and western blotting (section **2.5.4**).

2.5.4 Western blot

Protein samples denatured in 2X sample buffer were resolved using SDS-PAGE (sodium dodecyl sulphate – polyacrylamide gel electrophoresis) on 1mm thick polyacrylamide gels (**Table 2.8**) using the Mini-PROTEAN Tetra Cell (Bio-Rad)

apparatus. For optimal separation, resolving gels with polyacrylamide concentrations of 8%, 10% or 12% were used, depending on the size of the proteins being analysed.

Reagents	Resolving Gel (8%) (10ml)	Stacking Gel (5%) (50ml)
30% acrylamide/bis mix	2.7ml (8%) 3.3ml (10%) 5.4ml (12%)	8.5ml
1.5M Tris-HCl pH8.8	2.5ml	-
1.0M Tris-HCl pH6.8	-	6.25ml
dH_2O	Upto 10ml	34.75ml
10% SDS	100µl	500µl
10% Ammonium persulphate (APS)	100µl	5µl
TEMED	10µl	2µl

Table 2.8: Polyacrylamide gel composition

Electrophoresis was performed at 200V in 1X Running buffer (1L of 5X Running buffer: 30g Tris Base, 144g glycine, 5g SDS). PageRuler Prestained Protein Ladder (Thermo Scientific) was used as protein size marker. The proteins were then electroblotted onto Protran nitrocellulose membrane (0.45µm pore size, GE Healthcare) using a semi-dry blotter (Hoefer). For protein transfer, the gel was placed over the membrane and these were further sandwiched between three layers of 3MM Whatman filter paper on each side (all pre-wetted in ice-cold Blotting buffer (1X Running buffer containing 25% acetic acid and 10% methanol). Proteins were transferred for 1 hour at 65mA per gel. The transfer efficiency was determined by Ponceau Red (Sigma) staining for 5 minutes at room temperature (Ponceau staining was avoided for detection using Licor imaging due to the high non-specific background caused by staining). The membrane was blocked with Blocking buffer (5% dried milk, 0.1% Tween-20 in 1X PBS) for 1 hour at room temperature to prevent

non-specific antibody binding. It was then incubated overnight at 4°C in the presence of primary antibody diluted in 1X PBST (1X PBS with 0.1% Tween-20). Anti-GST (kind gift from the Hay Lab, Dundee) was used at 1:2500; anti-flag-HRP at 1:5000 and anti-flag/anti-c-myc/anti-GFP(Roche) at 1:1000 dilution. The membrane was then washed three times with 1X PBST at room temperature. For anti-flag-HRP conjugated primary antibody, this was followed by protein detection. Otherwise, the membrane was incubated for 1 hour at room temperature with peroxidase-conjugated (or IRDye-labelled secondary antibody for Licor) diluted 1:10,000 in Blotting buffer. After three washes with 1XPBST followed by a final wash with 1XPBS, immunoreactive bands were visualised by chemiluminescence using Pierce ECL Western Blotting Substrate kit (Thermo Scientific) based on manufacturer's protocol. The blot was exposed to Amersham Hyperfilm ECL (GE Healthcare) for different exposure times as required for optimal detection of signal. Near infra-red imaging was performed using an Odyssey imaging system (Licor)..

2.5.5 *In-vitro* binding assay

S^{35}-labelled proteins were produced from pGAD/pGBK plasmids (**Table 2.10**) carrying gene of interest, using the TNT T7 kit (Promega) according to the manufacturer's protocol. For *in vitro* binding, 4µg of GST-recombinant proteins were incubated with 10µl of S^{35}-labelled proteins on ice for 30minutes in binding buffer (1xPBS, 0.5mM DTT, 1mM PMSF, 0.5mg/ml BSA and 1X Protease inhibitors (EDTA-free, Roche)). The samples were then incubated with glutathione sepharose 4B beads for 1hr followed by four washes in binding buffer. Bound proteins were analysed by SDS-PAGE and fluorography. A schematic representation of the steps involved in *in-vitro* binding assays using GST-Stc1 is shown in Chapter 3, **Fig 3.3**. For *in-vitro* binding assays using biotinylated H3K9me2-peptides (Abcam), S^{35}-labelled

Clr4-CD mutant proteins were incubated with 20ng of biotinylated H3K9me2-peptides. Peptides were incubated with streptavidin sepharose beads. Binding reactions and peptide-bound protein were performed as described above. In all assays, input corresponds to 5% of sample used for the binding experiments.

2.5.6 Size Exclusion Chromatography (SEC)

S. pombe cultures were grown to a density of $1x10^8$ cells/ml in 200ml 4X YES medium and harvested and drop-frozen in liquid Nitrogen. Cell lysate for SEC was prepared following the mechanical lysis procedure using mortar and pestle (section **2.5.2**). The soluble fraction was filtered using a 0.22µm filter membrane (Millipore) and concentrated using a Vivaspin column with molecular weight cut off of 10kDa (Sartorius). Gel filtration was performed on a high resolution pre-packed Superose-6 10/300GL column using an AKTA purifier system (Edinburgh Protein Production Facility). Prior to sample injection, the column was washed once with ultrapure water (distilled and degassed) and calibrated using two column volumes of filtered and degassed Gel Filtration buffer (50mM HEPES-NaOH pH 7.4, 150mM NaCl, 5mM EDTA, 0.1% NP-40). A calibration curve was calculated using protein standards of known sizes ranging from 6.5kDa to 669kDa (Gel filtration calibration kit, GE Healthcare), prior to the column separation of yeast lysate (**Fig 2.2**). Sizes for the collected sample fractions were subsequently back-calculated using the calibration curve.

Gel filtration was conducted at 4°C with a flow rate of 0.5ml/minute using approximately 36ml of Gel Filtration buffer (one and half column volume). Eluates were collected in 96-well plates and stored at 4°C overnight. This was followed by

immunoprecipitation of flag tagged Stc1 from each fraction using dynabeads coupled to anti-flag antibody, and analysis of immunoprecipitated proteins by SDS-PAGE and western blotting.

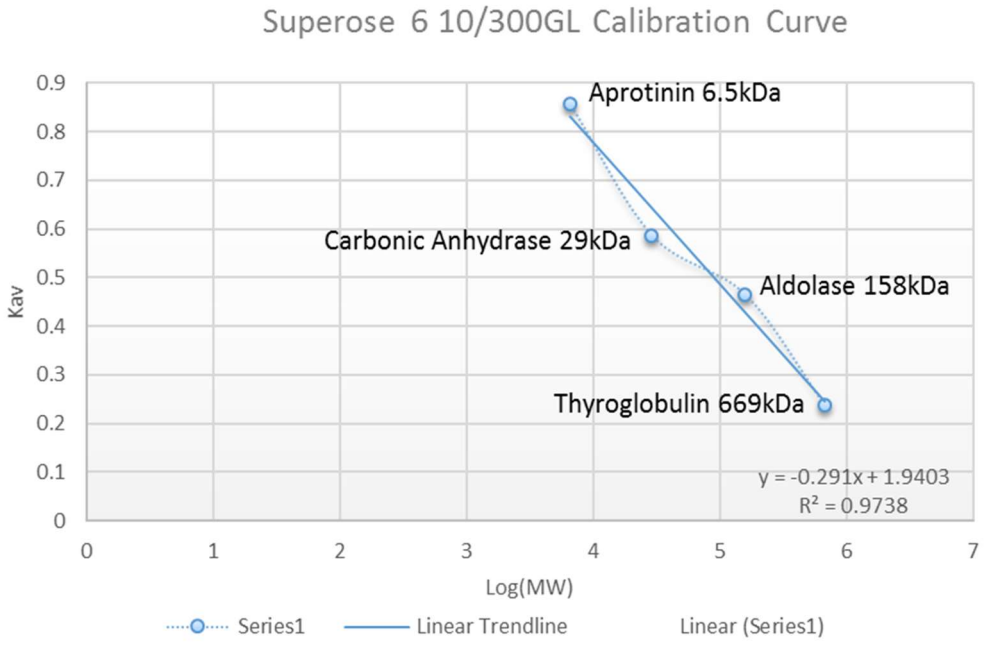


Fig 2.2: Calibration curve for Superose 6 10/300GL column used for SEC

Plot of partition coefficient (Kav) versus the logarithm of molecular weight of the protein standards (Log(MW). This plot was used to determine the unknown molecular weight of a sample using the formula Kav = [(Ve-V0)/(Vt-V0)] (Ve – elution volume, V0 - void volume of the column, Vt - column volume). The linear equation for the trendline was used to calculate sample concentration 'x'. R^2 represents correlation coefficient; a value closer to 1 indicates better fit.

2.6 DisEMBL protein disorder prediction analysis

Protein disorder prediction was performed utilising the online portal for DisEMBL (Intrinsic Protein Disorder Prediction 1.5) (Linding et al. 2003). Full length Stc1 protein sequence was provided as input, and prediction was run on default parameters

set by DisEMBL. DisEMBL utilises three distinct methods for disorder prediction, providing separate predictions based on each of these methods. In the output graph (Chapter 3, **Fig 3.6A**), each of the different predictions are shown in a different colour. For each method, the confidently predicted regions in the output are represented by parts of the plot above the respective baseline (colour-coded dotted lines).

The methods are:

1-Disordered by loops/coils (Blue graph): - this is based on the prediction of whether the residues lie within a defined a secondary structure or within poorly defined loops/coils

2- Disordered by hot-loops (Red graph): – this is based on the ability of loops to be highly mobile as determined from B-factors (Boltzmann's factors). Mobile loops are considered to be intrinsically disordered.

3- Disordered by Remark-465 (Green graph): – this attributes disorder to missing coordinates within the X-ray structure entries n PDB (Protein Data Bank))

Table 2.9 DNA Oligonucleotides

MOLECULAR CLONING

Name	Sequence 5' - 3'
Clr4-N_Nde_F	GTAACCCATATGATGTCGCCTAAACAAGAGGAG
Clr4-N_Xma_R	GCCTTTCCCGGGAGGGTTTCGCGGTTTTGT
Clr4-CD_Xma_R	GCCTTACCCGGGGGAGTTACTTCCCTTTAGTCT
Clr4-PCD_Nde_F	GTAACCCATATGGACTCGGATTCACCGCAC
Clr4-Cat_Nde_F	GTAACCCATATGTCCAAACTTGACTCTTATACCCA
Clr4-Cat_Xma_R	GTTTGCCCCGGGTTAACCGAAAAGCCAGCCACG
pGBK-Clr4-CD-M1-F	CAAGAGGAGTATGAAGTTGAAGCGATTGCGGCCGCAGCAG CGGCTGCTAATGGAGCTGTCAAACTATATCGC
pGBK-Clr4-CD-M1-R	GCGATATAGTTTGACAGCTCCATTAGCAGCCGCTGCTGCG GCCGCAATCGCTTCAACTTCATACTCCTCTTG
pGBK-Clr4-CD-M2-F	GGACGAAAAATTGGATCGTAATGGAGCTGCCGCAGCATAT GCCATTGCATGGTTGAACTATTCATCTAGAAG
pGBK-Clr4-CD-M2-R	CTTCTAGATGAATAGTTCAACCATGCAATGGCATATGCTGC GGCAGCTCCATTACGATCCAATTTTTCGTCC
pGBK-Clr4-CD-M3-F	GTCAAACTATATCGCATTCGATGGGCGGCCTATGCAGCTG CAGCTGCTACTTGGGAGCCCCCTGAGAACC
pGBK-Clr4-CD-M3-R	GGTTCTCAGGGGGCTCCCAAGTAGCAGCTGCAGCTGCATA GGCCGCCCATCGAATGCGATATAGTTTGAC
pGBK-Clr4-CD-M4-F	GATACTTGGGAGCCCGCTGCGAACCTTGCTGGATGTTCTGC CG
pGBK-Clr4-CD-M4-R	CGGCAGAACATCCAGCAAGGTTCGCAGCGGGCTCCCAAGT ATC
pGBK-Clr4-CD-M5-F	GAGAACCTTTCTGGAGCTGCTGCCGTTTTAGCAGCATGGA AAGCGCGGAAGAGGAGAC
pGBK-Clr4-CD-M5-R	GTCTCCTCTTCCGCGCTTTCCATGCTGCTAAAACGGCAGCA GCTCCAGAAAGGTTCTC
pGBK-Clr4-CD-M6-F	GAATGGAAAAGGCGGAAGGCGGCAGCAGCGGGAAGTAAC TCCGACTCGG
pGBK-Clr4-CD-M6-R	CCGAGTCGGAGTTACTTCCCGCTGCTGCCGCCTTCCGCCTT TTCCATTC
Swi6_Nde_F	GTAACCCATATGATGAAGAAAGGAGGTGTTCGA
Swi6_Xma_R	GCCTTTCCCGGGTTATTCATTTTCACGGAACGT
Swi6-CD_Nde_F	GTAACCCATATGTATGTTGTAGAAAAGGTTTTA
Swi6-CD_Xma_R	GCCTTTCCCGGGTTTAGAAGGTTCTGGTCTTCC
PacI-YFP_F	GCCACCTTAATTAACATGGTGAGCAAGGGCGAGGAGCTG
AscI-YFP_R	GATCAGGGCGCGCCTTATATAGATCCGGTGGATCCCGG

stc1-Ctag-F	AAGAATATGATGATGATGACTCTGATGAAGACAGGATGGA AGAGATATTT CAGCAATTCAAAAAAGAAAAACAAATTGTACGGATCCCCG GGTTAATTAA
stc1-Ctag-R	CACTAATGGACCTTTATAATTTATCATATAAAATGGGATCG TAAGCAAATC TTTACATAATTCAAATAAGGATGAATATAGAATTCGAGCT CGTTTAAAC
dos2-Ctag-F	AAGTTTATAT TTAAAACTCC TTCGAACAGT GTCCGATCGT TTACAAATAA ACTTGAACAC ATAATGTACA ACATAAAT CGG ATC CCC GGG TTA ATT AA
dos2-Ctag-R	ACATTAGCCATGTAAAATTGTAATCCTTTATAAACATACTA TATATTAACTTTAATGCAGTGTGTCTTTTGTTAATTACTAGT AAAAATCATCATATCTATCAGAATTCGAGCTCGTTTAAAC

GENOTYPING

Dos2_5'-120F	TCAGAAGCAGAGATCAGA
Dos2_3'+340R	GCTTGGTCTCAGTTATTATC
EYFP_1F	ATGGTGAGCAAGGGCGAGGA
EYFP_270R	ATGGCGGACTTGAAGAAGTC
EYFP_700R	TCGTCCATGCCGAGAGTGAT
otr_ade6_F	CTACTCTTCTCGATGATCCTGTA
otr_ade6_R	GGCCACCATAGACATAACTG
Dos2+631F	TGCTCGTCAA GAGCGAAAGA
Dos2+631R	TCTTTCGCTC TTGACGAGCA
Dos2+1606F	GCTCG ATTTAGCTTC CGCATC
Dos2_3’300F	TCCGCATCTTCCTGCTTATC
Dos2+650F	CAATGCTCGTCAAGAGCGAAAG
Stc1+50F	CGAAGGATTGCAAACTACATGC
Myc_R	CCGTTCAACTCTTCTTCTGAG

stc1-240F	CCTGATCAGCGTAAAGATCTC
clr4+80_R	GTTTGACAGCTCCATTACGATC
clr4-200_F	CTCTGAAATTGAACACATCGAC
clr4-400F	GCT CTG AAA TTG AAC ACA TCG AC
clr4+600R	CTT AGA TGG GTA TAA GAG TCA AG
ago1-250F	GTTTGCGTGCTCAGAGAAG
ago1+270R	CAATTCCAGTAACACTTGCACG
ago1_intR	CAATTCCAGTAACACTTGCACG
ago1_1850F	GCTGAACAGGTTGGTAATG
ago2300R	CGTTGTGGTTGTTGTTTAGTC
swi6_F	GAAGGTTATGACGA CCCAGT G
swi6_R	GTTTAACCGTCAGCTCTCTGTTG
T7-F	TAATACGACTCACTATAGGGCGA

Table 2.10 Plasmids

Name	Source
pGEX-4T1-GST	He et al. 2013
pGEX-4T1-GST-Stc1-WT	He et al. 2013
pGEX-4T1-GST-Stc1-ZF	He et al. 2013
pGEX-4T1-GST-Stc1-ΔC	Yunyu Shi Lab, China
pGEX-4T1-GST-Stc1-C	He et al. 2013
pGEX-4T1-GST-Stc1-n	Yunyu Shi Lab
pGEX-4T1-GST-Stc1-Δn	Yunyu Shi Lab
pGEX-4T1-GST-Stc1-n+zf2+C	He et al. 2013
pGEX-4T1-GST-Stc1-n+zf1+C	He et al. 2013
pGEX-4T1-GST-Stc1-n+C	Yunyu Shi Lab
pFA6a-13xmyc-kanMX6	Bahler et al. 1998

PFA6a-GFP-KanMX6	Bahler et al. 1998
PFA6a-GFP-NatMX6	Bahler et al. 1998
pFA6a-YFP-Kan	This study
pFA6a-Cerulean-Nat	This study
pCerulean-C1	Lamond Lab, Dundee
pEYFP-C1	Lamond Lab, Dundee
pREP41-swi6	Lab stock
pGAD-ago1	Lab stock
pGAD-tas3	Lab stock
pGAD-chp1	Lab stock
pGBK-clr4	Lab stock
pGAD-dos2	Lab stock
pGAD-dos1	Lab stock
pGBK-rik1	Lab stock
pGAD-cul4	Lab stock
pGBK-clr4-CD-M1	This study
pGBK-clr4-CD-M2	This study
pGBK-clr4-CD-M3	This study
pGBK-clr4-CD-M4	This study
pGBK-clr4-CD-M5	This study
pGBK-clr4-CD-M6	This study
pGBK-clr4-N	This study
pGBK-clr4-CD	This study
pGBK-clr4-PCD	This study
pGBK-clr4-Catalytic	This study
pGBK-clr4-CD+20	This study
pGBK-clr4-CD+60	This study
pGBK-clr4-CD+80	This study
pGAD-swi6	This study
pGAD-swi6-CD	This study

Table 2.11 Fission yeast strains

Strain No.	Genotype	Source
7	h+ ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	Lab stock
8	h- ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	Lab stock
209	h-natR-3xmyc-ago1-D580A ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	Lab stock
248	h- stc1-3xFLAG-NATR otr1R(*SphI*):*ade6*$^{+}$+ ade6-210 leu1-32 ura4-D18	Lab stock
264	stc1-flag-NatR 3xmyc-ago1 ura4-D18 leu1-32 ade6-210/704	Lab stock
278	h+ stc1-flag-NatR dos2-HA-NatR otr1R(*SphI*):*ade6*$^{+}$ ura4-D18 leu1-32 ade6-210	Lab stock
348	Flag-clr4 rik1-myc-KanR stc1Δ::NatR ade6-210/216 leu1-32 ura4-D18/DSE	Lab stock
396	dos2-Flag-NatR GFP-dos1-ura4+ stc1Δ::KanR otr1R(*SphI*):*ade6*$^{+}$ade6-210 leu1-32 ura4-D18	Lab stock
434	h+ stc1-3xflag-HygR Nmyc-clr4 ade6-210/216 leu1-32 ura4-D18/DSE	Lab stock
1446	h+ stc1-flag-HygR dos2-HA-NatR ura4-D18 leu1-32 ade6-210/216	Lab stock
1501	h+ stc1-cerulean-Nat ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	This study
1502	h+ dos2-yfp-Kan ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	This study
1679	h+ stc1-3xflag-HygR dos2-HA-NAtR cul4Δ::NatR ade6-210 leu1-32 ura4-D18	This study
1711	stc1-flag-NatR dos2-HA-NatR rik1Δ::KanR ade6-210 leu1-32 ura4-D18	Lab stock
1712	stc1-flag-NatR dos2-HA-NatR clr4Δ::KanR ade6-210 leu1-32 ura4-D18	Lab stock
1713	stc1-flag-NatR dos2-HA-NatR dos1Δ::KanR ade6-210 leu1-32 ura4-D18	Lab stock
1750	stc1-3xflag-HygR natR-3xmyc-ago1-D580A ade6-210/216 leu1-32 ura4-D18/DSE	This study

2270	h 3xFlag ago1- KAN chp1-myc-leu2 leu1-32	This study
2271	h+ 3xFlag ago1- KAN chp1-myc-leu2 stc1Δ::NatR leu1-32 ura4-D18/DSE	This study
2540	h+ raf2-D57A-FLAG-NAT leu1-32 ura4D18 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	Allshire lab
2541	h+ raf2-I98A-FLAG-NAT leu1-32 ura4D18 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	Allshire lab
2542	h+ raf2-E104A-FLAG-NAT leu1-32 ura4D18 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	Allshire lab
2543	h- dos2-ts3-Flag-NatR (raf2-S100F) ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	Allshire lab
2697	h+ raf2-D57A-FLAG-NAT stc1-myc-KanR leu1-32 ura4D18 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	This study
2699	h+ raf2-I98A-FLAG-NAT stc1-myc-KanRleu1-32 ura4D18 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	This study
2701	h+ raf2-E104A-FLAG-NAT stc1-myc-KanR leu1-32 ura4D18 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	This study
2703	h- dos2-ts3-Flag-NatR (raf2-S100F) stc1-myc-KanR ade6-210 leu1-32 ura4-D18 otr1R(*SphI*):*ade6*$^{+}$	This study
2795	stc1-myc dos2-flag-NatR ura4-D18 leu1-32 ade6-210 otr1R(*SphI*):*ade6*$^{+}$	This study

CHAPTER 3

Mode of Stc1-interaction:

Stc1 utilises its N and C-termini

to link

the histone modifying CLRC complex

and the RNAi-effector RITS complex

3.1 Introduction

At fission yeast centromeres, RNAi and chromatin modification machineries cooperate to implement transcriptional silencing of the repetitive DNA sequences at pericentromeres. The resulting heterochromatin is crucial in chromosome dynamics and ensuring proper chromosome segregation and hence cell survival (Goto & Nakayama 2012). As mentioned previously in Chapter 1, the tandem-zinc finger (ZF) domain containing protein Stc1 has been found to have a crucial role in mediating the inter-talk between the RNAi-effector complex RITS and the chromatin modification machinery, CLRC. A key observation that put Stc1 under the limelight was its association with the RNAi-effector complex, RITS and the chromatin modifying machinery, CLRC. Stc1 co-immunoprecipitated both RITS and CLRC components *in-vivo,* and was found to directly interact with Ago1 *in-vitro.* This led to the proposal that Stc1 could be acting as a physical link between RITS and CLRC.

A collaborating research group determined the structure of the tandem ZF domain within the N-terminus of Stc1 by NMR (He et al. 2013). A schematic representation of the tertiary structure of the ZF module is presented in **Fig. 3.1A**. Residues within the ZF domain have been found to be important in Stc1-Ago1 interaction. Indeed, single point mutations to residues within zf2, K100 and R116, found on the electrostatic potential surface of the solved structure (**Fig 3.1B**) affects Stc1-Ago1 association *in-vivo.* However, these mutations do not affect Stc1 association with CLRC *in-vivo* implying that other residues/region within Stc1 could be involved in CLRC association. More importantly, it has not been established whether the association of Stc1 with the CLRC complex is direct or indirect. In this chapter, I will

present data towards the goal of establishing the mode of CLRC association with Stc1 and will discuss the implications of the findings. Additionally, I will also provide evidence for how Stc1 might bridge the interaction between RITS and CLRC. Understanding the manner in which Stc1 mediates this interaction is key in obtaining a clearer picture of the complex interactions involved in the crosstalk between RNAi and chromatin modification.

3.2 GST-Stc1 recombinant proteins used in this study

To establish direct interactors and study the mode of Stc1-interactions, I generated wild type (WT) as well as a series of deletion mutants of Stc1 as GST-recombinant proteins in *E. coli* (**Fig 3.2**). In order to study the regions of Stc1 involved in direct interaction with RITS/CLRC partner proteins, I generated truncated Stc1 proteins lacking the C-terminal domain (Stc1-ΔC), or comprising only the C-terminal domain (Stc1-C) or only the ZF domain. Additional truncations targeted distinct structural features within the N-terminus of Stc1: - Deletion of individual zfs [Stc1-(n+zf2+C), Stc1-(n+zf1+C)], and deletion of the 3' N-terminal short stretch (Stc1-Δn), along with a truncation that encodes the N-terminal short stretch only (Stc1-n) (**Fig 3.2A-C**). Expression of the recombinant proteins was confirmed by coomassie staining (**Fig 3.2D**). Subsets of these GST-recombinant proteins have been used in various experiments throughout the entire study. A detailed protocol used for recombinant protein production can be found in Chapter 2, section **2.5.1**.

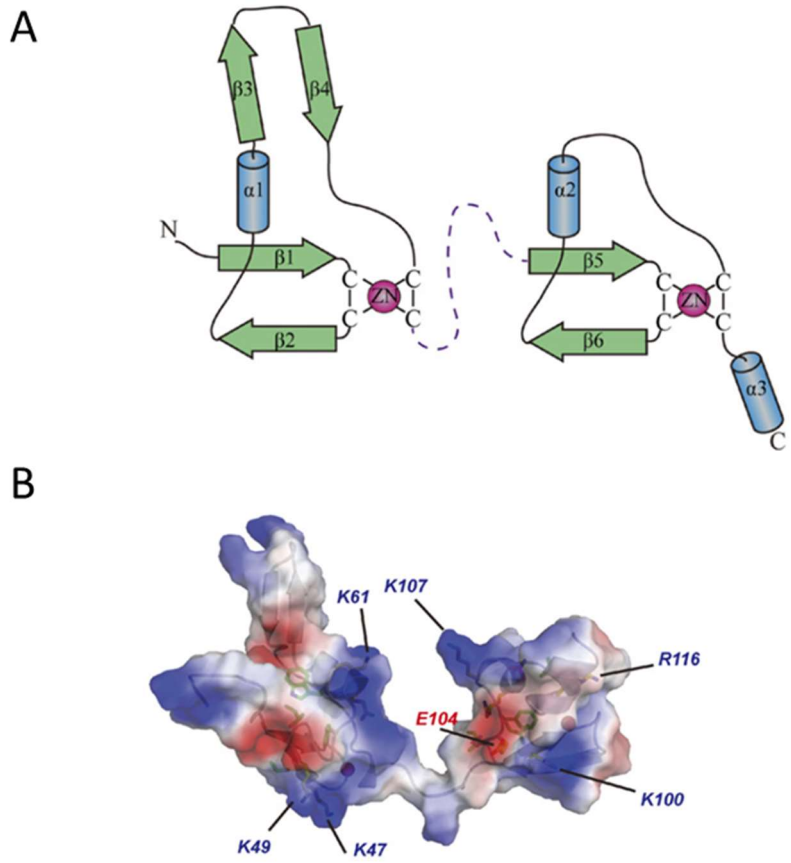


Fig 3.1 Stc1 tandem Zinc Finger Domain

(A) Schematic representation of the tertiary structure adopted by Stc1 in solution. Residues 32-126 within Stc1-N terminus fold into two distinct zinc fingers connected via a 6aa-long flexible linker. Each zinc finger can potentially fold separately without any preferred orientation with each other. Alpha helices and beta sheets are shown in blue and green respectively (He et al. 2013).

(B) Electrostatic potential surface of Stc1 Zinc Finger domain. Positive patches are shown in blue and negative in red. Highlighted are conserved charged residues that are found on the surface and hence potentially involved in interactions (He et al. 2013).

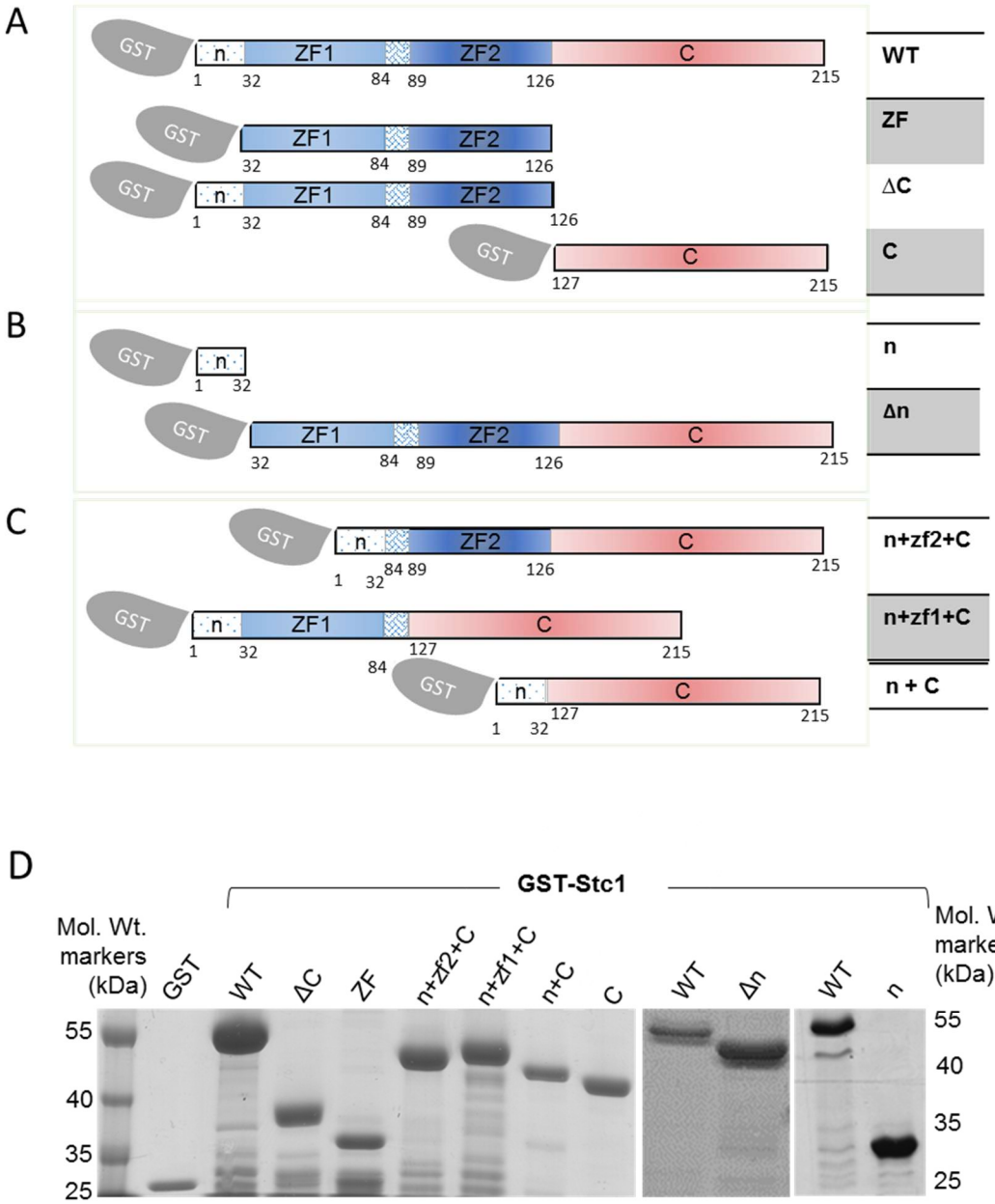


Fig 3.2 Gst-Stc1 constructs used in this study

(A)-(C) Schematic representation of GST-recombinant wild type Stc1 protein (Stc1-WT) and various truncations used throughout this study

(D) Coomassie gels showing bacterially expressed and purified GST-recombinant Stc1-proteins

3.3 Stc1 has multiple binding partners within CLRC and RITS

I utilised *in-vitro* binding assays to determine the proteins within CLRC and RITS that can directly associate with Stc1. A schematic representation of the method is shown in **Fig 3.3** with the detailed protocol described in Chapter 2, section **2.5.5**. All CLRC and RITS subunits were produced as S^{35}-labelled proteins by transcription and translation in rabbit reticulocyte lysate. Each of these proteins were allowed to bind individually to either GST alone or GST-Stc1-WT proteins. GST-pulldown assays revealed direct interaction of Stc1 with CLRC subunits Dos2 and Clr4, and RITS complex components Ago1 and Tas3 (**Fig 3.4A** and **3.4B**). None of these proteins bound GST alone, indicating that their interactions with GST-Stc1 are specific. Moreover, failure of GST-Stc1 to bind to the other CLRC and RITS subunits indicates that the interactions I observed are not just due to the GST-Stc1 recombinant protein being 'sticky'. Though the CLRC component Cul4 shows some binding to GST-Stc1, relatively strong binding to the GST control calls the specificity of this interaction into question. To further investigate the specificity of Stc1-Cul4 interaction, I performed further *in-vitro* binding assays using GST-Stc1 domain-deletion mutant proteins. I utilised GST constructs that code for the Stc1-N and C-termini individually (Stc1-ΔC and Stc1-C) to test whether S^{35}-labelled Cul4 might bind specifically to one region of Stc1 but not the other. However, Cul4 bound weakly to all recombinant proteins including the GST-only negative control, indicating that the interaction with GST-Stc1 is likely non-specific (**Fig. 3.4D**).

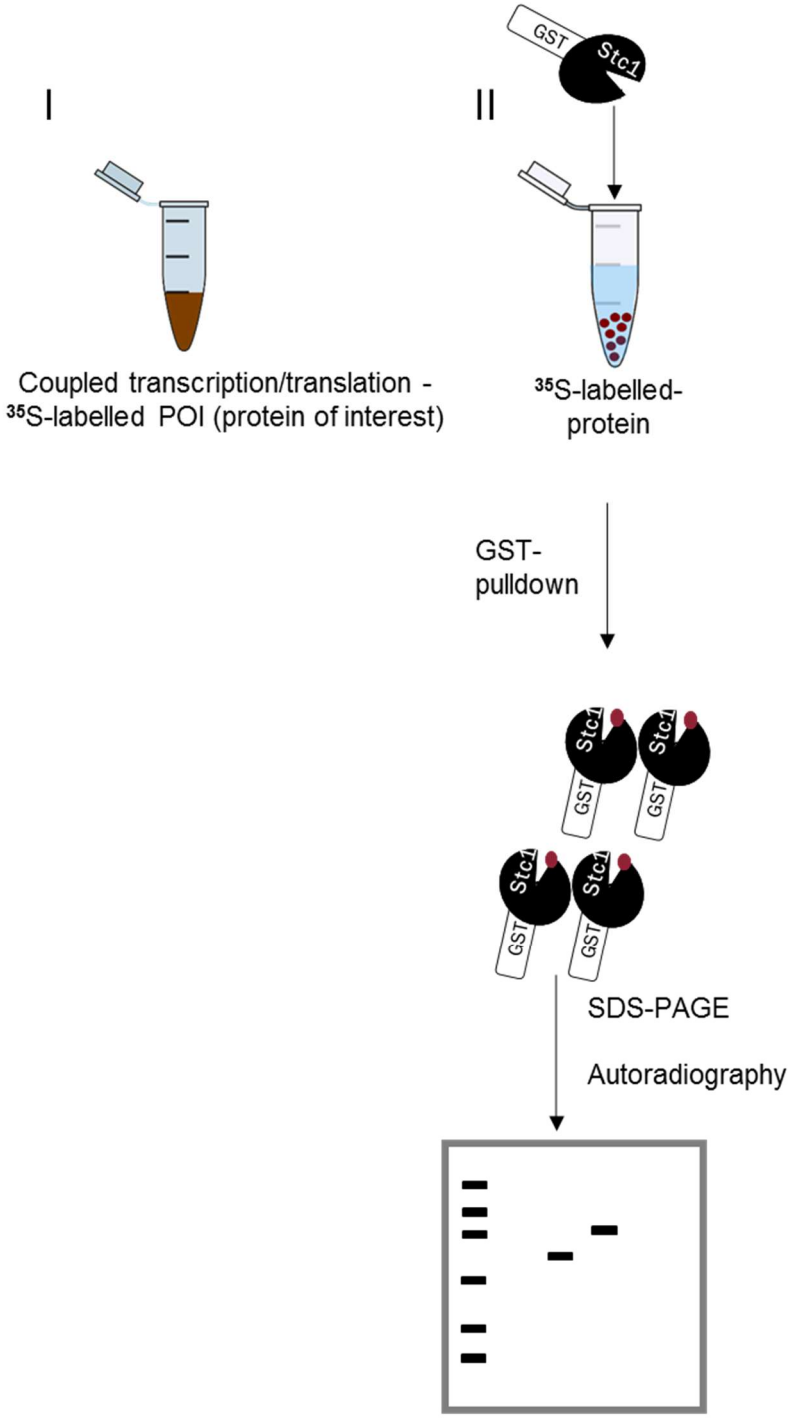


Fig 3.3 Outline of the GST-pulldown in-vitro assay

^{35}S-labelled protein transcribed and translated *in-vitro* using rabbit reticulolysate is allowed to bind with GST-recombinant Stc1 protein. Following a brief incubation, Stc1 along with any bound protein is pulled down via the GST-tag. The pulled down proteins are then subjected to SDS-PAGE and the ^{35}S-labelled protein is detected using autoradiography.

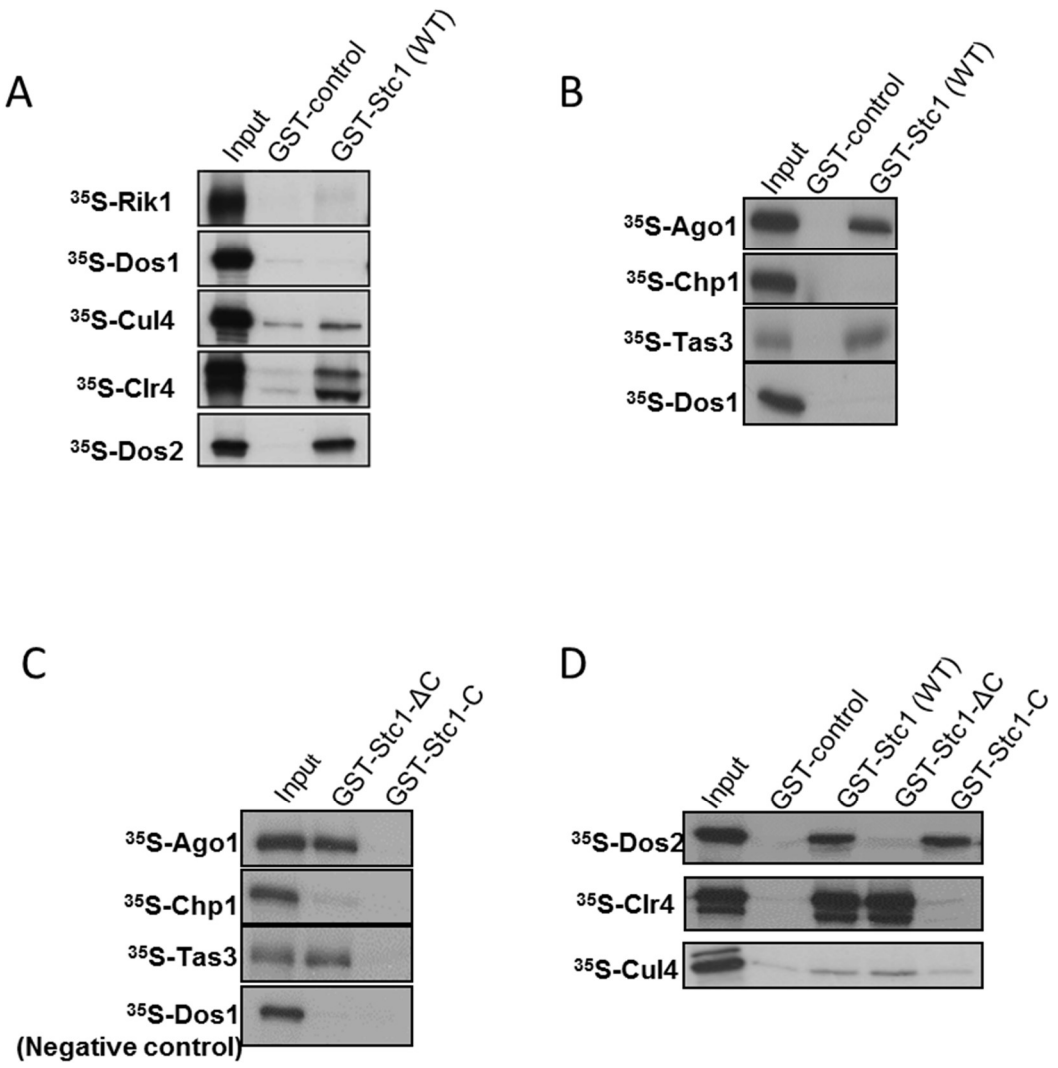


Fig 3.4 Stc1 directly interacts with RITS and CLRC via its N and C-terminal domains *in-vitro*

(A) and **(B) *In-vitro* assays showing CLRC subunits Dos2 and Clr4, and RITS subunits Ago1 & Tas3 interacting directly with Stc1 *in-vitro*.** CLRC and RITS components produced as ^{35}S-labelled proteins *in-vitro* were allowed to bind to either GST alone (negative control) or GST-Stc1-wild type (WT) recombinant proteins. Following GST-pulldown, bound proteins, along with input samples (input corresponds to 5% of sample used for the binding assay), were analysed by SDS-PAGE and autoradiography. The experiments were repeated three times.

(C) and **(D) *In-vitro* assays showing Stc1-N terminus binding to Clr4, Ago1 and Tas3, and Stc1-C terminus binding solely to Dos2**. CLRC and RITS components translated *in-vitro* were incubated with Stc1-truncated recombinant proteins harbouring either the ZF-containing N-terminal domain(ΔC) or the unstructured C-terminal domain. These assays were performed twice.

3.4 N-terminal tandem zinc finger domain acts as a platform for RITS and CLRC interaction while Stc1-C terminus is solely dedicated for CLRC association

The findings above indicating multiple binding partners for Stc1 demanded further investigation to understand the basis of these associations. S^{35}-labelled CLRC and RITS subunits were tested for binding to the two GST-Stc1 constructs that lack either the entire C-terminal domain (Stc1-ΔC) or the entire N-terminal tandem ZF domain (Stc1-C) (**Fig 3.2A**). Surprisingly, these assays revealed that the N-terminal ZF-containing domain of Stc1 can potentially bind to three different proteins – Ago1 and Tas3 within the RITS complex and Clr4 within the CLRC complex (**Fig 3.4 C & D**). On the other hand, Stc1-C terminus bound only Dos2 (**Fig 3.4D**).

The results above indicate that Stc1 can potentially bind CLRC via N-terminal interaction with Clr4 and/or C-terminal interaction with Dos2. To further explore the roles of the ZF module and C-terminus in Stc1 binding to CLRC, I performed a GST-Stc1 pulldown assay using yeast cell lysate. In this assay, lysate prepared from strains bearing tagged CLRC components (IP1:Rik1-myc & Clr4-flag, IP2: Dos2-flag & GFP-Dos1) and deleted for endogenous *stc1*$^{+}$, were used to test CLRC-association of various GST-Stc1 recombinants. Along with the full length Stc1 protein, truncated Stc1 proteins Stc1-ZF, Stc1-(n+zf2+C), Stc1-(n+zf1+C) and Stc1-C were tested (**Fig 3.2**). The full length recombinant Stc1 successfully co-precipitated all tested CLRC subunits. Interestingly, all Stc1-recombinants harbouring the C-terminus (Stc1-(n+zf2+C), Stc1-(n+zf1+C) and Stc1-C) also associated with the CLRC subunits while the ZF-module alone failed to bring down CLRC subunits (**Fig 3.5A**). This indicates that Stc1 interacts with CLRC mainly via its C-terminus and hence Stc1 directly

binding to Dos2 could be the major driving factor in Stc1-CLRC association *in-vivo*. The Stc1 interaction with Clr4 observed *in-vitro* could be a transient interaction *in-vivo,* possibly required for stabilising the CLRC-interaction. Another possibility is that Clr4 requires the N-terminal stretch within Stc1, designated as 'n', for its efficient association *in-vivo.* All Stc1-truncations used for GST-pulldown assays contain the 'n' region except for Stc1-ZF and Stc1-C and the truncation used in the *in-vitro* assays (Stc1-ΔC) also retains this region. Hence it is possible that the *in-vivo* Clr4 association with Stc1-C is via Dos2, and the failure to detect Stc1-Clr4 association with Stc1-ZF could be due to the lack of 'n' region.

To further probe the roles of the 'n' and ZF regions within Stc1 for Clr4 binding, I tested binding of recombinants encoding 'n' only and ZF modules *in-vitro*. Clr4 clearly shows binding to both the ZF module and the 'n' while failing to bind to the C-terminus of Stc1 (**Fig. 3.5B**). This potentially indicates that the 'n' region might be required for efficient Stc1-Clr4 interaction *in-vivo.* Since Ago1 also associates with the N-terminus of Stc1, I tested the requirement for the 'n' region for Stc1-Ago1 binding *in-vitro.* Ago1-binding followed a pattern similar to Clr4, showing interaction with Stc1-ZF and Stc1-n, but not Stc1-C (**Fig 3.5C**). I tested four additional Stc1-truncations (**Fig 3.2B** and **3.2C**) for Ago1 binding: - One of them lacks the 'n' region (Stc1-Δn), while the other three harbour both 'n' and C-terminal regions within Stc1 but lack one or both zinc fingers [Stc1-(n+C), Stc1-(n+zf1+C) and Stc1-(n+zf2+C)]. Interestingly, results from these assays indicated that lack of the 'n' region abolished Stc1-Ago1 interaction *in-vitro* (**Fig. 3.5D**). This raises the possibility that the 'n' region of Stc1 is required for the ZF domain to be in a favourable state, either for the correct folding or presenting residues for electrostatic interaction, in the presence of the Stc1-C terminus.

Additionally, the analyses demonstrated that although the C-terminus of Stc1 alone failed to bind Ago1, the addition of either one of the zinc fingers along with 'n' [Stc1-(n + zf1+C) or Stc1-(n + zf2+C)] was sufficient to bind Ago1 (**Fig. 3.5C**). However, addition of 'n' alone Stc1-(n+C) did not allow Ago1 binding. This could be either due to improper folding of Stc1-(n+C) recombinant protein or due to potential masking of the relatively small 'n' region by the C-terminus (89 residues).

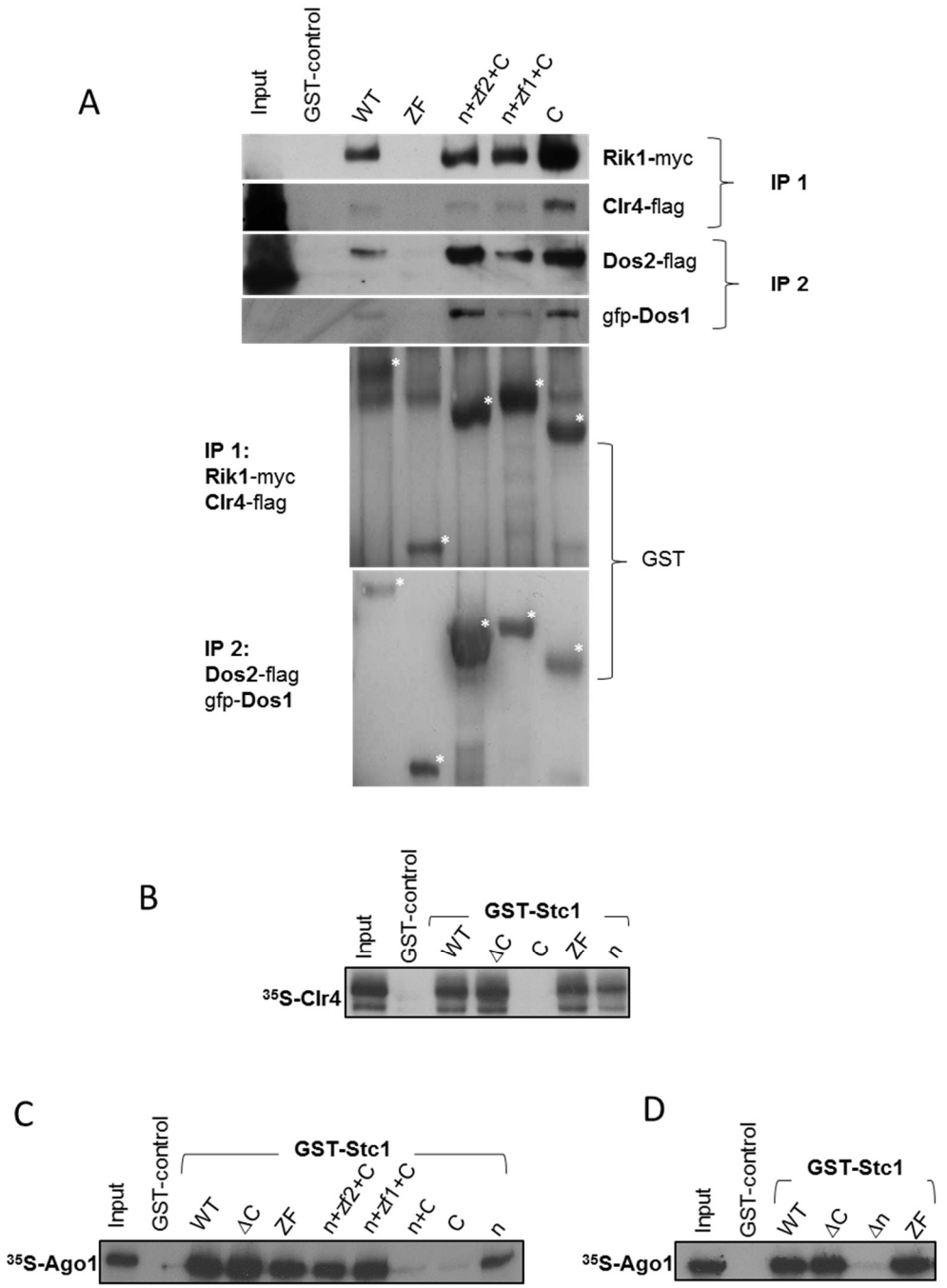


Fig. 3.5 Interactions via Stc1-N terminus are not a pre-requisite for Stc1-association with CLRC

(A) GST-Stc1 pulldown on yeast cell lysate showing that the Stc1-C terminal domain is sufficient for CLRC association. Recombinant GST-Stc1 was used for pulldown using yeast cell lysate from strains bearing tagged CLRC components (IP1:Rik1-myc, Clr4-flag; IP2: Dos2-flag, gfp-Dos1), and lacking endogenous *stc1*. CLRC components were detected by western blotting using antibodies against respective tags. Input corresponds

to 0.5% of lysate volume used for each IP. GST-Stc1 proteins are indicated by asterisks. The assays were performed three times.

(B) *In-vitro* assays using Stc1-truncated recombinants indicating that both the ZF module and the very N-terminal residues ('n') of Stc1 are involved in Clr4 binding.

(C) and (D) *In-vitro* assays utilising a range of Stc1-truncated recombinants showing that both the ZF module and the very N-terminal residues ('n') of Stc1 are involved in Ago1 binding. Assays (B)-(D) were performed two to three times.

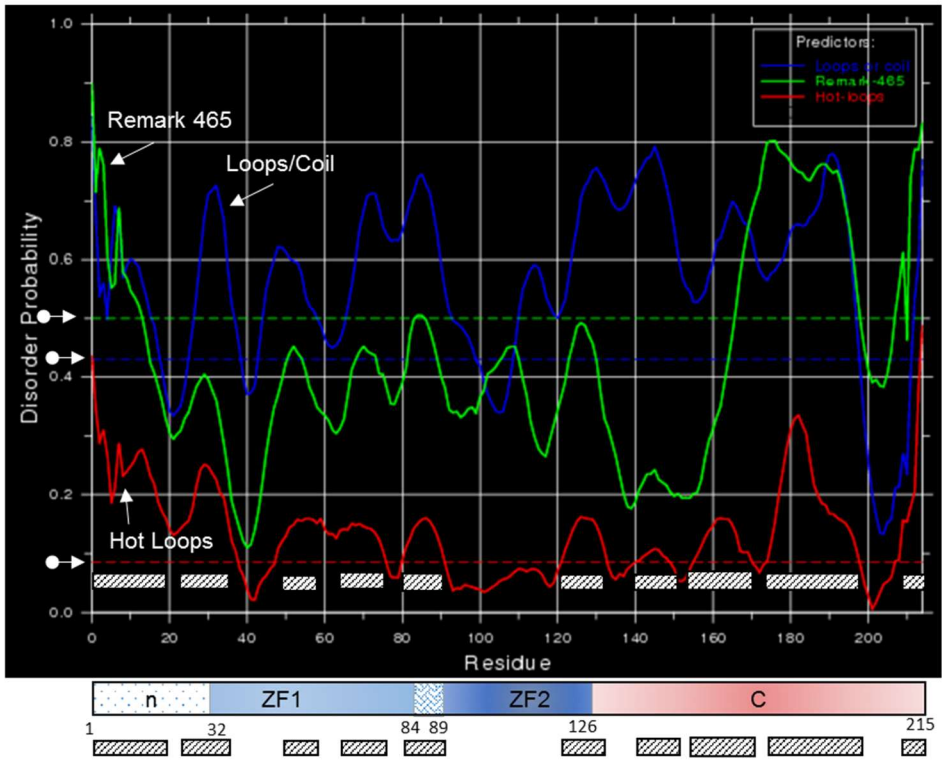
3.5 Disordered regions within Stc1

The 'one to many' feature displayed by Stc1 N terminus and the ability of Stc1-n to interact with Ago1 and Clr4 *in-vitro* prompted to look for intrinsic disorder within the N-terminus of Stc1. Disordered regions are known to either act as interaction interface in protein-protein interactions or contribute towards the specificity of binding to partner proteins (Wright & Dyson 1999). Additionally intrinsic disorder has been implicated in providing a protein with the plasticity needed to carry out multiple interactions (Hsu et al. 2013). NMR analysis indicated that the C-terminus of Stc1 is unstructured (He et al. 2013). However, further information regarding potential intrinsically disordered regions within the Stc1 protein is lacking. MoRF (Molecular Recognition Features) is a term proposed by (Mohan et al. 2006) for regions within long disordered peptide stretches/proteins that can function in recognition and binding of interacting partner protein(s). Prediction of MoRFs within Stc1 could assist in future investigation of the multi-protein binding ability of Stc1 since it is possible that the multiple binding events could be a feature regulated via MoRFs. Towards the goal of discovering potential MoRFs within Stc1, I employed the online tool DisEMBL(Linding et al. 2003) which predicts disordered regions based on three

different criteria: Disordered by loops/coil; Disordered by hot-loops; and Disordered by Remark-465 (for details see Chapter 2, section **2.6**).

Using default parameters, DisEMBL predicted disordered regions within Stc1 (**Fig 3.6**). The probability plot (**Fig. 3.6A**) indicates the probability of the prediction for a particular residue/region. For each prediction method, regions that fall above the baseline (represented as dashed lines – **Fig 3.6A**) are potential regions of disorder. Together the prediction indicates that the first ~18 residues within the 'n' region and a few stretches of residues within the C-terminus are disordered (**Fig 3.6A and 3.6B**). However, it is not clear as to why the prediction based on the method 'Remark-465' failed to confidently predict the whole C-terminus as disordered since this prediction method is based on the missing structural coordinates in PDB-deposited data. There is no structural data for Stc1 C-terminus in PDB since this region was shown to be unstructured in solution (He et al. 2013), and hence the false negative prediction by 'Remark-465' method is questionable. For this reason, I will not include prediction based on 'Remark-465' method. When disorder predictions based on the other two criteria are taken into account, this results in the prediction of additional stretches of residues within zf1 and interestingly, the linker between the zfs (zinc fingers) and a short stretch in zf2 leading to C-terminal residues (**Fig 3.6B**). Though the ZF is structured in solution as deduced by He *et al.* (2013), presence of short stretches of MoRFs within the N-terminus of Stc1 might provide clues towards Stc1's ability to bind more than one protein partner via its N-terminus.

A



B

Disordered by Loops/coils definition

>Stc1_LOOPS 1-18, 26-39, 44-99, 110-198

MDFKSRKYKI KKHPKDCKlh akkyrGTLNS KGKNDNDCLi mcmRCRKVKG IDSYSKTQWS

KTFTFVRGRT VSVSDPKVIC RTCQPKQHDS IWCTACQQTk ginefskaqR HVLDPRCQIC

VHSQRNDGDD NLESDKFVDP FIGDDSDLDD DIYIHDKQTI NSEYADDVSD NTDEERTESK

GQQESNSAEE YDDDDSDEdr meeifqqfkk ekqiv

Disordered by Hot-loops definition

>Stc1_HOTLOOPS 1-38, 49-76, 81-91, 123-133, 156-170, 175-198, 208-215

MDFKSRKYKI KKHPKDCKLH AKKYRGTLNS KGKNDNDCli mcmrcrkvKG IDSYSKTQWS

KTFTFVRGRT VSVSDPkvic RTCQPKQHDS iwctacqqtk ginefskaqr hvldprcqic

vhSQRNDGDD NLEsdkfvdp figddsdldd diyihDKQTI NSEYADDVSD ntdeERTESK

GQQESNSAEE YDDDDSDEdr meeifqqFKK EKQIV

Disordered by Remark-465 definition

>Stc1_REM465 1-14, 167-198, 208-215

MDFKSRKYKI KKHPkdcklh akkyrgtlns kgkndndcli mcmrcrkvkg idsysktqws ktftfvrgrt vsvsdpkvic

rtcqpkqhds iwctacqqtk ginefskaqr hvldprcqic vhsqrndgdd nlesdkfvdp figddsdldd diyihdkqti

nseyadDVSD NTDEERTESK GQQESNSAEE YDDDDSDEdr meeifqqFKK EKQIV

Fig 3.6 Disorder prediction for Stc1 using DisEMBL

(A) DisEMBL plot showing probability of a region being disordered based on the three models. Disorder probability plots are shown for each of the three prediction methods (blue, red and green; see materials and methods). Regions where probability is above cut

off lines (dashed lines; also indicated by double-ended arrows) for each model have significant probability of being disordered. Below the DisEMBL plot is a schematic illustration of Stc1 domains. The regions predicted as disordered based on the combined DisEMBL output are indicated by patterned boxes.

(B) DisEMBL output showing disorder predictions within the Stc1 sequence. Coloured residues indicate regions of disorder as predicted by the three different criteria used by DisEMBL. Boxed regions indicate Stc1 domains: 'n' is represented in a blue box, linker region in green, and the C-terminus in grey box. zf1 and zf2 regions are the unmarked regions between 'n' and linker, and linker and C, respectively.

3.6 Stc1 is dispensable for an intact RITS and is potentially capable of associating with RITS prior to centromeric localisation

Stc1 is considered an auxiliary component of CLRC (Bayne et al. 2010; Kuscu et al. 2014), and seems to be a key integral part of the RNAi pathway. This is largely based on the observations that Stc1 is absolutely essential for the processing of centromeric transcripts to siRNAs while H3K9-methylation is partially retained in the absence of Stc1. Moreover, the absence of Stc1 does not compromise CLRC integrity. However, the role of Stc1 in RITS integrity has not been investigated. As a preliminary step towards understanding the integral role of Stc1 within the RNAi pathway, I decided to examine whether RITS integrity requires Stc1.

To test the integrity of the RITS complex, I created *S. pombe* strains bearing tagged RITS components (flag-Ago1 and Chp1-myc) in *stc1*$^{+}$ and *stc1Δ* cells and performed CoIP experiments to investigate whether association of Ago1 and Chp1 is perturbed in the absence of Stc1. Interestingly, Ago1 still retains its association with Chp1 when Stc1 is not present indicating that Stc1 is dispensable for an intact RITS complex (**Fig 3.7A**). Next, I investigated if Stc1 is capable of interacting with an inactive RNAi-

complex. To test this, I utilised the catalytically-dead Ago1 mutant Ago1-D580A that fails to convert duplex siRNA to single stranded siRNA (ss-siRNA) which is thought to be required for activation of RITS. Intriguingly, Stc1 seems to associate with Ago1-D580A, albeit possibly at reduced levels, indicating that Stc1 can potentially interact with the ds-siRNA loaded RITS (**Fig 3.7B**). Since ss-siRNA-loaded Ago1 is crucial for localising RITS to the nascent centromeric transcripts via complementary base pairing, and Stc1-localisation to centromere is Ago1-dependent, these results potentially indicate that Stc1 has the ability to associate with Ago1/RITS prior to RITS localisation to centromeric transcripts (Bayne et al. 2010). In support of this, during live cell imaging of Stc1-CerFP (cerulean fluorescent protein), I have noted that Stc1-puncta can be observed in the cytoplasm as well as the nucleus (**Fig 3.8**). This is of interest since although Stc1 was previously reported to localise in the nucleus, other RNAi proteins including Ago1 have been reported to be present as distinct puncta in both the cytoplasm and the nucleus (Buker et al. 2007; Carmichael et al. 2006).

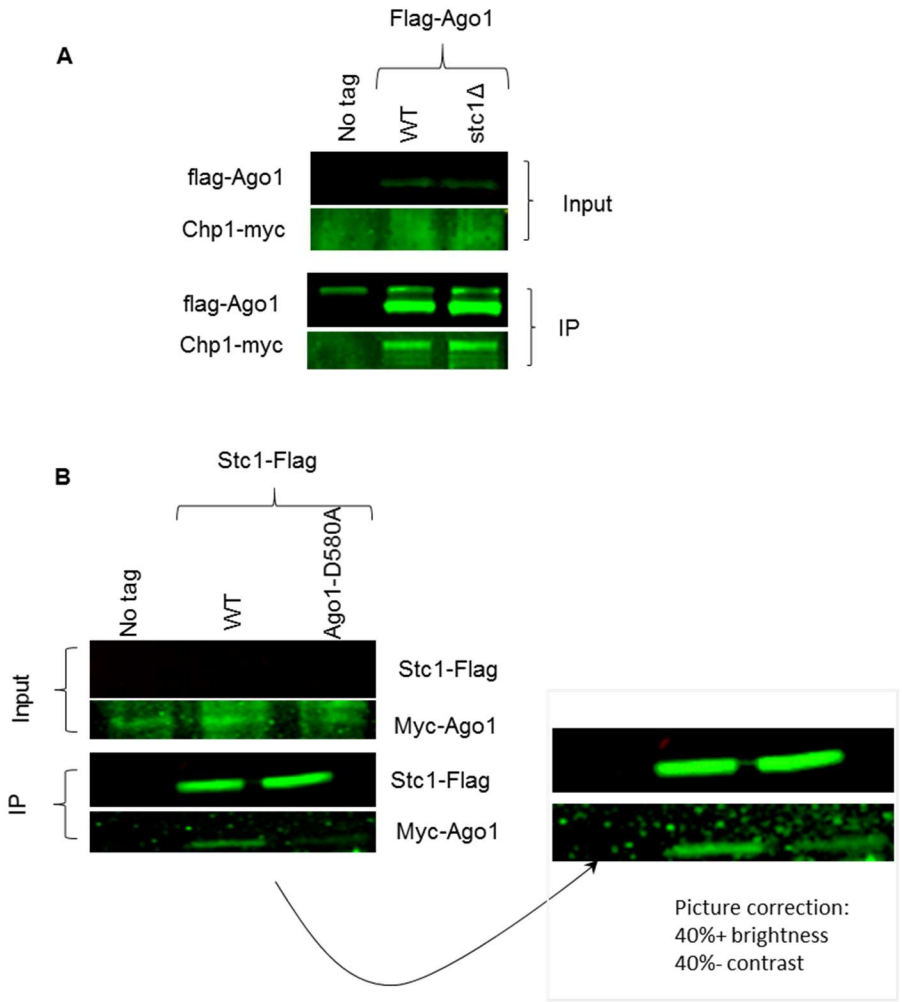


Fig 3.7 Stc1 potentially associates with Ago1 prior to centromeric localisation

(A) Stc1 is not a pre-requisite for RITS integrity. *In vivo* CoIP experiments in which flag-tagged Ago1 was immunoprecipitated to test for association with Chp1-myc in WT and *stc1Δ* backgrounds. Input and immunoprecipitated material was analysed by western blotting.

(B) Stc1 associates with catalytically dead Ago1. *In-vivo* CoIP experiments testing for association of myc-tagged WT or catalytically dead (D580A) Ago1 with immunoprecipitated Stc1-flag.

These *in-vivo* experiments were performed only once.

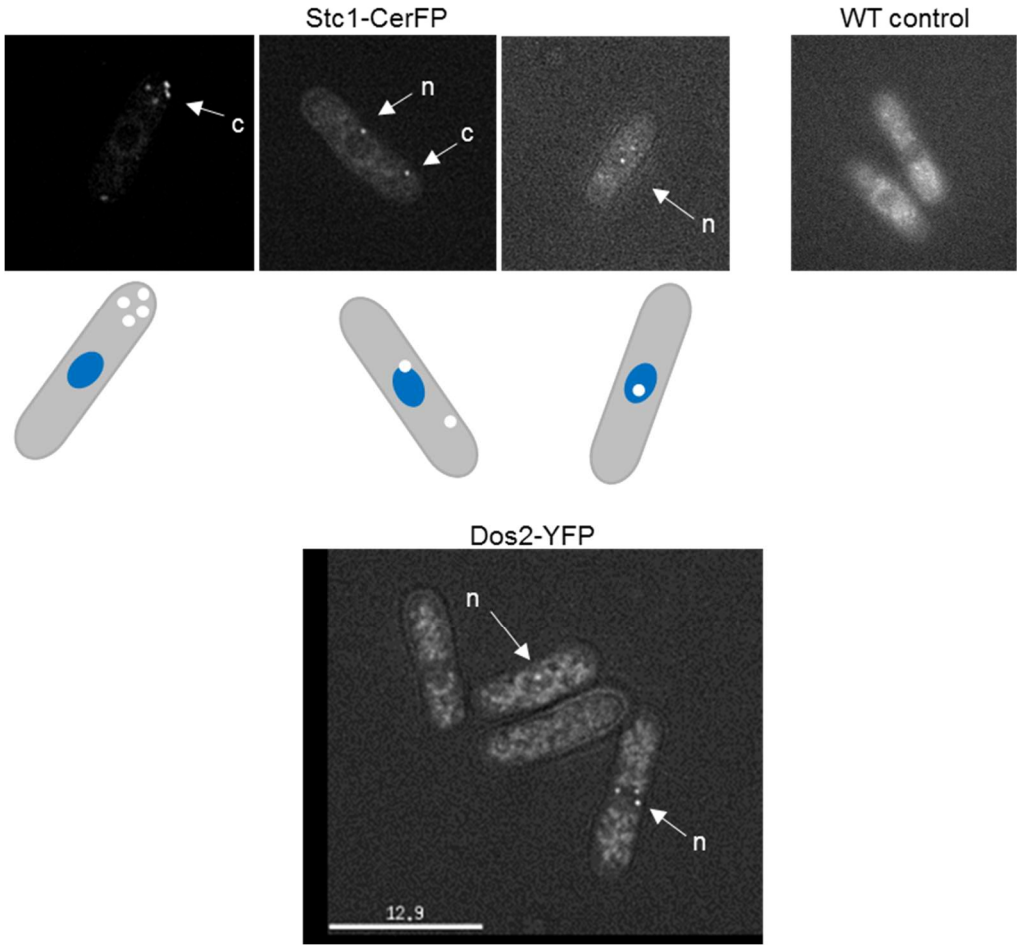


Fig 3.8 Stc1 potentially forms cytoplasmic as well as nuclear puncta

Live cell imaging of Stc1-CerFP (cerulean fluorescent protein) is shown along with schematic representations of the images indicating the position of the nucleus (blue) and Stc1 foci (white). WT serves as the untagged control and Dos2-YFP (yellow fluorescent protein) live cell images are additionally shown as an example of foci limited to the nucleus. 'n' and 'c' indicates nuclear and cytoplasmic foci respectively (scale bar represented in um for Dos2-YFP image).

3.7 Stc1 is potentially present in distinct complexes *in-vivo*

Based on the *in-vitro* data on Stc1-interactions presented in this chapter, Stc1 could potentially be present in multiple, distinct complexes. These could include Stc1-RITS and Stc1-CLRC complexes, as well as the predicted multi-complex structure where Stc1 bridges RITS and CLRC (RITS-Stc1-CLRC). Towards the goal of discovering Stc1-containing complexes, I utilised size exclusion chromatography (SEC) to distinguish distinct Stc1-complexes based on size. In order to be able to detect Stc1-containing complexes, I employed an *S. pombe* strain that harbours C-terminally flag-tagged Stc1, and also C-terminally tagged HA-tagged Dos2. Soluble yeast protein extract was separated on a high resolution Superose column with a broad fractionation range of 5kDa to 5MDa (**Fig 3.9**). Filtrates collected were subjected to flag-immunoprecipitation followed by western blotting to identify the presence of Stc1 across fractions (**Fig 3.10B**). Interestingly, Stc1 was detected in multiple fractions ranging from around 40kDa to 4MDa (**Fig 3.10A** and **3.10B**). Specifically, Stc1 was found in fractions ranging from 33kDa to 69kDa (C12-C9) and 392kDa to 3.7MDa (C2-B5), but not in fractions corresponding to 89kDa to 306kDa (C8-C3). The presence of Stc1 in higher molecular weight complexes does not appear to be the result of non-specific aggregation of proteins, as the absorbance profile at 400nm (A_{400}) does not show aggregation of proteins across Stc1-containing fractions except for fraction B5 (**Fig 3.9** – compare solid blue and red graphs). Unfortunately, attempts to detect the presence of Dos2 (a known Stc1-interactor) in Stc1-containing fractions failed and hence information on co-occurrence of CLRC subunits in Stc1-associated fractions is

currently lacking. Nevertheless, the size range distribution of Stc1 in SEC indicates the potential for several Stc1-containing complexes *in-vivo*.

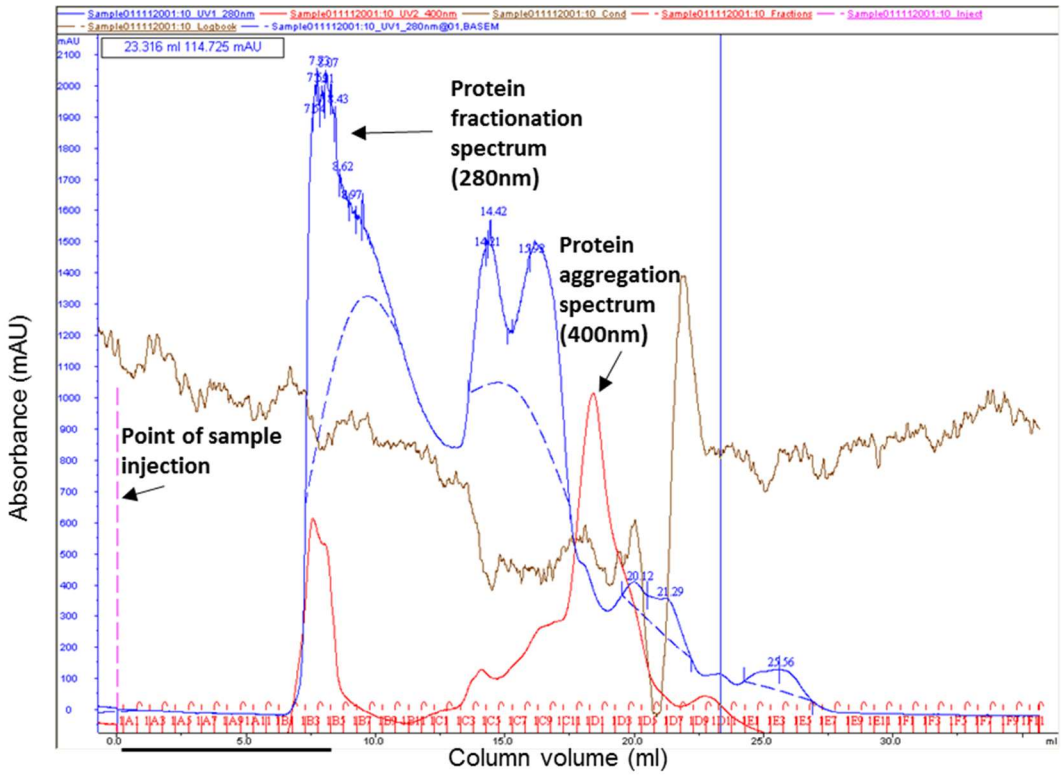


Fig 3.9 Chromatogram from Size Exclusion Chromatography (SEC) showing the fractionation profile of the soluble proteome of yeast cell lysate

Blue and red solid graphs indicate protein fractionation and protein aggregation profiles respectively. Protein absorption (mAU: milli-Absorbance Units) is plotted on the Y-axis against the eluted column volume (in ml) on the X-axis. The solid black line along the X-axis indicates the void volume of the column (corresponding to the elution volume of those molecules larger than the size exclusion limit of the gel filtration medium). Brown graph indicates the conductivity of the buffer used.

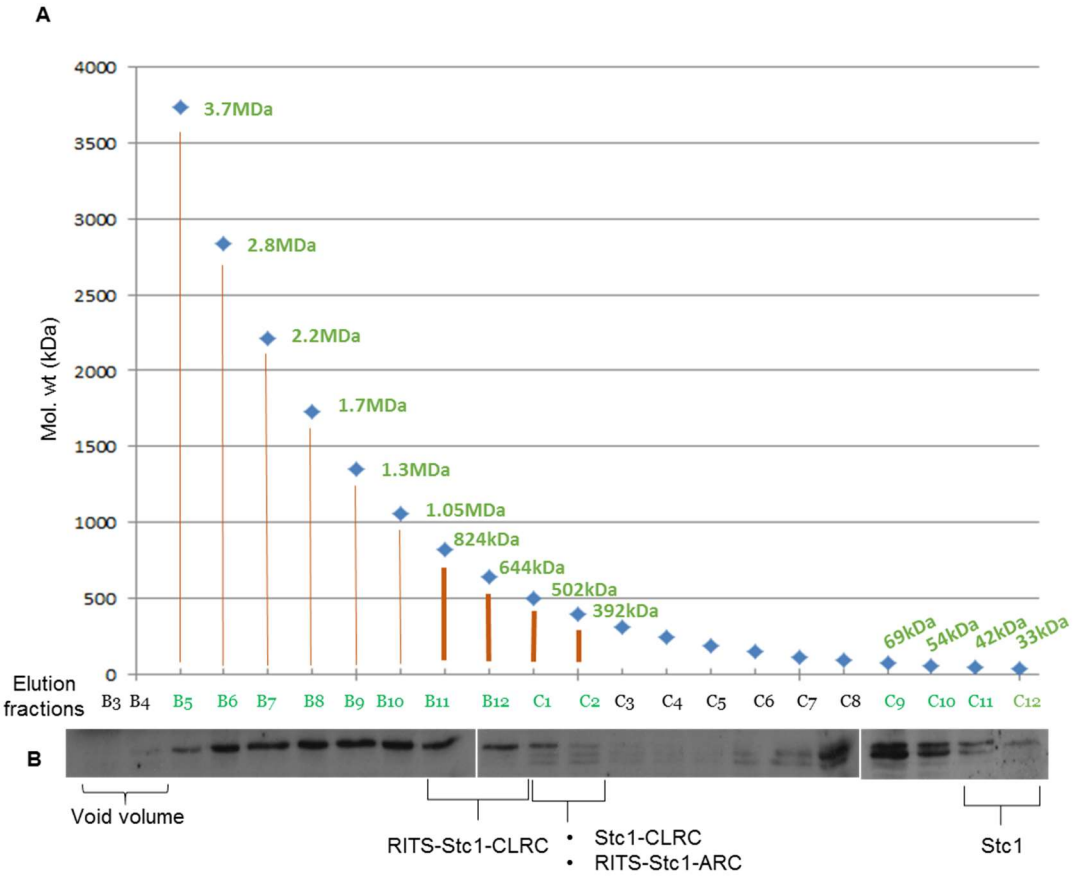


Fig 3.10 Stc1 is potentially present in distinct complexes *in-vivo*

(A) Graph showing SEC fractionation profile of total soluble proteins. Molecular weights of the eluted protein complexes were plotted on the Y-axis against the corresponding elution fractions (volume in ml) on the X-axis. See section 2.5.6 for details on calculation of the molecular weight of the eluted protein complexes in each fraction.

(B) Flag-IP profile of the SEC fractions. SEC fractions were subjected to flag-IP to immunoprecipitate Stc1-flag, followed by SDS-PAGE and western blotting. Stc1-containing fractions that potentially correspond to distinct Stc1-complexes based on their size are indicated using solid red bars on the graph.

3.8 Discussion

Stc1 acts as a crucial link between the RITS and CLRC complexes in RNAi-mediated heterochromatin pathway at *S. pombe* centromeres. Stc1 harbours a tandem zinc finger domain at its N-terminus while the C-terminus is unstructured in solution (He et al. 2013). Stc1 associates with both RITS and CLRC components *in-vivo* and has been shown to interact directly with the RITS subunit, Ago1, *in-vitro* (Bayne et al. 2010). In this chapter, I address questions regarding the mode of Stc1-interaction with the RITS and CLRC complexes, thereby providing further insight into the mechanism of RNAi-mediated chromatin modification at fission yeast centromeres.

My initial quest was to identify the direct binding partners of Stc1 within RITS and CLRC. *In-vitro* binding assays revealed that Clr4 and Dos2 (CLRC), and Ago1 and Tas3 (RITS) have the ability to directly bind Stc1 (**Fig 3.4A** and **3.4B**). Interestingly, during the course of my work, a study published by Kuscu *et al.* also reported Stc1 binding to Dos2 (Kuscu et al. 2014). Their study deduced the subunit arrangement within the CLRC complex through pairwise protein-interaction analyses using recombinant proteins purified from insect cells. Their observation of Stc1-Dos2 interaction using a different method gives further validation to my *in-vitro* assays. Interestingly, my work also identified an additional Stc1 interaction partner within CLRC, Clr4 – the sole H3K9 methyl-transferase. Kuscu *et al.* failed to detect any robust Clr4 interactions, possibly due to a generic problem with the purified protein. However, within the data they present as non-interacting pairs, I observe weak binding of Clr4 with Stc1. It is possible that protein production using the cell-free system-rabbit reticulolysate, utilised in my *in-vitro* assays, might have proven better than the

insect cell system for Clr4 folding and hence for the robust detection of Clr4 interactions.

From the NMR structural data, it can be hypothesised that the largely unstructured C-terminus could act as a separate entity to the ZF-containing N-terminus of Stc1. Indeed, it is widely accepted that intrinsically disordered regions within proteins can act as interaction-interfaces (Dyson & Wright 2005). Since Stc1 seems to have multiple binding partners, I hypothesised that the disordered C-terminus could potentially act as a multi-protein interaction domain. This idea is supported by many examples in the literature where a structurally disordered protein or region binds to multiple binding partners. A well-known example and one that is also relevant to the context is Histone H3 interactions. The N-terminus of H3 is known to undergo a myriad of post-translational modifications, all of which involve interactions with the relevant effector proteins. Intrinsically disorder protein (IDP) prediction tools predict the long N-terminal tail of H3 to be disordered. However, within this predicted disorder the presence of binding regions termed MoRFs (Molecular-Recognition-Features) is thought to impart the ability of the N-terminus of H3 to associate with nine different protein partners. This amazing one-to-many interaction of H3 results in the fine tuning of important cellular events (Hsu et al. 2013).

In-vitro assays carried out using truncated GST recombinants corresponding to the individual N and C terminal domains of Stc1 (ΔC and C) provided unexpected, yet interesting results. Contrary to my hypothesis, the unstructured Stc1-C terminus binds to a single protein partner amongst those tested – the CLRC subunit Dos2, while the structured ZF-containing N-terminus binds to multiple proteins – Ago1, Tas3 & Clr4 (**Fig 3.4C** and **3.4D)**. These results indicate that the structured N-terminus of Stc1 can

act as a multi-protein interaction platform. It is worth noting that although much of this region folds to form tandem zinc fingers (He et al. 2013), the 6aa-long flexible linker and regions within the first 31 residues of the N-terminus of Stc1 ('n') are predicted to be disordered using the DisEMBL prediction method (**Fig 3.6**). Interestingly, no structural data could be obtained for the linker due to conformational flexibility (He et al. 2013), consistent with the disorder prediction by DisEMBL tool.

Potentially, the fact that the N-terminus of Stc1 came out as the 'one-to-many' interaction platform while the comparatively more disordered C-terminus was identified as a 'one-to-one' interaction interface could be due to the limited number of potential binding proteins tested in this study. Hence the possibility of additional unknown interactors via the C-terminus cannot be ruled out. However, the rest of this discussion will focus on the implications of the specific Stc1 interactions observed within the RITS and CLRC complexes. From the data presented in this chapter, it is clear that Stc1 has multiple binding partners. While the C-terminus binds Dos2, the N-terminus associates with a second CLRC subunit, Clr4. This raises the question of whether Dos2 or Clr4 is the key player in Stc1's interaction with CLRC. Two lines of evidence suggest that Dos2 interaction with Stc1 is key to Stc1-CLRC association.

Firstly, I have shown that the C-terminus of Stc1 is sufficient to pulldown all tested CLRC components *in-vivo*, while the ZF domain is not. This indicates that Stc1 is capable of associating with CLRC via its C-terminus. Indeed, increased levels of binding to CLRC components can be observed for Stc1-C. Secondly, tethering assays done in parallel by F. Taglini in the lab supports the role of the Stc1-C terminus as the major recruiter of CLRC. The tethering assays utilise the ability of a chimeric Stc1 protein to be tethered at a euchromatic locus and ask if the tethered Stc1 is capable of

bringing in CLRC and silencing a reporter gene at the tethered locus (Bayne et al. 2010). These assays revealed that tethering of the Stc1 C-terminal domain alone is sufficient to trigger reporter gene silencing, while tethering of the ZF-module alone is not. In fact, in agreement with my pulldown data, methylation levels induced by chromatin tethering of Stc1-C are even higher than those induced by tethering of full-length Stc1 (He et al. 2013). It could be hypothesised that when Stc1 is present without its N-terminus, it solely acts as CLRC recruiter as without the N-terminus (which is involved in RITS binding), there is no competition for Stc1 to be sequestered for its potential role(s) in RNAi pathway. Together these observations clearly demonstrate that the C-terminus of Stc1 is capable and sufficient for CLRC-interaction and recruitment *in-vivo*. This indicates that Stc1-Dos2 interaction is the key mediator of CLRC recruitment, and additionally suggests that although I observed an interaction between the ZF-containing N-terminus of Stc1 and Clr4 *in-vitro*, this binding is not sufficient for CLRC association *in-vivo*.

Stc1 interaction with RITS could potentially be via Ago1 and/or Tas3 based on my *in-vitro* results. However, unlike Ago1, Tas3 has not been detected in any Stc1-co-immunoprecipitation (CoIP) experiments performed in the lab. This could indicate that the Stc1-Tas3 interaction might be a weak transient event occurring during Stc1 association with RITS via Ago1. Alternatively, it could be that Stc1-Tas3 interaction occurs in stoichiometrically less abundant (sub)complexes that are difficult to detect by CoIP-western blot methods. Hence I decided to focus on Ago1 and Clr4 interactions with the N-terminus of Stc1.

In-vitro assays using N-terminal truncations indicated that both the N-terminal 31 residues (‘n’) and the ZF regions of Stc1 are involved in Ago1/Clr4 binding (**Fig 3.5B-**

D). Clearly the lack of just the 'n' region (Stc1- Δn) abolishes Stc1-Ago1 binding *in-vitro.* This could be due to the requirement of 'n' for Ago1-binding. However, the fact that ZF alone has the ability to bind to Ago1 contradicts this. A possible explanation is that when ZF is presented along with the C-terminus (i.e. Stc1-Δn), it requires the N-terminal residues for proper folding. This could be true for all the Stc1 recombinants containing the C-terminus, since when either the whole ZF domain or the individual zfs are presented along with the C-terminus of Stc1, the 'n' region is required for successful Ago1 binding. Another possibility could be that Ago1 binding to Stc1-n stabilises the Ago1 interaction via the Stc1-ZF domain.

Most of my *in-vitro* data is in agreement with *in-vivo* studies performed in parallel by F. Taglini in the lab. Pericentromeric silencing assays where the endogenous *stc1*$^{+}$ was replaced with truncated versions showed that silencing is perturbed in the absence of the ZF module (*stc1-n+C* truncation) (He et al. 2013). The only surprising *in-vivo* data is that silencing is completely lost upon deletion of zf2 [Stc1-(n+zf1+C) in this study; denoted as Δzf2 in (He et al. 2013)] but only moderately affected upon deletion of zf1 [Stc1(n+zf2+C); denoted as Δzf1 in (He et al. 2013)]. Based on the *in-vitro* data presented in this chapter, Ago1 is capable of binding similarly to both the truncated proteins indicating that the binding is not specific to a particular zf within the Stc1-ZF-module. However, it seems that zf2 has a more crucial role *in-vivo*. It could be that zf1 is involved in stabilising Stc1-Ago1 interaction mediated via the rest of the N-terminal regions (i.e. n & zf2). Since the *in-vivo* study did not test the ability of Stc1-Δn to maintain pericentromeric silencing, the importance of the 'n' region for Stc1 interactions *in vivo* remains unknown. However, based on my *in-vitro* results, the 'n' could have some role towards Stc1 associations *in-vivo.*

The *in-vivo* experiments looked only at maintenance; hence it is possible that some of the Stc1 N-terminal interactions I observed *in-vitro* which seems to be dispensable *in-vivo* could have functional implications in the more sensitised stage of heterochromatin establishment. For example, although Stc1-Dos2 interaction appears to be sufficient to maintain CLRC association, it is possible that Stc1-Clr4 interaction could play a specific role during establishment of heterochromatin. Hence these interactions need careful study *in-vivo* to fully understand their roles within the pathway.

Stc1 is required for proper centromeric silencing and for RITS association with CLRC *in vivo*. However, Stc1 interaction with CLRC seems to have only a supporting role in heterochromatin maintenance since H3K9 methylation and Swi6 association are not completely abolished in the absence of Stc1. In contrast, Stc1 appears to be an integral part of the RNAi pathway, based on the complete absence of centromeric siRNA production from centromeric transcripts upon deletion of Stc1 (Bayne et al. 2010).

To better understand the role of Stc1 in the RNAi pathway, I investigated whether Stc1 is required for an intact RITS complex. CoIP analyses provided evidence that Stc1 is not required for interaction of the RITS components Ago1 and Chp1 (**Fig 3.7A**). My data indirectly indicates that Ago1-Tas3 interaction is also not compromised since Tas3 acts as the mediator protein linking Chp1 to Ago1 (Debeauchamp et al. 2008). Since Stc1 seems to be dispensable for an intact RITS, I further investigated if Stc1 can associate with an inactive Ago1/RITS that fails to localise at centromeres. The catalytic-dead Ago1-mutant (Ago1-D580A) fails to release the passenger strand from duplex siRNA, preventing the transition to ss-siRNA loaded Ago1 which predominantly associates with RITS (Buker et al. 2007). ss-siRNA loaded Ago1 is thought to be crucial for localising RITS to the nascent centromeric transcripts via

complementary base pairing. My data indicates that Stc1 can associate with the catalytic-dead Ago1-D580A (**Fig 3.7B**). This suggests that Stc1 association with Ago1 can occur upstream of RITS maturation and localisation to centromeres. Interestingly, a recent study used the Ago1-D580A mutant to investigate the role of Ago1-slicing in heterochromatin assembly beyond siRNA/RITS maturation function. The study reports that priRNAs, Dcr1-independent small RNAs that are loaded into Ago1 as ssRNAs, mediate silencing that is independent of Ago1 slicer activity, but depend on Stc1 (Jain et al. 2016). Taken together, these studies indicate that Stc1 can associate with Ago1 prior to slicing, but is required downstream of slicing, consistent with its proposed role in mediating recruitment of the CLRC complex to chromatin.

Two lines of evidence suggest a possibility for Stc1 to be involved in events that do not require chromatin association: (1) Stc1 association with Ago1 prior to slicing; and (2) Observed cytoplasmic localisation of Stc1. My observation of Stc1 puncta in both the cytoplasm and nucleoplasm **(Fig 3.8**) is interesting since ARC subunits, especially Arb2, have been reported to form distinct puncta in the cytoplasm and nucleus (Buker et al. 2007). Moreover, a study by Carmichael et al. reported presence of dynamic cytoplasmic and nuclear puncta for Ago1, Dcr1 and Rdp1 proteins and their co-localisation in these foci could imply functional significance (Carmichael et al. 2006). Hence future co-localisation studies for Stc1 and the key RNAi-proteins including Ago1 and Arb2 could help in deducing potential additional roles of Stc1 beyond heterochromatin formation.

A preliminary attempt to identify the predicted RITS-Stc1-CLRC multi-complex using SEC provided interesting results, as Stc1 was found to be present in complexes both larger and smaller than predicted for a single RITS-Stc1-CLRC entity (**Fig 3.9, 3.10A**

and 3.10B). Elution of Stc1 at an expected size for a complex with RITS alone (Stc1-RITS) was lacking. However, expected elution profiles for Stc1 in complex with CLRC (Stc1-CLRC) and Stc1 bridging RITS and CLRC complexes (RITS-Stc1-CLRC) could be observed (**Fig 3.10A and 3.10B**). Based on the *in-vivo* data presented in this chapter, the ability of Stc1 to associate with duplex-siRNA loaded Ago1 opens up the possibility of a role for Stc1 in Ago1 loading/RITS maturation along with the ARC complex. Stc1 can be observed in the expected elution size-range for RITS-Stc1-ARC; however, the profile overlaps with size range expected for Stc1-CLRC. The preliminary data is promising and worth following-up in future studies. I also attempted to fractionate immunoprecipitated Stc1-complexes, however this proved unsuccessful. This could be due to instability of the immunoprecipitated Stc1-complexes during fractionation or elution. Indeed, Dos2 is thought to be unstable when subjected to column chromatography (Personal communication – Dr. Thomas Schalch; Pombe 2013 Conference). Extensive attempts to determine *in vitro* whether Stc1 interactions with RITS and CLRC are competitive or simultaneous were also inconclusive. Protein separation techniques like sucrose gradient sedimentation could be used in future to study Stc1-containing complexes. However, data presented in this work hints at the existence of diverse classes of Stc1-containing complexes.

To conclude, the findings presented in this chapter show that Stc1, via its structured and unstructured domains, interacts directly with both RITS and CLRC components, providing further evidence that Stc1 acts as a physical link between RNAi and chromatin modification machineries. This study also provides insight into the basis of Stc1-interactions which will help in understanding Stc1's mode of action in the heterochromatin pathway. The C-terminal region of Stc1 interacts directly with Dos2,

the core protein within the CLRC complex, and several lines of evidence suggest that this is the key interaction mediating association of Stc1 with CLRC. The Stc1-N terminus acts as a multiprotein interaction interface, binding Ago1, Tas3 and Clr4. The potential role for intrinsically disordered regions within Stc1 in its binding dynamics is noteworthy and should be tested *in-vivo.* Key observations including the ability of Stc1 to bind inactive Ago1 and the presence of diverse classes of Stc1-containing complexes *in vivo* are interesting and call for further *in-vivo* studies to probe the key roles played by the various Stc1-interactions.

CHAPTER 4

Characterisation of Dos2 and Clr4 interaction with Stc1

4.1 Characterisation of Dos2-Stc1 interaction *in-vivo*

Based on the data presented in Chapter 3, it is possible that CLRC-Stc1 interaction can be mediated via Clr4 and/or Dos2. Results from my GST-Stc1 CoIP using yeast cell lysate (Chapter 3, **Fig 3.5A**) revealed that, in the absence of endogenous Stc1, CLRC subunits successfully associate with all GST-Stc1 constructs harbouring the C-terminus while the Stc1 ZF-domain alone is not sufficient for binding CLRC (**Fig 3.2**). Since, *in vitro,* Dos2 interacts with the Stc1-C terminus while Clr4 binds the ZF-containing N-terminus, this indicates that the Stc1-Dos2 interaction is likely a key event in Stc1-CLRC interaction and hence I decided to study Dos2-Stc1 interaction in detail.

As a first step towards understanding the basis of Dos2-Stc1 interaction, I tested the requirement for other CLRC subunits for Dos2 interaction with Stc1. To achieve this, I performed CoIP experiments, immunoprecipitating Stc1 and testing if Dos2 was co-immunoprecipitated in the absence of other CLRC subunits. Dos2-interaction with Stc1 was found to be lost in the absence Dos1, Rik1 or Clr4. Interestingly, Dos2-Stc1 interaction was retained in the absence of Cul4 (**Fig 4.1**).

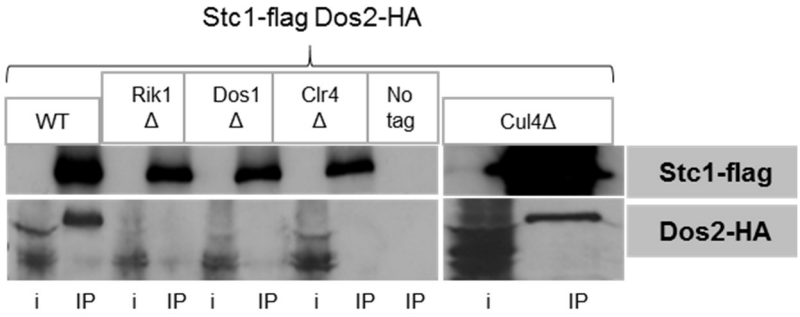


Fig 4.1 Stc1 interaction with Dos2 *in vivo* requires an intact CLRC

Western blot analysis of Stc1-flag immunoprecipitates from wild type and mutant cells to determine the requirements for Stc1 association with Dos2-HA *in-vivo.* Shown are input (i) and immunoprecipitate (IP). No tag indicates an absence of tagged versions of Stc1 and Dos2 and serve as a negative control. This experiment was performed only once.

A previous study (White et al. 2014) reported separation of function mutations within the RFTS domain of Dos2 which abolish its interaction with CLRC subunit Cul4, abolishing H3K9 methylation without affecting siRNA production. Amongst the three reported Dos2 mutants, two mutants, I98A and S100F, were found to compromise heterochromatin integrity at the centromeres, while the third mutant E104A had no silencing defect (White et al. 2014). Both I98 and E104 are conserved residues within other RFTS-containing fungal and mammalian proteins. Dos2-S100 on the other hand, is conserved within fungal RFTS-containing proteins (White et al. 2014). I decided to analyse these Dos2-mutants as well as Dos2-D57A, a mutation within the RFTS domain that hasn't been previously characterised (kind gift from the Allshire lab) to probe their ability to associate with Stc1. I reasoned that since Cul4 does not seem to be a prerequisite for Dos2-Stc1 association, the Dos2 temperature-sensitive mutants (I98A and S100F), which abolish Dos2-Cul4 interaction at 36ºC, might still be able to maintain Dos2-Stc1 interaction, and that this interaction could be important for the uninterrupted siRNA production observed in these Dos2 mutants. The following

experiments using Dos2 mutants were done as part of an Honours project by student Rona Lindsay.

Dos2-Stc1 interactions were examined in strains bearing the previously described single mutations within the RFTS domain of Dos2 – D57A, I98A, S100F and E104A (**Fig 4.2**). To study Dos2-Stc1 interactions, Stc1 was tagged with myc at its C-terminus in cells harbouring flag-tagged versions of the Dos2-point mutants. To determine if the Dos2-mutants affect Dos2-Stc1 interaction, I attempted to perform CoIP experiments on cells grown at restrictive temperature (36ºC), immunoprecipitating Stc1-myc and testing whether Dos2-flag was co-precipitated. Stc1-myc and Dos2-flag-mutant proteins were found to be stable at 32 ºC (**Fig 4.3A** and **4.3B**). At 36ºC, all the Dos2 mutants except Dos2-D57A have been shown previously to be stable (White et al. 2014). Stc1-myc protein was detectable at 36ºC in WT and Dos2 mutant backgrounds, however seemingly at reduced levels compared to at 32ºC (**Fig 4.3A** and **4.3C**). The reduced levels of Stc1-myc observed at 36ºC are not due to lower amounts of total protein since total protein levels appear broadly similar at 32ºC and 36ºC (**Fig 4.3A** and **4.3D**). Strangely, Stc1-myc was seen to migrate faster in the Dos2-I98A mutant background (denoted by asterisk - **Fig 4.3A** and **4.3C**). This could potentially be due to truncation or proteolysis; however, the exact reason remains unexplained.

Stc1-myc protein was efficiently immunoprecipitated in Dos2-D57A and Dos2-E104A mutant backgrounds at 36ºC; however, this was not the case for WT or the other Dos2-mutant backgrounds (**Fig. 4.3C**). The observed weak immunoprecipitation of Stc1-myc in WT and Dos2-I98A and S100F mutant backgrounds was not due to a marked decrease in protein loading, since the ponceau staining indicated broadly

similar loading across all the samples (**Fig 4.3D;** also compare input panels in **Fig 4.3C**). Although Stc1-myc was immunoprecipitated in the Dos2-D57A and E104A backgrounds, I could not detect association of mutant Dos2 proteins. Failure to detect co-immunoprecipitated Dos2 could be due to the low amounts of Stc1-myc protein immunoprecipitated to begin with, in these mutant backgrounds (**Fig 4.3C**).

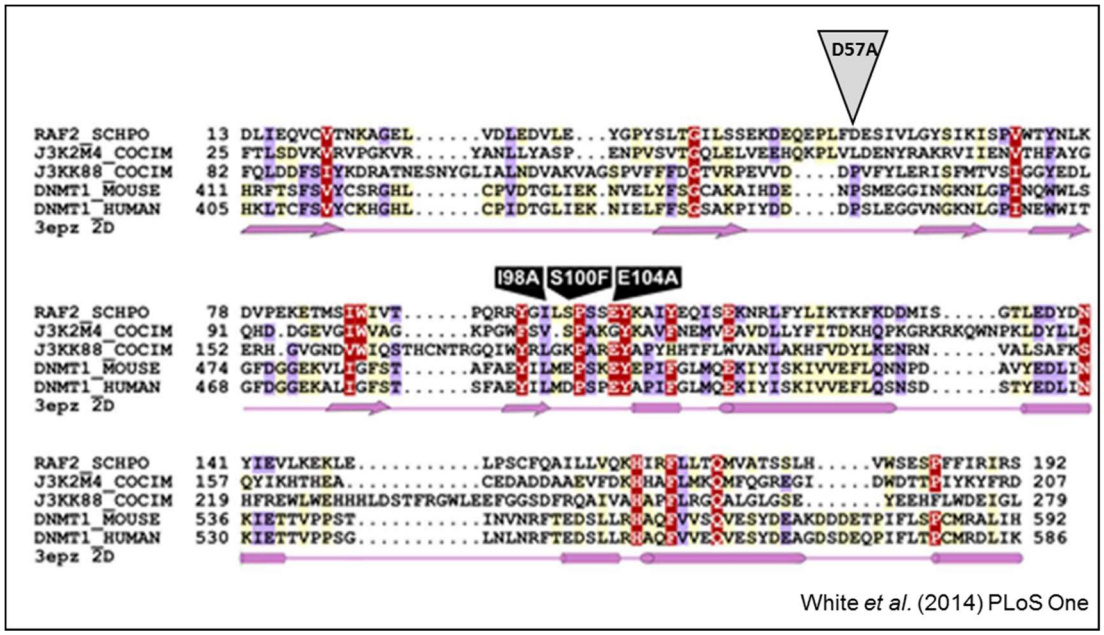


Fig 4.2 Multiple sequence alignment of RFTS domains

Figure adapted from White *et al.* (2014) showing conserved residues within RFTS-containing proteins. The previously described Dos2 RFTS domain mutations are indicated by solid black markers (I98A, S100F and E104A). The position of the uncharacterised mutation in Dos2 RFTS domain (D57A) is indicated by grey marker. Full species names are: RAF2_SCHPO, *Schizosaccharomyces pombe*; J3K2M4_COCIM, *Coccidioides immitis*; J3KK88_COCIM, *Coccidioides immitis*; DNMT1_ARATH, *Arabidopsis thaliana*; DNMT1_HUMAN, *Homo sapiens*, DNMT1_MOUSE, *Mus musculus*. The colouring scheme indicates amino acid conservation using average BLOSUM62 scores.

To further investigate the functionality of Stc1-myc protein in the different Dos2 mutant backgrounds, I created strains with the reporter gene, *ade6*$^{+}$, inserted at the pericentric repeats of centromere one (otr1R(*SphI*):*ade6*$^{+}$). Intact heterochromatin at the centromere results in *ade6*+ silencing rendering the cells red in

colour, while compromised heterochromatin results in white cells (see Chapter 2, **Fig 2.1**). Spotting assays were performed at 32ºC and 36ºC as two of the mutants – Dos2-I98A and Dos2-S100A - display temperature sensitive phenotypes at 36ºC (White *et al* (2014)) (**Fig. 4.4A** and **4.4B**). Unexpectedly, the addition of the myc-tag at the Stc1-C terminus was found to disrupt *ade6*$^{+}$ silencing at 36ºC (**Fig 4.4A**, compare rows 1 and 2). This suggests that myc-tagging of Stc1 at its C-terminus could be impairing the structural stability of the protein which could explain the near failure to immunoprecipitate Stc1-myc at 36ºC in the Dos2-flag (WT) strain (**Fig 4.3C**, lane – WT). Strains bearing Stc1-myc along with either of the temperature-sensitive Dos2-mutants, Dos2-I98A-flag or Dos2-S100A-flag, also displayed disruption of silencing at 36ºC as expected. However, surprisingly, the other two Dos2 mutants, D57A and E104A, appear to suppress the Stc1-myc silencing defect at 36ºC (**Fig. 4.4A** (compare rows #6 with #2) and **4.4B** (compare row #13 with #9). This phenotypic rescue may indicate that the Dos2-D57A and E104A mutant proteins can impart structural stability on the Stc1-myc tagged protein, and provides indirect evidence for Dos2-Stc1 direct interaction *in-vivo*.

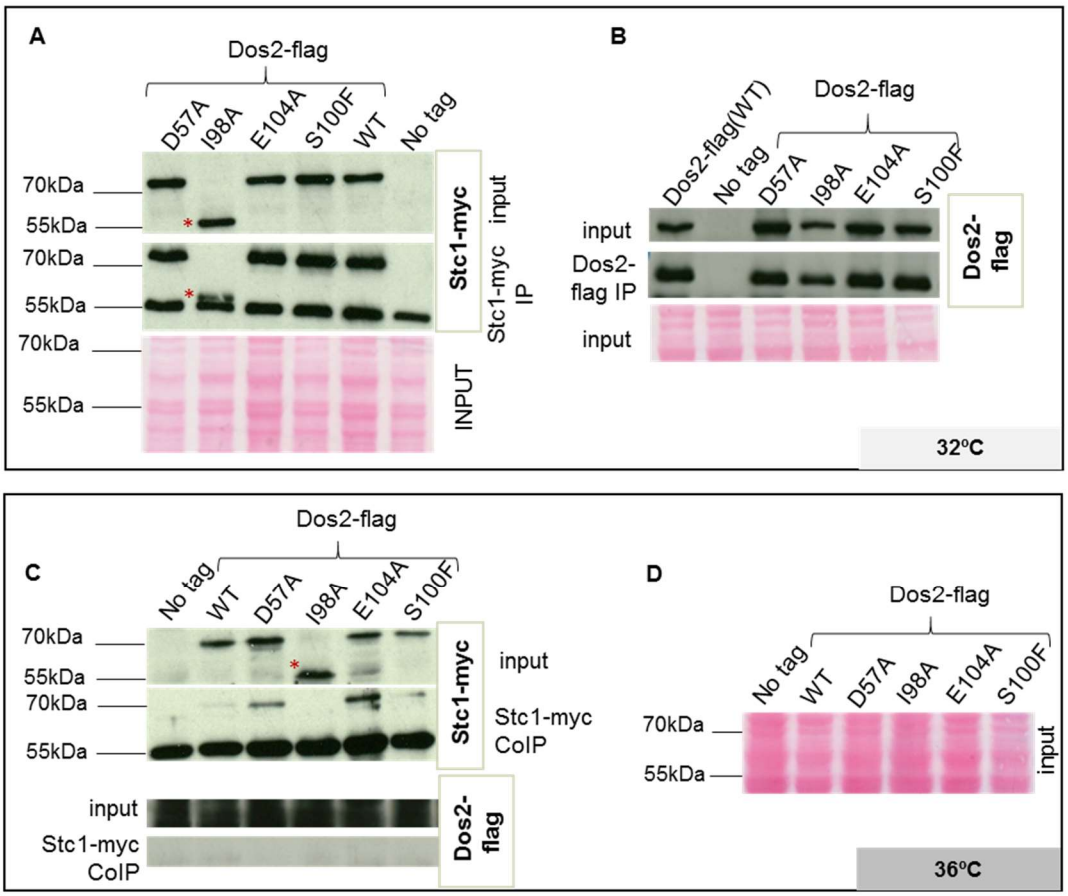


Fig 4.3 Analysis of Stc1-Dos2 association in Dos2 RFTS domain mutants

(A) and (B) Stc1-myc and Dos2-flag are stable at 32ºC. Stc1-myc **(A)** and Dos2-flag **(B)** were immunoprecipitated from strains bearing flag-tagged Dos2-RFTS mutants grown at 32ºC, and analysed by western blot. Note that Stc1-myc in Dos2-I98A mutant background migrates faster than normal (denoted by asterisk). Shown below is ponceau staining of the input blot indicating amount of protein loaded. Experiments performed at least twice by Rona Lindsay.

(C) and (D) Stc1-myc shows reduced stability at 36ºC. Stc1-myc was immunoprecipitated from cells bearing flag tagged Dos2-RFTS mutants grown at 36ºC, and the immunoprecipitates analysed by western blot to study Stc1-Dos2 association **(C)**. Ponceau staining is shown to indicate protein loading **(D)**. Experiments performed at least twice by Rona Lindsay.

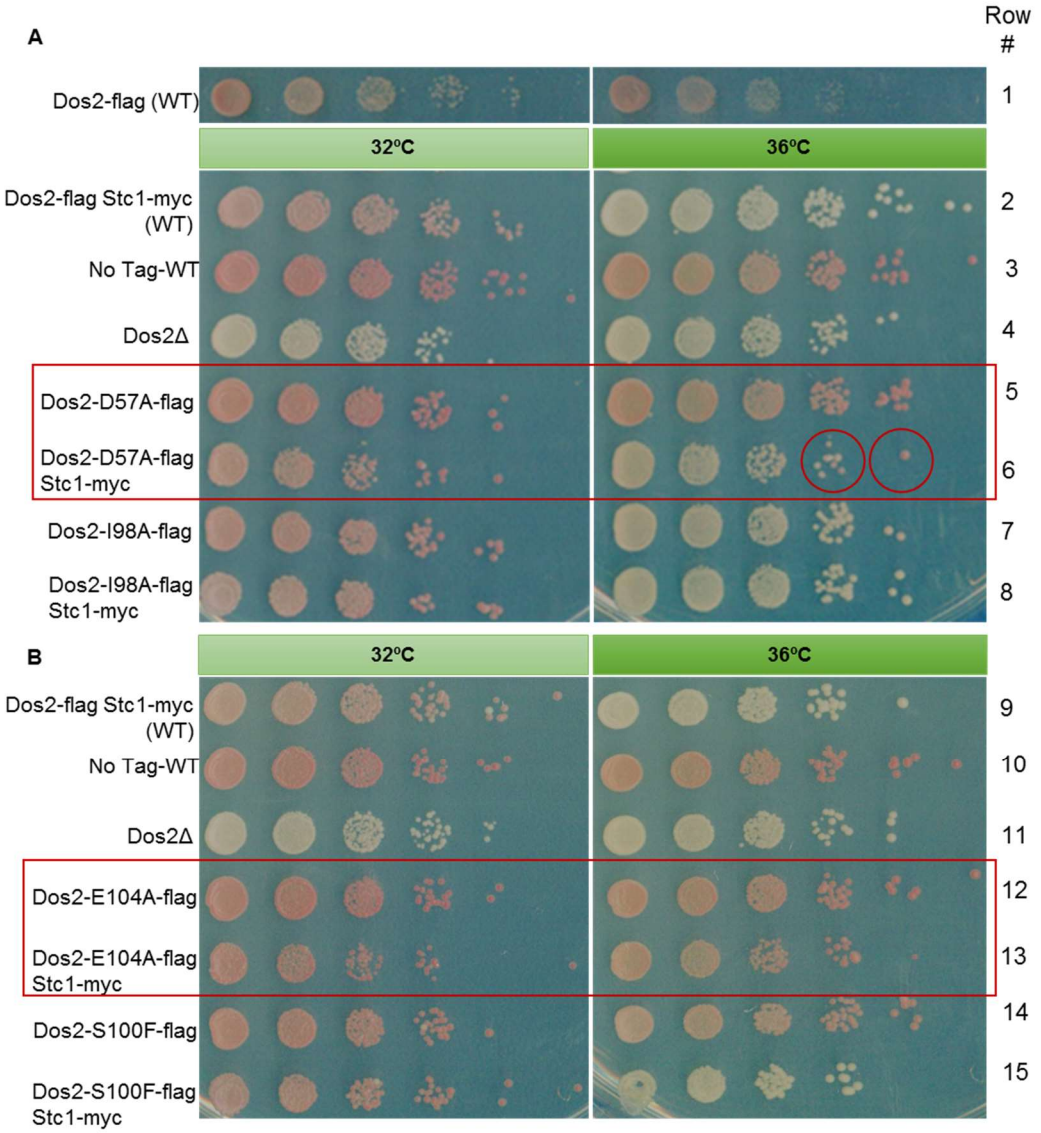


Fig 4.4 Dos2-point mutants rescues the thermosensitive phenotype of Stc1-myc

(A) and (B) Assays for centromeric silencing. Wild type and mutant cells carrying the centromeric reporter otr1R(*SphI*):*ade6*$^{+}$ were plated in serial dilutions on low adenine media and grown at either 32ºC or 36ºC. Wild type cells with silenced otr1R(*SphI*):*ade6*$^{+}$ form red colonies on low adenine media; impaired silencing leads to pale/white colonies. Rescue phenotype displayed by two of the Dos2 mutants are boxed. Circled colonies show variegated phenotype. Experiments performed by Rona Lindsay.

4.2 Characterisation of Clr4-Stc1 interaction

In vitro analyses performed in Chapter 3 identified Clr4 as one of the CLRC components potentially in direct interaction with Stc1. A previous study indicates that Stc1 retains its interaction with the core CLRC components to some degree even upon deletion of Clr4 (Bayne et al. 2010). In addition, observations made in this study and work by Kuscu et al. (Kuscu et al. 2014) confirm a direct interaction of Stc1 with Dos2. All these data together indicate that Dos2 may be the major CLRC-interaction partner for Stc1. However, it is worth noting that the Kuscu (Kuscu et al. 2014) study failed to detect any reliable interactions for Clr4 in their pairwise analyses. Hence detection of Clr4 as a direct interactor of Stc1 in my *in vitro* analysis is interesting and worth following up as this might help elucidate any potential additional ways in which Clr4 could be regulated by components of the RNAi-mediated chromatin modification pathway.

Clr4 belongs to the Su(var)3-9 family of methyltransferases, possessing a well-conserved Chromo-Domain (CD) at its N-terminus, and a catalytic domain at the C-terminus consisting of pre-SET, SET and post-SET domains (Horita et al. 2001). Between the CD and the catalytic domains is a 122aa region that lacks any known domain assignment or sequence conservation with other Su(var)3-9 proteins and will be referred to as the Post-Chromo-Domain (PCD) in this study (**Fig 4.5A**).

4.2.1 Clr4 utilises its non-catalytic N-terminal domain to associate with Stc1

In order to characterise the region in Clr4 involved in binding to Stc1, I initially constructed two truncation mutants of Clr4 – Clr4-N, which consists of the CD and the PCD, and Clr4-Cat, comprising the catalytic domain (**Fig 4.5A**). I employed *in-vitro* binding assays using recombinant GST-Stc1-WT protein, to determine the Stc1-binding ability of the *in-vitro* translated S^{35}-labelled Clr4 truncation mutants. All S^{35}-labelled Clr4-proteins were successfully produced *in-vitro*. Although I detected bands corresponding to the expected size for all of the S^{35}-Clr4 proteins, in each case I also observed slower migrating species corresponding to higher molecular weight (**Fig. 4.5B**). For S^{35}-Clr4-Cat the additional species ran at approximately 4kDa above the expected size, while in the other cases the size difference was approximately 10kDa. It could be that the higher molecular weight species represents the actual Clr4 proteins while the faster migrating species could be the products of proteolytic cleavage of Clr4 proteins *in-vitro.* However, the fact that in each case the faster migrating species is of the expected size makes this unlikely. Another interesting possibility is potential post-translational modification of S^{35}-Clr4-proteins *in-vitro.* Rabbit reticulo-lysate systems have been previously reported to be capable of post-translationally modifying proteins *in-vitro* (Promega n.d.).

As expected, Clr4-WT protein clearly associates with GST-Stc1-WT protein but not GST alone in the GST-pulldown assay. Interestingly, Clr4-N also associates strongly with Stc1, while Clr4-Cat does not (**Fig 4.5B**). As shown in Chapter 3 (**Fig 3.4D**), Clr4-WT protein associates with the N-terminal domain of Stc1. In order to verify the finding that Stc1 binds to the N-terminal domain of Clr4, I utilised GST-Stc1-(ΔC) recombinant protein to test for interaction with the non-catalytic and catalytic domains

of Clr4 *in-vitro*. The results indicate that the non-catalytic domain of Clr4 indeed associates with the N-terminal tandem zinc-finger domain of Stc1 (**Fig 4.5C**), backing up the results observed with GST-Stc1-WT protein (**Fig 4.5B**). The very weak signal observed for the catalytic domain of Clr4 can be attributed to non-specific binding. However, weak interactions within this region during Stc1-binding cannot be ruled out.

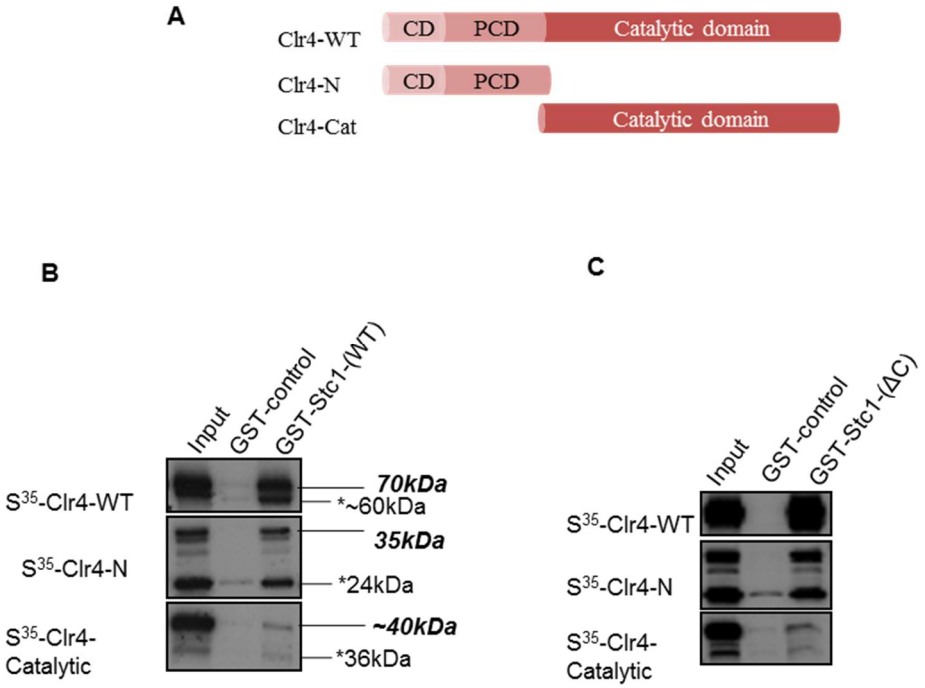


Fig 4.5 Non-catalytic domain of Clr4 responsible for binding to Stc1-N terminus

(A) Schematic representation of wild type and truncation mutants of Clr4

(B) and (C) *In-vitro* assays showing Clr4-N terminus binding to Stc1. Clr4 wild type and truncated proteins translated *in-vitro* were incubated with wild type Stc1 recombinant protein **(B)** or Stc1 harbouring the ZF-containing N-terminal domain **(C)**. Following GST-pulldown, bound proteins, along with input samples (input corresponds to 5% of sample used for the binding assay), were analysed by SDS-PAGE and autoradiography. Asterisk indicates the expected molecular weight for each form of Clr4. These assays were performed at least twice.

4.2.2 Clr4 Chromo-Domain displays separable functions in Stc1 and H3K9-methylation binding *in-vitro*

As observed from assays shown in section **4.2.1**, the CD-containing N-terminal region of Clr4 is involved in binding Stc1. Chromo-domains can be found in a variety of chromatin-associating proteins, including other proteins involved in heterochromatin formation in *S. pombe,* eg. the HP1-homolog Swi6. Clustal Omega multiple alignment of a selection of CD-containing proteins including Clr4 clearly shows that the nature of the core residues of chromodomains is conserved across species (**Fig 4.6**) (McWilliam et al. 2013; Sievers et al. 2014; Li et al. 2015) .

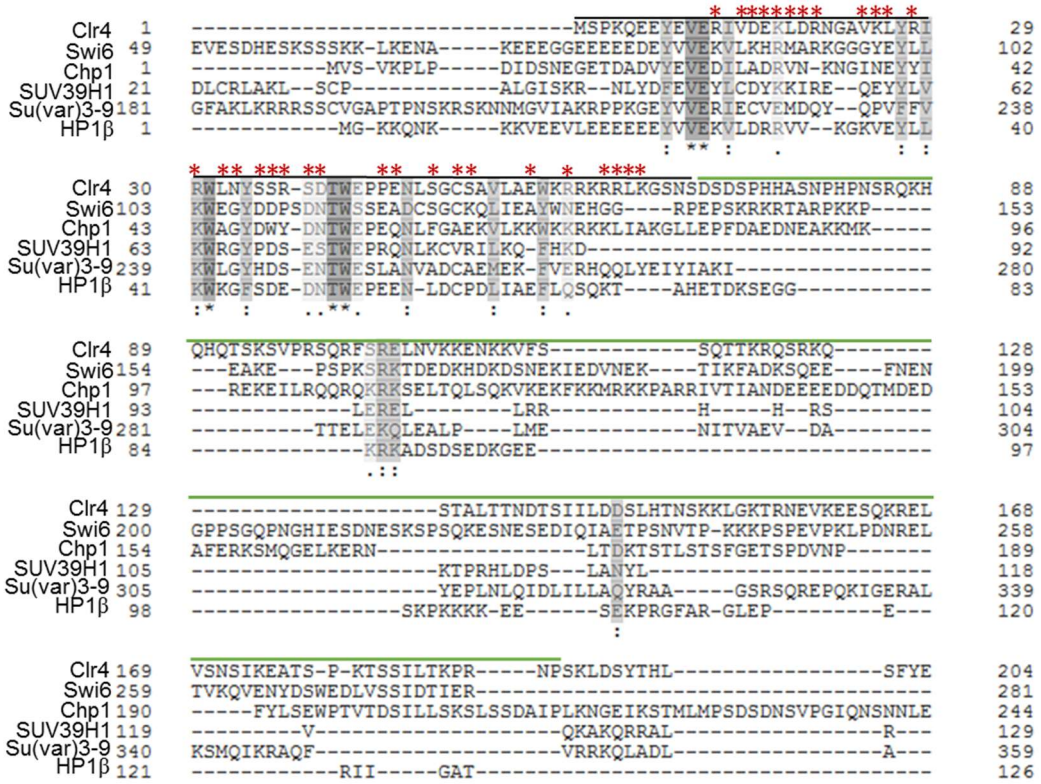


Fig 4.6 Multiple sequence alignment of chromodomain-containing proteins

Clr4, Swi6 and Chp1 are fission yeast CD-containing proteins; SUV39H1, Su(var)3-9 and HP1β are Clr4-homologs in human, *Drosophila* and mouse respectively. Sequences were obtained from Pombase and GenBank databases. The black line indicates the conserved CD region with red asterisks indicating the less conserved residues mutated in the alanine-scanning mutagenesis experiment. The green line corresponds to the PCD region of Clr4 showing no significant conservation amongst the representative Clr4 homologs.

In order to rule out the possibility of non-specific binding of Stc1 to any CD-containing protein, I tested Stc1 binding to the full length Swi6 protein as well as the Swi6-CD alone. Swi6-WT and Swi6-CD proteins were transcribed and translated *in-vitro* as S^{35}-labelled proteins. Neither the Swi6-WT protein nor the Swi6-CD bound to Stc1 *in-vitro,* confirming the specificity of the Stc1-Clr4 interaction (**Fig 4.7A**). Since CD-proteins Swi6 and Chp1(Chapter 3**, Fig 3.4B** and **3.4C**) failed to bind to Stc1, I reasoned that highly conserved CD residues might not be involved in Stc1-Clr4 interaction. Hence I next tested the involvement of residues within the CD of Clr4 that are not strongly conserved in other CD-proteins in *S. pombe* (**Fig 4.6**). I employed alanine-scanning methodology, whereby a set of adjacent residues are mutated to alanine in order to study the functional role of those residues. Six different sets of alanine scanning mutations (M1-M6) within the CD of full-length Clr4 were generated using site-directed mutagenesis (**Fig 4.7B**).

Clr4-CD mutants were successfully expressed as S^{35}-labelled proteins *in-vitro* and tested for their ability to bind to Stc1. All Clr4-CD mutant proteins were found to bind to Stc1 at comparable levels to Clr4-WT protein, while there was almost no binding to GST-only protein, confirming the specificity of Clr4 binding to Stc1 (**Fig 4.7C**). Though the residues mutated within the CD are not thought to be directly involved in binding to methylated H3K9 (Horita et al. 2001), I reasoned that, owing to their proximity to the key residues involved in H3K9-methylation binding, these residues might have an effect on the efficiency of binding to H3K9-methyl marks. In order to investigate the effect of these CD mutations on Clr4 binding to methylated H3K9, I utilised a biotinylated H3K9me2-peptide pulldown assay. Clr4-CD mutants and Clr4-

WT expressed as S^{35}-labelled proteins *in-vitro* were incubated with biotinylated H3K9me2 peptides. Binding was analysed by pulling down the biotinylated peptides and detecting the bound proteins by SDS-PAGE and autoradiography. The results indicate that the CD-mutants M1-M3 have largely lost H3K9me2 binding *in-vitro* (**Fig 4.7D**). Note that the *in-vitro* translated S^{35}-labelled Clr4 protein shows some non-specific background binding to the sepharose beads used for pulling down biotinylated peptides. However, it is evident that *in-vitro,* CD-mutants M4-M6 retain H3K9me2-binding ability. Taken together, the results from these *in-vitro* binding assays indicate that Clr4-CD binding of Stc1 and H3K9-methylation are potentially separable activities.

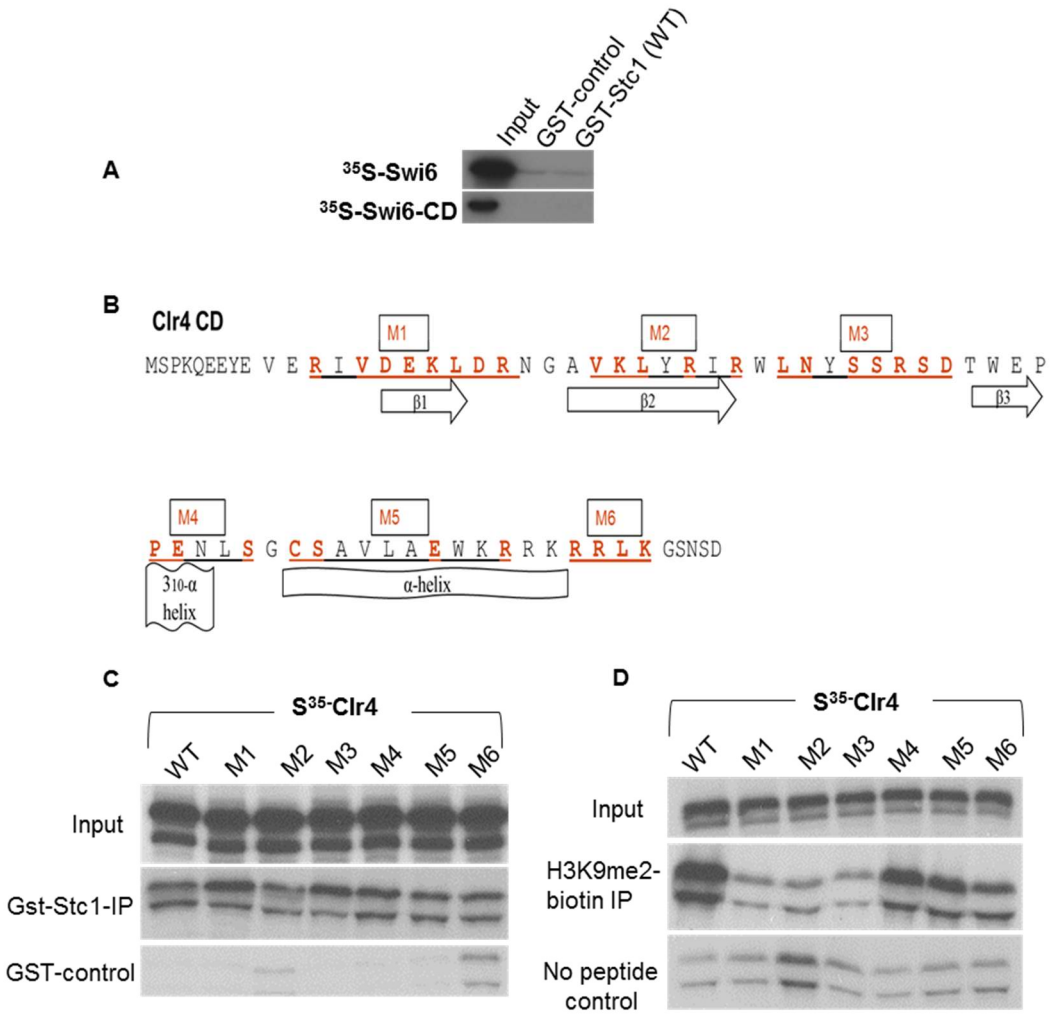


Fig 4.7 Stc1-Clr4 association is specific, and the Clr4 chromo-domain has potentially separate functions in Stc1 and H3K9-methylation binding

(A) Stc1 does not interact with the CD-containing protein Swi6. *In-vitro* translated Swi6 (full length or CD alone) was allowed to bind to either GST alone (negative control) or GST-Stc1-WT recombinant proteins. Bound proteins were recovered via GST-pulldown and analysed by SDS-PAGE and autoradiography.

(B) Clr4 CD sequence showing residues mutated in Clr4-CD mutants. M1-M6 indicates each of the Clr4-CD mutants, with coloured residues representing those mutated to alanine in each of the respective Clr4-CD mutants (M1-M6: underlined). Secondary structure illustration is based on the solution structure of Clr4 CD by Horita *et al.*(2001)

(C) and (D) Clr4-CD mutants indicate potential separation of function. Clr4-WT and CD mutants were translated *in-vitro* and incubated with either GST-Stc1-WT protein **(C)** or H3K9me2-biotinylated peptides **(D)**. Incubation of Clr4-proteins with GST alone or beads only ('no peptide') served as negative controls. Following GST pulldown **(C)** or biotin pulldown **(D)**, bound proteins, along with input samples (input corresponds to 5% of sample used for the binding assay), were analysed by SDS-PAGE and autoradiography. *In-vitro* assays were performed only once

4.2.3 Residues within the Post-Chromo-Domain of Clr4 are involved in Stc1 binding *in-vitro*

Since mutations within the CD of Clr4 did not affect Stc1-binding *in-vitro,* I went on to further characterise the region within the N-terminal non-catalytic domain of Clr4 responsible for interacting with Stc1. Clr4 constructs that code for CD and PCD individually (**Fig 4.8A**) were translated *in-vitro* as S35-labelled proteins and incubated with GST-Stc1-WT protein to investigate which of these domains is responsible for the observed direct binding of Clr4 with Stc1. GST-pulldown experiments surprisingly revealed that neither of the domains when present alone was sufficient for Stc1 binding (**Fig. 4.8B**).

Since Stc1 failed to bind to the individual domains within the Clr4-N terminus, I constructed a further series of Clr4-segments containing CD plus increasing portions of the PCD (**Fig. 4.8C**). These Clr4-truncations were again produced as S^{35}-labelled proteins *in vitro* and tested for Stc1 binding activity in GST-Stc1 pulldown assays. The results revealed that the CD plus the first 60 residues within PCD (Clr4-CD+60aa) is necessary and sufficient for successful Stc1-binding (**Fig. 4.8D**). The CD along with the first 20residues within PCD (Clr4-CD+20aa) failed to bind Stc1. This could indicate that the region between 20-60 residues into PCD [ie, residues 90-129 within full-length Clr4] might be involved in direct interaction with Stc1; alternatively, it might be that the shorter protein (Clr4-CD+20aa) does not achieve the proper folding required to bind Stc1 *in-vitro*.

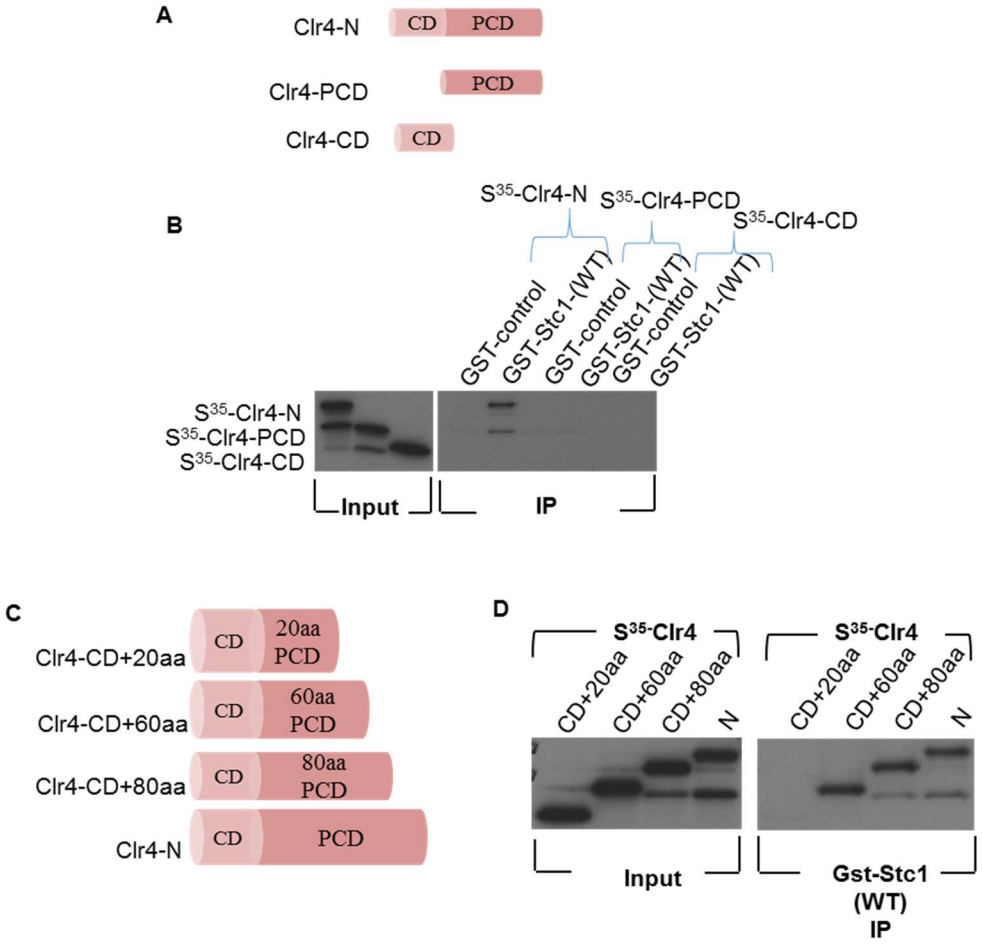


Fig 4.8 The Chromo-domain of Clr4 alone is not sufficient for Stc1 binding

(A) Schematic representation of Clr4 constructs encoding Clr4-N and domain-deletion mutants

(B) Clr4 N-terminal domains alone are not sufficient for Stc1-binding. Clr4 truncation mutants were translated *in-vitro* and incubated with either GST alone or GST-Stc1. Following GST-pulldown, bound proteins were analysed by SDS-PAGE and autoradiography to study association of Stc1 with the individual domains within Clr4 N-terminus.

(C) Schematic representation of Clr4 constructs encoding CD plus additional residues from PCD

(D) Clr4-CD requires residues within the PCD for Stc1-binding. *In-vitro* translated Clr4 constructs were tested for binding to GST-Stc1 by GST-pulldown followed by autoradiography.

Assays were performed at least twice.

4.3 Discussion

In-vitro data on Stc1-CLRC interaction presented in Chapter 3 identified Dos2 and Clr4 as direct binding partners of Stc1. Dos2 was found to bind to the C-terminus of Stc1, while Clr4 binds to the Stc1 N-terminus. In this chapter I attempted to further investigate the molecular basis and functional significance of these interactions. Stc1 interaction with Dos2 was found to depend on CLRC subunits Rik1, Dos1 and Clr4 but not Cul4. Dos2 has been reported as the hub protein within CLRC, interacting with all CLRC subunits except Clr4 (Kuscu et al. 2014), and Stc1 has been envisaged as an auxiliary component of CLRC (Bayne et al. 2010; Kuscu et al. 2014).

Although I found that Stc1-Dos2 interaction was lost in the absence of Clr4, a previous study (Bayne et al. 2010) reported low levels of Dos2 interaction with Stc1 in the absence of Clr4. This difference could be partly due to the low exposure time I used for detection, meaning any weak signal could have escaped detection. Nevertheless, the drastic reduction in Dos2-Stc1 association in the absence of Clr4 raises the possibility that Clr4 is required for the efficient association of Dos2 with Stc1. My observations are consistent with Dos2, Dos1, Rik1 and Clr4 all being required for the efficient association of Stc1 with CLRC, with Stc1-Dos2 interaction a crucial event in CLRC-Stc1 association.

Attempts to study Stc1 interactions using yeast two hybrid (Y2H) failed. This approach seems to be unsuitable for studying Stc1-interactions as White *et al.* also failed to see Stc1-Dos2 interaction using Y2H. The fact that Dos2-Stc1 interaction was observed using a different method to mine, by another group (Kuscu et al. 2014) provides confidence in the findings of the *in-vitro* approach I have utilised in studying Stc1-interactions. However, attempts to study the role of distinct domains of Dos2 in Stc1-

binding *in-vitro* failed due to the unstable nature of *in-vitro* translated Dos2-truncated proteins. Instead I decided to analyse existing Dos2 mutants that carry point mutations within the RFTS domain (White et al. 2014). It was shown previously that the thermosensitive Dos2 mutants Dos2-I98A and Dos2-S100F fail to bind to Cul4 and compromise heterochromatin integrity at 36ºC (White et al. 2014). Unexpectedly, in CoIP experiments I failed to efficiently immunoprecipitate Stc1-myc at 36ºC in WT and mutant backgrounds (**Fig 4.3C**). Stc1-myc was detectable in the whole cell extract (See 'input' panels **Fig 4.3C**) but at reduced levels compared to 32ºC (compare 'input' panels in **Fig 4.3A** and **4.3C,** and ponceau staining in **Fig 4.3A** and **4.3D**). This could indicate that, at 36ºC, Stc1-myc fails to fold properly so that the myc epitope is not available for binding by the myc-antibody and hence fails to immunoprecipitate. Consistent with this, I observed compromised heterochromatin integrity at 36ºC when Stc1 was tagged with myc at its C-terminus. Stc1-myc does not perturb heterochromatin integrity at 32ºC; and, moreover, Stc1-myc can be successfully immunoprecipitated at 32ºC (**Fig 4.4A**, **4.4B** and **4.3A**), suggesting that folding of tagged Stc1 is impaired at 36ºC but not at 32ºC. Thus C-terminal epitope tagging of Stc1 with myc renders the Stc1 protein thermosensitive.

Interestingly, Stc1-myc could still be immunoprecipitated with reasonable efficiency in Dos2-mutants D57A and E104A at 36ºC suggesting that Stc1-myc was less impaired in these mutant backgrounds (**Fig 4.3C**). This agrees with the observation that these mutants have the ability to rescue the temperature-sensitive loss of silencing phenotype of Stc1-myc-bearing strains either partially (Dos2-D57A) or fully (Dos2-E104A) (**Fig 4.4A** and **4.4B**). This suggests that either the Dos2 residues D57 and E104 are involved in direct interaction with Stc1, or mutating these residues renders

Dos2 in a more favourable state for Stc1-binding *in-vivo*. Either way, the data provides *in-vivo* evidence for direct Dos2-Stc1 interaction. The Dos2-D57A and Dos2-E104 mutants do not display any silencing defect at either 32ºC or 36ºC; in the future it would be interesting to test if the Dos2-D57A, E104A double mutant perturbs silencing. Additionally, Dos2-Stc1 *in-vivo* association could be tested using a different tagging strategy.

While my attempts to translate Dos2-truncations *in-vitro* failed, similar attempts on Clr4 proved to be fruitful. *In-vitro* assays revealed that the N-terminus of Clr4 interacts with Stc1 (**Fig 4.5B and 4.5C**). This led me to hypothesise that the well-defined CD of Clr4 might be responsible for Clr4-Stc1 interaction. Two lines of evidence indicate that Stc1 binding is not a general property of CD-containing proteins: (1) Stc1 fails to bind Swi6 protein (full-length or CD alone) (**Fig 4.7A**); and (2) Stc1 does not exhibit interaction with the RITS complex component Chp1, another CD-containing protein (Chapter 3, **Fig 3.4B**). In fact, Stc1 failed to bind to either the Clr4-CD or PCD when these were expressed individually (**Fig 4.8A** and **4.8B**). Rather, I found that the Clr4-CD along with the first 60 residues within PCD is necessary and sufficient for Clr4-Stc1 interaction *in-vitro* (**Fig 4.8C**). For the purposes of this study, the Clr4-CD was defined according to Horita *et al.* (Horita et al. 2001). Although Zhang *et al.* used a larger segment to demonstrate the H3K9-methyl binding ability of the Clr4 chromo-domain (Zhang et al. 2008), a recent study showed that the region defined as Clr4 CD by Horita *et al.* is sufficient for binding to the H3K9 methylation mark (Schalch et al. 2009). Their study indicates that Clr4-PCD is not required for the function of the CD in direct recognition of H3K9-methyl mark. This additionally suggests that the failure of Stc1 to bind to the Clr4-CD only or Clr4-CD+20aa proteins might not be due to

improper folding of the CD in these proteins, and could potentially be due to the absence of PCD residues required for Stc1-binding. However, the study by Schalch *et al.* was not conducted using *in-vitro* translated Clr4 CD, and hence the failure of Stc1 to bind Clr4-CD in this system might still be due to misfolding.

The requirement for both CD plus part of the PCD for Clr4 binding to Stc1 raises the question as to the extent of CD residues required for Stc1-binding. Surprisingly, following *in-vitro* alanine scanning mutagenesis, none of the Clr4-CD mutants tested showed impaired Clr4-Stc1 binding *in-vitro* (**Fig 4.7C**). The dispensability of the tested residues within Clr4-CD implies that these residues might not be directly involved in Stc1-binding. So, if Stc1 indeed interacts with the residues within the PCD stretch, there is a possibility that the PCD requires the CD as a structural support for its proper folding to be able to bind to Stc1. It could be that combining some of the CD alanine mutations could prove detrimental to Stc1-binding either due to effects on CD structure or involvement of these residues in direct Stc1-association. Interestingly, three of the alanine-scanning Clr4 mutants (M1-M3) lost the ability to bind to H3K9me2 peptide (**Fig 4.7D**). The mutations in these mutants reside across the first three beta-strands of Clr4-CD (**Fig 4.7B**). The relatively high number of mutations within each of these Clr4-CD-alanine-mutants could affect the core structure of the Clr4 CD required for recognising the H3K9me mark. However, if there is indeed structural perturbation in these cases, it does not affect Clr4-Stc1 binding *in-vitro*. It is also possible that these mutations affect H3K9me-binding without having a major effect on Clr4 structure. In either scenario, it can be inferred that residues that form part of the antiparallel strands β1 and β2 and the disordered loop between β2 and β3

are dispensable for Clr4-Stc1 interaction *in-vitro*; on the contrary these residues are required for the recognition of H3K9me marks.

The *in-vivo* data on Dos2 and *in-vitro* data for Clr4 presented in this chapter further validates the direct association with Stc1 observed for these proteins. Dos2 *in-vivo* mutagenesis studies have not been successful in determining the residues involved in direct interaction with Stc1. This could be partly due to Dos2 being the hub protein within CLRC, binding to almost all of the CLRC subunits. It is possible that the same binding interfaces could be involved in multiple protein interactions making it difficult to identify residues specifically involved in binding to one protein partner. The molecular basis of Stc1-Dos2 interaction could in the future be elucidated using structural studies. Currently, there are no obvious homologues of Dos2 based on sequence. Identification of a functional homologue based on the structural properties of Dos2 could significantly help in understanding how hub proteins similar to Dos2 might act in RNAi-mediated chromatin modification mechanisms in higher organisms. Stc1-Clr4 interaction, on the other hand, could be further studied *in-vivo* and *in-vitro* by dissecting out the residues involved in the interaction, and using mutational analyses to understand the role of this specific interaction in the pathway. Attempts to use bacterially produced Clr4 in *in-vitro* Stc1-binding assays were unsuccessful, as were several attempts to co-purify pairs of interacting proteins co-expressed in bacteria (for example, Stc1 and Dos2; Stc1 and Clr4). Hence further studies using recombinant proteins will likely require expression in different host systems. Such studies could help address whether Stc1-bound Clr4 retains the ability to be catalytically active and/or to bind to H3K9-methyl marks, questions that are important in understanding the functional significance of the Stc1-Clr4 interaction. Another possible approach

would be to design synthetic peptides corresponding to the Stc1-binding region of Clr4, and use these in competition assays to investigate, for example, whether Stc1 bound to Clr4-peptides can be displaced by H3K9-methyl peptides *in-vitro*. Since Stc1 interaction with Dos2 appears to be the major driving factor in Stc1-CLRC interaction, studying Stc1-Clr4 interaction *in-vivo* may prove difficult. Nevertheless, *in-vivo* studies using mutational analysis could provide clues to possible transient roles of this interaction (for example in establishment but not maintenance of heterochromatin) which cannot be deduced from *in-vitro* analyses.

CHAPTER 5

Discussion

5.1 Introduction

The RNA interference (RNAi) machinery plays a key role in effecting centromeric chromatin modifications resulting in heterochromatin formation in the fission yeast *Schizosaccharomyces pombe* (Lejeune et al. 2010). Stc1 is a critical pathway component that mediates the RNAi-dependent recruitment of CLRC to chromatin (Bayne et al. 2010). However, little is known about how Stc1 mediates this interaction, apart from it being a scaffold protein that connects CLRC to Ago1. In this work, I presented evidence that Stc1 interacts directly with CLRC and that this association is largely mediated by the unstructured C-terminus of Stc1 binding to the CLRC subunit Dos2. My study additionally indicated that the structured zinc finger domain of Stc1 is capable of carrying out multiple interactions and hints towards a possible role for disordered residues for the observed plasticity. Here, I will initially discuss aspects related to the functional significance of intrinsic disorder in proteins. Following this, I will discuss possible models for the order of events involved in Stc1 mediated CLRC recruitment, based on the evidence presented in this thesis. Further discussions will be focussed on representative examples of Stc1-like proteins in other organisms, and I will conclude with suggestions for future work for elucidating the significance of the Stc1-interactions presented in this thesis.

5.2 Conformational versatility – A protein perspective of socialising

In the past, proteins were thought to possess a rigid structure and hence protein-protein interactions were seen as the result of 'lock and key' precision joining. However, it is now clear that protein-protein interactions are seldom the outcome of rigid protein

structures interlocking. It is widely accepted that conformational flexibility of proteins is a key feature in protein-protein interactions.

Intrinsically disordered proteins (IDPs) have gained attention in the post-genomic era as various studies have indicated structural plasticity imparted by such proteins as a key factor for many vital cellular events including regulation of transcription, translation and cell cycle. Intrinsically disordered regions (IDRs) lack a defined three dimensional structure and have little or no secondary structure elements. The presence of protein entities that exist without a predetermined structure in physiological conditions is quite a common occurrence across various kingdoms of life (Wright & Dyson 1999). Interestingly, the prevalence of IDPs increases hierarchically, with eukaryotic proteomes typically consisting of over 30% of proteins with IDR greater than 30 residues in length (Ward et al. 2004). The significance of the presence of IDRs within a protein is demonstrated by the functional promiscuity displayed by such proteins through their ability to associate with multiple protein partners. Importantly, many of these proteins have been implicated in diseases. A well-documented example is that of the tumour suppressor p53. The various signal transduction events carried out by p53 requires it to bind to multiple proteins including activators and inhibitors of p53 as well as downstream targets such as transcription factors. Greater than 70% of these interactions are mediated by the intrinsically disordered regions of p53 which constitute 29% of the protein. Interestingly, the occurrence of post translational modifications (PTMs) in p53 is highly biased towards the IDRs (Oldfield et al. 2008). Thus IDRs within p53 appear to be crucial in mediating and regulating its various functions. The idea that IDRs can act as a platform for various PTMs is inviting as the

regulation of protein function via modification of the more accessible IDRs would be an excellent way for the cell to tightly regulate molecular events.

In the context of IDPs, Stc1 seems to be a good candidate for a protein displaying functionally significant disorder, similar to p53. Stc1 harbours a defined tandem zinc finger (ZF) domain within its N-terminus, but the C-terminus and the first 31 N-terminal residues ('n') are intrinsically disordered. Based on the data presented in this thesis, Stc1 displays 'one to many' interactions mediated via the N-terminal domain of Stc1. The unique nature of the Stc1 ZF domain means that the individual zinc fingers have the ability to assume different orientations with respect to each other, a property thought to be conferred by the flexible linker joining them (He et al. 2013). The multiple interactions via Stc1 N-terminus could therefore arise as a result of the zinc fingers being able to act as separate units and hence bind different protein partners. However, *in-vitro* data presented in this thesis and *in-vivo* data obtained from the lab hints that these interactions are not restricted to the ZF domain alone and rather involve the whole of the N-terminus. The ability of the disordered N-terminal residues ('n') to bind to both Ago1 and Clr4 is interesting and could be key in facilitating the multiple interactions via the ZF domain. It is possible that the residues/regions upstream and downstream of an interaction interface contribute to the specificity of a particular interaction. An excellent example of such fine tuning is the ability of Heterochromatin Protein 1 (HP1) to bind with different protein partners via the same domain. A detailed study on two of the interactions mediated by the chromo shadow domain (CSD) of HP1a (HP1 isoform in *D. melanogaster*) indicated the importance of disordered residues adjacent to the binding interface in modulating binding affinities (Mendez et al. 2011). HP1a homodimer utilises the same binding interface in the CSD to bind to

similar consensus motifs within Heterochromatin Protein 2 (HP2) and the nuclear RNAi component PIWI in separate binding events. However, HP1a displays a higher affinity for HP2 compared to PIWI. Using domain-swap experiments, Mendez *et al.* demonstrated that the higher affinity of HP1-HP2 binding is not due to the differences in the consensus motifs within HP2 and PIWI; but rather is determined in part by HP1a residues adjacent to the CSD in the disordered extended C-terminus (CTE). Absence of these adjacent residues does not abolish HP1a-HP2 interaction, but causes a significant drop in the binding affinity. A similar involvement of residues in proximity to the HP1a-binding pentapeptide motif of HP2 was also observed. Strikingly, these residues are absent from the regions flanking the HP1a-binding motif in PIWI (Mendez et al. 2013). This study thus demonstrates the potential role of residues adjacent to a binding interface in forming distinct complexes. Differential expression patterns have been observed for HP2 and PIWI: PIWI is abundant in the female germline and in the initial few hours of embryo formation while HP2 is absent in the germline but present abundantly throughout embryogenesis (Mendez et al. 2011). This additionally highlights the need to consider the cellular environment in which protein interactions occur, since co-localisation and local protein concentrations could all be driving factors in deciding binding preferences and order of events.

Intrinsic disorder, while thought to impart the much needed structural plasticity for protein interactions, cannot be the only driving factor in protein-protein interactions. The role for electrostatic interactions in protein-protein binding events is well documented. Compared to their rather hydrophobic interiors, proteins typically have largely polar and charged surfaces, and these surfaces are often involved in short and long-distance interactions. Based on many studies, it has been proposed that these short

and long range electrostatic interactions between proteins can increase rates of association, impart specificity to particular binding events, and stabilise interactions (Sheinerman et al. 2000). Moreover, electrostatic interactions have also been reported to display promiscuity (Friedler et al. 2005). Thus, intrinsic disorder in concert with electrostatic interactions could potentially influence the 'many to one' binding displayed by proteins including Stc1. Interestingly, more than 50% of residues in Stc1 are disordered, and high proportions of charged residues are found in the 5' N terminal region and the C-terminus. A relatively high degree of disorder and highly charged termini are characteristics of Stc1 that are shared with 'hub' proteins. Recent studies have placed 'hub' proteins in the limelight as they seem to play central roles in intricate protein-interaction networks and have been implicated in diseases (Patil et al. 2010). These hub proteins stand out in protein-protein interaction networks as they take part in a large number of interactions. IDRs have been shown to provide the plasticity required by hub proteins to undergo local and global conformational flexibility. Of particular interest are a subset of hubs classified as 'sociable hub proteins' by Higurashi *et al.* These proteins, found to act as transient hubs in intricate protein-interaction networks, display 'one to many' interactions with dynamic interchanging of protein partners. Their apparent 'sociable' nature is thought to be linked to conformational flexibility provided by IDRs, along with a potential role for charged interfaces (Higurashi et al. 2008). Another study also reports the prevalence of high surface charge in hub proteins (Patil & Nakamura 2006). This study supports the idea that disorder and high surface charge are complementary factors that likely equip hubs with their ability to interact with multiple protein partners. However, Patil and Nakamura report that small hubs (proteins of less than 250 residues) generally display

little or no disorder but possess particularly high surface charge that drives the multiple binding events. It is worth considering that most of these studies utilise structural data deposited in PDB (Protein Data Bank) for their analyses and hence interactions that are difficult to study using structural tools will be missing, which could affect the interpretation. Nevertheless, these studies have helped in appreciating the important role played by IDRs in conformational flexibility of proteins, especially those that are involved in 'hub'-like interactions. Stc1 could potentially be 'socialising' with its interactors utilising its extensive intrinsic disorder along with distinctively charged regions driving promiscuity with some specificity.

5.3 Stc1-complexes: A Model

Based on the data presented in this thesis, I propose models for potential roles for Stc1-containing complexes in the RNAi-mediated heterochromatin pathway.

5.3.1 Stc1 in the ARC-RITS pathway

Ago1 is thought to be a part of two different RNAi-complexes involved in heterochromatin assembly – the Argonaute-siRNA-Chaperone (ARC) complex and the RNA-Induced Transcriptional Silencing (RITS) complex (Verdel et al. 2004; Buker et al. 2007). While RITS associates predominantly with ss-siRNAs (single stranded siRNAs), ARC mostly contains duplex siRNAs, and assembly in ARC was shown to inhibit the slicer activity of Ago1 *in-vitro* (Buker et al. 2007). Recent evidences suggest a requirement for ARC subunits Arb1 and Arb2, and sRNA loading of Ago1 for RITS integrity (Holoch & Moazed 2015b; Jain et al. 2016). This indicates that Ago1 when not loaded with siRNA, fails to assemble into RITS. *In-vivo* data presented in this thesis (Chapter 3, **Fig 3.7B**) indicate that Stc1 is able to associate with

slicer-deficient Ago1, suggesting that Stc1 may associate with duplex siRNA-loaded Ago1. However, the observation that centromeric siRNAs are markedly reduced in slicer-deficient Ago1 (Jain et al. 2016) suggests an alternative possibility that the Stc1-associated Ago1-mutant might not be loaded with duplex siRNAs at all (**Fig 5.1**). This raises the question as to whether Stc1-Ago1 interaction is RNA-dependent or not. Stc1 successfully associates with Ago1 produced *in-vitro* (Chapter 3, **Fig 3.4B**), providing some clue that this interaction is not dependent on siRNAs (since the rabbit reticulo-lysate cell-free system only consists of molecular machineries required for transcription/translation). However, spurious sRNA generation in recombinant protein production systems might be a possibility since Nakanishi *et al.* reported the autonomous incorporation of guide RNA during recombinant expression of a yeast-Argonaute, KpAGO (*Kluyveromyces polysporus)* for crystallographic purposes. Most of the unexpected sRNA that co-purified with the recombinant KpAGO originated from the source plasmid used for KpAGO expression production, in a manner suggestive of siRNA-like duplex loading onto the recombinant KpAGO with subsequent release of the passenger strand (Nakanishi et al. 2012). Such fortuitous studies highlight the need for caution in interpreting analyses of protein-protein interactions involving RNA or DNA binding proteins.

A future goal for the study of Stc1-Ago1 interaction could be to understand the nature of siRNA associated with Stc1-Ago1 complexes. High throughput RNA sequencing of Stc1-associated sRNAs in wild type Ago1 and slicer-deficient Ago1 mutant backgrounds could potentially reveal whether Stc1 binds to the siRNA-deficient Ago1 fraction or the duplex loaded Ago1 fraction. To further investigate the role of Stc1 in the RNAi pathway, it would be interesting to test whether Stc1 interacts with the ARC

subunits Arb1 or Arb2. Though Stc1 is not required for an intact RITS, a role for Stc1 in the shuttling of Ago1 between ARC and RITS is plausible if Stc1 has the ability to interact with Ago1 independent of Tas3, which should also be tested. Additionally, it would be interesting to study whether Stc1 is required for ARC assembly. Data presented in Chapter-3 on Stc1-Ago1 interaction, together with the proposed future studies could help in understanding whether Stc1 could be part of the proposed shuttling of Ago1 between ARC and RITS.

Of course interacting proteins must be present in the same cellular compartment prior to interaction. In this regard, Stc1-puncta observed in the cytoplasm in addition to the nuclear pool, are interesting as they are reminiscent of the cytoplasmic ribonucleoprotein particles referred to as P-bodies (PBs). RNAi components including Ago1 have been reported to be localised to such P-bodies in *S. pombe* (Carmichael et al. 2006). The PB-like puncta observed for Stc1 might hint towards potential additional roles for Stc1 in the RNAi pathway as suggested by Bayne *et al.* (Bayne et al. 2010). Interestingly, ARC subunits, especially Arb2, also form cytoplasmic foci and thus steps in siRNA biogenesis involving RITS and ARC complexes in *S. pombe* have been proposed to occur in both cytoplasm and nucleoplasm (Buker et al. 2007). Based on the localisation patterns of Ago1, Arb2 and Stc1, cytoplasmic Stc1 could aid in the initial steps of siRNA biogenesis. Localisation of RNAi factors is somewhat controversial, since an alternative large scale localisation study of the *S. pombe* proteome reported the presence of ARC subunits in both cytoplasm and nucleus, but Ago1 and Stc1 as solely cytoplasmic and nuclear respectively (Matsuyama et al. 2006). Surprisingly, that study also reported Dos2 localisation as mitochondrial, suggesting that caution must be taken while utilising

information from such proteome-wide localisation studies. Erroneous localisation of proteins could occur as a result of the methodology used, for example, over-expression of tagged proteins or immunofluorescence studies on fixed cells, or due to an effect of peptide sequences used for protein tagging.

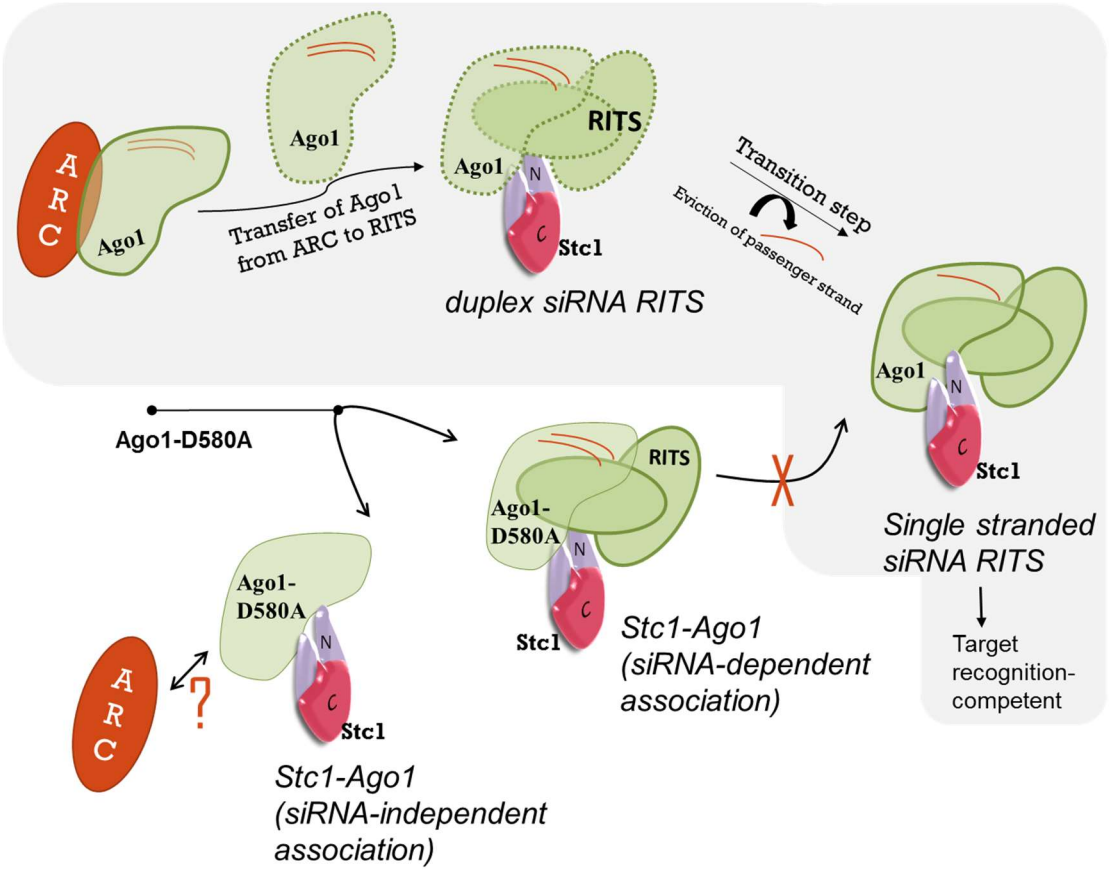


Fig 5.1 Stc1 in the ARC-RITS pathway

Diagram in the shaded portion depicts the transfer of siRNA-loaded Ago1 from ARC to RITS and eventually RITS association with centromere. This involves the catalytic activity of Ago1, essential for evicting the passenger strand from the duplex siRNA leading to the maturation of RITS. Ago1-D580A catalytic-dead mutant, however, fails to release the passenger strand and hence RITS trapped with duplex siRNAs. This mutant additionally affects siRNA generation resulting in substantially low pool of siRNA-associated Ago1. Association of Stc1 with WT and mutant Ago1 indicates a potential for Stc1 to associate with Ago1 prior to RITS maturation. siRNA-independent association of Stc1 with Ago1 also raises the possibility for Stc1 to function in the ARC pathway.

5.3.2 Order of events

5.3.2.1 Open and Closed Stc1-complexes

I have presented data indicating that the Stc1-N terminus has multiple protein partners, while the C-terminus is dedicated to Dos2-binding (Chapter 3). Based on this data, possibilities exist for different Stc1 complexes representing 'open' and 'closed' conformations. In the case of a Stc1-RITS complex, if RITS subunits Ago1 and Tas3 interacts with the N-terminus of Stc1 leaving the C-terminus of Stc1 unbound, the resulting complex would have an open conformation (**Fig 5.2A**). However, for Stc1-CLRC complexes, open and closed structures are possible (**Fig 5.2B** and **5.2C**). An open Stc1-CLRC complex would result if Stc1 is bound by one of either Clr4 at the N-terminus or Dos2 at the C-terminus (**Fig 5.2B**). Alternatively, a closed Stc1-CLRC conformation would arise when Stc1 simultaneously binds Clr4 at the Stc1 N-terminus and Dos2 at the C-terminus. Finally, the potential multi-complex structure where Stc1 interacts with RITS via the N-terminus and CLRC via the C-terminus (**Fig 5.2D**) would be key for the RNAi-dependent recruitment of chromatin modification machinery.

5.3.2.2 Models for Stc1-mediated CLRC recruitment

The functional significance of the multiple interactions of Stc1 needs to be addressed. If indeed, two interacting proteins can bind to the same Stc1-interface, the outcome could be determined by factors like co-localisation and the concentration of the interacting proteins at that space and time. Additionally, this could also be true for Stc1 interactions via different domains, and the order and nature of binding events could act as the decision making step for commitment to downstream processes. If Stc1 is

an integral part of RNAi, one can envision Stc1-Ago1 binding as the first step in the pathway. Although the previous observation that Stc1-Ago1 interaction cannot be observed in the absence of Clr4 seems to contradict this (Bayne et al. 2010), it is possible that Stc1-Ago1 interaction might still be occurring but at reduced levels (below the limit of detection by western blot), given the fact that reduced levels of siRNAs can be detected in the absence of Clr4. Therefore, Stc1-Ago1 interaction independent of Clr4 is still a possibility.

A few lines of evidence suggest the possibility that Stc1 could be localised to the nucleus in association with Ago1 as part of ARC, and subsequently transferred to RITS during the ARC-RITS transition step in the nucleus: (1) Arb1-Ago1 interaction potentially occurs in the cytoplasm (Holoch & Moazed 2015b) (2) Stc1 also shows cytoplasmic puncta (Chapter 3, **Fig 3.8**), similar to Arb1 and Arb2 localisation (Buker et al. 2007) and hence the cytoplasmic Stc1 could possibly associate with the Arb1/2-Ago1 sub-complex (3) Stc1-Ago1 interaction potentially occurs prior to nuclear/centromeric localisation ((Bayne et al. 2010) and Chapter 3, **Fig 3.7B**). Although a preliminary bioinformatic analysis indicated the possible presence of a bipartite nuclear localisation signal at the very N-terminal region of Stc1, this was predicted with only low confidence (data not shown). If indeed Stc1 lacks a putative nuclear localisation signal, it is tempting to speculate that Stc1 localisation to the nucleus may be Ago1-dependent. Association of Ago1 within ARC does not require sRNA-loading; however, RITS assembly is dependent on ARC and sRNA loaded Ago1 (Holoch & Moazed 2015a). Given that nuclear localisation of Dcr1 indicates that siRNA biogenesis occurs in the nucleus (Emmerth et al. 2010), RITS assembly

most likely occurs in the nucleus. Therefore, it is likely that Stc1 association with RITS occurs in the nucleus.

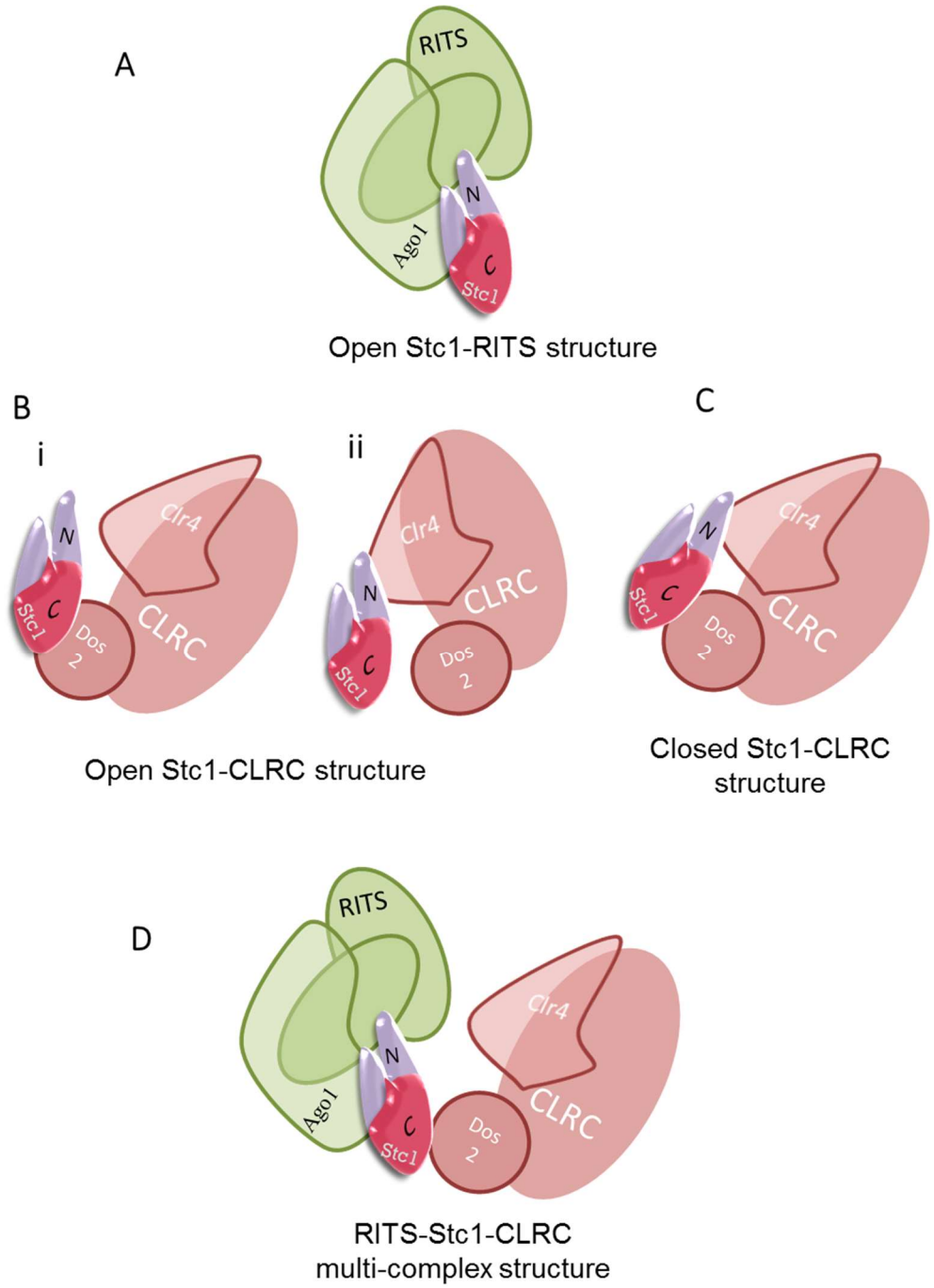


Fig 5.2 Potential Stc1-complexes

(A)– (D): Various possibilities for Stc1-containing complexes based on *in-vitro* data from Chapter 3. The open and closed conformation complexes are based on whether Stc1 is bound by its interacting partners at one or both ends.

Interestingly, Dos2 has been shown to localise to both cytoplasm and nucleus (Li et al. 2005). Given that Dos1 localisation to the nucleus is dependent on Dos2 (Li et al. 2005), it can be assumed that Dos2-Dos1 interaction also occurs in the cytoplasm. Although there is a possibility for the Dos2-Dos1-Rik1 core CLRC sub-complex to be assembled in the cytoplasm, existing data indicates the ability of Rik1 to localise to the nucleus, potentially independently of Dos2, Dos1 and Clr4 (Matsuyama et al. 2006; Zhang et al. 2008). Considering the requirement for the core CLRC subunits for Stc1-CLRC association (Bayne et al. 2010), it is possible that Ago1-bound Stc1 binds to Dos2 (as part of the core CLRC sub-complex) and this bridging could potentially occur in the nucleus. Additionally, CLRC complex assembly is not dependent on Stc1 (Bayne et al. 2010; He et al. 2013) and there is a possibility that CLRC complex already exists in the vicinity due to Clr4 binding to existing H3K9methyl marks. Based on this scenario, I propose a model (**Model I**) for Stc1-interaction where Stc1 is localised to nucleus by Ago1 via Stc1-Ago1 interaction as a potential first step in the pathway (**Fig 5.3**). Once localised in the nucleus, Stc1-Ago1 as part of RITS (open Stc1-RITS complex) recruits CLRC via Stc1-Dos2 interaction (formation of RITS-Stc1-CLRC multi-complex) and gets recruited to nascent centromeric transcript via complementary base pairing of siRNA loaded onto Ago1, thereby ensuring the maintenance of heterochromatin at the centromere. Once localised to nascent transcript, it is possible that Stc1 dissociates from CLRC whilst maintaining its interaction with RITS and plays a role in secondary siRNA biogenesis, thereby ensuring efficient feedback to maintain heterochromatin via RNAi at the

centromere. This would be consistent with the previous observation that RITS-RDRC interaction is dependent on Stc1 (Bayne et al. 2010).

An alternative possible model exists (**Model II a**) whereby Stc1-Dos2 association acts as the first step leading to localisation of Stc1-Dos2 to the nucleus (along with Dos1 and Rik1; this could happen regardless of the potential ability of Rik1 to be localised to centromere independent of Dos2 mentioned in the previous section) (**Fig 5.4**). In support of this, Stc1-CLRC interaction has been found to be RNAi-independent (Bayne et al. 2010). Dos2-bound Stc1 (open Stc1-CLRC) could eventually bind to nuclear RITS (containing siRNA-loaded Ago1), leading to localisation of the RITS-Stc1-CLRC multi-complex to nascent centromeric transcripts via complementary base pairing. As mentioned in Model I, Stc1 could then remain as part of RITS and function in the secondary siRNA biogenesis pathway. The observation that slicer deficient Ago1 can still be assembled into RITS and that the slicer activity of Ago1 is not required for secondary siRNA generation (Jain et al. 2016) are noteworthy. Along with the finding that Stc1 is able to interact with slicer-deficient Ago1, these results are supportive of the proposed integral role for Stc1 in RNAi beyond CLRC recruitment. Though Model I and II depict scenarios where the first Stc1-binding step takes place in the cytoplasm, it is equally plausible that the initial Stc1-interaction could occur in the nucleus following nuclear import of Stc1 through binding of an unknown protein partner or due to the presence of a yet to be identified non-consensus nuclear localisation signal residing in Stc1. Currently, it is unclear whether an open Stc1-CLRC conformation formed via Stc1-Clr4 interaction as the first step is possible since there is no evidence for or against this scenario.

5.3.2.3 Models for possible involvement of the closed Stc1-CLRC complex

Amongst the various possibilities of open and closed Stc1-complexes, although the closed Stc1-CLRC complex conformation is possible, a lack of clear evidence for any effect of Stc1-Clr4 interaction on Clr4 enzymatic activity or Clr4 binding to H3K9-methyl mark renders it difficult to functionally place this complex in the pathway. Nevertheless, the closed Stc1-CLRC complex could act at two possible stages: (1) early in the pathway (**Model II b**) :- Once the Stc1-Dos2 (along with Dos1 and Rik1) sub-complex is imported to the nucleus, CLRC could be fully assembled leading to Clr4-Stc1 interaction and thus the formation of closed Stc1-CLRC complex (**Fig 5.5**); or (2) late in the pathway (**Model I/IIa**):- once the RITS-Stc1-CLRC multi-complex is recruited to nascent transcript, instead of remaining associated with RITS, Stc1 could dissociate from RITS and interact with Clr4 either associated with the same CLRC or with another CLRC in the vicinity (Model I/IIa; **Fig 5.3 and 5.4**). Based on my observations that Clr4 binding to Stc1 and the H3K9-methyl mark are separable events (Chapter 4, **Fig 4.6C and 4.6D**), and that Clr4-CD is required in addition to part of the PCD for Stc1 binding (Chapter 4, **Fig 4.7**), it can be inferred that Clr4 present in the closed Stc1-CLRC complex is not able to bind to H3K9-methyl mark. In the case of Model II b where closed Stc1-CLRC acts as the initial step in the pathway, inability to bind to the existing low levels of H3K9-methylation could prevent Stc1-associated CLRC complex localising to centromeres via H3K9-methylation recognition by Clr4, and hence impose recruitment via RNAi and Stc1. When the closed Stc1-CLRC complex encounters siRNA-loaded nuclear RITS, competition between Clr4 and Ago1 for Stc1 N-terminal binding could lead to the transition to the RITS-Stc1-CLRC multi-complex structure. Thus Clr4 in the multi-

complex would no longer be bound to Stc1 and could bind to existing H3K9-methyl marks and/or become enzymatically active.

Regardless of the stage at which the closed Stc1-CLRC complex forms, this structure presents an interesting way in which CLRC could be sequestered and/or recycled via RNAi components for heterochromatin maintenance. This could help to explain why RNAi is required for maintenance of heterochromatin specifically at centromeres but not at telomeres or mating type locus (Hall 2002; Kanoh et al. 2005; Bayne et al. 2010). However, further studies on Stc1-interaction kinetics and structural analyses will provide a better picture of the order of events. It needs to be empirically determined whether Stc1 undergoes binding-induced conformational changes that could determine the specificity of subsequent binding events. Additionally, *in-vivo* studies using separation of function mutants could help in understanding whether any of the Stc1-interactions or the proposed Stc1-conformational structures play any specific role in the establishment phase of heterochromatin assembly.

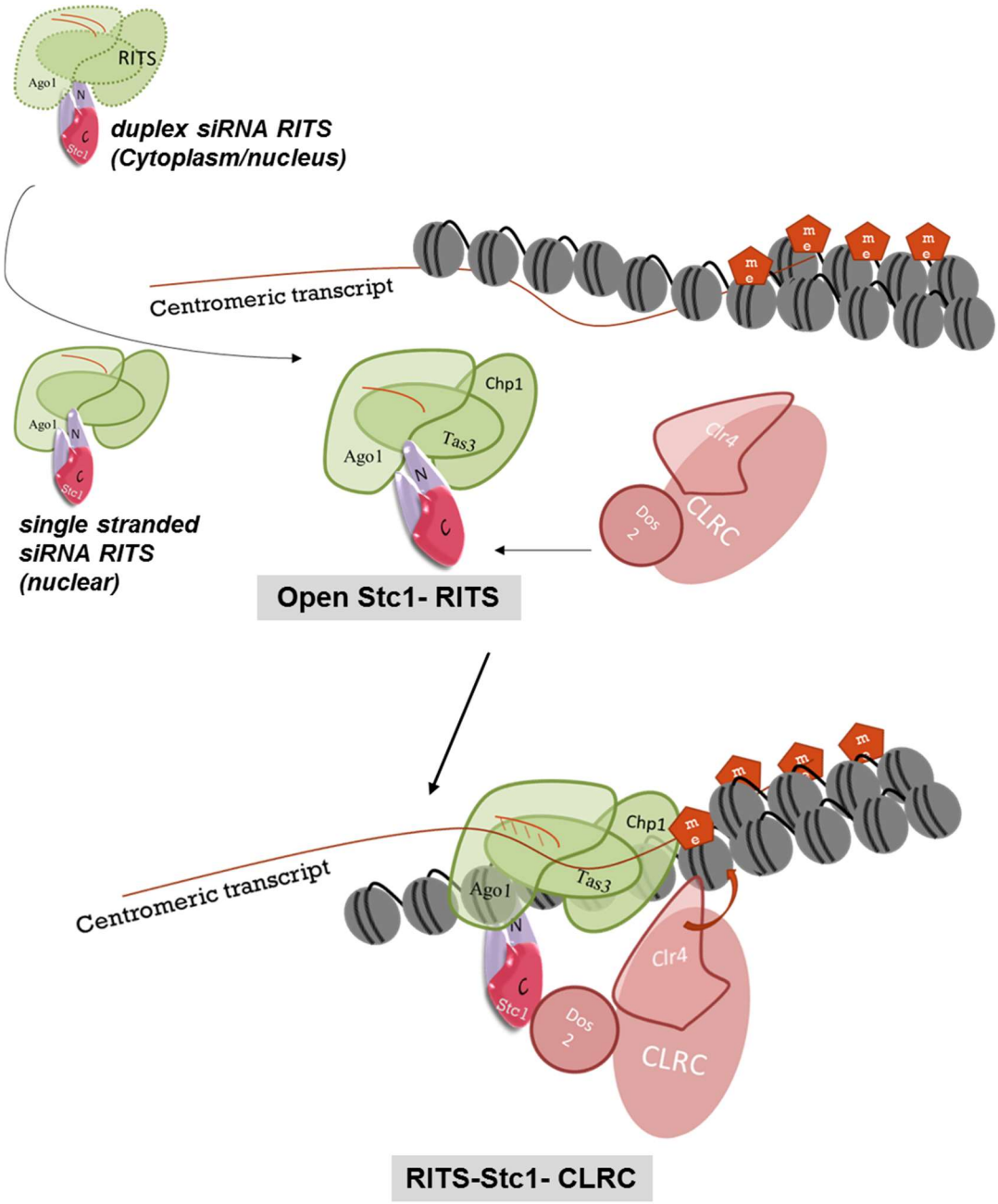


Fig 5.3 Stc1-mediated CLRC recruitment: Model I

In this model, Stc1 first associates with siRNA-loaded RITS and gets recruited to the nucleus. The free C-terminus of Stc1 within the 'open Stc1-RITS' structure then binds to Dos2, eventually forming the RITS-Stc1-CLRC mutli-complex structure. This complex localises to centromeres via siRNA-transcript complementary base pairing. Localised RITS is further stabilised via Chp1 binding to H3K9-methyl marks already deposited through RNAi-independent mechanism.

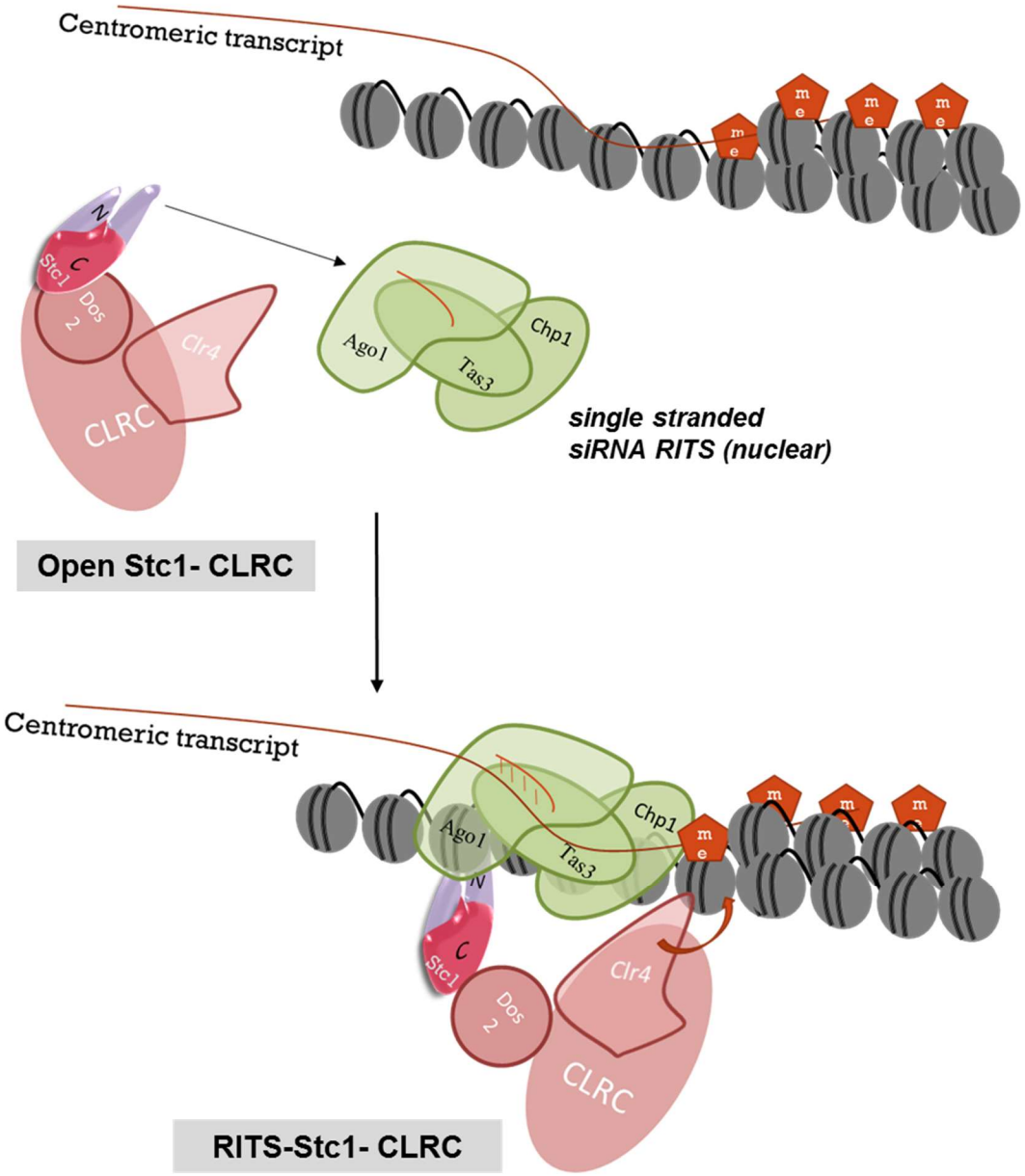


Fig 5.4 Stc1-mediated CLRC recruitment: Model IIa

In this model, Stc1 first interacts with Dos2 to form an 'open Stc1-CLRC' complex. Nuclear pool of siRNA-loaded RITS eventually binds to Stc1 N-terminus, eventually forming the RITS-Stc1-CLRC mutli-complex structure. This complex localises to centromeres via siRNA-transcript complementary base pairing. Localised RITS is further stabilised via Chp1 binding to H3K9-methyl marks already deposited through RNAi-independent mechanism.

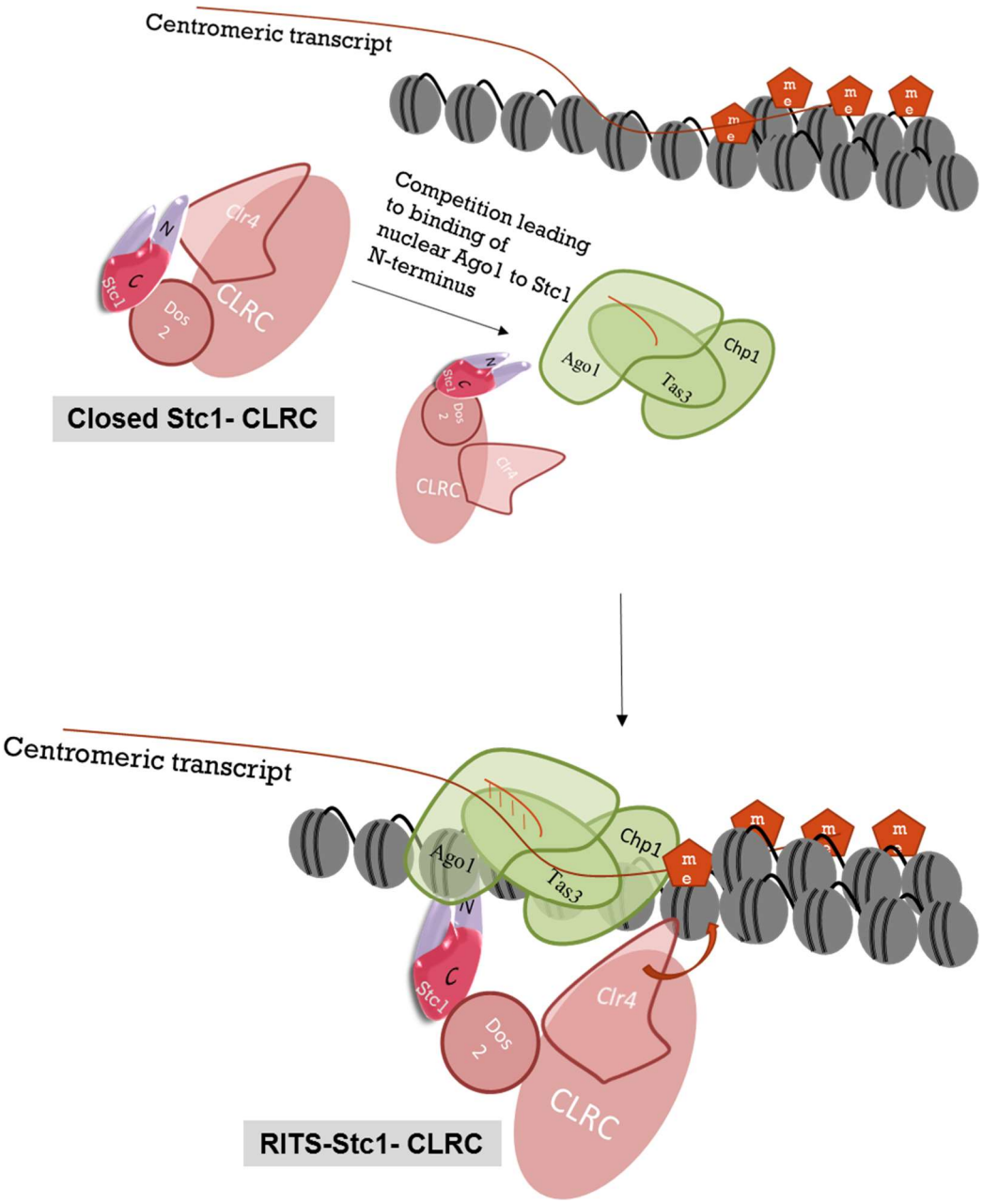


Fig 5.5 Stc1-mediated CLRC recruitment: Model IIb

In this model, the first step in the pathway involves simultaneous binding of Clr4 and Dos2 to Stc1 N and C termini respectively, resulting in the formation of the 'closed Stc1-CLRC' complex. Ago1 as part of siRNA-loaded RITS then competes with Clr4 for Stc1 N-terminal binding leading, to the formation of the RITS-Stc1-CLRC mutli-complex structure. Through siRNA-transcript complementary base pairing, this multi-complex is targeted to centromeres, where Chp1 binding to H3K9-methyl marks already deposited through RNAi-independent mechanism, stabilise RITS onto the chromatin.

5.4 Stc1-like proteins: A glimpse

The findings from my study could help in understanding how RNAi can function in chromatin modification in other organisms. Although Stc1 lacks a homologue in higher eukaryotes based on sequence, it is possible that structural homologs are present. Considering the high level of disordered residues in Stc1, it is possible that through evolution, these regions could have acquired mutations at a high rate and hence display unusually high sequence divergence. This could explain why it is difficult to find sequence homologues. It is also possible that an ancestral gene could have diverged to two different genes performing the function or vice versa.

Following the discovery of Stc1, clues for the existence of Stc1-like proteins in other organisms came from studies in *Drosophila* which reported Gtsf1 (gametocyte specific factor 1)/Asterix as a Piwi-interacting protein required for proper transposon silencing (Dönertas et al. 2013; Ohtani et al. 2013; Muerdter et al. 2013). In *Drosophila*, germline cells are protected from the deleterious effects of transposable elements by an RNAi-related mechanism that silences these invasive elements. piRNAs (Piwi-interacting RNAs) derived from transposon sequences act as guide RNAs to effect post-transcriptional and transcriptional gene silencing in the *Drosophila* germline. Transcriptional gene silencing is mediated by the nuclear Argonaute called Piwi in complex with the guide piRNA. Piwi-piRNA complex mediated silencing involves recruitment of chromatin modification machinery which deposits H3K9-methylation marks (Czech & Hannon 2016). Interestingly, Gtsf1 domain architecture resembles that of Stc1 in that Gtsf1 harbours a zinc finger domain at its N-terminus followed by an unstructured C-terminus. However, the study by Dönertas *et al.* suggested

involvement of the unstructured C-terminus of Gtsf1 in binding to Piwi which is in contrast to the observed involvement of the zinc finger domain of Stc1 in Ago1 binding. However, it has to be noted that the zinc fingers present in Stc1 and Gtsf1 are quite different. While the Stc1 tandem zinc fingers are related to LIM domains, which are thought to be involved in protein-protein interactions (Bayne et al. 2010; He et al. 2013), Gtsf1/Asterix harbours CHHC zinc fingers that are predicted to be involved in RNA-binding (Dönertas et al. 2013). Hence, although Stc1 and Gtsf1/Asterix appears to be similar, their mode of action seems to be different. Silencio/Panoramix, another piRNA-pathway protein in *Drosophila* appears to be functionally more similar to Stc1. When tethered to RNA or DNA, Silencio/Panoramix is capable of eliciting silencing by recruiting the chromatin modification machinery required for heterochromatic silencing (Sienski et al. 2015; Yu et al. 2015). This is reminiscent of Stc1, which when tethered to DNA, also results in silencing of the tethered locus via recruitment of CLRC (Bayne et al. 2010). Sienski *et al.* additionally demonstrated that in contrast to Silencio/Panoramix , Gtsf1/Asterix failed to induce silencing upon tethering. This indeed suggests a strong functional similarity between Stc1 and Silencio/Panoramix. Interestingly, Silencio/Panoramix harbours a large unstructured N-terminus and predicted secondary structure elements at its C-terminus, and does not contain any domains of known function. Future studies on how Silencio/Panoramix acts as a linker between Piwi and the chromatin modification machinery in the *Drosophila* germline are required to understand whether the mode of action is similar to Stc1. piRNA mediated transcriptional silencing occurs in a wide range of organisms including mammals (Czech & Hannon 2016), and likely involves a similar mode of action;

therefore, discovery of adaptor proteins like Stc1 that connects Argonaute and chromatin modification machinery in metazoans awaits.

5.5 Stc1: future prospects

Given the importance of adaptor proteins like Stc1, further studies are warranted to understand the precise role of Stc1 in the RNAi-mediated heterochromatin pathway. Parallel *in-vivo* and *in-vitro* studies are essential for elucidating the basis of Stc1-interactions and to determine the order of events. Numerous attempts to express Stc1-interactors in bacteria failed and hence the goal to obtain crystallographic structures for Stc1 bound to its interaction partner(s) could not be achieved. It is plausible that large scale protein production could be carried out using insect cells. However, structural analysis of Stc1-Clr4 interaction will require other host systems for protein production since the study by Kuscu *et al.* failed to detect a robust Stc1-Clr4 interaction in their pairwise interaction analyses using proteins produced using insect cells (Kuscu et al. 2014). Additionally, it is not known if Ago1 expressed in baculovirus can interact with Stc1. Nevertheless, this expression method could prove useful for elucidating Stc1-Dos2 structure based on the Kuscu *et al.* study. An interesting alternative host system for recombinant protein production would be a mammalian expression system (for example, ExpiCHO expression system – ThermoFisher) as this allows better folding and post translational modifications of the recombinant proteins. In addition, cross linking coupled mass spectrometry (XL-MS) has emerged as a powerful alternative to study the structural basis of interactions (Holding 2015). While I attempted the strategy of XL-MS on Stc1 co-immunoprecipitated material, preliminary results indicated that the methodology

requires optimisation as I failed to obtain Stc1 crosslinked peptides. XL-MS should be a future priority and additionally this method can be applied to recombinant protein mixtures (for example, Stc1 and Dos2; Stc1 and Ago1; Stc1, Ago1 and Dos2) as an alternative to X-ray crystallography.

Attempts to study binding kinetics using the *in-vitro* cell free reticulolysate system proved difficult due to the inconsistency in expression levels displayed by this system during competition assays. An alternative approach would be to study Stc1-binding kinetics *in-vivo* using FRET (Fluorescence resonance energy transfer) and FRAP (Fluorescence recovery after photo-bleaching). Labelling Stc1 and the known Stc1-interactors (based on the *in-vitro* data presented in this work) using fluorescent tags would enable the interaction dynamics to be quantified *in-vivo*. These techniques could reveal, for example, whether Stc1 stably associates with Ago1 or exhibits dynamic interactions.

Another area of future research could be *in-vivo* characterisation of Stc1-interactions in the RNAi pathway. Data presented in this thesis indicates that assembly of RITS occurs independently of Stc1 (Chapter 3, **Fig 3.7A**) and that Stc1 has the ability to bind to an inactive RITS (Chapter 3, **Fig 3.7B**). However, it is not clear whether Stc1 interaction with Ago1 requires Ago1 to be assembled into RITS. To address this question, available separation-of-function Tas3 mutants can be utilised in Stc1-CoIP experiments. The $Tas3_{WG}$ mutant has been shown to disrupt Ago1-Tas3 interaction while retaining Tas3-Chp1 interaction. It has been shown that the Tas3-Chp1 sub-complex can maintain silencing at centromere; however, it cannot establish *de novo* centromeric heterochromatin efficiently (Partridge et al. 2007). Another Tas3 mutant, $Tas3_{\Delta 10\text{-}24}$ retains Ago1-Tas3 interaction but not Chp1-binding. Unlike Tas3-Chp1

sub-complex, the Ago1-Tas3 sub-complex fails to maintain pre-established heterochromatin due to inability to localise to centromeres (Debeauchamp et al. 2008). Analysis of Stc1-Ago1 interaction in these Tas3-mutant backgrounds will help in determining whether Stc1 can interact with Ago1 independent of the Tas3-Chp1 sub-complex, or whether Stc1 requires a minimal Ago1-Tas3 sub-complex for association, which could indicate a role for potential direct interaction of Stc1 and Tas3. These *in-vivo* CoIP experiments could be complemented with localisation studies to better understand the localisation pattern of Stc1 and associated proteins. Either live cell or fixed cell imaging techniques could be employed for the same purpose. For example, it would be interesting to see if Stc1 co-localises with Ago1 cytoplasmic puncta reported by Carmichael *et al.* (Carmichael et al. 2006), as this could provide clues as to whether Stc1-Ago1 interaction requires co-localisation into P-body-like ribonucleoprotein structures. Interestingly, in human cells, the GW-protein TNRC6A (trinucleotide repeat-containing gene 6A) has been suggested to use its NLS for importing the AGO clade proteins into the nucleus for transcriptional gene silencing (Meister 2013). Whether a similar localisation event involving Tas3 (also a GW-protein) occurs in *pombe* needs to be tested, and a role for Stc1-Tas3 interaction prior to RITS assembly awaits future study. Detailed localisation and co-localisation studies of ARC, RITS and CLRC components along with Stc1 could significantly contribute towards better understanding of the order of events for Stc1-interactions.

Attention is additionally required to follow-up the observed *in-vitro* association of Stc1 and Clr4. To understand the functional significance of this interaction, it is inevitable that *in-vivo* approaches are required. *In-vivo* Clr4 mutants, in which key polar charged residues within the minimal- Stc1-binding region of Clr4 are substituted to alanine,

can be utilised to study effects on the major heterochromatic loci using chromatin immunoprecipitation (ChIP). Preliminary ChIP experiments using Clr4 mutants bearing mutations within the PCD region involved in Stc1-binding showed unexpected results. However, due to the incomplete nature of experiments, I did not include this data as part of this thesis. Hence further analysis of these Clr4 point-mutants is required and could open up new areas of future research into Clr4-interactions.

Finally, the role of the very N-terminal residues of Stc1 in binding events requires future attention. Extensive mutational analyses based on the predicted disordered residues/regions could be a way to study the involvement of these residues in various Stc1-interactions. Of particular interest is the presence of oppositely charged residues in equal proportions at either end of the Stc1 protein (Stc1-n and Stc1-C bear 37% positively and negatively charged residues, respectively). Whether this results in a closed conformation for Stc1 via intra-protein interactions when it is not bound to interaction partner(s) awaits future study. Thus a balanced combination of *in-vivo* and *in-vitro* studies could help to shed light on the exact mechanisms underlying Stc1-interactions and to deduce additional roles or new interaction partners for highly disordered proteins like Stc1.

References

Allshire, R.C. et al., 1994. Position effect variegation at fission yeast centromeres. *Cell*, 76(1), pp.157–169.

Allshire, R.C. & Ekwall, K., 2015. Epigenetic Regulation of Chromatin States in *Schizosaccharomyces pombe*. *Cold Spring Harbor Perspectives in Biology*, 7(7), p.a018770.

Allshire, R.C. & Karpen, G.H., 2008. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nature Reviews Genetics*, 9(12), pp.923–937.

Ayoub, N. et al., 2003. A novel jmjC domain protein modulates heterochromatization in fission yeast. *Molecular and cellular biology*, 23(12), pp.4356–70.

Bähler, J. et al., 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast*, 14(10), pp.943–951.

Barnhart, M.C. et al., 2011. HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *The Journal of Cell Biology*, 194(2), pp.229–243.

Baum, M., Ngan, V.K. & Clarke, L., 1994. The centromeric K-type repeat and the central core are together sufficient to establish a functional Schizosaccharomyces pombe centromere. *Molecular Biology of the Cell*, 5(7), pp.747–761.

Bayne, E., Portoso, M. & Kagansky, A., 2008. Splicing factors facilitate RNAi-directed silencing in fission yeast. *Science*, (October), pp.602–606.

Bayne, E.H. et al., 2014. A systematic genetic screen identifies additional factors influencing centromeric heterochromatic integrity in fission yeast. *Submitted*, pp.1–16.

Bayne, E.H. et al., 2010. Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. *Cell*, 140(5), pp.666–77.

Belmont, a S. et al., 1999. Large-scale chromatin structure and function. *Current opinion in cell biology*, 11(3), pp.307–11.

Bernard, P. et al., 2010. Splicing factor Spf30 assists exosome-mediated gene silencing in fission yeast. *Molecular and cellular biology*, 30(5), pp.1145–57.

Blower, M.D. et al., 2002. Conserved Organization of Centromeric Chromatin in Flies and Humans. *Developmental Cell*, 2(3), pp.319–330.

Bühler, M. et al., 2007. RNAi-Dependent and -Independent RNA Turnover Mechanisms Contribute to Heterochromatic Gene Silencing. *Cell*, 129(4), pp.707–721.

Bühler, M., Verdel, A. & Moazed, D., 2006. Tethering RITS to a Nascent Transcript

Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. *Cell*, 125(5), pp.873–886.

Buker, S. et al., 2007. Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. *Nature structural & ...*, 14(3).

Burgess, R.J. & Zhang, Z., 2013. Histone chaperones in nucleosome assembly and human disease. *Nature Structural & Molecular Biology*, 20(1), pp.14–22.

Buscaino, A. et al., 2012. Raf1 is a DCAF for the Rik1 DDB1-like protein and has separable roles in siRNA generation and chromatin modification. *PLoS Genetics*, 8(2), pp.1–14.

Cam, H.P. et al., 2005. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature genetics*, 37(8), pp.809–19.

Canzio, D. et al., 2011. Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Molecular Cell*, 41(1), pp.67–81.

Capelson, M. & Corces, V.G., 2005. The Ubiquitin Ligase dTopors Directs the Nuclear Organization of a Chromatin Insulator. *Molecular Cell*, 20(1), pp.105–116.

Carmichael, J.B. et al., 2006. RNA interference effector proteins localize to mobile cytoplasmic puncta in schizosaccharomyces pombe. *Traffic*, 7(8), pp.1032–1044.

Chen, C.-C. & Mellone, B.G., 2016. Chromatin assembly: Journey to the CENter of the chromosome. *The Journal of Cell Biology*, 214(1), pp.13–24.

Chen, E.S. et al., 2008. Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature*, 451(7179), pp.734–7.

Cheutin, T. et al., 2004. In Vivo Dynamics of Swi6 in Yeast: Evidence for a Stochastic Model of Heterochromatin. *Molecular and Cellular Biology*, 24(8), pp.3157–3167.

Cheutin, T. et al., 2003. Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science (New York, N.Y.)*, 299(5607), pp.721–5.

Clapier, C.R. & Cairns, B.R., 2009. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, 78(1), pp.273–304.

Colmenares, S.U. et al., 2007. Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Molecular cell*, 27(3), pp.449–61.

Cowieson, N.P. et al., 2000. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Current Biology*, 10(9), pp.517–525.

Czech, B. & Hannon, G.J., 2016. One Loop to Rule Them All: The Ping-Pong Cycle

and piRNA-Guided Silencing. *Trends in Biochemical Sciences*, 41(4), pp.324–337.

Debeauchamp, J.L. et al., 2008. Chp1-Tas3 interaction is required to recruit RITS to fission yeast centromeres and for maintenance of centromeric heterochromatin. *Molecular and cellular biology*, 28, pp.2154–2166.

Djupedal, I. et al., 2009. Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *The EMBO journal*, 28(24), pp.3832–44.

Djupedal, I. et al., 2005. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes & development*, 19(19), pp.2301–6.

Dönertas, D., Sienski, G. & Brennecke, J., 2013. Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes and Development*, 27(15), pp.1693–1705.

Dyson, H.J. & Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology*, 6(3), pp.197–208.

Earnshaw, W.C. & Migeon, B.R., 1985. Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma*, 92(4), pp.290–296.

Ekwall, K. et al., 1997. Transient Inhibition of Histone Deacetylation Alters the Structural and Functional Imprint at Fission Yeast Centromeres. *Cell*, 91(7), pp.1021–1032.

Elgin, S.C.R. & Reuter, G., 2013. Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology*, 5(8), pp.a017780–a017780.

Emmerth, S. et al., 2010. Nuclear Retention of Fission Yeast Dicer Is a Prerequisite for RNAi-Mediated Heterochromatin Assembly. *Developmental Cell*, 18(1), pp.102–113.

Fachinetti, D. et al., 2015. DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Developmental Cell*, 33(3), pp.314–327.

Festenstein, R. et al., 2003. Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells. *Science (New York, N.Y.)*, 299(5607), pp.719–21.

Filion, G.J., 2015. 3D genome structure. Organization of the nucleus in space and time. *FEBS Letters*, 589(20), pp.2867–2868.

Fischer, E.S. et al., 2011. The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell*, 147(5), pp.1024–39.

Fischer, T. et al., 2009. Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22), pp.8998–9003.

Fisher, C.R., 1969. Enzymology of the pigmented adenine-requiring mutants of Saccharomyces and Schizosaccharomyces. *Biochemical and Biophysical Research Communications*, 34(3), pp.306–310.

Forsburg, S.L. & Rhind, N., 2006. Basic methods for fission yeast. *Yeast*, 23(3), pp.173–183.

Friedler, A. et al., 2005. Binding of Rad51 and Other Peptide Sequences to a Promiscuous, Highly Electrostatic Binding Site in p53. *Journal of Biological Chemistry*, 280(9), pp.8051–8059.

Fussner, E., Ching, R.W. & Bazett-Jones, D.P., 2011. Living without 30nm chromatin fibers. *Trends in Biochemical Sciences*, 36(1), pp.1–6.

Gerace, E.L., Halic, M. & Moazed, D., 2010. The methyltransferase activity of Clr4Suv39h triggers RNAi independently of histone H3K9 methylation. *Molecular cell*, 39(3), pp.360–72.

Gerasimova, T.I., Byrd, K. & Corces, V.G., 2000. A Chromatin Insulator Determines the Nuclear Localization of DNA. *Molecular Cell*, 6(5), pp.1025–1035.

Goto, D.B. & Nakayama, J.I., 2012. RNA and epigenetic silencing: Insight from fission yeast. *Development Growth and Differentiation*, 54(1), pp.129–141.

Grewal, S.I.. & Elgin, S.C.., 2002. Heterochromatin: new possibilities for the inheritance of structure. *Current Opinion in Genetics & Development*, 12(2), pp.178–187.

Grewal, S.I.S. & Elgin, S.C.R., 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature*, 447(7143), pp.399–406.

Grewal, S.I.S. & Jia, S., 2007. Heterochromatin revisited. *Nature reviews. Genetics*, 8(1), pp.35–46.

Gullerova, M. & Proudfoot, N.J., 2008. Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S. pombe. *Cell*, 132(6), pp.983–995.

Halic, M. & Moazed, D., 2010. Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell*, 140(4), pp.504–16.

Hall, I.M., 2002. Establishment and Maintenance of a Heterochromatin Domain. *Science*, 297(5590), pp.2232–2237.

Halverson, D. et al., 1997. A centromere DNA-binding protein from fission yeast affects chromosome segregation and has homology to human CENP-B. *The Journal of cell biology*, 136(3), pp.487–500.

Hansen, K.R., Ibarra, P.T. & Thon, G., 2006. Evolutionary-conserved telomere-linked helicase genes of fission yeast are repressed by silencing factors, RNAi components and the telomere-binding protein Taz1. *Nucleic acids research*, 34(1), pp.78–88.

He, C. et al., 2013. Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification. *Proceedings of the National Academy of*

Sciences of the United States of America, 110(21), pp.E1879-88.

Heitz, E., 1928. Das Heterochromatin der Moose. *Jahrb Wiss Botanik*, 69, pp.762–818.

Henikoff, S., Ahmad, K. & Malik, H.S., 2001. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science*, 293(5532).

Higurashi, M., Ishida, T. & Kinoshita, K., 2008. Identification of transient hub proteins and the possible structural basis for their multiple interactions. *Protein science : a publication of the Protein Society*, 17(1), pp.72–8.

Hiriart, E. et al., 2012. Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. *The EMBO Journal*, 31(10), pp.2296–2308.

Holding, A.N., 2015. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods*, 89, pp.54–63.

Holoch, D. & Moazed, D., 2015a. Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. *Nature Structural & Molecular Biology*, 22(4), pp.328–335.

Holoch, D. & Moazed, D., 2015b. Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. *Nature Structural & Molecular Biology*, (April 2014), pp.1–10.

Hong, E.-J.E., Villén, J. & Moazed, D., 2005. A Cullin E3 Ubiquitin Ligase Complex Associates with Rik1 and the Clr4 Histone H3-K9 Methyltransferase and is Required for RNAi-Mediated Heterochromatin Formation. *RNA Biology*, 2(3), pp.106–111.

Horita, D.A. et al., 2001. Solution structure, domain features, and structural implications of mutants of the chromo domain from the fission yeast histone methyltransferase Clr4. *Journal of molecular biology*, 307(3), pp.861–870.

Horn, P.J., Bastie, J.-N. & Peterson, C.L., 2005. A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes & development*, 19(14), pp.1705–14.

Hsu, W. et al., 2013. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Science*, 22, pp.258–273.

Huisinga, K.L. & Elgin, S.C.R., 2009. Small RNA-directed heterochromatin formation in the context of development: What flies might learn from fission yeast. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1789(1), pp.3–16.

Irelan, J.T. et al., 2001. Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics*, 157(3), pp.1191–203.

Jain, R., Iglesias, N. & Moazed, D., 2016. Distinct Functions of Argonaute Slicer in siRNA Maturation and Heterochromatin Formation. *Molecular Cell*, 63(2),

pp.1–15.

Jenuwein, T. & Allis, C.D., 2001. Translating the Histone Code. *Science*, 293(5532), pp.814–816.

Jia, S., Kobayashi, R. & Grewal, S.I.S., 2005. Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin. *Nature Cell Biology*, 7(10), pp.1007–1013.

Jia, S., Noma, K. & Grewal, S.I.S., 2004. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science (New York, N.Y.)*, 304(5679), pp.1971–6.

Kanoh, J. et al., 2005. Telomere Binding Protein Taz1 Establishes Swi6 Heterochromatin Independently of RNAi at Telomeres. *Current Biology*, 15(20), pp.1808–1819.

Karpen, G.H. & Allshire, R.C., 1997. The case for epigenetic effects on centromere identity and function. *Trends in Genetics*, 13(12), pp.489–496.

Kato, H. et al., 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science (New York, N.Y.)*, 309(5733), pp.467–9.

Ketting, R.F., 2011. The Many Faces of RNAi. *Developmental Cell*, 20(2), pp.148–161.

Kim, H.S. et al., 2004. Regulation of Swi6/HP1-dependent Heterochromatin Assembly by Cooperation of Components of the Mitogen-activated Protein Kinase Pathway and a Histone Deacetylase Clr6. *Journal of Biological Chemistry*, 279(41), pp.42850–42859.

Kloc, A. et al., 2008. RNA Interference Guides Histone Modification during the S Phase of Chromosomal Replication. *Current Biology*, 18(7), pp.490–495.

Kornberg, R.D., 1977. Structure of Chromatin. *Annual Review of Biochemistry*, 46(1), pp.931–954.

Kurzynska-Kokorniak, A. et al., 2015. The many faces of Dicer: The complexity of the mechanisms regulating Dicer gene expression and enzyme activities. *Nucleic Acids Research*, 43(9), pp.4365–4380.

Kuscu, C. et al., 2014. CRL4-like Clr4 complex in Schizosaccharomyces pombe depends on an exposed surface of Dos1 for heterochromatin silencing. *Proceedings of the National Academy of Sciences of the United States of America*, 111(5), pp.1795–800.

Lam, A.L. et al., 2006. Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11), pp.4186–91.

Lee, J.K., Huberman, J.A. & Hurwitz, J., 1997. Purification and characterization of a CENP-B homologue protein that binds to the centromeric K-type repeat DNA of Schizosaccharomyces pombe. *Proceedings of the National Academy of Sciences of the United States of America*, 94(16), pp.8427–32.

Lejeune, E., Bayne, E.H. & Allshire, R.C., 2010. On the connection between RNAi and heterochromatin at centromeres. *Cold Spring Harbor Symposia on Quantitative Biology*, 75(Hannon 2002), pp.275–283.

Lejeune, E., Bayne, E.H. & Allshire, R.C., 2010. On the connection between RNAi and heterochromatin at centromeres. *Cold Spring Harbor symposia on quantitative biology*, 75(Hannon 2002), pp.275–83.

Li, F. et al., 2005. Two novel proteins, Dos1 and Dos2, interact with Rik1 to regulate heterochromatic RNA interference and histone modification. *Current Biology*, 15(16), pp.1448–1457.

Li, F., Martienssen, R. & Cande, W.Z., 2011. Coordination of DNA replication and histone modification by the Rik1-Dos2 complex. *Nature*, 475(7355), pp.244–8.

Li, G. & Zhu, P., 2015. Structure and organization of chromatin fiber in the nucleus. *FEBS Letters*, 589(20), pp.2893–2904.

Li, W. et al., 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*, 43(W1), pp.W580–W584.

Linding, R. et al., 2003. Protein Disorder Prediction : Implications for Structural Proteomics. , 11, pp.1453–1459.

Malecová, B. & Morris, K. V, 2010. Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. *Current opinion in molecular therapeutics*, 12(2), pp.214–22.

Malik, H.S. & Henikoff, S., 2001. Adaptive Evolution of Cid, a Centromere-Specific Histone in Drosophila. *Genetics*, 157(3).

Martienssen, R. & Moazed, D., 2015. RNAi and Heterochromatin Assembly. *Cold Spring Harbor Perspectives in Biology*, 7(8), p.a019323.

Masumoto, H. et al., 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *The Journal of Cell Biology*, 109(5), pp.1963–1973.

Matsuyama, A. et al., 2006. ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nature Biotechnology*, 24(7), pp.841–847.

McKinley, K.L. & Cheeseman, I.M., 2015. The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology*, 17(1), p.16–29McKinley, K.L. & Cheeseman, I.M., 2015. The m.

McWilliam, H. et al., 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, 41(W1), pp.W597–W600.

Meister, G., 2013. Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics*, 14(7), pp.447–459.

Mendez, D.L. et al., 2011. The HP1a Disordered C Terminus and Chromo Shadow Domain Cooperate to Select Target Peptide Partners. *ChemBioChem*, 12(7),

pp.1084–1096.

Mendez, D.L., Mandt, R.E. & Elgin, S.C.R., 2013. Heterochromatin Protein 1a (HP1a) Partner Specificity Is Determined by Critical Amino Acids in the Chromo Shadow Domain and C-terminal Extension. *Journal of Biological Chemistry*, 288(31), pp.22315–22323.

Mendiburo, M.J. et al., 2011. Drosophila CENH3 is sufficient for centromere formation. *Science (New York, N.Y.)*, 334(6056), pp.686–90.

Mizuguchi, T., Barrowman, J. & Grewal, S.I.S., 2015. Chromosome domain architecture and dynamic organization of the fission yeast genome. *FEBS Letters*, 589(20), pp.2975–2986.

Mohan, A. et al., 2006. Analysis of Molecular Recognition Features (MoRFs). *Journal of Molecular Biology*, 362(5), pp.1043–1059.

Moreno, S., Klar, A. & Nurse, P., 1991. [56] Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. *Methods in Enzymology*, 194, pp.795–823.

Motamedi, M.R. et al., 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell*, 119(6), pp.789–802.

Muerdter, F. et al., 2013. A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in Drosophila. *Molecular Cell*, 50(5), pp.736–748.

Murakami, Y., Huberman, J.A. & Hurwitz, J., 1996. Identification, purification, and molecular cloning of autonomously replicating sequence-binding protein 1 from fission yeast Schizosaccharomyces pombe. *Proceedings of the National Academy of Sciences of the United States of America*, 93(1), pp.502–7.

Muro, Y. et al., 1992. Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *The Journal of cell biology*, 116(3), pp.585–96.

Nakagawa, H. et al., 2002. Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. *Genes & development*, 16(14), pp.1766–78.

Nakanishi, K. et al., 2012. Structure of yeast Argonaute with guide RNA. *Nature*, 486(7403), pp.368–374.

Nakayama, J. et al., 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science (New York, N.Y.)*, 292(5514), pp.110–3.

Noma, K. et al., 2004. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature genetics*, 36(11), pp.1174–1180.

Nurse, P., 1975. Genetic control of cell size at cell division in yeast. *Nature*, 256(5518), pp.547–551.

Ohtani, H. et al., 2013. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. *Genes and Development*, 27(15), pp.1656–1661.

Oldfield, C.J. et al., 2008. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC genomics*, 9 Suppl 1(Suppl 1), p.S1.

Partridge, J.F. et al., 2002. cis-Acting DNA from Fission Yeast Centromeres Mediates Histone H3 Methylation and Recruitment of Silencing Factors and Cohesin to an Ectopic Site. *Current Biology*, 12(19), pp.1652–1660.

Partridge, J.F. et al., 2007. Functional Separation of the Requirements for Establishment and Maintenance of Centromeric Heterochromatin. *Molecular Cell*, 26, pp.593–602.

Partridge, J.F., Borgstrøm, B. & Allshire, R.C., 2000. complex centromere Distinct protein interaction domains and protein spreading in a complex centromere. , pp.783–791.

Patil, A., Kinoshita, K. & Nakamura, H., 2010. Hub promiscuity in protein-protein interaction networks. *International Journal of Molecular Sciences*, 11(4), pp.1930–1943.

Patil, A. & Nakamura, H., 2006. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Letters*, 580(8), pp.2041–2045.

Pidoux, A.L. & Allshire, R.C., 2004. Kinetochore and heterochromatin domains of the fission yeast centromere. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 12(6), pp.521–34.

Promega, Chapter 8 - Post translational modifications. *Promega In-vitro Resource*, pp.39–44. Available at: https://www.promega.com/-/media/files/resources/product-guides/in-vitro-protein-expression/ivex_chp8.pdf?la=en.

Reinhart, B.J. et al., 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science (New York, N.Y.)*, 297(5588), p.1831.

Richmond, T.J. et al., 1997. Crystal structure of the nucleosome core particle at 2.8|[thinsp]||[Aring]| resolution. *Nature*, 389(6648), pp.251–260.

Sadaie, M. et al., 2004. A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. *The EMBO journal*, 23(19), pp.3825–35.

Schalch, T. et al., 2009. High-Affinity Binding of Chp1 Chromodomain to K9 Methylated Histone H3 Is Required to Establish Centromeric Heterochromatin. *Molecular Cell*, 34(1), pp.36–46.

Schirle, N.T. et al., 2012. The crystal structure of human Argonaute2. *Science (New York, N.Y.)*, 336(6084), pp.1037–40.

Scott, K.C., Merrett, S.L. & Willard, H.F., 2006. A Heterochromatin Barrier Partitions the Fission Yeast Centromere into Discrete Chromatin Domains. *Current Biology*, 16(2), pp.119–129.

Scrima, A. et al., 2011. Detecting UV-lesions in the genome: The modular CRL4 ubiquitin ligase does it best! *FEBS letters*, 585(18), pp.2818–25.

Shankaranarayana, G.D. et al., 2003. *Sir2 Regulates Histone H3 Lysine 9 Methylation and Heterochromatin Assembly in Fission Yeast*,

Sheinerman, F.B., Norel, R. & Honig, B., 2000. Electrostatic aspects of protein–protein interactions. *Current Opinion in Structural Biology*, 10(2), pp.153–159.

Sienski, G. et al., 2015. Silencio / CG9754 connects the Piwi – piRNA complex to the cellular heterochromatin machinery. *Genes and Development*, 29, pp.1–14.

Sievers, F. et al., 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), pp.539–539.

Stimpson, K.M. & Sullivan, B.A., 2010. Epigenomics of centromere assembly and function. *Current Opinion in Cell Biology*, 22(6), pp.772–780.

Stunnenberg, R. et al., 2015. H3K9 methylation extends across natural boundaries of heterochromatin in the absence of an HP1 protein. *The EMBO Journal*, 34(22), pp.2789–2803.

Sugiyama, T. et al., 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proceedings of the National Academy of Sciences of the United States of America*, 102(1), pp.152–7.

Sugiyama, T. et al., 2007. SHREC, an effector complex for heterochromatic transcriptional silencing. *Cell*, 128(3), pp.491–504.

Swygert, S.G. & Peterson, C.L., 2014. Chromatin dynamics: interplay between remodeling enzymes and histone modifications. *Biochimica et biophysica acta*, 1839(8), pp.728–36.

Thon, G. et al., 2005. The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast Schizosaccharomyces pombe. *Genetics*, 171(4), pp.1583–1595.

Till, S. et al., 2007. A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nature structural & molecular biology*, 14(10), pp.897–903.

Verdaasdonk, J.S. & Bloom, K., 2011. Centromeres: unique chromatin structures that drive chromosome segregation. *Nature Reviews Molecular Cell Biology*, 12(5), pp.320–332.

Verdel, A. et al., 2009. Common themes in siRNA-mediated epigenetic silencing pathways. *International Journal of Developmental Biology*, 53(2–3), pp.245–257.

Verdel, A. et al., 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science (New York, N.Y.)*, 303(5658), pp.672–6.

Volpe, T. et al., 2003. RNA interference is required for normal centromere function in fission yeast. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 11(2), pp.137–46.

Volpe, T. a et al., 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science (New York, N.Y.)*, 297(5588), pp.1833–7.

Wang, J., Jia, S.T. & Jia, S., 2016. New Insights into the Regulation of Heterochromatin. *Trends in Genetics*, 32(5), pp.284–294.

Ward, J.J. et al., 2004. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *Journal of Molecular Biology*, 337(3), pp.635–645.

White, S. a et al., 2014. The RFTS domain of Raf2 is required for Cul4 interaction and heterochromatin integrity in fission yeast. *PloS one*, 9(8), p.e104161.

van Wolfswinkel, J.C. & Ketting, R.F., 2010. The role of small non-coding RNAs in genome stability and chromatin organization. *Journal of cell science*, 123(Pt 11), pp.1825–1839.

Wood, V. et al., 2002. The genome sequence of Schizosaccharomyces pombe. *Nature*, 415(6874), pp.871–80.

Wright, P.E. & Dyson, H.J., 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2), pp.321–331.

Yamada, T. et al., 2005. The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. *Molecular cell*, 20(2), pp.173–85.

Yamanaka, S. et al., 2013. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature*, 493(7433), pp.557–60.

Yu, Y. et al., 2015. Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science (New York, N.Y.)*, 350(6258), pp.339–42.

Zhang, K. et al., 2008. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nature structural & molecular biology*, 15(4), pp.381–8.

Zhang, P. et al., 2016. An Overview of Chromatin-Regulating Proteins in Cells. *Current protein & peptide science*, 17(5), pp.401–10.

Zofall, M. et al., 2012. RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science (New York, N.Y.)*, 335(6064), pp.96–100.

Zofall, M. & Grewal, S.I.S., 2006. Swi6/HP1 Recruits a JmjC Domain Protein to

Facilitate Transcription of Heterochromatic Repeats. *Molecular Cell*, 22(5), pp.681–692.