BIOCHEMICAL AND STRUCTURAL ANALYSES OF BUDDING YEAST TELOMERE ASSOCIATED CST COMPLEX

by

Jia Sun

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Ming Lei, Chair
Professor David R. Engelke
Professor Janet L. Smith
Associate Professor Zhaohui Xu

This dissertation is dedicated to my parents: Suping Sun and Shuqin Tao. Without their unconditional love and support, I would not be able to achieve what I have accomplished.

Acknowledgements

First of all, I would like to thank my advisor Dr. Ming Lei. During my days in the Lei lab, Ming advised me on not just how to be a good scientist, but also how to be a responsible and wise person. In additional to the scientific knowledge and techniques, I benefited tremendously from Ming's valuable insights he has gained during his own career. I could not ask for a better mentor. Without his directions in science and help in life, I would not have gone this far. I thank Ming from the bottom of my heart.

I would like to thank all the other members of my dissertation committee: Drs. David R. Engelke, Janet L. Smith, and Zhaohui Xu, for their valuable guidance and suggestions every step along my journey.

I would like to thank all the members in Lei lab. I want to thank Dr. Yuting Yang for her tremendous help with crystallography in each of my project. I want to thank Drs. Feng Wang and Yong Chen for offering me valuable suggestions and discussions during my research. I want to thank Laura A. Confer and Ke Wan for their direct help in different stages of my research. I would like to thank Drs. Zhixiong Zeng, Wei Deng, and Bingbing Wan for being supportive throughout my days in lab.

I want to thank all my collaborators: Drs. Neal F.N. Lue and Eun Young Yu at Weil Cornell Medical College. Our collaboration was rewarding and fruitful.

At last but not least, I want to thank all the faculty, staff and students in the Chemical Biology Doctoral Program and Department of Biological Chemistry, University of Michigan. Especially, I want to thank Laura Howe and Justine Altman, for their assistance throughout my stay in Ann Arbor.

Table of Contents

List of Figures

Figure

List of Tables

Table

Abstract

Telomeres are specialized protein-DNA complexes that compose the natural termini of linear chromosomes. Telomeres prevent chromosome ends from deleterious degradation and fusion events and ensure the complete replication of chromosomes.

In *Saccharomyces cerevisiae*, Cdc13, Stn1 and Ten1 are essential for both chromosome capping and telomere length homeostasis. These three proteins have been proposed to fulfill their roles at chromosome termini as a telomere-dedicated RPA (Replication Protein A, including Rpa70, Rpa32 and Rpa14) complex on the basis of several parallels with the conventional RPA. However, no direct evidence has been provided for this hypothesis. Here I provided the first direct evidence based on our crystal structures. Structural and functional analyses of *Candida albicans* Stn1-Ten1 revealed striking similarities with Rpa32-Rpa14 and critical roles for these proteins in suppressing aberrant telomerase activities at telomeres. All proved that Stn1-Ten1 is an Rpa32-Rpa14-like complex at telomere. However, the relationship between Cdc13 and Rpa70 remained unclear. The crystal structures of multiple OB (oligonucleotide/oligosaccharide binding)-folds at the N- and C-terminal ends of Cdc13 established an Rpa70-like domain organization, although the structures of Cdc13 OB-folds are significantly different from their Rpa70 counterparts. Furthermore, our structural and biochemical analyses revealed unexpected Cdc13 dimerization by either N- or C-terminal OB-fold and showed that

homodimerization is probably a conserved feature of all Cdc13s. We also uncovered the versatility of Cdc13 dimerization in mediating interaction with different targets. The structural characterization of the interaction between the Cdc13 N-terminal OB-fold and Pol1, the catalytic subunit of DNA polymerase α, demonstrated a role for N-terminal dimerization in Pol1-binding. The discovery of *Candida spp.* Cdc13 dimerization through its OB4 domain revealed its important role in high affinity telomere DNA binding. Collectively, our findings provided novel insights into the mechanisms and evolution of Cdc13. Additionally, we have shown Cdc13's role in regulating the synthesis of telomere by interacting with telomerase subunit Est1. The interaction involves the second OB-fold in addition to the previously recognized recruitment domain of Cdc13. The finding significantly furthered our understandings about the synthesis of leading and lagging strands of chromosome and the essential role of Cdc13 in solving the end-replication problem.

CHAPTER 1

INTRODUCTION

1.1 Early Development of Telomere Biology

Telomeres, the termini of linear chromosomes, primarily serve the function of providing integrity at each end of chromosomes in dividing and resting cells [1]. The presence of telomeres ensures the complete succession of genetic information from parental to daughter chromosomes. Therefore, the stability and integrity of the telomeres are crucial to living organisms. Since the discovery of chromosomes and cytokinesis, great contributions have been made by geneticists and botanists, including Hermann J. Muller (1890–1967, Nobel laureate 1946) and Barbara McClintock (1902–1992, Nobel laureate 1983), towards understanding the nature of chromosomes. McClintock and Muller, using maize and flies, respectively, functionally defined capping as protection from chromosomal fusion (end-to-end joining) and its deleterious consequence-genomic instability [2]. In the early 1920s, using X-ray as the primary method, Muller observed different kinds of chromosome breaks such as inversions, translocations and deficiencies. He was able to recover some of them using the genetic techniques he developed but failed to recover the chromosome terminal deficiencies [3]. He explained that the recovered chromosomes were usually the result of the "rejoining of two broken ends" and such rejoining could not occur between "originally free ends" or between "originally free

ends" and broken ends [3, 4]. This promoted Muller to realize the possibly more important function of the chromosome ends. Even though he did not have a clear idea about the nature of telomeres, his hypothesis was that each telomere contains an indispensible gene exclusively located at the end of chromosome [3].

Meanwhile, using maize as model organism, Barbara McClintock developed new microscopes that allowed her to visualize individual maize chromosomes. McClintock showed that soon after fertilization, in specific cell types, a broken end can heal in a genetically determined process [5]. Based on her observation, McClintock concluded that the intact chromosome ends have a unique function that is different from broken chromosome ends caused by X-ray irradiation, as the broken ends never fused with the "natural ends" [6]. Furthermore, McClintock hypothesized that there must exist some mechanism that could heal a single broken end "during the reproductive cycle of the chromosome" [7]. Both Muller and McClintock's insights served as foundation for the field of modern telomere biology and provided the first evidence that telomerase is actively involved in healing broken ends during S-phase. The modern studies of telomere DNA structure which started in the 1970s identified the DNA that confers stability to a newly created chromosomal ends, with the help of binding protein. When telomerase was discovered in the 1980s, it demonstrated the enzymatic mechanism by which such DNA can be acquired, and maintained at chromosomal termini [8, 9]. By then, many basic concepts of modern telomere biology in the molecular era had emerged.

1.2 Telomeric DNA

In the 1950s and 1960s, it became clear that all eukaryotic chromosomes are made of linear DNA molecules. The essential telomeric DNA sequences at each end of the eukaryotic linear chromosomes are, in most species, tandem repeats of a specific short sequence. Telomerase, the enzyme responsible for the replication of telomeres, adds multiple copies of this DNA sequence to the terminal region. The identification of telomeric DNA sequences is essential for understanding telomere replication and chromosome end structure. However, this was not an easy task due to the low abundance of chromosomes in somatic cells. Also, most chromosomes are very long, making direct analysis of telomeric DNA technically challenging. The discovery of mini-chromosomes that comprise the amplified ribosomal RNA genes (rDNAs) in some simple eukaryotes, such as the ciliated protozoan *Tetrahymena*, made it possible to sequence the telomeric DNA [10, 11]. They were chosen because of their relative shortness (<100 kb) and high abundance, with occasional observation of sticky ends.

The first such sequence, determined by Engberg, Karrer and Gall in 1976, was that of the amplified rDNA minichromosomes of the somatic nucleus of the ciliated protozoan *Trtrahymena thermophila* [12, 13]. Later, Blackburn and Gall determined the end sequence by combining *in vitro* labeling with restriction endonuclease digestion and fingerprinting analyses. Great heterogeneity was observed in the length of digested fragments, ranging from 120bp to 400bp. They all contained tandem repeats of the hexanucleotide unit CCCCAA/GGGGTT, with the G-rich strand bearing the 3'-OH end of each end of the linear chromosome [14].

As time went by, more and more evidence emerged that similar, extremely simple and tandemly repeated DNA sequences comprised the ends of linear DNAs in other

eukaryotic nuclei. Szostak and Blackburn designed a linear plasmid that allowed them to clone out the budding yeast *Saccharomyces cerevisiae* telomeres and mapped the sequence [15, 16]. It's now widely accepted that telomeric DNA of eukaryotic chromosomes in general consists of simple tandemly repeated sequences characterized by clusters of G residues in one strand and C residues in the other strand. The G-rich strand is oriented 5' to 3' towards the chromosome terminus. Each species has its own characteristic telomeric repeat sequence common to all ends of its chromosomes, and usually different from species to species [17]. The telomeric DNA sequence in different species is summarized in TABLE 1.1. Notably, the DNA end is not blunt. There is a 3' overhang of the G-rich strand extending at the DNA termini [18-20].

1.3 The End-Replication Problem

The discovery of the DNA replication mechanism at the molecular level raised another problem that telomeres must solve. Linear DNA molecules such as those of eukaryotic chromosomes require additional mechanisms besides the conventional DNA polymerase to complete the replication of their ends. It was predicted that terminal attrition of chromosomal DNA would lead to loss of genetic information and eventually prevent cells from replicating (aka, cellular "senescence"), if left unattended [21].

The replication of the double-stranded DNA is semi-conservative and each strand of the double helix is used as the template for the new stand synthesis [22]. This semi-conservative replication presents a unique challenge: the process only works in the 5' to 3' direction [23]. Newly synthesized DNA strand that is synthesized in the 5' to 3' direction is defined as the leading stand while the strand running 3' to 5' is called the lagging

strand (Fig. 1.1). The leading strand can be synthesized continuously while the lagging strand is synthesized in short "Okazaki fragments" [24, 25]. All known DNA polymerases require a polynucleotide primer (either DNA or RNA) bearing a 3'-OH group. This primer is removed, if it's RNA, once synthesis has been initiated. At each round of DNA replication, after the last RNA primer is removed, a gap at the 5' end of the chromosome is left as the terminal DNA cannot be synthesized by conventional DNA polymerase [26-28]. This problem was first raised by James Watson in 1972 and subsequently referred to as the "end replication problem" [29]. The problem has to be solved by telomeres before the cells lose too much genetic information. It was predicted that the average attrition rate of telomere in human cells would be about 63 base pairs in each cell division cycle [30, 31]. If there were no mechanisms to compensate this telomere attrition, chromosome ends would eventually lose protection, and the "bald" chromosome ends would be recognized as DNA damage sites. Consequently, the cells would trigger downstream DNA repair pathways. Thus, different models for telomere replication have been proposed, including the ones proposed by Cavalier-Smith [32], Bateman [33], Dancis and Holmquist [34, 35], but none of them held true for the majority of chromosomal DNAs.

1.4 A Specialized Reverse Transcriptase, Telomerase, that Synthesizes Telomeric DNA

The DNA end-replication problem of telomeres was solved by the discovery of telomerase by Carol Greider and Elizabeth Blackburn [36]. In 1985, Greider and Blackburn first successfully demonstrated the enzymatic activity of the extracts of mating

Tetrahymena. The enzyme they isolated, now known as telomerase, could add the correct *Tetrahymena* telomeric repeats TTGGGG to the 3' end of an oligonucleotide [36]. More importantly, telomerase can specifically recognize the ending sequence of the oligonucleotide that it is elongating without any added template [37]. For example, if the oligonucleotide ends with TTG, telomerase will first add GGG and then start the next round of TTGGGG addition. In other words, the sequence at the 3' end determines what telomerase will first add to the primer [8]. This special feature implied that telomerase might use an internal nucleic acid template to mediate the nucleotide addition. This model was later confirmed by the isolation of telomerase RNA, which was used by telomerase as the replication template [37]. The RNA template of *Tetrahymena* telomerase contains one and a half telomeric repeats (5'-CAACCCCAA-3') and helps telomerase recognize and synthesize TTGGGG repeats [37]. When site-specific mutagenesis of nucleotides were introduced into the RNA template, it led to the deposition of complementary nucleotides in the end of *Tetrahymena* telomeres [38].

This discovery confirmed early speculation that telomerase is an RNA-dependent DNA polymerase [9]. Since then, telomerase was identified in many other organisms. A genetic screen in budding yeast *Saccharomyces serevisiae* performed by Vicki Lundblad in Jack Szostak's lab identified three genes (*EST1, EST2* and *EST3*) whose deletion resulted in an EST phenotype (Ever Shorter Telomeres, progressively shorter telomeres and a senescence phenotype, the same phenotypes as for a defective telomerase RNA gene) [39]. Meanwhile, telomerase was also purified from the hypotrichous ciliate *Euplotes aediculatus* by using an antisense oligonucleotide bait that was complimentary to the telomerase RNA template by Cech and coworkers [40]. The telomerase enzyme is

a ribonucleoprotein (RNP) that consists of two essential core components: a catalytic protein component, telomerase reverse transcriptase (TERT), and an essential RNA component, telomerase RNA (TR) [37, 41-44]. As the name implies, TERT is homologous to the catalytic motifs of reverse transcriptase [45, 46]. This explains how telomerase copies the RNA template to synthesize telomere DNA repeats. TR contains a short region that is complementary to the telomeric repeat sequence [37, 47]. It specifies the sequence that is added to the chromosome end using conventional Watson-Crick base-pairing.

1.4.1 Telomerase RNAs

Telomerase is currently viewed as composed of an RNA molecule with a well-defined secondary structure, a conserved TERT catalytic subunit and a number of additional protein subunits, only some of which are conserved phylogenetically [48]. TRs have been identified from a wide array of species from ciliated protozoa, several yeast species and a large number of vertebrates including human [49, 50]. Comparison of the TR sequences from different phyla demonstrates that TRs are only conserved among closely related organisms. Different secondary structure models were developed by the different laboratories by phylogenetic comparison [51-55].

Fig. 1.2a shows the secondary structure models of TRs from the ciliate *T. thermophila* [51], *Homo sapiens* [53], and the yeast *Saccharomyces cerevisiae* [54, 55]. Surprisingly, despite the difference in sequence and size, several structural features are conserved across species. The template region of all telomerases is single stranded. It allows the Watson-Crick base-pairing with the 3' end of telomere while residing in the

active site of the TERT. The length of the template is approximately 1.5-2 times the telomeric repeat length, enabling both annealing of the 3' end of telomere with the template and addition of one repeat per replication cycle. Although the essential template function of TR was discovered more than 20 years ago [37], the TR contains more than just a template. Across different species, TRs include a large loop containing the template, a 5′ template boundary element, a pseudoknot, a loop-closing helix, and a stem terminus element. These regions of TR are involved in species-specific roles in telomerase biogenesis, RNA processing, localization, and accumulation [56-60].

1.4.2 TERT, the Telomerase Reverse Transcriptase

Human RNA (hTR) and the catalytic subunit hTERT form the core of human telomerase. Two different genes in the human genome code for TR and hTERT separately. TERTs contain a C-terminal reverse transcriptase (RT) domain that is similar to RTs from retroelements and retroviruses. Three-dimensional structural modeling results suggested that similar to the RT domains from human immunodeficiency virus (HIV) and murine leukemia virus (MLV), the seven RT motifs from TERT form a "right hand"-like structure [61-65]. Besides the RT motif as the active site for catalysis, several conserved motifs in the N-terminal half "rivet" the RNA component to the protein, assuring maintenance of a stable RNA while allowing the template to move through the active site [66].

After the correct assembly, TERT forms a special "mitten" structure to wrap the chromosome end in order to favor the telomeric repeats addition [44]. It takes four steps for telomerase to add nucleotides to the end of telomere: annealing, elongation,

translocation and further elongation in a processive manner [67, 68]. First, telomerase localizes to the chromosome end and anneals to the 3' G-overhang via its internal RNA template domain of TR (Fig. 1.3). Then the catalytic component TERT functions as the reverse transcriptase and adds nucleotides to the end of the G-overhang complementary to the RNA template. After the initial round of nucleotide extension, the RNA template/telomere DNA hybrid duplex is disassembled and the telomere end realigns to the 3' end region of the template (the "Translocation" step) before the next round of "GGTTAG" nucleotide addition (the "Elongation" step again). Human telomerase repeats the translocation and elongation steps, and add "GGTTAG" repeats to telomeres continuously. Telomerase adds telomeric repeats to telomere without falling off the DNA and this is called repeat addition processivity of telomerase, which is defined as "the number of the bases synthesized when the cumulative probability of dissociation is ½" [67]. After telomerase finishes the terminal extension, the 5' C-strand needs to be properly processed in order to generate the 3' G-overhang. The 3' G-overhang could end with any nucleotide within the "TTAGGG" sequence [69]. However, when telomerase is present, TAG-3' is most likely to be observed. However, the exact mechanism of C-strand processing is still unknown.

In *S. cerevisiae*, five genes are required for the telomerase pathway [39, 42, 70, 71]. *TLC1* and *EST2* encode the RNA and reverse transcriptase subunits of telomerase, respectively. The two encoded subunits are essential for catalysis and telomerase activity is absent in extracts from strains defective in *EST2* or *TLC1* [72]. In contrast, mutations in *EST1*, *EST3*, and *CDC13* do not diminish enzyme activity *in vitro*, although they result in similar severe telomere replication defects as Δest2 or Δtlc1. The interaction of the

Est2 with telomeres is mediated by the RNA binding protein Est1, which interacts with the essential single-stranded telomeric DNA binding protein Cdc13 to recruit telomerase to the telomere [21, 73].

1.5 Telomere Binding Proteins

Soon after the discovery of the telomeric sequence of *Tetrahymena*, attempts were initiated to identify what proteins are associated with the unusual DNA sequence. The somatic nuclear DNA and rDNA telomeres of *Tetrahymena* were found to be protected from nuclease attack in a very different manner from that of nucleosomally packaged DNA [74]. However, no covalently attached proteins were discovered on *Tetrahymena* rDNA telomeres, even using methods that could have detected small amount of proteins at the ends of these minichromosomes [75]. A few years later, a few tightly but noncovalently bound proteins, in several different species, were discovered to protect the short telomeric tracts, such as *Oxytricha nova* telomere end binding protein (TEBP), the metazoan, fission yeast, and plant protection of telomeres 1 (Pot1) proteins, and *Saccharomyces cerevisiae* Cdc13 [76-80]. Binding of telomere-associated proteins is necessary for adequate maintenance of telomeric DNA. These telomere-binding proteins generally carry out their functions in two ways: structurally, they form a protective cap and functionally, they regulate telomere length [81]. Upon association of telomere binding proteins with telomeric DNA, the nucleoprotein complexes distinguish natural chromosome ends from double-stranded breaks and therefore shield chromosome termini from hazardous end-to-end fusion. In addition, telomere binding proteins recruit and regulate telomerase to ensure an appropriate length of structural DNA that is maintained

as a buffer against loss of genetic information stored in genes close to the chromosome terminus. Based on the DNA binding specificity, telomere binding proteins can be broadly classified into two classes: double-stranded and single-stranded DNA-binding proteins. Specifically, different recognition motifs are utilized by each class to confer the specificity: the OB-fold recognizes the single-stranded G-overhang while the Myb motif designates double-stranded telomere DNA association.

1.5.1 Telomeric proteins that recognize the single-stranded G-overhang

Most eukaryotic organisms preserve a special structural feature: the 3' end single-stranded G-overhang that protrudes outside of the duplex region of the telomere. It is seen and conserved from ciliated protozoa, to yeasts and mammals [82-84]. The first protein to be identified that specifically recognizes and caps the single-stranded G-overhang was the ciliate *O. nova* TEBP (telomere end binding protein) [77, 85]. It is composed of two subunits, α and β. These two proteins can form two alternative complexes, a α–β heterodimer and a α–α homodimer, that both bind specifically but differently to the chromosome overhangs. While the α–β heterodimer inhibits the action of telomerase, the α–α homodimer does so to a lesser extent [86]. These telomere-specific structural protein complex recognize and bind to the single-stranded G-overhang in a sequence-specific fashion, also protecting the neighboring duplex telomeric DNA [87].

In budding yeast, the single-stranded G-overhang is bound specifically by Cdc13 [70, 88]. Cdc13 has two separate functions: both in telomere protection and telomerase recruitment, hence telomere replication [39, 88]. After a long search, POT1 (protecting of

telomeres 1), a widespread protein, has been found in fission yeast, humans, and other species. POT1 proteins across different species only share a weak sequence similarity to the amino-terminal region of the *O. nova* TEBPα subunit [89-94]. Interestingly, both POT1 and Cdc13 appear to be the homologs of TEBPα, with which they share the existence of oligonucleotide/oligosaccharide binding domains (OB folds) that mediate the interaction with the G-tail [95, 96]. Deletion of the *POT1* gene in fission yeast leads to rapid loss of telomeric DNA and chromosome circularization, suggesting POT1 has a crucial role in telomere protection [89]. In human cells, POT1 plays a crucial role in telomere length homeostasis, through its interaction with both the G-tail and human TRF1 [97, 98]. Overexpression of a hPOT1 mutant incompetent for DNA binding led to rapid telomere lengthening in telomerase-positive cells [98]. Thus, POT1 is a negative regulator of telomere length. POT1 also contributes to the protection of chromosome ends, preventing telomeres from initiating inappropriate telomere-telomere recombination events, such as homologous recombination (HR) and non-homologous end joining (NHEJ) [93, 99].

Structural biology has been particularly important for inferring evolutionary relationships among telomere binding proteins. Crystal structure of the DNA-binding domain from *O. nova* in complex with cognate single-stranded telomere DNA provided the first high-resolution view of telomere overhang binding protein in action [100]. Since then the structures of several other single-stranded DNA-binding domains from other species have been determined, including Cdc13 and POT1 [100-104]. All these proteins bind the single-stranded G-overhang using a conserved DNA-binding motif, the OB fold

[105]. The basic structural elements are discussed so as to provide a foundation for understanding the mechanisms of telomere G-overhang binding protein action.

The OB fold was first described as an example of a homologous protein family which shares common three-dimensional structures without much similarity on the amino-acid level [105]. OB-fold is a structural domain of 70 – 180 amino acids in length with diverse functions, and has been found in many proteins including human replication protein A (RPA), the B subunit of heat-labile enterotoxin and *E. coli* single-stranded DNA-binding protein (SSB) [105]. The OB fold comprises two orthogonally packed anti-parallel β sheets with β1: β4: β5 strand topology in one sheet and β1: β2: β3 topology in the other. The N-terminal strand β1 extends as the cap for both sheets. Strands β4 and β5 often fold over onto the other sheet and closes the whole β-barrel-like structure (Fig. 1.4) [105]. Most telomere-binding OB-folds are further characterized by a C-terminal α-helix. The loops connecting β strands of the OB-fold are variable in length and these insertions/deletions account for the unreliability of current bioinformatics tools for positively identifying OB-folds.

The first well-characterized structure of the single-stranded G-overhang binding protein is *O. nova* TEBP [100]. The X-ray cocrystal structure of the heterotrimeric TEBPα-TEBPβ-ssDNA complex revealed a total of four OB folds, with three OB folds devoted to recognition of DNA and a fourth OB fold involved in protein-protein interactions between TEBPα and TEBPβ [100]. More recently, the NMR structure of this well-characterized domain from *S.cerevisiae* Cdc13p is revealed to be composed of a single OB fold [95, 106, 107]. As expected from sequence comparisons, the crystal structure of the amino-terminal region of *S. pombe* POT1 complexed with single-stranded

telomeric repeat GGTTAC confirmed the presence of an OB fold [103]. However, as seen in crystal structures, two OB folds make up the DNA-binding domain in human POT1 bound with the minimum binding sequence (TTAGGGTTAG) that is more than one telomere repeat [104]. The two OB folds pack together and form a continuous DNA binding cleft. The superposition of the crystal structures of *O. nova* TEBPα, *S. pombe* POT1 and human POT1 shows that the three single-stranded G-overhang-binding proteins bind to their cognate single-stranded telomere DNAs in a similar manner (Fig. 1.5). The three share a very similar central core consisting of a curved five-stranded antiparallel β–barrel and a helical extension at their carboxyl termini. The single-stranded DNA primarily binds in a groove formed by one face of the β–barrel and two flanking loops. In all three, the DNA strands bind with the same polarity and take up a more or less extended and irregular conformation, with the DNA ribose-phosphate backbones exposed and the bases buried by the proteins. The interaction between the DNA and protein is predominantly aromatic and hydrophobic [102, 104, 108]. The most evident difference is the length of G-overhang bound to each OB fold, which varies from six to twelve nucleotides. In conclusion, the structural information has revealed not only that single-stranded G-overhang recognition is achieved via a conserved OB fold, but also that the number of OB folds involved in binding and the length of G-overhangs varies greatly.

1.5.2 Telomeric proteins that recognize the double-stranded telomeric DNA

S. cerevisiae Rap1p (repressor-activator protein 1) was the first protein to be discovered that binds specifically to the double-stranded telomeric DNA [109]. It was originally

identified as a regulator of transcription and was later discovered to bind the irregular telomeric sequence d(GTG_{1-3}) of *S. cerevisiae*. The double-stranded d(TTAGGG) repeats of mammals are found to be bound by TRF1 (telomere repeat binding factor 1) [110] and TRF2 [111, 112]. The two TRFs have different functions at telomeres: TRF1 negatively regulates telomere length [113] while TRF2's primary role appears to be in capping and protecting chromosome ends [111]. In a similar fashion, fission yeast telomeres are protected from inappropriate fusions by TAZ1 (telomere-associated protein in *Schizocaccharomyces pombe*), a TRF-like protein [114].

Similar to the OB folds in single-stranded G-overhang binding proteins, the homeodomain myb-like motif is found in all telomere binding proteins that recognize double-stranded telomere DNAs. The myb motif is named after the transcription factor, c-Myb, a proto-oncogene that regulates differentiation and proliferation during hematopoiesis [115]. The Myb motif consists of three α-helices arranged in an orthogonal bundle around a hydrophobic core (Fig. 1.6). The third helix contains residues that make sequence-specific contacts with bases in the major groove of B-form DNA [116-118]. For telomere proteins, these DNA recognition residues are especially well conserved and define the so-called telobox sequence feature [119, 120]. In budding yeast, two imperfect, tandem, myb-like repeats bind to the irregular telomeric sequence (GTG_{1-3}) directly [109, 121]. TRF1 binds double-stranded telomeric DNAs as a dimer, with flexible loops connecting the conserved TRF-homology (TRFH) domain with a single myb-like domain [110, 122]. The dimerization is mediated via the central TRFH domain [123]. Shortly after the discovery of TRF1, a novel protein TRF2, which shares high homology with TRF1, was found in human and mouse [111, 113]. Taz1 in fission yeast

and TRF2 share the same TRFH-myb domain architecture seen in TRF1 [112]. The crystal structures of the TRFH domains from both human TRF1 and TRF2 reveal that they have almost identical and entirely α-helical dimeric structures [123]. After comparing of domain architecture of Rap1, TRF1, TRF2 and Taz1, it is believed that telomere double-stranded DNA-binding proteins are required to use two myb-like motifs to recognize telomeric DNA. Thus, dimerization may have been a result of adaptive feature during evolution.

1.5.3 The shelterin complex

Several proteins have been implicated in the control of telomerase accessibility to telomere ends and chromosome end protection. In mammals, the core complex is called telosome or shelterin, which contains six telomere-specific proteins, TRF1, TRF2, TPP1, TIN2, RAP1 and POT1 [124-126]. It is delivered to telomeres by two of its components, TRF1 and TRF2. These proteins are abundant at chromosome ends in the nucleus but do not accumulate elsewhere. Both TRF1 and TRF2 bind to double-stranded telomeric DNA while POT1 binds to single-stranded TTAGGG repeats. TRF2 recruits an additional factor, RAP1 [127]. A sixth factor, TIN2 (TRF1-interacting nuclear factor 2) interacts with three shelterin proteins, TRF1, TRF2 and TPP1, and thus has an important architectural role in bridging the duplex binding proteins, TRF1 and TRF2-RAP1, with the POT1-TPP1 complex at the single-stranded G-overhang (Fig. 1.7). Shelterin is implicated in the formation of t-loops, affects the structure of the telomere terminus and controls the synthesis of telomeric DNA by telomerase [128, 129]. Sheltein is at telomeres throughout the cell cycle. This exclusivity is the major difference between

shelterin and other DNA damage-processing factors and various other proteins that are found at telomeres. These telomere-associated proteins can have crucial roles at telomeres, but also accumulate elsewhere in the cell.

Shelterin is proposed to have a fundamental role in regulation of telomere length [93, 130, 131]. As longer telomeres load more shelterin complexes onto telomeric DNA ends, these shelterin complexes function in telomere length control by providing a length-sensing mechanism [132]. It also plays other important roles in keeping telomeres away from DNA damage checkpoints and therefore protects chromosome ends from inappropriate DNA repair pathways [133]. Shelterin complex dysfunction or telomerase mutations will result in excessive telomere shortening, which in turn triggers a DNA damage response at chromosome ends and are then recognized as double-strand breaks. In addition, the emerging view is that shelterin changes the structure of the telomeric DNA in order to control the synthesis of telomeric DNA by telomerase via limited access.

Similarly, a multi-protein telomeric complex with a shelterin-like architectural organization has been revealed in fission yeast *S. pombe* [134]. There are seven components in this complex, and many of them are the structural and functional homologues of the mammalian shelterin proteins (Fig. 1.8). The double-stranded telomeric repeats bind Taz1, the only known ortholog of the mammalian telomere proteins TRF1 and TRF2 [114, 123]. Like its higher eukaryotic counterparts, the Taz1 protein contains a carboxyl-terminal myb-like motif and a TRFH domain [127]. Fission yeast Rap1 is recruited to telomeres by binding to Taz1. Until now, the only known single-stranded telomere specific binding protein is Pot1, which contains an OB fold

homologous to those found in the *O. nova* TEBP and budding yeast Cdc13. Poz1, a small protein with no obvious sequence similarity to any components of the mammalian shelterin, interacts with both Tpz1 and Rap1, and thus connects the single-stranded and double-stranded binding proteins together. This bridging function of Poz1 closely resembles the role of TIN2 in mammalian shelterin complex, raising the possibility that Poz1 might be a TIN2 homolog. The last component is Ccq1 (coiled-coil quantitatively enriched protein 1). It interacts with Tpz1 and plays a key role in recruiting telomerase to telomeres ([134], reviewed in [135]).

1.6 Budding Yeast Telomeres

The 2009 Nobel Prize in Physiology or Medicine was awarded jointly to Drs. Elizabeth H. Blackburn, Carol W. Greider and Jack W. Szostak for their pioneer research on "how chromosomes are protected by telomeres and the enzyme telomerase" (http://www.nobelprize.org/nobel_prizes/medicine/laureates/2009/press.html). Their seminal study in 1982 also started the field of yeast telomere biology [15]. Since then, studies from numerous groups have greatly expanded our knowledge of the details of yeast telomere structure, as well as the gene products that maintain yeast chromosome ends.

1.6.1 Yeast telomeres

Over the last couple of decades, several model organisms have been used as model systems to study the protection and maintenance of telomere, including *Tetrahymena thermophila*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Arabidopsis*

thaliana. They all have irregular telomere repeat sequences [83]. The degeneracy is most pronounced in *S. cerevisiae* with the telomere consensus sequences often described as $G_{1-3}(TG)_{1-6}$ [16, 80]. This divergence from the more common theme of homogenous telomeric repeat sequences is due to degenerate copying of the template region of the yeast telomerase RNA (TLC1) [42]. Although telomere and telomere-associated factors assume evolutionarily conserved functions, the subphylum of budding yeast that includes *S. cerevisiae* (Saccharomycotina) exhibits arguably the greatest evolutionary diversity. Besides the well-studied *S. cerevisiae*, Saccharomycotina also includes *Kluyveromyces*, and *Candida* spp [136, 137]. Based on comparison of the whole genome, *S. pombe*, the only member of the Schizosaccharomycetes, sits outside Saccharomycotina phyla clade [136]. The branches of budding yeast exemplified by *Candida albicans* have apparently undergone rapid evolutionary divergence with respect to its telomere sequence and telomere-related proteins. For instance, unlike *S. cerevisiae*, many *Candida* spp. have long (up to 25-base-pair), distinct and regular telomere repeat units. Moreover, the putative telomere maintenance proteins of Candida spp. (e.g., Rap1) have been observed to exhibit significant structural divergence from their *Saccharomyces* counterparts. Analysis of telomeres cloned from multiple yeast species revealed that telomeric repeat variation is a characteristic shared by many fungal species [138-140]. Examination of TLC1 genes from multiple yeast species shows the source of this sequence variation. A conserved core sequence of about 6 nucleotides can be detected in the alignment of TLC1 templates from both *Kluyveromyces* and *Saccharomyces* genus [141]. It has been argued that this core preserves a binding site for the essential duplex telomere-binding protein Rap1 and there is evidence that the Rap1 consensus corresponds closely to the conserved

telomeric repeat core [140]. It could also serve as a selection for the binding site of other telomere-binding proteins, such as Cdc13. Localization of the Cdc13 protein to single-stranded telomeric DNAs also relies on sequence-specific recognition [102].

1.6.2 The CST complex

Yeast chromosomes terminate with a single-strand 3' extension of the G-rich strand. Intriguingly, budding yeast telomeres are not protected by a shelterin-like complex. Although the double-stranded region of the telomere is bound by Rap1 and two associated factors, Rif1 and Rif2, these proteins are not involved in chromosome end protection. Instead, this function is fulfilled by a heterotrimer named CST (Cdc13-Stn1-Ten1, Fig. 1.9) that is more closely related to Replication Protein A (RPA) [142, 143]. The CST complex was initially discovered in budding yeast; however, recent data show that CST is also present in a wide range of multicellular organisms in which its function seems to be telomere capping. None of the CST components show obvious sequence identity to POT1, TPP1/Tpz1 or other shelterin components [144]. *S. cerevisiae* CST plays a dual role in telomere protection and regulation of telomere replication. Although Cdc13 is the main DNA-binding subunit, all three proteins function in the process and depletion of any subunit results in degradation of the telomeric C-strand, accumulation of long G-overhangs and activation of a DNA-damage response.

Cdc13 was first discovered in Hartwell's legendary screen for *cdc* (cell division cycle) mutants and was later shown to play a central role in coordinating multiple events at yeast chromosome termini [145]. Its essential function at telomeres is to protect chromosome ends from degradation. Cdc13 associates with the long single-stranded G-

overhang through its ability to recognize the telomeric extension with high sequence specificity and high affinity during the late S to G2 phase, as well as to short telomeric overhangs in the rest of the cell cycle [88]. Loss of Cdc13 exposes these termini to immediate extensive resection of the C-strand of telomeres, leading to a *RAD9*-mediated arrest [146-149]. Two essential Cdc13-associated proteins, Stn1 and Ten1, also contribute to this capping activity, based on the increased resection that also occurs in strains impaired for either *STN1* or *TEN1* function [150, 151]. *STN1* was originally identified as a high copy suppressor of *cdc13-1* temperature sensitivity [151], and *TEN1* was similarly isolated as a dosage suppressor of *stn1-13 [152]*. The *Saccharomyces cerevisiae* Stn1 and Ten1, when overexpressed, are capable of mediating Cdc13-independent protection of telomeres [153]. As a matter of fact, the lethality of a *cdc13*-Δ strain can be rescued if the Stn1 protein is delivered to telomeres though a fusion of the Stn1 protein and the Cdc13 DNA-binding domain (DBD). Neither the DBD alone nor the DBD fused to other telomere-specific proteins is sufficient to mitigate the same defect. Thus, it is speculated that Cdc13's primary role in end protection is to deliver Stn1 to telomeres [154]. In addition to these genetic interactions, Stn1 and Ten1 proteins interact with each other both *in vivo* and *in vitro* [152, 155], and each associates with Cdc13 in the yeast two-hybrid assay [150, 151, 156]. Besides, Stn1 and Ten1 can make contributions to capping independent of Cdc13 [153]. Based on these data, Cdc13, Stn1, and Ten1 are suggested to function as a single complex that mediates chromosome end protection in *S. cerevisiae*. Recent bioinformatics analysis pointed to potential structural similarities between Stn1 and Rpa2 [155]. In Chapter 2, I will discuss in more detail some similarities between CST and the single-stranded DNA binding complex RPA.

The ability of Cdc13 to associate with the telomeric overhang is mediated through the use of multiple OB folds (Fig. 1.10). The first characterized OB fold of Cdc13, known as $Cdc13_{DBD}$, was described by the Wuttke laboratory in 2002 [95]. This centrally located domain dictates the high affinity and high specificity binding of Cdc13 for single-stranded telomeric DNAs [154]. It also facilitates the localization of full-length Cdc13 to the chromosome ends. A second OB fold at the N-terminus of Cdc13 ($Cdc13_{OB1}$) was recently identified by the Lei and Skordalakes laboratories independently [157, 158]. Although $Cdc13_{OB1}$ is structurally most similar to $Cdc13_{DBD}$ and also contains a basic cleft that corresponds to the canonical nucleic acid-binding pocket of OB folds, recent research showed that $Cdc13_{OB1}$ does not possess DNA-binding activity, which is contradictory to earlier results. Instead, it has been reported to mediate protein-protein interactions at telomeres [159, 160]. One of the $Cdc13_{OB1}$-binding proteins is Pol1, the catalytic subunit of DNA polymerase α-primase complex and this interaction will be discussed in more detail in Chapter 3. Most recently in 2011, a third OB fold was discovered by the Lei laboratory [161]. The C-terminus OB fold was found to be important for high affinity DNA binding through self-association and represent a novel mechanism of OB fold dimerization. It will be covered in more detail in Chapter 4 of this dissertation.

Cdc13 also plays two functionally distinct roles as a regulator of telomere length [162]. On one hand, Cdc13 acts as a positive regulator of telomere elongation by recruiting telomerase to the telomeres via its interaction with the telomerase subunit Est1 [39]. An important separation-of-function allele of *CDC13*, called *cdc13-2*, confers a severe telomere replication defect but leaves the end-capping capability of Cdc13 intact

[39, 70]. The analysis based on *cdc13-2* showed that Cdc13 uses a 15-kD domain at the amino terminus to interact with the Est1 sununit of telomerase. On the other hand, Cdc13 also has a separate, less well-characterized role as a negative regulator of telomere replication [163]. Interestingly, the discovery was made on the analysis of a new mutation, *cdc13-5*, which exhibited extensive elongation of the G strand of the telomere by telomerase, the same phenotype as DNA polymerase α mutations. The phenotype can be suppressed by Stn1, suggesting that Stn1 coordinates the action of the lagging strand replication complex with the regulatory activity of Cdc13. Combing the two lines of evidence, Cdc13 appears to participate in a two-step pathway: first, recruit telomerase to chromosome ends by interacting with Est1; then, a negative regulatory mechanism involving Cdc13, Stn1 and DNA polymerase α to regulate the extent of elongation and limit G-strand synthesis by telomerase in response to C-strand replication.

Because of its key roles in cell viability, Cdc13 has been the focus of intense study in many laboratories. The following chapters will detail my efforts towards understanding the structure and function of Cdc13 and its binding partners Stn1 and Ten1.

1.6.3 The yeast telomerase RNP

Contrary to the biochemical approaches scientists took to uncover the components of the ciliate telomerase complex, genetic strategies were proven crucial to unravel the components of the yeast telomerase complex. The gene encoding the RNA subunit, TLC1, was first identified by Singer and Gottschling [42]. The protein components were identified from genetic screening based on the Est (for ever shorter telomeres) phenotype

in budding yeast *Saccharomyces cerevisiae* [39, 164]. The two characteristics of the Est phenotype include continuous telomere shortening and decrease in cell viability. The catalytic core of the budding yeast telomerase comprises of the Est2 protein and TLC1 telomerase RNA. Besides Est2 and TLC1, there are two accessory factors, Est1, which binds to a bulged stem in TLC1, and Est3, that are not essential for catalysis in either *S. cerevisiae* [165] or *C. albicans* [166]. Mutations of Est1 or Est3 lead to progressive telomere shortening. The highly basic 82-kD Est1 protein possesses three distinct biochemical functionalities. One, it associates with telomerase RNP by way of TLC1 RNA. Two, Est1 can be recruited to the telomere region by Cdc13, thus localize the telomerase catalytic core to the chromosome ends. Finally, it also interacts with single-stranded telomeric DNA.

Regulation of telomerase could take place at three levels: at the level of recruitment to the telomere terminus, at the initiation of elongation, or at the rate and processivity of the elongation cycles. Most notably, the Est1 and Cdc13 interaction in the recruitment stage has been the focus of multiple research groups. The primary evidence for the recruitment model stems from a number of gene fusion experiments in which Cdc13 or its DNA-binding domain were fused to Est2, Est1, or Est3. The chimeric proteins could mitigate or even completely rescue the telomere maintenance defects of *cdc13–2* and *est1Δ* strains [73, 154]. For example, a $Cdc13_{DBD}$-Est2 fusion can bypass Est1 in telomere maintenance. Consequently, these experiments suggest that the recruitment step is essential for telomere maintenance and the Cdc13-Est1 interaction is central to recruit telomerase to the very end of the chromosomes. Additionally, another "activation" model has been proposed by the Zakian group [167]. Contrary to the

expectations of a recruitment model, the *cdc13-2* protein can interact with Est1 normally by both *in vitro* and *in vivo* criteria, suggesting that the functional interaction between Cdc13 and Est1 that is lost in a *cdc13-2* strain occurs at a step other than recruitment [159, 168].

Are there any additional components of the yeast telomerase RNP yet to be discovered? Neither the genetic defect screening nor the *EST* genes was exhaustive, which left the possibilities to identify more components wide open. McEachern and colleagues uncovered more than 150 nonessential genes that had not been previously studied for their effects on telomere length [169]. Many of these "new" genes are involved in cellular processes such as DNA replication, nucleotide metabolism and chromatin remodeling. It would be interesting to study how they contribute to the regulation of telomeres.

1.7 Telomere, Telomerase and Cancer

Human cancers are invariably associated with activation of some mechanism to maintain telomere length [48]. Telomerase is highly expressed in cells that need to divide regularly, such as cancer cells and stem cells. Approximately 85%-90% cancer cells show reactivation of telomerase while the rest maintain telomeres by ALT (alternative lengthening of telomeres), which occurs by exchange of sequences between telomeres [170]. Many precancerous tissues have critically shortened telomeres prior to telomerase detection, suggesting that short telomeres may limit the growth of precancerous cells and that only when telomerase is up-regulated or reactivated do additional cancerous changes happen [171-175]. Besides, many cell types have been immortalized by the introduction

of hTERT, without the complication of cancerous changes [176, 177]. This makes telomerase an attractive target for pharmaceutical development of anti-cancer chemotherapeutics. Currently, multiple telomerase targeting approaches are in development in both preclinical and clinical trials [178-181]. They include direct telomerase enzyme inhibitors such as oligonucleotide, small-molecule inhibitors and gene therapy [182], or vaccination via direct injection of a plasmid containing hTERT into lymphocytes [183]. Other creative approaches include inhibition of telomerase assembly [184], hammerhead ribozymes directed against hTR [185], mutant template RNA gene therapy [186] and reverse transcriptase inhibitors [187, 188]. Although telomerase activity is greatly reduced or undetectable in most normal tissues [189, 190], inhibition of telomerase could have detrimental effects on normal cells that do express telomerase, such as germ-line cells and renewable tissues [190-193]. Another issue that has been raised about telomerase inhibitors is that alternative mechanisms for telomerase maintenance (ALT) have been found in other organisms and in some rare human cancers [194-196]. Thus, telomerase inhibition may promote the drug resistance of telomerase-independent cancer cells. At present, there is still a lot of basic research needs to be done, but it is encouraging that there are already clinical studies under way with hTERT vaccines and immunotherapy. It is reasonable to expect greater progress after valid animal models are established.

1.8 Outline of the Thesis

From the extensive studies of the function of telomere-binding proteins during the last two decades, we now understand how important many of the proteins are in protecting

telomeres from deleterious events and in telomere replication. However, the structural characterization of these proteins lagged behind. In particular, only one centrally located OB domain in the CST complex of budding yeast has been characterized. Learning more about the molecular architecture of this complex and determining its overall structure will greatly enrich our understanding of its contribution in telomere protection and regulation.

I have been working for the past three years on understanding the biochemistry and structures of the CST complex using different biochemical and biophysical methods, including X-ray crystallography. In Chapter 2, by cocrystallizing the N-terminal fragment of Stn1 and full-length Ten1, I determined the crystal structure of the complex, which revealed striking structural similarity between Stn1-Ten1 and Rpa32-Rpa14. To understand the degree of resemblance between Cdc13 and Rpa70, the structure of multiple domains of Cdc13 has been described in Chapter 3 and 4. From the structure, I also discovered a potentially conserved feature of dimerization in Cdc13 proteins and its implication in mediating interaction with DNA polymerase catalytic subunit (Pol1) and DNA binding. Finally, Chapter 5 describes my most recent efforts to characterize the interaction between Cdc13 and telomerase subunit Est1.

Figure 1.1 End replication problem
The leading strand (green) is continuously synthesized from 5' to 3' by polymerase. The lagging strand (red) synthesis is initialized by RNA primers (purple) from 3' to 5'. When RNA primers are removed, a gap will be left at the 3' end.

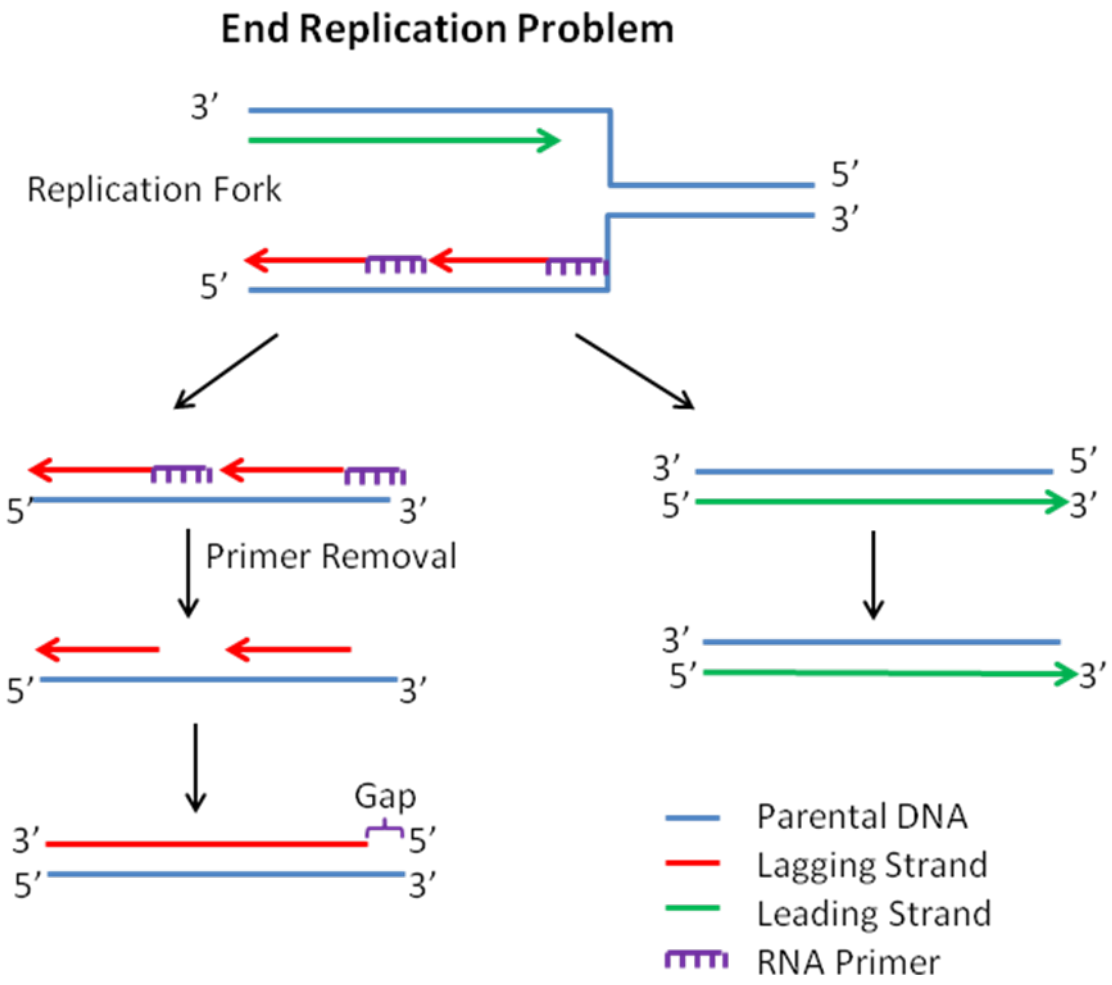


Figure 1.2 Structures of TERT and TR

Structures of TRs. Secondary structures of *T. thermophila*, *H.sapiens*, and *S. cerevisiae* TRs are illustrated on the basis of published studies [51-55]. Template regions (yellow), main TERT-binding regions (boxed in purple boxes), and template boundary regulating elements (highlighted in blue) are indicated. The template boundary element (TBE) in *T. thermophila* overlaps with the main TERT-binding region. Low-affinity TERT-binding sites in helix IV and the template recognition element (TRE) in *Tetrahymena* and in the pseudoknot/template domain in humans have also been identified. These regions are illustrated in light brown. Several structures have been proposed for the yeast telomerase RNA pseudoknot region, only one of which is presented here. Dimethyl sulphate-based footprinting analysis suggests that the yeast pseudoknot structure may be in equilibrium with other conformational state(s) [197]. The yeast RNA is unusually large and contains, in addition to the central core presented in the figure, several arms that interact with Est1, Ku, and other proteins. These remaining parts are schematically represented by lines interrupted with slashes. (Figure 1.2 is adopted from Figure 3 in [198].)

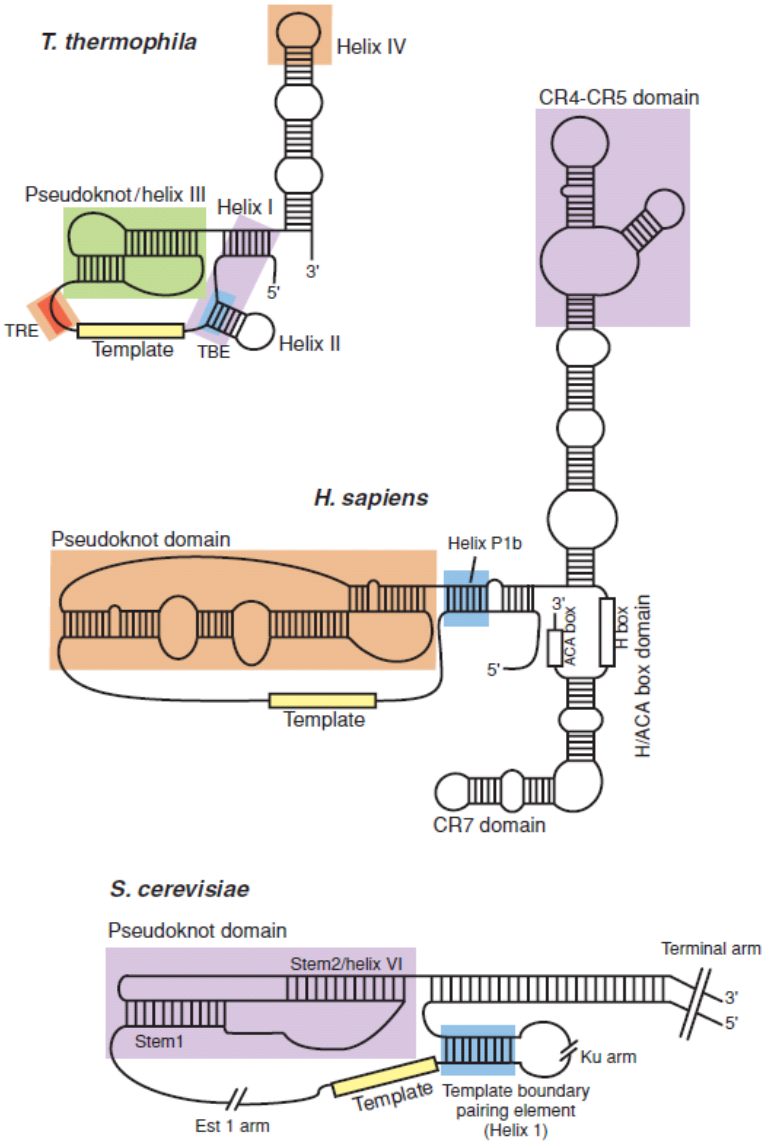
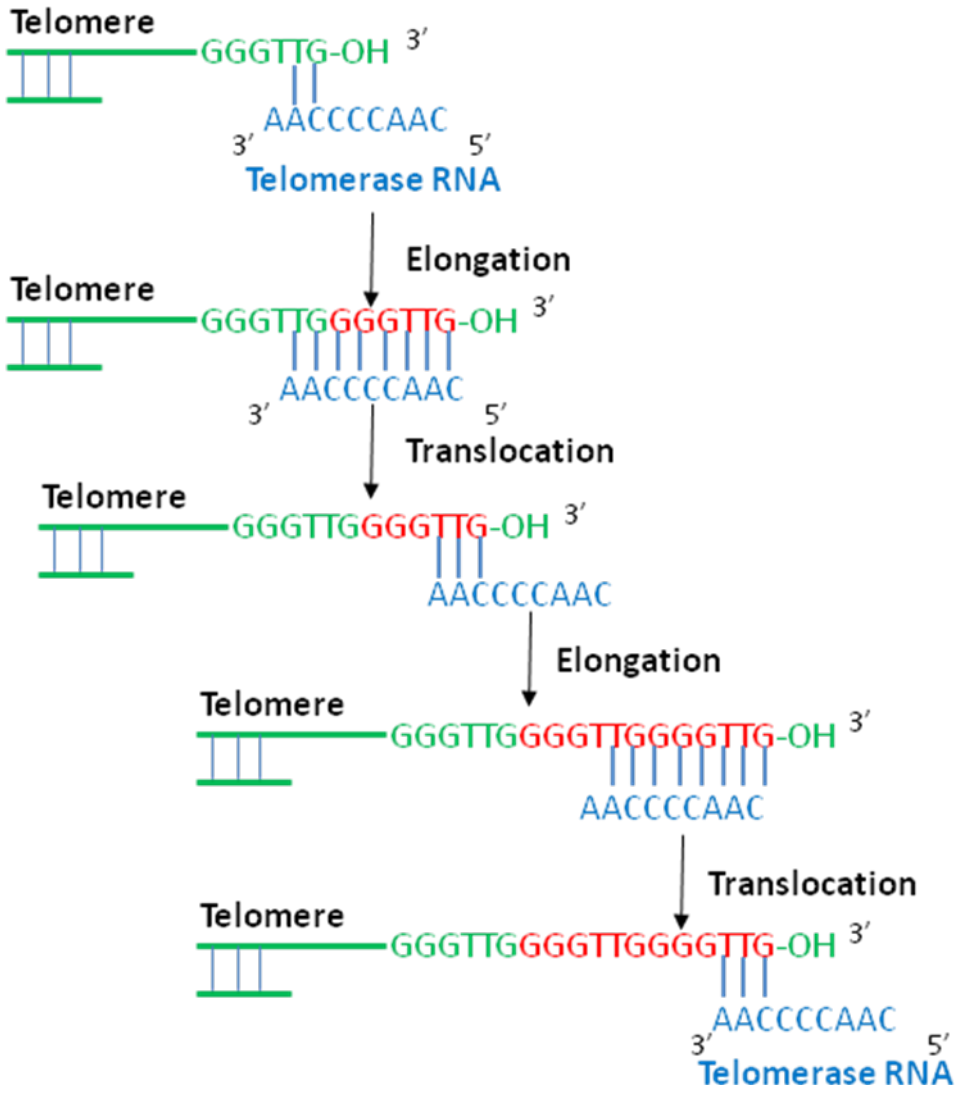


Figure 1.3 Telomerase elongates telomere in a processive manner.

Mechanism of the G-rich strand synthesis of telomeric DNA. The RNA template of telomerase is shown in blue and the nucleotides added to the G-rich strand of the primer are shown in red.



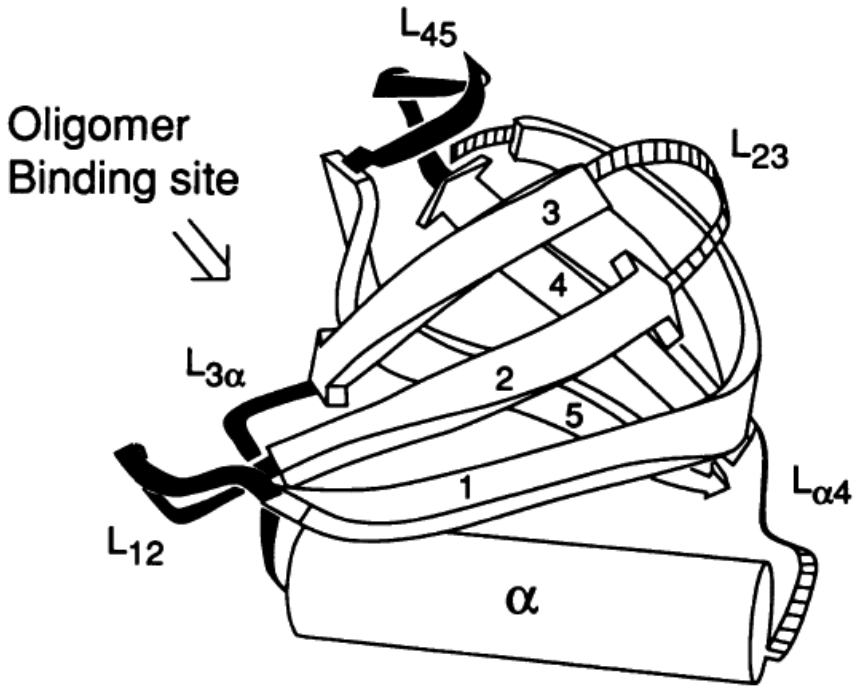
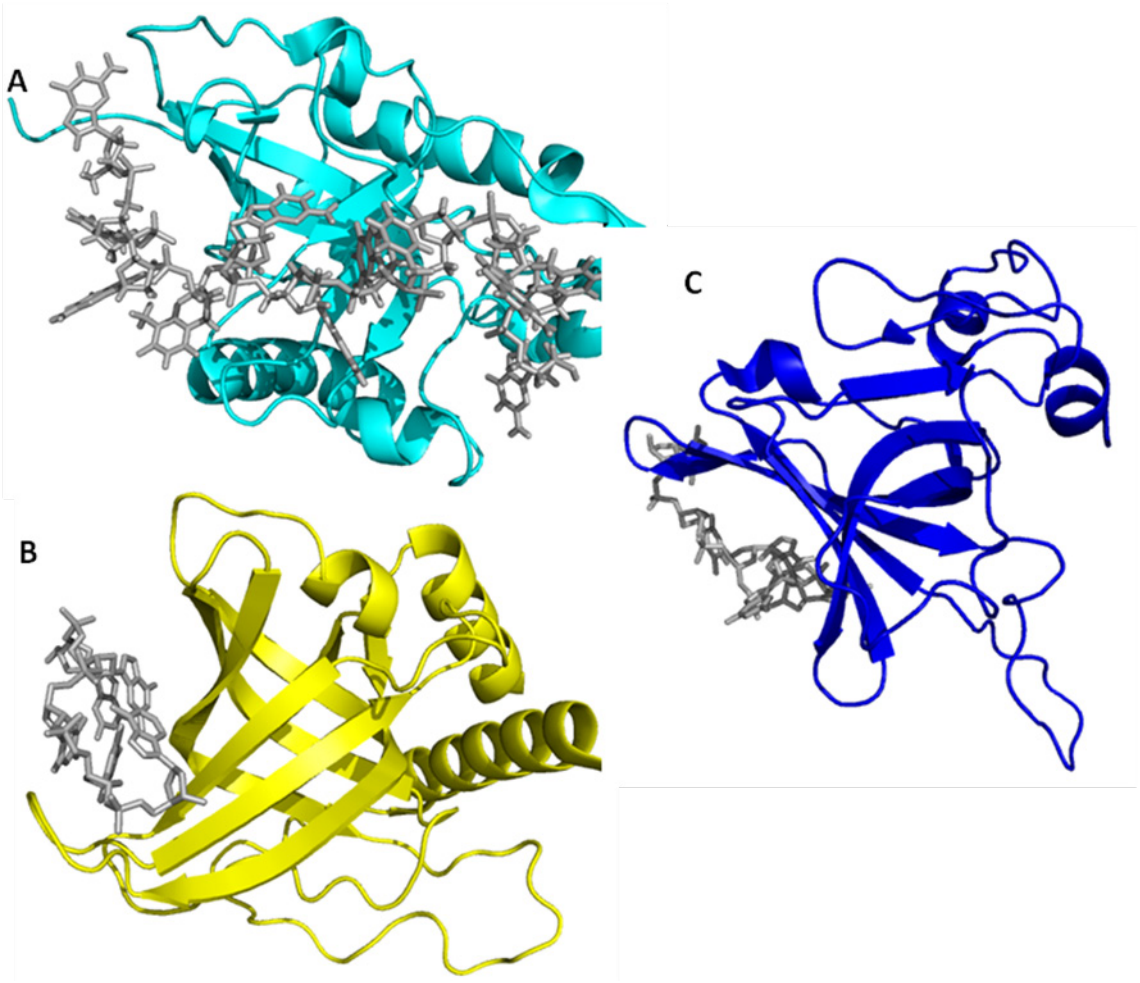


Figure 1.4 Description of the OB-fold, based on its smallest representative, the B subunit of verotoxin-1 (VT1B).

Structural architecture of a typical OB-fold [105], Fig. 1); protein loops are shown as ribbons, connecting the structural segments and numbered accordingly. The arrow indicates the oligomer binding site.

Figure 1.5 Comparison of the structures of $Cdc13_{DBD}$, *S. pombe* $Pot1_{OB1}$, and *O. nova* $TEBP\alpha_{OB1}$.

(A) $Cdc13_{DBD}$ is shown in cyan, (B) *S. pombe* $Pot1_{OB1}$ in gold, and (C) *O. nova* $TEBP\alpha_{OB1}$ in blue (D).



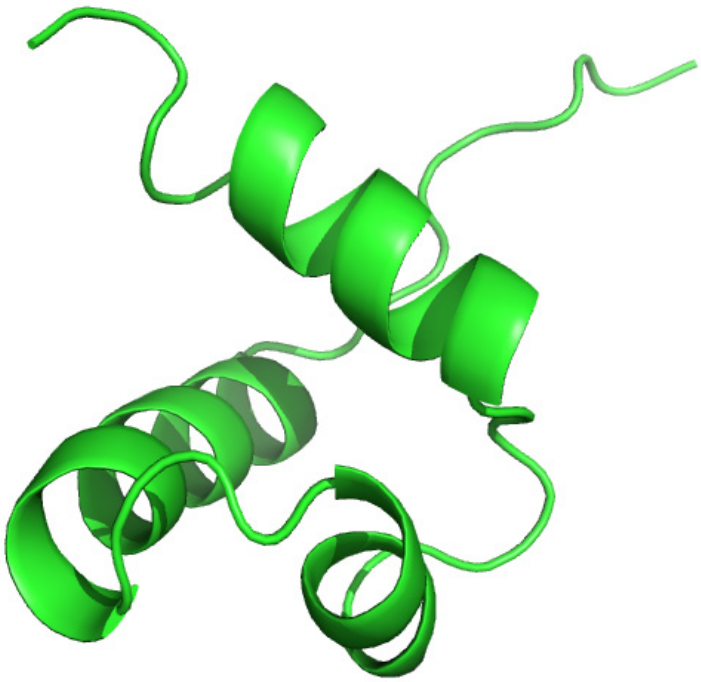


Figure 1.6 Structure of a typical c-Myb motif
Cartoon illustration of the mouse c-Myb DNA-binding domain repeat 3 (PDB #: 1idy).

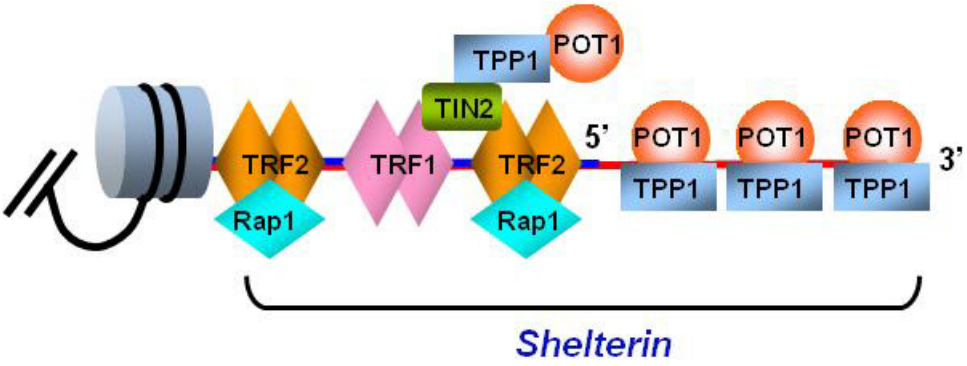


Figure 1.7 Cartoon illustration of mammalian shelterin complex
(Courtesy of M. Lei)
Six shelterin complex components: TRF1 and TRF2 on the double-stranded telomeric DNA; TIN2 binds to both TRF1 and TRF2; Rap1 interacts with TRF2; TPP1 interacts with both TIN2 and POT1. POT1 binds to single-stranded telomeric DNA.

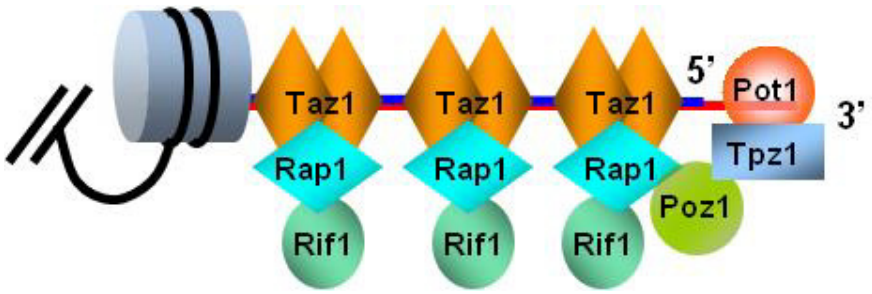


Figure 1.8 Cartoon illustration of the shelterin-like complex in fission yeast *S. pombe* (Courtesy of M. Lei)
There are seven components in this complex: Taz1 on the duplex telomeric DNA, Rif1 and Rap1 interact with Taz1. Pot1 binds to single-stranded telomeric DNA and interacts with Tpz1. Ccq1 (not shown in this picture) interacts with Tpz1. Poz1 connects Tpz1 and Rap1. Some of them are the structural and functional homologues of the mammalian shelterin proteins.

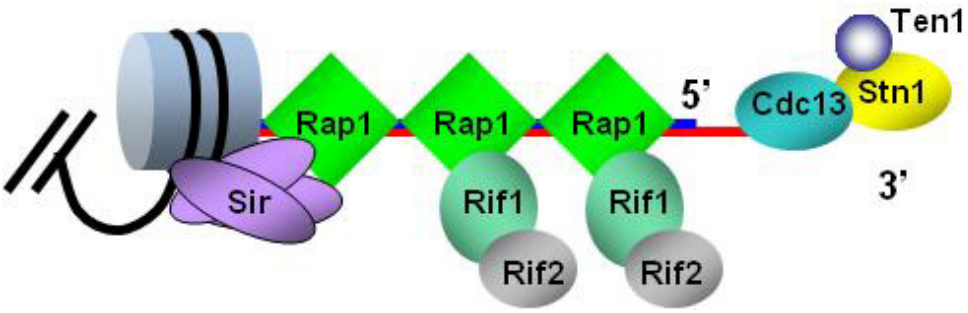


Figure 1.9 Cartoon illustration of the telomere binding proteins in *Saccharomyces cerevisiae* (Courtesy of M. Lei)
There are six components in this complex: Rap1 on the duplex telomeric DNA, Rif1 and Rif2 interact with Rap1. Cdc13 binds to single-stranded telomeric DNA and interacts with Stn1 and Ten1.

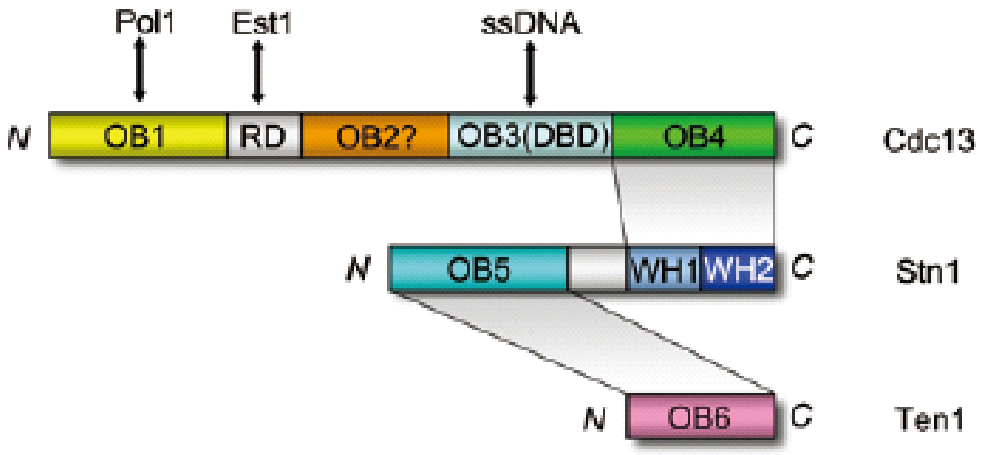


Figure 1.10 Domain organization of *Saccharomyces cerevisiae* CST complex

Table 1.1 Telomeric DNA Sequence

Species	**Organism**	**5' - Telomeric DNA Repeat Sequence - 3'**
Vertebrates	Human, mouse, *Xenopus*	TTAGGG
Higher plants	*Arabidopsis thaliana*	TTTAGGG
Budding yeast	*Saccharomyces cerevisiae*	TGTGGGTGTGGTG (from RNA template) or G(2-3)(TG)(1-6)T (consensus)
	Candida albicans	GGTGTACGGATGTCTAACTTCTT
	Candida glabrata	GGGGTCTGGGTGCTG
	Candida tropicalis	GGTGTA[C/A]GGATGTCACGATCATT
	Kluyveromyces lactis	GGTGTACGGATTTGATTAGGTATGT
Fission yeast	*Schizosaccharomyces pombe*	$TTAC(A)(C)G_{(1-8)}$
Ciliate protozoa	*Tetrahymena*	TTGGGG
	Oxytricha	TTTTGGGG
Kinetoplastid protozoa	*Trypanosoma*	TTAGGG

Reference:

1. de Lange, T., *Protection of mammalian telomeres.* Oncogene, 2002. **21**(4): p. 532-40.
2. Levis, R.W., *Viable deletions of a telomere from a Drosophila chromosome.* Cell, 1989. **58**(4): p. 791-801.
3. Muller, H.J., *The Production of Mutations by X-Rays.* Proc Natl Acad Sci U S A, 1928. **14**(9): p. 714-26.
4. Muller, H.J., *The Measurement of Gene Mutation Rate in Drosophila, Its High Variability, and Its Dependence upon Temperature.* Genetics, 1928. **13**(4): p. 279-357.
5. McClintock, B. and H.E. Hill, *The Cytological Identification of the Chromosome Associated with the R-G Linkage Group in ZEA MAYS.* Genetics, 1931. **16**(2): p. 175-90.
6. McClintock, B., *The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis.* Proc Natl Acad Sci U S A, 1939. **25**(8): p. 405-16.
7. McClintock, B., *The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion.* Proc Natl Acad Sci U S A, 1942. **28**(11): p. 458-63.
8. Greider, C.W. and E.H. Blackburn, *The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity.* Cell, 1987. **51**(6): p. 887-98.
9. Blackburn, E.H., et al., *Recognition and elongation of telomeres by telomerase.* Genome, 1989. **31**(2): p. 553-60.
10. Gall, J.G., *Free ribosomal RNA genes in the macronucleus of Tetrahymena.* Proc Natl Acad Sci U S A, 1974. **71**(8): p. 3078-81.
11. Prescott, D.M. and K.G. Murti, *Chromosome structure in ciliated protozoans.* Cold Spring Harb Symp Quant Biol, 1974. **38**: p. 609-18.
12. Karrer, K.M. and J.G. Gall, *The macronuclear ribosomal DNA of Tetrahymena pyriformis is a palindrome.* J Mol Biol, 1976. **104**(2): p. 421-53.
13. Engberg, J., et al., *Free ribosomal DNA molecules from Tetrahymena pyriformis GL are giant palindromes.* J Mol Biol, 1976. **104**(2): p. 455-70.
14. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.* J Mol Biol, 1978. **120**(1): p. 33-53.
15. Szostak, J.W. and E.H. Blackburn, *Cloning yeast telomeres on linear plasmid vectors.* Cell, 1982. **29**(1): p. 245-55.
16. Shampay, J., J.W. Szostak, and E.H. Blackburn, *DNA sequences of telomeres maintained in yeast.* Nature, 1984. **310**(5973): p. 154-7.
17. Blackburn, E.H., *Telomeres and their synthesis.* Science, 1990. **249**(4968): p. 489-90.
18. Oka, Y., et al., *Inverted terminal repeat sequence in the macronuclear DNA of Stylonychia pustulata.* Gene, 1980. **10**(4): p. 301-6.
19. Klobutcher, L.A., et al., *All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3′ terminus.* Proc Natl Acad Sci U S A, 1981. **78**(5): p. 3015-9.

20. Pluta, A.F., B.P. Kaine, and B.B. Spear, *The terminal organization of macronuclear DNA in Oxytricha fallax.* Nucleic Acids Res, 1982. **10**(24): p. 8145-54.
21. Blackburn, E.H., *Switching and signaling at the telomere.* Cell, 2001. **106**(6): p. 661-73.
22. Meselson, M. and F.W. Stahl, *The Replication of DNA in Escherichia Coli.* Proc Natl Acad Sci U S A, 1958. **44**(7): p. 671-82.
23. Drummond, M.W., et al., *Concise review: Telomere biology in normal and leukemic hematopoietic stem cells.* Stem Cells, 2007. **25**(8): p. 1853-61.
24. Okazaki, R., et al., *Mechanism of DNA replication possible discontinuity of DNA chain growth.* Jpn J Med Sci Biol, 1967. **20**(3): p. 255-60.
25. Ogawa, T. and T. Okazaki, *Discontinuous DNA replication.* Annu Rev Biochem, 1980. **49**: p. 421-57.
26. Ohki, R., T. Tsurimoto, and F. Ishikawa, *In vitro reconstitution of the end replication problem.* Mol Cell Biol, 2001. **21**(17): p. 5753-66.
27. Baird, D.M., *Telomere dynamics in human cells.* Biochimie, 2008. **90**(1): p. 116-21.
28. Berg, J.M., et al., *Biochemistry*. 2002, W.H. Freeman; NCBI: New York [Bethesda, MD].
29. Watson, J.D., *Origin of concatemeric T7 DNA.* Nat New Biol, 1972. **239**(94): p. 197-201.
30. Harley, C.B., et al., *The telomere hypothesis of cellular aging.* Exp Gerontol, 1992. **27**(4): p. 375-82.
31. Forsyth, N.R., et al., *Developmental differences in the immortalization of lung fibroblasts by telomerase.* Aging Cell, 2003. **2**(5): p. 235-43.
32. Cavalier-Smith, T., *Palindromic base sequences and replication of eukaryote chromosome ends.* Nature, 1974. **250**(5466): p. 467-70.
33. Bateman, A.J., *Letter: Simplification of palindromic telomere theory.* Nature, 1975. **253**(5490): p. 379-80.
34. Holmquist, G.P. and B. Dancis, *Telomere replication, kinetochore organizers, and satellite DNA evolution.* Proc Natl Acad Sci U S A, 1979. **76**(9): p. 4566-70.
35. Dancis, B.M. and G.P. Holmquist, *Telomere replication and fusion in eukaryotes.* J Theor Biol, 1979. **78**(2): p. 211-24.
36. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts.* Cell, 1985. **43**(2 Pt 1): p. 405-13.
37. Greider, C.W. and E.H. Blackburn, *A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis.* Nature, 1989. **337**(6205): p. 331-7.
38. Yu, G.L., et al., *In vivo alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs.* Nature, 1990. **344**(6262): p. 126-32.
39. Lendvay, T.S., et al., *Senescence mutants of Saccharomyces cerevisiae with a defect in telomere replication identify three additional EST genes.* Genetics, 1996. **144**(4): p. 1399-412.
40. Lingner, J. and T.R. Cech, *Purification of telomerase from Euplotes aediculatus: requirement of a primer 3′ overhang.* Proc Natl Acad Sci U S A, 1996. **93**(20): p. 10712-7.

41. Feng, J., et al., *The RNA component of human telomerase.* Science, 1995. **269**(5228): p. 1236-41.
42. Singer, M.S. and D.E. Gottschling, *TLC1: template RNA component of Saccharomyces cerevisiae telomerase.* Science, 1994. **266**(5184): p. 404-9.
43. Shippen-Lentz, D. and E.H. Blackburn, *Functional evidence for an RNA template in telomerase.* Science, 1990. **247**(4942): p. 546-52.
44. Cohen, S.B., et al., *Protein composition of catalytically active human telomerase from immortal cells.* Science, 2007. **315**(5820): p. 1850-3.
45. Nakamura, T.M., et al., *Telomerase catalytic subunit homologs from fission yeast and human.* Science, 1997. **277**(5328): p. 955-9.
46. Lingner, J., et al., *Reverse transcriptase motifs in the catalytic subunit of telomerase.* Science, 1997. **276**(5312): p. 561-7.
47. Shippen-Lentz, D. and E.H. Blackburn, *Telomere terminal transferase activity from Euplotes crassus adds large numbers of TTTTGGGG repeats onto telomeric primers.* Mol Cell Biol, 1989. **9**(6): p. 2761-4.
48. Cech, T.R., *Beginning to understand the end of the chromosome.* Cell, 2004. **116**(2): p. 273-9.
49. Chen, J.L. and C.W. Greider, *An emerging consensus for telomerase RNA structure.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14683-4.
50. Chen, J.L. and C.W. Greider, *Telomerase RNA structure and function: implications for dyskeratosis congenita.* Trends Biochem Sci, 2004. **29**(4): p. 183-92.
51. Romero, D.P. and E.H. Blackburn, *A conserved secondary structure for telomerase RNA.* Cell, 1991. **67**(2): p. 343-53.
52. Lingner, J., L.L. Hendrick, and T.R. Cech, *Telomerase RNAs of different ciliates have a common secondary structure and a permuted template.* Genes Dev, 1994. **8**(16): p. 1984-98.
53. Chen, J.L., M.A. Blasco, and C.W. Greider, *Secondary structure of vertebrate telomerase RNA.* Cell, 2000. **100**(5): p. 503-14.
54. Dandjinou, A.T., et al., *A phylogenetically based secondary structure for the yeast telomerase RNA.* Curr Biol, 2004. **14**(13): p. 1148-58.
55. Zappulla, D.C. and T.R. Cech, *Yeast telomerase RNA: a flexible scaffold for protein subunits.* Proc Natl Acad Sci U S A, 2004. **101**(27): p. 10024-9.
56. Mitchell, J.R., E. Wood, and K. Collins, *A telomerase component is defective in the human disease dyskeratosis congenita.* Nature, 1999. **402**(6761): p. 551-5.
57. Lukowiak, A.A., et al., *The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus.* RNA, 2001. **7**(12): p. 1833-44.
58. Jady, B.E., E. Bertrand, and T. Kiss, *Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal.* J Cell Biol, 2004. **164**(5): p. 647-52.
59. Theimer, C.A., et al., *Structural and functional characterization of human telomerase RNA processing and cajal body localization signals.* Mol Cell, 2007. **27**(6): p. 869-81.
60. Cristofari, G., et al., *Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation.* Mol Cell, 2007. **27**(6): p. 882-9.

61. O'Reilly, M., S.A. Teichmann, and D. Rhodes, *Telomerases.* Curr Opin Struct Biol, 1999. **9**(1): p. 56-65.
62. Das, D. and M.M. Georgiadis, *The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus.* Structure, 2004. **12**(5): p. 819-29.
63. Leeper, T., N. Leulliot, and G. Varani, *The solution structure of an essential stem-loop of human telomerase RNA.* Nucleic Acids Res, 2003. **31**(10): p. 2614-21.
64. Leeper, T.C. and G. Varani, *The structure of an enzyme-activating fragment of human telomerase RNA.* RNA, 2005. **11**(4): p. 394-403.
65. Theimer, C.A., C.A. Blois, and J. Feigon, *Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function.* Mol Cell, 2005. **17**(5): p. 671-82.
66. Bryan, T.M., K.J. Goodrich, and T.R. Cech, *Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase.* Mol Cell, 2000. **6**(2): p. 493-9.
67. Greider, C.W., *Telomerase is processive.* Mol Cell Biol, 1991. **11**(9): p. 4572-80.
68. Sun, D., et al., *Regulation of catalytic activity and processivity of human telomerase.* Biochemistry, 1999. **38**(13): p. 4037-44.
69. Sfeir, A.J., et al., *Telomere-end processing the terminal nucleotides of human chromosomes.* Mol Cell, 2005. **18**(1): p. 131-8.
70. Nugent, C.I., et al., *Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance.* Science, 1996. **274**(5285): p. 249-52.
71. Lundblad, V. and E.H. Blackburn, *RNA-dependent polymerase motifs in EST1: tentative identification of a protein component of an essential yeast telomerase.* Cell, 1990. **60**(4): p. 529-30.
72. Lingner, J., et al., *Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity.* Proc Natl Acad Sci U S A, 1997. **94**(21): p. 11190-5.
73. Evans, S.K. and V. Lundblad, *Est1 and Cdc13 as comediators of telomerase access.* Science, 1999. **286**(5437): p. 117-20.
74. Blackburn, E.H. and S.S. Chiou, *Non-nucleosomal packaging of a tandemly repeated DNA sequence at termini of extrachromosomal DNA coding for rRNA in Tetrahymena.* Proc Natl Acad Sci U S A, 1981. **78**(4): p. 2263-7.
75. Blackburn, E.H., et al., *DNA termini in ciliate macronuclei.* Cold Spring Harb Symp Quant Biol, 1983. **47 Pt 2**: p. 1195-207.
76. Gottschling, D.E. and T.R. Cech, *Chromatin structure of the molecular ends of Oxytricha macronuclear DNA: phased nucleosomes and a telomeric complex.* Cell, 1984. **38**(2): p. 501-10.
77. Gottschling, D.E. and V.A. Zakian, *Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA.* Cell, 1986. **47**(2): p. 195-205.
78. Berman, J., C.Y. Tachibana, and B.K. Tye, *Identification of a telomere-binding activity from yeast.* Proc Natl Acad Sci U S A, 1986. **83**(11): p. 3713-7.
79. Buchman, A.R., et al., *Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae.* Mol Cell Biol, 1988. **8**(1): p. 210-25.

80. Conrad, M.N., et al., *RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability.* Cell, 1990. **63**(4): p. 739-50.
81. Smogorzewska, A. and T. de Lange, *Regulation of telomerase by telomeric proteins.* Annu Rev Biochem, 2004. **73**: p. 177-208.
82. Henderson, E.R. and E.H. Blackburn, *An overhanging 3' terminus is a conserved feature of telomeres.* Mol Cell Biol, 1989. **9**(1): p. 345-8.
83. McElligott, R. and R.J. Wellinger, *The terminal DNA structure of mammalian chromosomes.* EMBO J, 1997. **16**(12): p. 3705-14.
84. Wright, W.E., et al., *Normal human chromosomes have long G-rich telomeric overhangs at one end.* Genes Dev, 1997. **11**(21): p. 2801-9.
85. Price, C.M. and T.R. Cech, *Telomeric DNA-protein interactions of Oxytricha macronuclear DNA.* Genes Dev, 1987. **1**(8): p. 783-93.
86. Froelich-Ammon, S.J., et al., *Modulation of telomerase activity by telomere DNA-binding proteins in Oxytricha.* Genes Dev, 1998. **12**(10): p. 1504-14.
87. Sheng, H., et al., *Identification and characterization of a putative telomere end-binding protein from Tetrahymena thermophila.* Mol Cell Biol, 1995. **15**(3): p. 1144-53.
88. Lin, J.J. and V.A. Zakian, *The Saccharomyces CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13760-5.
89. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**(5519): p. 1171-5.
90. Baumann, P., E. Podell, and T.R. Cech, *Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing.* Mol Cell Biol, 2002. **22**(22): p. 8079-87.
91. Tani, A. and M. Murata, *Alternative splicing of Pot1 (Protection of telomere)-like genes in Arabidopsis thaliana.* Genes Genet Syst, 2005. **80**(1): p. 41-8.
92. Wei, C. and C.M. Price, *Cell cycle localization, dimerization, and binding domain architecture of the telomere protein cPot1.* Mol Cell Biol, 2004. **24**(5): p. 2091-102.
93. Wu, L., et al., *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres.* Cell, 2006. **126**(1): p. 49-62.
94. Hockemeyer, D., et al., *Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres.* Cell, 2006. **126**(1): p. 63-77.
95. Mitton-Fry, R.M., et al., *Conserved structure for single-stranded telomeric DNA recognition.* Science, 2002. **296**(5565): p. 145-7.
96. Theobald, D.L. and D.S. Wuttke, *Prediction of multiple tandem OB-fold domains in telomere end-binding proteins Pot1 and Cdc13.* Structure, 2004. **12**(10): p. 1877-9.
97. Colgin, L.M., et al., *Human POT1 facilitates telomere elongation by telomerase.* Curr Biol, 2003. **13**(11): p. 942-6.
98. Loayza, D. and T. De Lange, *POT1 as a terminal transducer of TRF1 telomere length control.* Nature, 2003. **423**(6943): p. 1013-8.

99. Churikov, D., C. Wei, and C.M. Price, *Vertebrate POT1 restricts G-overhang length and prevents activation of a telomeric DNA damage checkpoint but is dispensable for overhang protection.* Mol Cell Biol, 2006. **26**(18): p. 6971-82.
100. Horvath, M.P., et al., *Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA.* Cell, 1998. **95**(7): p. 963-74.
101. Hughes, T.R., et al., *The Est3 protein is a subunit of yeast telomerase.* Curr Biol, 2000. **10**(13): p. 809-12.
102. Mitton-Fry, R.M., et al., *Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13.* J Mol Biol, 2004. **338**(2): p. 241-55.
103. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.
104. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nat Struct Mol Biol, 2004. **11**(12): p. 1223-9.
105. Murzin, A.G., *OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.* EMBO J, 1993. **12**(3): p. 861-7.
106. Mitton-Fry, R.M. and D.S. Wuttke, *1H, 13C and 15N resonance assignments of the DNA-binding domain of the essential protein Cdc13 complexed with single-stranded telomeric DNA.* J Biomol NMR, 2002. **22**(4): p. 379-80.
107. Hughes, T.R., et al., *Identification of the single-strand telomeric DNA binding domain of the Saccharomyces cerevisiae Cdc13 protein.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6457-62.
108. Horvath, M.P., *Structural anatomy of telomere OB proteins.* Crit Rev Biochem Mol Biol. **46**(5): p. 409-35.
109. Longtine, M.S., et al., *A yeast telomere binding activity binds to two related telomere sequence motifs and is indistinguishable from RAP1.* Curr Genet, 1989. **16**(4): p. 225-39.
110. Chong, L., et al., *A human telomeric protein.* Science, 1995. **270**(5242): p. 1663-7.
111. Bilaud, T., et al., *Telomeric localization of TRF2, a novel human telobox protein.* Nat Genet, 1997. **17**(2): p. 236-9.
112. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. **17**(2): p. 231-5.
113. Broccoli, D., et al., *Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains.* Hum Mol Genet, 1997. **6**(1): p. 69-76.
114. Cooper, J.P., et al., *Regulation of telomere length and function by a Myb-domain protein in fission yeast.* Nature, 1997. **385**(6618): p. 744-7.
115. Friedman, A.D., *Runx1, c-Myb, and C/EBPalpha couple differentiation to proliferation or growth arrest during hematopoiesis.* J Cell Biochem, 2002. **86**(4): p. 624-9.
116. Konig, P., et al., *The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA.* Cell, 1996. **85**(1): p. 125-36.

117. Ogata, K., et al., *Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices.* Cell, 1994. **79**(4): p. 639-48.
118. Court, R., et al., *How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures.* EMBO Rep, 2005. **6**(1): p. 39-45.
119. Bilaud, T., et al., *The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human.* Nucleic Acids Res, 1996. **24**(7): p. 1294-303.
120. Vassetzky, N.S., et al., *Taz1p and Teb1p, two telobox proteins in Schizosaccharomyces pombe, recognize different telomere-related DNA sequences.* Nucleic Acids Res, 1999. **27**(24): p. 4687-94.
121. Shore, D., *RAP1: a protean regulator in yeast.* Trends Genet, 1994. **10**(11): p. 408-12.
122. Bianchi, A., et al., *TRF1 binds a bipartite telomeric site with extreme spatial flexibility.* EMBO J, 1999. **18**(20): p. 5735-44.
123. Fairall, L., et al., *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Mol Cell, 2001. **8**(2): p. 351-61.
124. Liu, D., et al., *Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins.* J Biol Chem, 2004. **279**(49): p. 51338-42.
125. O'Connor, M.S., et al., *The human Rap1 protein complex and modulation of telomere length.* J Biol Chem, 2004. **279**(27): p. 28585-91.
126. Ye, J.Z., et al., *TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres.* J Biol Chem, 2004. **279**(45): p. 47264-71.
127. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-83.
128. Griffith, J.D., et al., *Mammalian telomeres end in a large duplex loop.* Cell, 1999. **97**(4): p. 503-14.
129. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3′ telomeric overhang.* EMBO J, 2001. **20**(19): p. 5532-40.
130. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
131. de Lange, T., *How telomeres solve the end-protection problem.* Science, 2009. **326**(5955): p. 948-52.
132. Hug, N. and J. Lingner, *Telomere length homeostasis.* Chromosoma, 2006. **115**(6): p. 413-25.
133. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. **19**(18): p. 2100-10.
134. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.
135. Bianchi, A. and D. Shore, *How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex.* Mol Cell, 2008. **31**(2): p. 153-65.
136. Fitzpatrick, D.A., et al., *A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis.* BMC Evol Biol, 2006. **6**: p. 99.

137. Lue, N.F., *Plasticity of telomere maintenance mechanisms in yeast.* Trends Biochem Sci, 2010. **35**(1): p. 8-17.
138. McEachern, M.J. and J.B. Hicks, *Unusually large telomeric repeats in the yeast Candida albicans.* Mol Cell Biol, 1993. **13**(1): p. 551-60.
139. McEachern, M.J. and E.H. Blackburn, *A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3453-7.
140. Cohn, M., M.J. McEachern, and E.H. Blackburn, *Telomeric sequence diversity within the genus Saccharomyces.* Curr Genet, 1998. **33**(2): p. 83-91.
141. Konig, P. and D. Rhodes, *Recognition of telomeric DNA.* Trends Biochem Sci, 1997. **22**(2): p. 43-7.
142. Sun, J., et al., *Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres.* Genes Dev, 2009. **23**(24): p. 2900-14.
143. Gelinas, A.D., et al., *Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19298-303.
144. Linger, B.R. and C.M. Price, *Conservation of telomere protein complexes: shuffling through evolution.* Crit Rev Biochem Mol Biol, 2009. **44**(6): p. 434-46.
145. Wood, J.S. and L.H. Hartwell, *A dependent pathway of gene functions leading to chromosome segregation in Saccharomyces cerevisiae.* J Cell Biol, 1982. **94**(3): p. 718-26.
146. Weinert, T.A. and L.H. Hartwell, *Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint.* Genetics, 1993. **134**(1): p. 63-80.
147. Garvik, B., M. Carson, and L. Hartwell, *Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint.* Mol Cell Biol, 1995. **15**(11): p. 6128-38.
148. Krauss, D., et al., *Diastolic intraventricular pressure gradient in symmetric left ventricular hypertrophy.* Am Heart J, 1995. **130**(1): p. 178-80.
149. Booth, C., et al., *Quantitative amplification of single-stranded DNA (QAOS) demonstrates that cdc13-1 mutants generate ssDNA in a telomere to centromere direction.* Nucleic Acids Res, 2001. **29**(21): p. 4414-22.
150. Grandin, N., C. Damon, and M. Charbonneau, *Cdc13 prevents telomere uncapping and Rad50-dependent homologous recombination.* EMBO J, 2001. **20**(21): p. 6127-39.
151. Grandin, N., S.I. Reed, and M. Charbonneau, *Stn1, a new Saccharomyces cerevisiae protein, is implicated in telomere size regulation in association with Cdc13.* Genes Dev, 1997. **11**(4): p. 512-27.
152. Grandin, N., C. Damon, and M. Charbonneau, *Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13.* EMBO J, 2001. **20**(5): p. 1173-83.
153. Petreaca, R.C., et al., *Chromosome end protection plasticity revealed by Stn1p and Ten1p bypass of Cdc13p.* Nat Cell Biol, 2006. **8**(7): p. 748-55.
154. Pennock, E., K. Buckley, and V. Lundblad, *Cdc13 delivers separate complexes to the telomere for end protection and replication.* Cell, 2001. **104**(3): p. 387-96.
155. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.

156. Petreaca, R.C., H.C. Chiu, and C.I. Nugent, *The role of Stn1p in Saccharomyces cerevisiae telomere capping can be separated from its interaction with Cdc13p.* Genetics, 2007. **177**(3): p. 1459-74.
157. Sun, J., et al., *Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase alpha.* Cell Res, 2010. **21**(2): p. 258-74.
158. Mitchell, M.T., et al., *Cdc13 N-terminal dimerization, DNA binding, and telomere length regulation.* Mol Cell Biol. **30**(22): p. 5325-34.
159. Qi, H. and V.A. Zakian, *The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein.* Genes Dev, 2000. **14**(14): p. 1777-88.
160. Hsu, C.L., et al., *Interaction of Saccharomyces Cdc13p with Pol1p, Imp4p, Sir4p and Zds2p is involved in telomere replication, telomere maintenance and cell growth control.* Nucleic Acids Res, 2004. **32**(2): p. 511-21.
161. Yu, E.Y., et al., *Analyses of Candida Cdc13 orthologues revealed a novel OB fold dimer arrangement, dimerization-assisted DNA-binding, and substantial structural differences between Cdc13 and RPA70.* Mol Cell Biol, 2011.
162. Evans, S.K. and V. Lundblad, *Positive and negative regulation of telomerase access to the telomere.* J Cell Sci, 2000. **113 Pt 19**: p. 3357-64.
163. Chandra, A., et al., *Cdc13 both positively and negatively regulates telomere replication.* Genes Dev, 2001. **15**(4): p. 404-14.
164. Lundblad, V. and J.W. Szostak, *A mutant with a defect in telomere elongation leads to senescence in yeast.* Cell, 1989. **57**(4): p. 633-43.
165. Cohn, M. and E.H. Blackburn, *Telomerase in yeast.* Science, 1995. **269**(5222): p. 396-400.
166. Singh, S.M., et al., *Analysis of telomerase in Candida albicans: potential role in telomere end protection.* Eukaryot Cell, 2002. **1**(6): p. 967-77.
167. Taggart, A.K., S.C. Teng, and V.A. Zakian, *Est1p as a cell cycle-regulated activator of telomere-bound telomerase.* Science, 2002. **297**(5583): p. 1023-6.
168. Wu, Y. and V.A. Zakian, *The telomeric Cdc13 protein interacts directly with the telomerase subunit Est1 to bring it to telomeric DNA ends in vitro.* Proc Natl Acad Sci U S A, 2011.
169. Askree, S.H., et al., *A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length.* Proc Natl Acad Sci U S A, 2004. **101**(23): p. 8658-63.
170. Dunham, M.A., et al., *Telomere maintenance by recombination in human cells.* Nat Genet, 2000. **26**(4): p. 447-50.
171. Kitada, T., et al., *Telomere shortening in chronic liver diseases.* Biochem Biophys Res Commun, 1995. **211**(1): p. 33-9.
172. Wiemann, S.U., et al., *Hepatocyte telomere shortening and senescence are general markers of human liver cirrhosis.* FASEB J, 2002. **16**(9): p. 935-42.
173. De Marzo, A.M., et al., *Pathological and molecular mechanisms of prostate carcinogenesis: implications for diagnosis, detection, prevention, and treatment.* J Cell Biochem, 2004. **91**(3): p. 459-77.

174. Meeker, A.K. and P. Argani, *Telomere shortening occurs early during breast tumorigenesis: a cause of chromosome destabilization underlying malignant transformation?* J Mammary Gland Biol Neoplasia, 2004. **9**(3): p. 285-96.
175. Wu, X., et al., *Telomere dysfunction: a potential cancer predisposition factor.* J Natl Cancer Inst, 2003. **95**(16): p. 1211-8.
176. Jiang, X.R., et al., *Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype.* Nat Genet, 1999. **21**(1): p. 111-4.
177. Morales, C.P., et al., *Absence of cancer-associated changes in human fibroblasts immortalized with telomerase.* Nat Genet, 1999. **21**(1): p. 115-8.
178. Minev, B., et al., *Cytotoxic T cell immunity against telomerase reverse transcriptase in humans.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4796-801.
179. Nair, S.K., et al., *Induction of cytotoxic T cell responses and tumor immunity against unrelated tumors using telomerase reverse transcriptase RNA transfected dendritic cells.* Nat Med, 2000. **6**(9): p. 1011-7.
180. Vonderheide, R.H., *Telomerase as a universal tumor-associated antigen for cancer immunotherapy.* Oncogene, 2002. **21**(4): p. 674-9.
181. Vonderheide, R.H., et al., *Equivalent induction of telomerase-specific cytotoxic T lymphocytes from tumor-bearing patients and healthy individuals.* Cancer Res, 2001. **61**(23): p. 8366-70.
182. Norton, J.C., et al., *Inhibition of human telomerase activity by peptide nucleic acids.* Nat Biotechnol, 1996. **14**(5): p. 615-9.
183. Mavroudis, D., et al., *A phase I study of the optimized cryptic peptide TERT(572y) in patients with advanced malignancies.* Oncology, 2006. **70**(4): p. 306-14.
184. Holt, S.E., et al., *Functional requirement of p23 and Hsp90 in telomerase complexes.* Genes Dev, 1999. **13**(7): p. 817-26.
185. Yokoyama, Y., et al., *Attenuation of telomerase activity by a hammerhead ribozyme targeting the template region of telomerase RNA in endometrial carcinoma cells.* Cancer Res, 1998. **58**(23): p. 5406-10.
186. Kim, M.M., et al., *A low threshold level of expression of mutant-template telomerase RNA inhibits human tumor cell proliferation.* Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7982-7.
187. Strahl, C. and E.H. Blackburn, *Effects of reverse transcriptase inhibitors on telomere length and telomerase activity in two immortalized human cell lines.* Mol Cell Biol, 1996. **16**(1): p. 53-65.
188. Murakami, J., N. Nagai, and K. Ohama, *[Telomerase activity and telomere length as diagnostic tumor marker for ovarian tumors]*. Nihon Rinsho, 1998. **56**(5): p. 1310-5.
189. Shay, J.W. and W.E. Wright, *Telomerase activity in human cancer.* Curr Opin Oncol, 1996. **8**(1): p. 66-71.
190. Forsyth, N.R., W.E. Wright, and J.W. Shay, *Telomerase and differentiation in multicellular organisms: turn it off, turn it on, and turn it off again.* Differentiation, 2002. **69**(4-5): p. 188-97.
191. White, L.K., W.E. Wright, and J.W. Shay, *Telomerase inhibitors.* Trends Biotechnol, 2001. **19**(3): p. 114-20.

192. Elmore, L.W., et al., *Real-time quantitative analysis of telomerase activity in breast tumor specimens using a highly specific and sensitive fluorescent-based assay.* Diagn Mol Pathol, 2002. **11**(3): p. 177-85.
193. Granger, M.P., W.E. Wright, and J.W. Shay, *Telomerase in cancer and aging.* Crit Rev Oncol Hematol, 2002. **41**(1): p. 29-40.
194. Lundblad, V. and E.H. Blackburn, *An alternative pathway for yeast telomere maintenance rescues est1- senescence.* Cell, 1993. **73**(2): p. 347-60.
195. Bryan, T.M., et al., *The telomere lengthening mechanism in telomerase-negative immortal human cells does not involve the telomerase RNA subunit.* Hum Mol Genet, 1997. **6**(6): p. 921-6.
196. Henson, J.D., et al., *Alternative lengthening of telomeres in mammalian cells.* Oncogene, 2002. **21**(4): p. 598-610.
197. Forstemann, K. and J. Lingner, *Telomerase limits the extent of base pairing between template RNA and telomeric DNA.* EMBO Rep, 2005. **6**(4): p. 361-6.
198. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase.* Annu Rev Biochem, 2006. **75**: p. 493-517.

CHAPTER 2

STN1-TEN1 IS AN RPA32-RPA14-LIKE COMPLEX AT TELOMERES

2.1 Attributions

This chapter contains the manuscript "Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres" by J. Sun, E.Y. Yu, Y. Yang, L.A. Confer, S.H. Sun, K. Wan, N.F. Lue, and M. Lei published in *Genes & Development* (2009) 23: 2900-2914. Constructs were designed by J. Sun and M. Lei. Mutagenesis was performed by J. Sun and K. Wan. Protein expression, purification and crystallization ware performed by J. Sun. X-ray data collection and structure determination were done by Y. Yang and J. Sun. *In vivo* yeast telomere assays were performed by E.Y. Yu. The manuscript was written by M. Lei, J. Sun and N.F. Lue.

2.2 Abstract

In budding yeast, Cdc13, Stn1, and Ten1 form a heterotrimeric complex (CST) that is essential for telomere protection and maintenance. Previous bioinformatics analysis revealed a putative OB fold at the N terminus of Stn1 (Stn1N) that shows limited sequence similarity to the OB fold of Rpa32, a subunit of the eukaryotic ssDNA-binding protein complex replication protein A (RPA). Here I present functional and structural analyses of Stn1 and Ten1 from multiple budding and fission yeast. The crystal structure

of the *Candida tropicalis* Stn1N complexed with Ten1 demonstrates an Rpa32N–Rpa14-like complex. In both structures, the OB folds of the two components pack against each other through interactions between two C-terminal helices. The structure of the C-terminal domain of *Saccharomyces cerevisiae* Stn1 (Stn1C) was found to comprise two related winged helix–turn–helix (WH) motifs, one of which is most similar to the WH motif at the C terminus of Rpa32, again supporting the notion that Stn1 resembles Rpa32. The crystal structure of the fission yeast *Schizosaccharomyces pombe* Stn1N–Ten1 complex exhibits a virtually identical architecture as the *C. tropicalis* Stn1N–Ten1. Functional analyses of the *Candida albicans* Stn1 and Ten1 proteins revealed critical roles for these proteins in suppressing aberrant telomerase and recombination activities at telomeres. Mutations that disrupt the Stn1–Ten1 interaction induce telomere uncapping and abolish the telomere localization of Ten1. Collectively, the structural and functional studies illustrate that, instead of being confined to budding yeast telomeres, the CST complex may represent an evolutionarily conserved RPA-like telomeric complex at the 3' overhangs that works in parallel with or instead of the well-characterized POT1–TPP1/TEBPα–β complex.

2.3 Introduction

Telomeres, the specialized nucleoprotein structures located at linear eukaryotic chromosomal termini, are essential for chromosome stability and are maintained by the special reverse transcriptase named telomerase [1-3]. Telomeric DNAs are typically repetitive in nature and terminate in 3' overhangs (G-tails) that are bound by distinct protein complexes in different organisms. In ciliated protozoa, a dimeric protein complex

(TEBPα and TEBPβ) is responsible for G-tail recognition and protection [4]. In fission yeast and humans, the TEBPα homologue POT1 provides the major G-tail binding activity and associates with the respective TEBPβ homologue (Tpz1 in *S. pombe* and TPP1 in humans) [5-8]. Interestingly, the G-tails of budding yeast telomeres are apparently protected by an altogether distinct, nonhomologous complex named CST (Cdc13-Stn1-Ten1) [9-12]. Nevertheless, all of these proteins appear to contain one or more OB folds, testifying to the versatility of this domain in single-strand nucleic acid recognition [13]. While many of the G-tail interacting proteins are essential for cell viability, hypomorphic alleles of genes encoding these proteins have been shown to induce a variety of telomere aberration, including catastrophic telomere loss, uncontrolled telomere elongation, telomere C-strand degradation, and telomere fusions, thus underscoring their fundamental importance in telomere protection [3, 14].

Initially, components of the CST complex were thought to be unique to budding yeast, and in particular to organisms without POT1 homologues. In other words, the POT-TPP1 and CST complex are postulated to represent two alternative means of G-tail protection. However, recent studies have uncovered Stn1 and Ten1 homologues in a multitude of POT1-containing organisms, and implicated the *S. pombe* Stn1 and Ten1 as well as *A. thaliana* Stn1 in telomere capping [15, 16]. Moreover, the *S. pombe* Stn1 and Ten1 proteins exhibit no evident interaction with Pot1, suggesting that they can function independently of the major G-tail binding activity [15]. Indeed, the *S. cerevisiae* Stn1 and Ten1, when over-expressed, are known to be capable of Cdc13-independent protection of telomeres [12]. Even though Stn1 or Ten1 alone apparently recognizes telomere G-tails with low affinity, available evidence suggests that they can be recruited to telomeres

through an interaction with Pol12 (a subunit of Polα) [12, 17]. Altogether, these observations hint at a far more prevalent role for Stn1 and Ten1, possibly as components of an alternative telomere end protective complex that functions in parallel to the POT1-containing complex.

Recent bioinformatic analysis points to potential structural similarities between Stn1 and Rpa32, as well as between Ten1 and Rpa14 [11]. The validity of the Stn1-Rpa32 analogy was supported by a domain swapping experiment, in which the N-terminal OB-fold-like domain of Stn1 was shown to function in place of the RPA32 OB fold. In addition, similar to Rpa32 and Rpa14, the N-terminus of Stn1 interacts with Ten1 *in vitro* and *in vivo* [11, 12]. Both Rpa32 and Rpa14 are subunits of a trimeric, non-specific single strand DNA binding complex (RPA) that mediates critical and diverse DNA transactions throughout the genome [13, 18]. Their potential similarities to Stn1 and Ten1 thus raise the intriguing possibility that the CST complex represents a chromosome locus-specific RPA complex. While highly provocative, this hypothesis awaits experimental confirmation. In addition, many questions with regard to the structure, function, and conservation of the CST complex remain unresolved. In this chapter, I provide structural and functional analyses of the Stn1 and Ten1 protein from multiple budding and fission yeast. My atomic resolution structures of several complexes and a protein domain provide direct confirmation of structural similarity between components of the CST and the RPA complexes, and reveal a detailed molecular view of the Stn1-Ten1 interaction interface. My functional studies and results from my collaborator, Dr Neal Lue of Weill Medical College of Cornell University, underscore the importance of Stn1-Ten1 interaction in telomere protection, and reveal critical functions

for these proteins in suppressing aberrant telomerase and recombination activities at telomeres.

2.4 Identification of the CST Complex Genes in Budding Yeast *Candida* and *Saccharomyces* Genomes

The branches of budding yeast exemplified by *Candida albicans* have apparently undergone rapid evolutionary divergence with respect with its telomere sequence and telomere related proteins [19, 20]. For instance, unlike *S. cerevisiae*, many *Candida spp.* have long (up to 25 base-pair [bp]), distinct and regular telomere repeat units. Moreover, the putative telomere maintenance proteins of *Candida spp*. (e.g., Rap1) have been observed to exhibit significant structural divergence from their *Saccharomyces* counterparts [21]. Indeed, until recently, homologues of the CST complex were difficult to identify in these genomes, raising interesting questions concerning their telomere protection mechanisms [20].

To initiate a comparative analysis of telomere end protection mechanisms in this unusual group of budding yeast, we systemically searched the NCBI and Broad Institute databases for homologues of Cdc13, Stn1, and Ten1 using available sequences as queries. This exercise resulted in the identification plausible homologues of each CST component in all completely sequenced *Candida* and *Saccharomyces* genomes (Figure 2.1). In keeping with the theme of rapid evolutionary divergence, we found that many Cdc13 homologues in *Candida spps*. are considerably smaller and evidently lack the N-terminal half of their *S. cerevisiae* counterpart, thus partly accounting for the prior difficulties in their detection. To ascertain the functions of these homologues in telomere regulation,

we attempted to generate *C. albicans* strains that are null for *CDC13*, *STN1* or *TEN1* by sequential deletion of the two alleles [22, 23]. Perhaps not surprisingly, we were unable to generate a *cdc13* null strain, suggesting that this gene, like its *S. cerevisiae* homologue, is essential for cell viability [9]. In contrast, we were able to obtain multiple isolates of *stn1* and *ten1* null strains, indicating that these genes are not essential in *C. albicans*. The availability of the null strains allowed us to investigate in detail the functions and mechanisms of Stn1 and Ten1 in *C. albicans*.

2.5 *C. albicans* Stn1 and Ten1 are Important for Telomere Maintenance

Both the *stn1* and *ten1* null mutant grow more slowly than the parental *BMP17* strain (Figure 2.2A). Microscopic examination revealed an abundance of filamentous cells in liquid cultures; quantitation indicated a ~20 fold increase in the percentage of such cells (data not shown). Though the reasons for this aberrant growth morphology are not understood, similar aberrations have been described for other *C. albicans* DNA repair mutants, suggesting a shared underlying mechanism [24, 25]. Consistent with a role for Stn1 and Ten1 in telomere regulation, we observed extremely long and heterogeneous telomeres in multiple isolates of both null mutants (Figure 2.2B and data not shown). Long and heterogeneous telomeres were detected at the earliest time point following the derivation of the mutants (~100 generations) and were stably maintained for at least 150 generations thereafter. In contrast to the parental BWP17, whose telomeres range in size from ~1-5 kb, the *stn1* and *ten1* mutants possess extremely long (>20 kb) and short (<1 kb) telomeres, consistent with loss of the homeostatic mechanism that normally regulates telomere length.

The extremely long and heterogeneous telomeres suggest that telomeres in *stn1* and *ten1* are de-protected. Two other frequent consequences of de-protection are the accumulation of G-tails and extra-chromosomal telomeric circles (t-circles), which can be detected by in-gel hybridization and 2D gel electrophoresis, respectively. Interestingly, we found no evidence of G-tail accumulation, but rather high levels of t-circles in the mutants (Figure 2.2C and 1D). Quantitative analysis indicates that ~10% of telomeric hybridization signals in the mutants reside in circular DNAs. In comparison, much less than 1% of the telomeric DNA in the parental BWP17 strain is in circular form. Notably, all of the growth and telomere abnormalities in the *stn1* and *ten1* mutants are suppressed by the re-integration of a wild type copy of the respective genes, confirming that these phenotypes are due to loss of Stn1 and Ten1 (data not shown). We conclude that both the *STN1* and *TEN1* genes in *C. albicans* are necessary for the maintenance of proper telomere length and structure.

Our observations with regard to the function of the Stn1-Ten1 complex in *C. albicans* echo earlier findings in other budding yeast. Specifically, hypomorphic CST mutations have been shown to result in abnormal telomerase and recombination activities at telomeres in both *S. cerevisiae* and *K. lactis* [10, 12, 26, 27]. A point mutant allele of *STN1* in *K. lactis*, in particular, exhibits extremely long and heterogeneous telomeres that are (at least partly) telomerase-independent [26]. The close phenotypic resemblance of this *K. lactis* mutant to the *C. albicans Stn1* and *ten1* mutant argues for a substantial degree of mechanistic conservation in budding yeast. On the other hand, some features of the *C. albicans* systems are clearly unique. For example, both *STN1* and *TEN1* are dispensable for cell viability, allowing the consequences of complete gene deletions to be

analyzed in the absence of other genetic changes that were often necessary to maintain viability of *stn1* or *ten1* mutant in other organisms. Also unusual was our failure to observed G-tail accumulation, which is a frequent consequence of hypomorphic CST mutations in budding yeast. Yet these differences do not necessarily imply fundamentally different mechanisms of telomere protection by the CST in *Candida*. Most prior studies of the CST complex were conducted in haploid yeast, which differs physiologically from the obligate diploid *Candida albicans* employed in our analysis. Similarly, failure to observe G-tails may be due to their transience rather than absence. One can imagine, for instance, that G-tails were generated by C-strand degradation in the *C. albicans stn1* and *ten1* mutant, but were more efficiently repaired by recombination or fill-in synthesis. Further studies will be necessary to determine if the apparent differences between *C. albicans* and other budding yeast reflect some fundamental mechanistic divergence.

In many respects, the phenotypes of the *C. albicans stn1* and *ten1* mutant mimic those of ALT cancer cells, which are also characterized by telomere length heterogeneity, elevation of t-circles, and telomere maintenance through recombination [28, 29]. Thus, our findings suggest that one possible pathway for attaining the ALT status was through de-protection of G-tails. Interestingly, a recent study in *K. lactis* argues that deficiency of Rap1 (the major double strand telomere binding protein in budding yeast) can lead to similar phenotypes [30]. It is tempting to speculate that aberrations in some telomere protein component may be a necessary condition for the activation of the ALT pathway.

2.6 Structure Determination of the *Candida tropicalis* Stn1-Ten1 Complex

Sequence alignment and secondary structure predictions of Stn1 proteins have previously revealed in members of this conserved family a putative N-terminal OB-fold domain that is most similar to the OB-fold of Rpa32 [11]. Notably, the predicted OB-fold of budding yeast *S. cerevisiae* Stn1 can replace the equivalent region of *S. cerevisiae* Rpa32, resulting in a chimeric protein that rescued the lethal phenotype of an *rpa2-Δ* yeast strain [11]. Stn1 interacts with Ten1 both *in vivo* and *in vitro* [11], and sequence analysis supports the existence of an OB fold in Ten1 as well (Figure 2.1). These results led to the hypothesis that Stn1 binds to Ten1 to form an Rpa32-Rpa14-like complex at telomeres [11]. However, there is no detectable sequence similarity between Ten1 and Rpa14 protein families. Furthermore, it is unknown how Stn1 interacts with Ten1 and whether this interaction resembles that between Rpa32 and Rpa14. Thus, validation of the hypothesis that Stn1-Ten1 represents a telomere-specific Rpa32-Rpa14 complex depends on structural characterization of the Stn1-Ten1 complex.

Complexes consisting of Ten1 and the N-terminal domain of Stn1 (Stn1N) from several different budding yeast species including *S. cerevisiae, C. albicans,* and *C. tropicalis* were prepared and used in the crystallization trials (Figure 2.3A). After extensive screening, the *C. tropicalis* Stn1N-Ten1 complex was found to generate crystals suitable for structural determination. The complex was crystallized in space group $P4_12_12$ with two complexes per asymmetric unit (Table 1). The structure was solved by single-wavelength anomalous dispersion (SAD) with mercury (MeHgAc) derivative crystals, and refined to 2.4 Å resolution. The high-quality composite omit electron density map enabled us to fit and refine most of the complex except several N- and C-terminal residues of Stn1N.

2.7 The Stn1N-Ten1 Complex Structure

The Stn1N-Ten1 complex structure reveals a 1:1 stoichiometry between Stn1N and Ten1, consistent with the observed molecular weight of the complex as determined by gel filtration (~37.5 kDa, Figure 2.3). The crystal structure (Figure 2.4A) shows that each protein indeed comprises a single OB fold, consisting of a highly curved five-stranded β-barrel, as expected from previous primary sequence analysis (Figure 2.4B). In addition to the central β-barrel, there are several structural features common to the OB folds of Stn1N and Ten1. First, both proteins contain a C-terminal helix αC, which contributes most of the contact interface between Stn1N and Ten1 (Figure 2.4B). Second, short α helices (αB in Stn1N, and αB' and αB in Ten1) that cover the bottom of the β-barrels of the OB folds are found between strands β3 and β4 (Figure 2.4B). Third, an N-terminal helix αA closes the other end of the β-barrel, and the position of this helix is stabilized by a short strand β0, which interacts with strand β1 in an anti-parallel orientation (Figure 2.4B).

Besides helix αA and strand β0, Stn1N contains a unique segment N-terminal to the core of the OB fold (Figure 2.4B). This segment, which consists of resides 1-45 (located N-terminal to β0), folds into a β hairpin (βA and βB) and a short helix α1 (Figure 2.4B). Another unique feature of Stn1N is the connection between helix αA and the β-barrel, which contains a 27-residue insertion (residues 57-83) that comprises two short helices (α2 and α3) and another short β hairpin (βD and βE). These two extra elements fold together into a unique motif to cap the top of the OB fold of Stn1 (henceforth referred to as the 'cap' motif of Stn1) (Figure 2.4B and Figure 2.5). Notably,

the C-terminal tails following helix αC (residues 204-213) of both Stn1N molecules in the asymmetric unit are well ordered and make hydrophobic contacts with the cap region (Figure 2.5). In particular, the aromatic side chain of W208 is nested in a hydrophobic pocket formed by Y32, L36, F37 and Y80 (Figure 2.5). All these residues are highly conserved in the Stn1 family members (Figure 2.4C). Consistent with this observation, efforts to prepare an Stn1N fragment without the C-terminal tail (residues 2-205) yielded little soluble protein, suggesting that this tail is important for the correct folding of Stn1 (data not shown).

2.8 The structural Conservation between Stn1N-Ten1 and Rpa32N-Rpa14

The crystal structure of Stn1N-Ten1 closely resembles that of the Rpa32N-Rpa14 complex (Figure 2.6A). An unbiased search for structurally homologous proteins using the Dali server [31] revealed that the structure of Stn1 OB fold is most similar to that of the OB fold of Rpa32, consistent with previous sequence alignment predictions (Figure 2.4C) [11]. The two OB folds can be superimposed with a root-mean-square deviation (RMSD) of 2.4 Å for 119 equivalent C_α pairs (Figure 2.6A). Notably, the structurally highly conserved region includes not only the central β-barrel of the OB fold, but also peripheral α helices (αA and αC) and β strands (βD βE and β0) in the N- and C-terminal extension regions, suggesting that Stn1 and RPA32 are structurally homologous proteins (Figure 2.6A). Unlike Stn1 and Rpa32, bioinformatics analysis failed to detect any substantial similarity between Ten1 and Rpa14 (Figure 2.4C). However, comparison of the structures of Ten1 and Rpa14 clearly reveals a high degree of structural similarity (Figure 2.6B). In fact, Rpa14 is one of the top solutions revealed by Dali that are

structurally most similar to Ten1 with an RMSD of 2.8 Å for 96 equivalent C_α atoms. This close structural similarity is rather unexpected given that the sequences of the OB folds of Ten1 and Rpa14 are substantially divergent and share only 7% identity (Figure 2.4C). In addition to similarities between the individual components, the Stn1N-Ten1 and the Rpa32-Rpa14 complexes share another unique feature; in both cases, the two subunits heterodimerize mainly through hydrophobic contacts mediated by the two C-terminal αC helices (Figure 2.6A). Taken together, these findings strongly support the notion that Stn1-Ten1 is structurally similar to and evolutionarily related to the Rpa32-Rpa14 complex.

Notwithstanding the high degree of overall structural conservation, there are substantial differences between the Stn1-Ten1 and the Rpa32-Rpa14 complexes. Most notably, the relative orientations between the two components are different in the two complexes. When both complex structures are overlaid based on the OB folds of Stn1 and Rpa32, Ten1 has a ~ 15° rotation relative to the position of Rpa14 (Figure 2.6A). Second, compared to Rpa32, Stn1 contains an extra N-terminal extension (βA, βB and α1) and a 12-residue insertion before strand βD (α2 and α3) (Figure 2.6C). Additionally, significant sequence and structural variances are evident in most of the connecting loop regions. For example, Stn1 has a long loop (12 residues) L_{45} between strands β4 and β5, which packs on helix α2 in the N-terminal cap motif (Figure 2.5). In contrast, strands β4 and β5 of Rpa32 are connected by a short two-residue turn.

These structural differences provide a plausible explanation for the published findings on domain exchange between Stn1 and Rpa32 [11]. As noted before, the specific interactions between Stn1 and Ten1 and between Rpa32 and Rpa14 primarily involve the

hydrophobic contacts between the two αC helices C-terminal to the OB-folds (Figure 2.4B and 6A, [32]). Thus, the chimeric Rpa32-OB^{Stn1} protein, which carries the OB-fold of Stn1 in place of the Rpa32 OB fold, and which still contains helix αC of Rpa32 retains the ability to bind Rpa14 and rescue the inviability of an *rpa2-Δ* yeast strain [11]. In contrast, due to the incompatibility between the two αC helices of Stn1 and Rpa14, the *rpa32-Δ* strain could not be rescued by high level expression of Stn1 [11] . For the same reason, the chimeric Rpa32-OB^{Stn1} protein could not interact with Ten1 to rescue a *stn1-Δ* strain [11]. Furthermore, the N-terminal cap motif of Stn1 (βA, βB and αA) is expected to collide with strands βD and βE of Rpa32 if the OB-fold of Stn1 is replaced with that of Rpa32 (Figure 2.6C). Hence, the chimeric Stn1-OB^{Rpa32} is unlikely to fold into a stable and functional protein, explaining the failure of Stn1-OB^{RPA32} to rescue the *stn1-Δ* mutant [11].

2.9 The Stn1N-Ten1 Interaction

The interface between Stn1N and Ten1 in the crystal structure is relatively flat and hydrophobic (Figure 2.7A). The interactions are mediated primarily by the amphipathic αC helices of both proteins and one side of the Ten1 β-barrel (Figure 2.6A), burying 1060 and 1128 Å^2 of solvent-accessible surface on Stn1N and Ten1, respectively. The angle between the axes of the two αC helices of Stn1 and Ten1 is ~ 60°. As a consequence, only the crossover regions of the helices make extensive contacts with each other; hydrophobic residues from Stn1 (F190, W193, and M197) and Ten1 (L111 and M115) interdigitate with one another to form the core of the hydrophobic interface (Figure 2.6C). At the N-terminal end of the αC helix of Stn1, the side chains of Stn1

L186 is positioned in a hydrophobic pocket of Ten1 formed by residues from helix αC, loop L_{5C} (between β5 and αC), and strands β0, β1, and β4 (Figure 2.7B). The β-barrel of Stn1 makes much less direct contact with Ten1 and contributes only one hydrogen-bonding interaction between Stn1 K90 and Ten1 Y97(Figure 2.7A and D).

In addition to hydrophobic contacts, hydrogen-bonding interactions appear also to strengthen the interface and contribute to the specificity of the Stn1-Ten1 complex. There are six intermolecular hydrogen bonds at the Stn1-Ten1 interface, all located at the periphery. Specifically, at the N-terminal end of the Stn1 αC helix, the carboxylate side chain of E189 makes two salt bridge interactions with the amino group of R27 in the Ten1 β1 strand (Figure 2.7D). The R27 side chain also makes two intra-molecular hydrogen-bonding interactions with D83 and Y97 of Ten1 (Figure 2.7D). Moreover, the side chain amino group of K90 of Stn1 donates another hydrogen bond to Y97 of Ten1 (Figure 2.7D). Together, this elaborate electrostatic interaction network extends the contact interface area and helps to stabilize the relative orientation of Stn1 and Ten1 in the complex. Notably, both E189 of Stn1 and R27 of Ten1 are highly conserved in both families of proteins (Figure 2.4C), consistent with their important roles in Stn1-Ten1 complex formation as revealed by the crystal structure.

To corroborate my structural analysis, I examined whether missense mutations on the interface residues of Stn1 and Ten1 could weaken or disrupt the Stn1-Ten1 interaction using yeast two-hybrid assay. Consistent with the crystal structure, I found that substitution of a hydrophobic residue (Leu186, Phe190, or Trp193) of Stn1 on the interface with alanine was sufficient to abolish its interaction with Ten1 (Figure 2.7E). Similarly, Ten1 mutation M115Ala on the other side of the interface also impaired the

interaction (Figure 2.7E). In contrast, Ala substitutions of Met197 of Stn1 and L111 of Ten1 on the interface still maintained the interaction (~ 30-40% of the wild type level), suggesting that the side chains of these two residues are not crucial for the Stn1-Ten1 complex formation (Figure 2.7E). Notably, disruption of the electrostatic interactions between E189 of Stn1 and R27 of Ten1 by alanine substitution of either residue was sufficient to abolish the Stn1-Ten1 interaction (Figure 2.7E). Collectively, we conclude that both the hydrophobic and the electrostatic contacts observed in the crystal structure are necessary for the interaction between Stn1 and Ten1.

2.10 Functional analysis of the Stn1N-Ten1 interaction

To assess the *in vivo* roles of the Stn1-Ten1 interaction in telomere regulation, we introduced several site-specific mutations in *C. albicans STN1* and *TEN1* designed to disrupt their contact interface based on the *C. tropicalis* Stn1N-Ten1 complex structure, and analyzed the phenotypes of the resulting mutants. To facilitate biochemical and genetic studies, each mutant allele was fused at its C-terminus to a GSCP (Gly_6-SBP-CBP-protein A) tag, which had little effect on the function of the wild type gene in telomere regulation (Figure 2.8A [cf. lanes 4–6 and 7–9], B [cf. lanes 4–6 and 7–12]). All three *ten1* mutant (R27A, I115A, and L119A (equivalent to *C. tropicalis* Ten1 R27A, L111A, and M115A)) as well as two of the *stn1* mutant (F208A and M212A ([equivalent to *C. tropicalis* Stn1 L186A, and F190]) proteins were expressed at near wild type levels, suggesting that in general, residues at the Stn1-Ten1 interface are not required for protein stability. The only exception was Stn1-E211A (equivalent to *C. tropicalis* Stn1 E189A), which was detected at ~ 20 % of the wild type level (Figure 2.8A). On the other hand,

most mutants except Stn1-M212A exhibited significant loss of function with regard to telomere length regulation (Figure 2.8B). Indeed, three of these mutants (*ten1-R27A*, *ten1-L119A*, and *stn1-E211A*) manifested phenotypes that were severe as the respective null mutant. Thus, the interaction between Stn1 and Ten1 are evidently critical for telomere length regulation.

Next, the effects of mutations on the telomere association of Ten1-GSCP were assessed using ChIP by our collaborator, Dr. Neal Lue's group (data not shown). Their findings reinforced the notion that the interaction between Stn1 and Ten1 is necessary both for Ten1 recruitment and telomere regulation. Further studies will be necessary to determine how the interactions between Stn1 and Ten1 influence the localization of Stn1.

2.11 Structural Conservation between the C-terminal Domains of Stn1 and RPA32

Besides the N-terminal OB fold, sequence alignment revealed another conserved domain at the C-terminus of Stn1 (henceforth referred to as Stn1C) (Figure 2.1 and 2.2A). Stn1C interacts with both Cdc13 and the B subunit of the DNA polymerase α-primase complex, Pol12 [17, 27]. Notably, the C-terminal region of Rpa32 is also known to be a globular domain that contains the winged helix-turn-helix (WH) motif [33]. This motif is composed of three α helices flanked by three β strands [33]. $Rpa32_{WH}$ interacts with a myriad of protein factors essential for DNA replication, recombination and repair [34]. Based on the observation that both Stn1C and $RPA32_{WH}$ are located at the C-termini and both mediate protein-protein interactions, we hypothesized that Stn1C might adopt an $RPA32_{WH}$-like WH fold conformation. However, no obvious sequence similarity could be detected between Stn1C and $RPA32_{WH}$ (Figure 2.10E). In addition, the size of Stn1C

(~200 amino acids) is almost three times that of $RPA32_{WH}$ (~70 amino acids) (Figure 2.4A) [33]. Thus, it is unclear whether the structural similarity between Stn1 and RPA32 could be extended to their C-terminal regions. To address this question, various Stn1C constructs from *S. cerevisiae, C. albicans,* and *C. tropicalis*, were expressed and purified for structural characterization. After optimization by limited proteolysis and mass spectrometry analysis, I succeeded in purifying and crystallizing *S. cerevisiae* Stn1C (residues 311 – 494) (Figure 2.9A and 2.9B) and determining its structure by multiple-wavelength anomalous dispersion (MAD) at a resolution of 2.1 Å using selenomethionine-containing crystals (Figure 2.10A; Table 2). The calculated electron density map allowed unambiguous tracing of most of Stn1C except a disordered loop (residues 472 – 479).

Unexpectedly, Stn1C is composed of two topologically similar WH motifs that are related to each other by a pseudo-dyad, although no such similarity was expected from its primary sequence (Figure 2.10A). Notably, the folding of the first WH motif, $Stn1_{WH1}$, is indeed structurally similar to $RPA32_{WH}$ (Figure 2.10B). The rmsd between the two WH motifs is 1.8 Å for 58 C_α atom pairs (Figure 2.10B). One unique feature of $Stn1_{WH1}$ is a large insertion (a 17-residue α2' helix and an eight-residue $L_{2'3}$ loop) between helices α2 and α3 (Figure 2.10B). In contrast, α2 and α3 of $Rpa32_{WH}$ are connected by a short five-residue loop (Figure 2.10B). This marked local variance explains the failure to detect the similarity between the WH motifs of Stn1 and Rpa32 by bioinformatics analysis. Nevertheless, the striking structural similarity between $Stn1_{WH1}$ and $Rpa32_{WH}$ further support the notion that Stn1 is an RPA32-like telomeric protein.

Other than sharing a similar topology, the structure of $Stn1_{WH2}$ is rather different from that of $Stn1_{WH1}$ (Figure 2.10A). $Stn1_{WH2}$ is most similar to the DNA-binding WH motifs of the pur operon repressor [35] and RepE replication initiator [36]. Nevertheless, comparison of the crystal structures of Stn1C and the RepE-DNA complex indicates that $Stn1_{WH2}$ would be unlikely to bind DNA due to the occlusion of its putative targets site by $Stn1_{WH1}$ (compare Figure 2.10A and C). This is further confirmed by an Electrophoretic Mobility Shift Assay (EMSA), in which Stn1C failed to exhibit binding to double stranded telomeric DNAs even at a very high protein concentration (100 μM) (data not shown).

Several features of Stn1 appear to fix the relative orientation between the WH1 and WH2 motif and allow Stn1C to adopt a compact and globular structure resembling a single folded unit. First, the N-terminus of $Stn1_{WH2}$ is immediately adjacent to the end of $Stn1_{WH1}$; there is no linker residue between β3 of WH1 and α1 of WH2 (Figure 2.10A). In addition, Leu401 and Phe405 in WH2 make contacts with a hydrophobic surface formed by the WH1 helix α1 (Figure 2.10D). Finally, the side chain of Leu398 in WH2 inserts into a deep hydrophobic pocket of WH1, further stabilizing the relative disposition of the two motifs (Figure 2.10D). The twisted architecture of Stn1C gives rise to a large surface area for potential interactions with other proteins such as Cdc13 and Pol12, as suggested by earlier genetic studies [17, 27].

2.12 Crystal structure of fission yeast *S. pombe* Stn1N-Ten1 complex

The budding yeast CST complex has long been considered an evolutionary exception, as most other eukaryotic organisms use the POT1–TPP1 or a POT1–TPP1-like complex to

bind G tails and protect telomeres [6, 37-42]. Recent studies have challenged this view. Putative Stn1 and Ten1 orthologs have been identified in a plethora of organisms ranging from fission yeast and plants to humans [15, 16, 43-45]. This suggests that the CST complex may be another conserved complex at the telomere G tails besides the well-characterized POT1–TPP1 complex. However, the sequences of the *S. pombe* Stn1 and Ten1 proteins are only weakly similar to those of the budding yeast proteins [15]. Thus, whether *Sp*Stn1 and *Sp*Ten1 represent true homologs of the budding yeast proteins is unclear. To resolve this question, we reconstituted, purified and crystallized the complex between full-length *Sp*Ten1 and the N-terminal putative OB fold of *Sp*Stn1 (*Sp*Stn1N, residues 2–186) (Figure 2.11) and determined its structure at 1.65 Å resolution by SAD method using Se-Met substituted proteins (Table 3).

The crystal structure of the *Sp*Stn1N–*Sp*Ten1 complex reveals that both *Sp*Stn1N and *Sp*Ten1 are indeed made of an OB fold, and the complex adopts a three-dimensional architecture similar to the *C. tropicalis* Stn1N–Ten1 complex (Figure 2.12A). The OB folds are closely conserved, with a Cα RMSD value of 2.2 Å between the OB folds of *Sp*Stn1N and *Ct*Stn1N and 2.0Å between the OB folds of *Sp*Ten1 and *Ct*Ten1. Given *C. tropicalis* Stn1–Ten1 is an Rpa32–Rpa14-like complex, it is not unexpected that the structure of the *Sp*Stn1–*Sp*Ten1 complex also closely resembles that of Rpa32N–Rpa14. In fact, *Sp*Stn1N is structurally more similar to Rpa32N than to *Ct*Stn1; the Cα RMSD is only 1.6 Å between *Sp*Stn1N and Rpa32N.

The interface between *Sp*Stn1N and *Sp*Ten1 involves both hydrophobic and electrostatic interactions (Figure 2.12B). Compared with the *C. tropicalis* Stn1N–Ten1 complex, the most conserved feature is the hydrophobic packing between the two αC

helices of *Sp*Ten1N and *Sp*Ten1, which appears to be the major driving force for complex formation (Figure 2.12B). Unlike the *C. tropicalis* Stn1N–Ten1, electrostatic interactions contribute more to the *Sp*Stn1N–*Sp*Ten1 interface. There are a total of nine intermolecular electrostatic interactions between *Sp*Stn1N and *Sp*Ten1 (Figure 2.12B). Except for the one between the side chains of *Sp*Stn1N E132 and *Sp*Ten1 R22, most of these electrostatic interactions are not present in the *C. tropicalis* Stn1N–Ten1 complex (Figures 2.4C, 2.6D, 2.12B). Thus, the weak similarities between the Stn1 and Ten1 protein of fission yeast and budding yeast at the primary sequence level can be explained in part by the evolution of distinct interacting residues at the subunit interface.

Similar to budding yeast Stn1, *Sp*Stn1 also contains a C-terminal domain (*Sp*Stn1C). We performed a secondary structure prediction for *Sp*Stn1C using the program PredictProtein [46], which accurately predicted the positions of the α helices and β strands in the two WH motifs of *Sc*Stn1C (data not shown). The putative secondary structural elements in *Sp*Stn1C were then aligned with those present in *Sc*Stn1 and Rpa32 (Figure 2.10E). This analysis identified two presumed WH motifs in *Sp*Stn1C (Figure 2.10E). However, unlike budding yeast Stn1, both WH motifs in *Sp*Stn1 show a similar distribution of α helices and β strands that coincides well with *Ct*Stn1$_{WH1}$ and Rpa32$_{WH}$, suggesting that *Sp*Stn1 has two similar Rpa32-like WH motifs (Figure 2.10E). Nonetheless, the detection of Rpa32-like WH motifs in *Sp*Stn1, together with the overall structural similarity between the fission yeast and budding yeast Stn1N–Ten1 complexes, strongly supports the notion that an Rpa32–Rpa14-like complex is also conserved at fission yeast telomeres.

2.13 Discussion

Our structural analyses demonstrate that both the budding yeast and the fission yeast Stn1–Ten1 complexes share the same three-dimensional architecture as the Rpa32–Rpa14 complex despite minimal sequence similarity, thus providing the first direct confirmation of structural similarity between components of the CST and the RPA complexes. The reliability of our structures was further corroborated by mutational analyses of Stn1 and Ten1, which underscored the importance of functional heterodimerization between Stn1 and Ten1 for telomere localization of Ten1 and telomere length regulation. Thus, our findings provide a foundation for leveraging insights from the analysis of RPA to the study of the CST complex.

Budding yeast was believed to have evolved a very different set of telomeric proteins to protect and maintain chromosome ends. Hence, the budding yeast CST complex has been considered to serve as the functional equivalent of the POT1–TPP1 complex in fission yeast and other POT1-containing organisms. However, putative homologs of the CST proteins have been identified recently in both plants and humans [43, 44, 47], suggesting that this telomere regulatory complex is probably more widespread in nature than previously believed, even in organisms that use POT1 for telomere protection. On the other hand, the almost complete lack of sequence similarity between the CST components from budding yeast and POT1-containing organisms raised serious doubts concerning the structural and functional conservation of these proteins in these two groups of organisms. These doubts are now substantially alleviated by our structural data showing that the budding and fission yeast Stn1–Ten1 complexes share similar three-dimensional structures. As a consequence, insights from our structural

studies are expected to provide a platform for functional studies of at least two components of the CST complexes in a wide range of organisms, including humans. In support of this notion, multiple sequence alignment indicates that the critical Glu–Arg interactions that we uncovered in the budding and fission yeast Stn1–Ten1 complexes (Stn1 E189–Ten1 R27 in *C. tropicalis* and Stn1 E132–Ten1 R22 in *S. pombe*) are likely to be conserved in both plants and mammals. Nevertheless, it would be premature to extrapolate from the current findings to other features of the CST complexes. In particular, whether the remaining components of the CST complexes in different organisms (i.e., Cdc13 in yeast and Ctc1 in plants and human) [43, 44] resemble one another and whether they exhibit similarities to Rpa70 are largely unresolved. Clarifying these and other key issues in CST structure, assembly, and mechanisms will require detailed structural and functional analyses of the entire complex.

2.14 Methods and materials

Strains and plasmids

The *C. albicans* strain BWP17 (*ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG*) were used as the parental strains [48]. The derivations of mutant strains are described below.

Construction of mutant Candida strains

The deletion strain *stn1*-ΔΔ was generated by subjecting *BWP17* to two rounds of transformation and 5-FOA selection using a *stn1::hisG-URA3-hisG* cassette (containing ~700 bp of *STN1* upstream and ~700 bp of downstream sequence). Similarly, the

deletion strain *ten1*-ΔΔ was generated by subjecting *BWP17* to two rounds of transformation and 5-FOA selection using a *ten1::hisG-URA3-hisG* cassette (containing ~900 bp of *TEN1* upstream and ~900 bp of downstream sequence). The reconstituted strains *stn1*-ΔΔ/*STN1* and *ten1*-ΔΔ/*TEN1* were obtained by transforming the deletion strains with the pGEM-URA3-STN1 and pGEM-URA3-TEN1 integrating plasmid linearized by *Hpa*I and *Hind*III digestion, respectively. The pGEM-URA3-STN1 plasmid contains a 3.1 kb fragment spanning the *STN1* gene, while the pGEM-URA3-TEN1 plasmid contains a 2.1 kb fragment spanning the *TEN1* gene, each cloned into the *Sal*I and *Sac*I site of pGEM-URA3 [48]. Derivatives of the plasmids were used to introduce epitope-tagged *STN1* and *TEN1* into the deletion strains, as follows. The C-terminus of each gene was mutated by QuikChange to introduce an *Avr*II and a *Bsp*EI restriction site, thus allowing the introduction of the GSCP tag, which contains a Gly$_8$ linker, a Streptavidin binding peptide, a Calmodulin binding peptide, and a Protein A tag (the complete sequence is available upon request). Alanine substitution mutants of tagged *STN1* and *TEN1* were generated by the same mutagenesis protocol.

The *tert*-ΔΔ/*ten1*-ΔΔ and *tert*-ΔΔ/*ten1*-ΔΔ/*rad50*-ΔΔ mutants were constructed sequentially starting with a *tert*-ΔΔ mutant [49] using the aforementioned *ten1::hisG-URA3-hisG* cassette and a *rad50::hisG-URA3-hisG* cassette (containing ~750 bp of *RAD50* upstream and ~700 bp of downstream sequence). *C. albicans* transformations and 5-FOA selections were carried out as previously described [23]. Correct integrations of all disruption and reconstitution cassettes were confirmed by Southern analysis.

Analysis of telomeres and G-strand overhangs

Chromosomal DNAs were isolated by Smash and Grab as previously described except that the initial aqueous phase was subjected to one additional round of PCI (Phenol/Chloroform/Isoamyl alcohol (25:24:1)) extraction to minimize nuclease contamination [50]. Standard telomere Southern analysis and the in-gel hybridization analysis were performed using established protocols [51, 52]. The two-dimensional gel analysis was performed according to the protocol of Brewer and Fangman as modified by Cohen and Lavi [53, 54]. Briefly, the first dimension (0.5% agarose) was run at 0.5 V/cm for 16 h in the absence of ethidium bromide (EtBr), while the second dimension (1.2% agarose) was run at 5 V/cm for 5 h in the presence of 0.3 μg/ml EtBr. The DNAs in the gels were transferred to nylon membrane and probed with labeled CaC2 oligonucleotides as in the case of standard telomere Southern blots.

Chromatin IP

Chromatin immunoprecipitation using a combination of previously described protocols with some additional modifications [55, 56]. Cells were fixed with 1 % formaldehyde for 30 min at 30 °C and crosslinking was quenched with 125 mM glycine for 5 min at 30 °C. Formaldehyde-fixed or untreated cells were re-suspended in lysis buffer (50 mM HEPES pH 7.5, 1 mM EDTA, 150 mM NaCl, and protease inhibitors) and broken by glass beads. The lysates were sonicated ten times for 5s each (constant duty cycle, 35-40 % output) to shear DNAs to a mean length of ~600 base pairs. Extracts were adjusted to 1.6 mg/ml protein in 600 μl Lysis buffer and then diluted with 600 μl of IP dilution buffer (0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 450 mM NaCl, and protease inhibitors). 5 % of each cell extract was set aside and used as the input

sample. The remainder was subjected to immunoprecipitation using 20 µl of IgG-Sepharose beads at 4°C for 2 hrs. IP samples were washed for 5 min with rotation in the following buffers; one time with Buffer A (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, and 400 mM NaCl), four times with Buffer B (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, and 600 mM NaCl), one time with Buffer C (0.25 M LiCl, 1 % NP-40, 1% Na-Deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl pH 8.0) and one time with TE. All wash buffers contain protease inhibitors. IP samples were eluted in 500 µl of 1% SDS, 0.1 M $NaHCO_3$ and crosslinks were reversed at 65 °C for 5 hrs. Samples were treated with RNase A and proteinase K, extracted with phenol/chloroform, precipitated with ethanol, and re-suspended in 100 µl of water. The DNA samples were then applied to Hybond-N using a dot blot apparatus, and the membrane probed with ^{32}P-labeled CaC2 (CATCCGTACACCAAGAAGTTAGACATCCGTACACCAAGAAGTTAGA) corresponding to two copies of the *C. albicans* telomeric repeat. Signals were quantified using ImageQuant software (Molecular Dynamics Inc.).

Western and IP-Western

These were performed as previously described using antibodies directed against protein A [51].

Yeast-two-hybrid assay

The yeast-two-hybrid assays were performed using L40 strain harboring pBTM116 and pACT2 (Clontech) fusion plasmids. The colonies containing both plasmids were selected on –Leu –Trp plates. β-galactosidase activities were measured by liquid assay [57].

Protein expression and purification

The N-terminal domains of *C. tropicalis* Stn1 (residues 2-217) and *S. pombe* Stn1N (residues 2-186) were cloned into a GST fusion protein expression vector, pGEX6p-1 (GE healthcare). *C. tropicalis* Ten1 (residues 2-217), *S. pombe* Ten1 (residues 2–123), and *S. cerevisiae* Stn1 C-terminal domain (residues 311-493) into a modified pET28b vector with a Sumo protein fused at the N-terminus after the His_6 tag [42].

The *C. tropicalis* Stn1N-Ten1 complexes was coexpressed in E. *coli* BL21(DE3). After induction for 16 hours with 0.1 mM IPTG at 25°C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 6 hours at 4˚C before elution with 250 mM imidazole. Then Ulp1 protease was added to remove the His_6-Sumo tag. The complex was then mixed with glutathione sepharose beads (GE Healthcare) and rocked for 8 hours at 4˚C before elution with 15 mM glutathione. Protease 3C was added to remove the GST-tag. Finally, the Stn1N-Ten1 complex was further purified by passage through Mono-Q ion-exchange column and by gel-filtration chromatography on Hiload Superdex200 equilibrated with 25 mM Tris-HCl

pH 8.0, 150 mM NaCl and 5 mM dithiothreitol (DTT). The purified Stn1-Ten1 complex was concentrated to 15 mg/ml and stored at -80˚C.

S. cerevisiae Stn1C and the *S. pombe* Stn1N-ten1 complex were expressed in *E. coli* and purified following the same procedure as described above except for only one affinity chromatography step (Ni-NTA agarose) was used for *Sc*Stn1C.

Crystallization, data collection and structure determination

***C. tropicalis* Stn1N-Ten1:** Crystals were grown by sitting drop vapor diffusion method at 4ºC. The precipitant/well solution contained 1 M $MgSO_4$ and 0.1 M sodium Citrate pH 5.6 and 10 mM DTT. Heavy atom derivatives were obtained by soaking crystals in a solution containing 1.5 M $MgSO_4$ and 0.3 mM of MeHgAc for 2-3 hr and backsoaking for 1hr in 1.25 M $MgSO_4$, 1.4 M $NaHCO_3$, and 0.1 M sodium citrate pH 5.6. Both native and heavy atom derivative crystals were gradually transferred into a harvesting solution (0.25 M $MgSO_4$, 5.25 M $NaHCO_3$, and 0.1 M sodium critrate pH 5.6) before flash-cooling in liquid nitrogen for storage and data collection under cryogenic conditions (100K). Native and Hg-SAD (at Hg peak wavelength) datasets were collected at beamline 21ID-D at APS and processed using HKL2000. Crystals belong to space group $P4_12_12$ and contain two Stn1N-Ten1 complexes per asymmetric unit. Native crystals diffracted to 2.4 Å resolution with cell parameters a = b = 92.072 Å and c = 200.909 Å. Six mercury sites were located and refined, and the SAD phases calculated using SHARP [58]. The initial SAD map was significantly improved by solvent flattening. A model was automatically built into the modified experimental electron density using ARP/WARP [59]; the model was then further refined using simulated-annealing and positional

refinement in CNS [60] with manual rebuilding using program O [61]. The majority (86%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

***S. cerevisiae* Stn1C:** Crystals were grown by sitting drop vapor diffusion method at 4ºC. The precipitant/well solution contained 80 mM HEPES pH 7.0, 8% PEG6K, and 1.6 M NaCl and 10 mM DTT. Heavy atom derivatives were obtained by soaking crystals in a solution containing 25% PEG6K, 0.5 M NaCl, 0.1 M HEPES pH7.1 and 0.3 mM MeHgAc for 2-3 hr and back soaking for 1hr in 25% PEG6K, 10% Glycerol, 0.5 M NaCl and 0.1 M HEPES pH7.1. Both native and heavy atom derivative crystals were gradually transferred into a harvesting solution (25% PEG6K, 25% glycerol, 0.5 M NaCl and 0.1 M HEPES pH 7.1) before flash-cooling in liquid nitrogen for storage and data collection under cryogenic conditions (100K). Native and Hg-SAD (at Hg peak wavelength) datasets were collected at beamline 21ID-D at APS and processed using HKL2000 [62]. *Sc*Stn1C crystal belongs to space group $P4_32_12$ and contains one molecule in asymmetric unit. Native crystals diffracted 2.1 Å resolution with cell parameter a = b = 52.957 Å, c = 186.397 Å and contains one molecule in asymmetric unit. Two mercury sites were located and refined, and the SAD phases calculated using SHARP [63]. Model building and refinement were carried out following the same procedure as those for the *C. tropicalis* Stn1N-Ten1 complex. The majority (92%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

***S. pombe* Stn1N-Ten1:** The native and the Se-Met-substituted *S. pombe* Stn1N-Ten1 complex crystals were obtained using hanging drop vapor diffusion method by at 4ºC. The precipitant/well solution contained 12% PEG4K, 12% isopropanol, 0.1 M sodium citrate pH 5.6, and 5 mM DTT. Crystals were gradually transferred into a harvesting solution containing 25% PEG 4K, 16% isopropanol, 0.1 M sodium citrate pH 5.6 and 25% glycerol before flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100K). Native and Se-Met-SAD (at Se peak wavelength) datasets were collected at beam line 21ID-F at APS and processed using HKL2000 [62]. *S. pombe* Stn1N-ten1 complex crystals belong to space group $P4_12_12$ and contain one complex per asymmetric unit. Native dataset diffracted to 1.65 Å resolution with unit cell parameters a = b = 93.871 Å and c = 56.273 Å. Seven selenium atoms were located and refined, and the MAD phases calculated using SHARP [64]. Model building and refinement were carried out following the same procedure as those for the *C. tropicalis* Stn1N-Ten1 complex. The majority (95%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

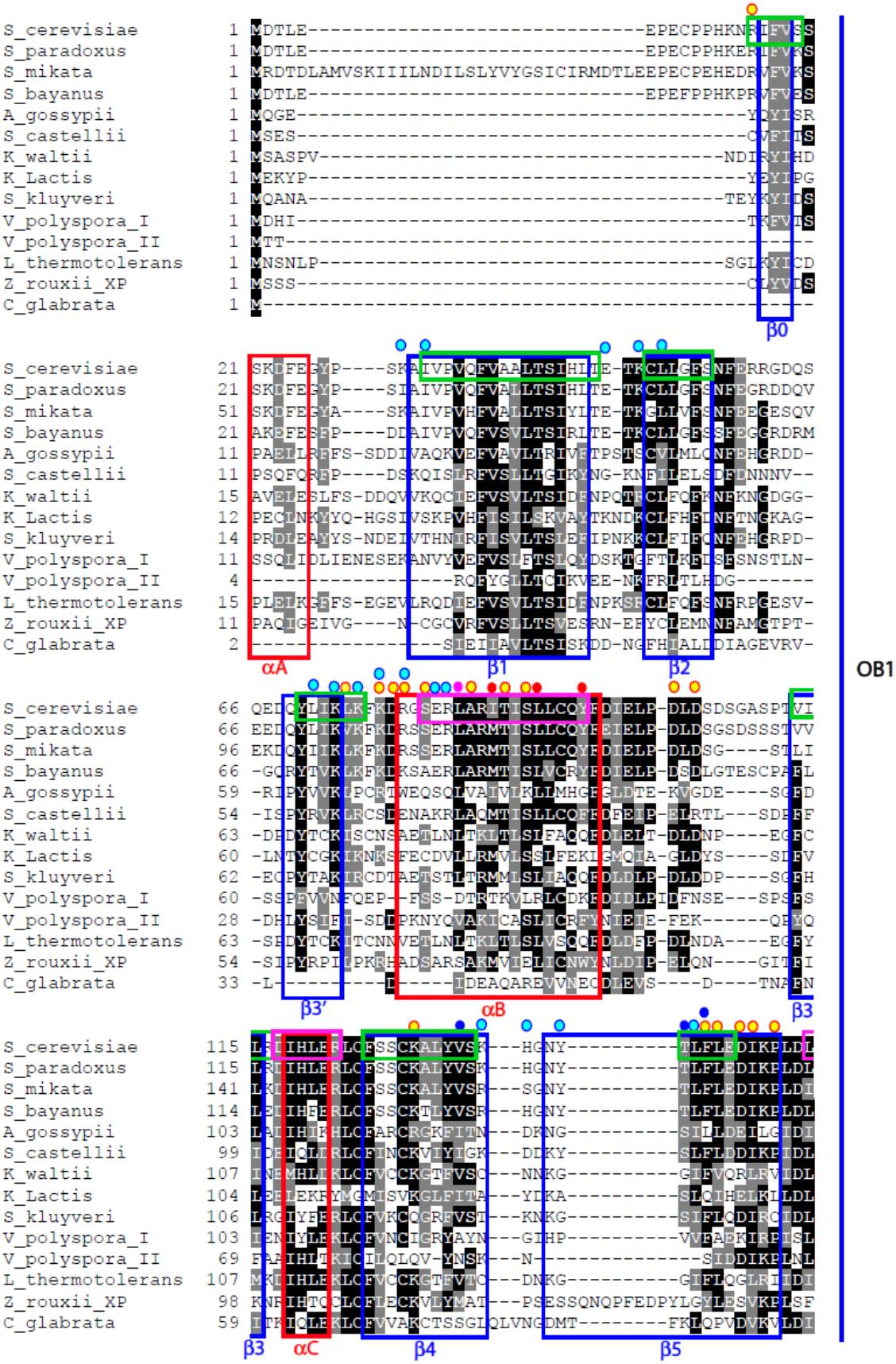
Accession numbers

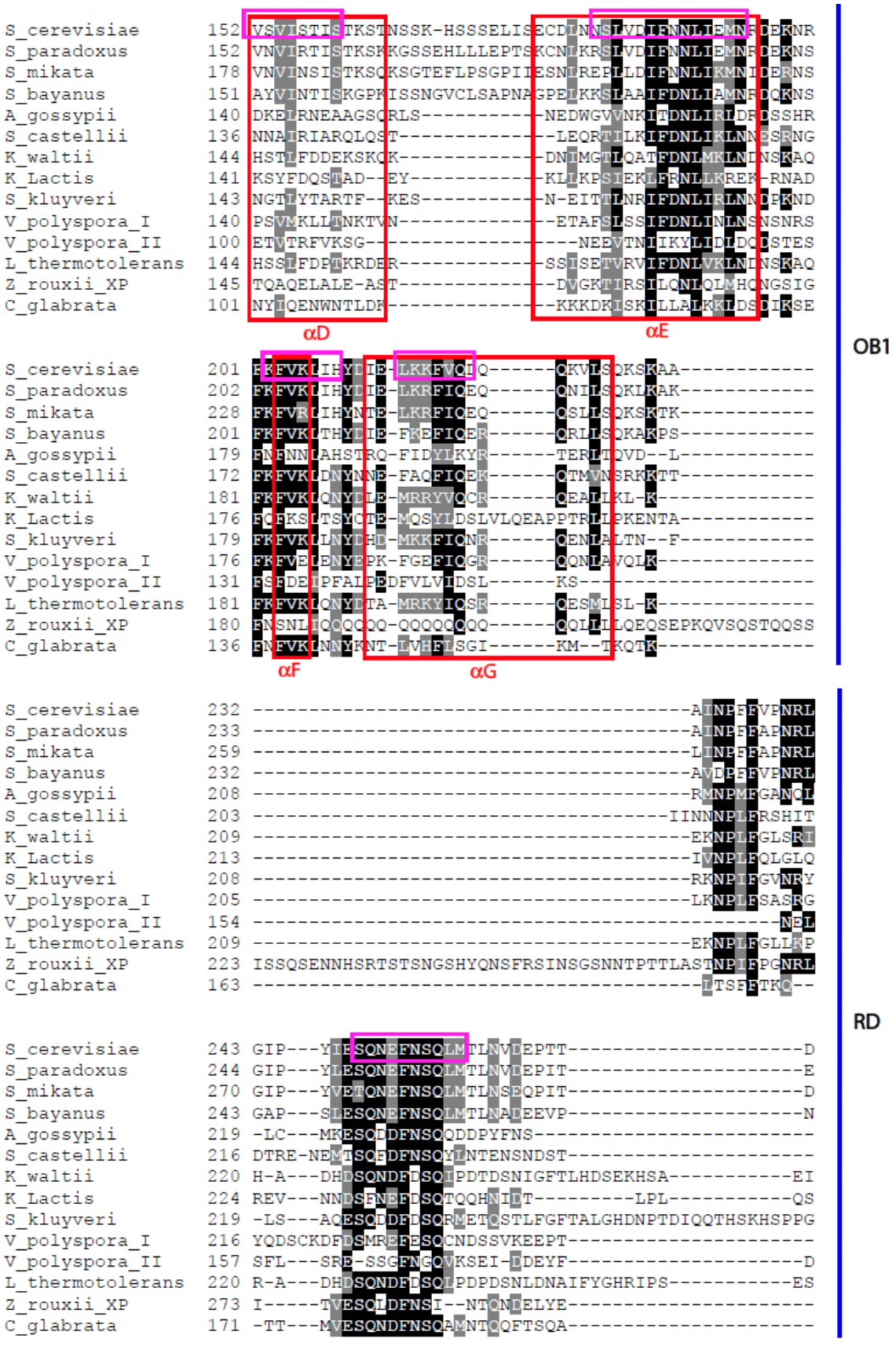
The coordinates and structure factors of the *C. tropicalis* Stn1N-Ten1 complex, *S. cerevisiae* Stn1C and the *S. pombe* Stn1N-Ten1 complex have been deposited in the RCSB Protein Data Bank under accession codes 3KF8, 3KEY, and 3KF6, respectively.

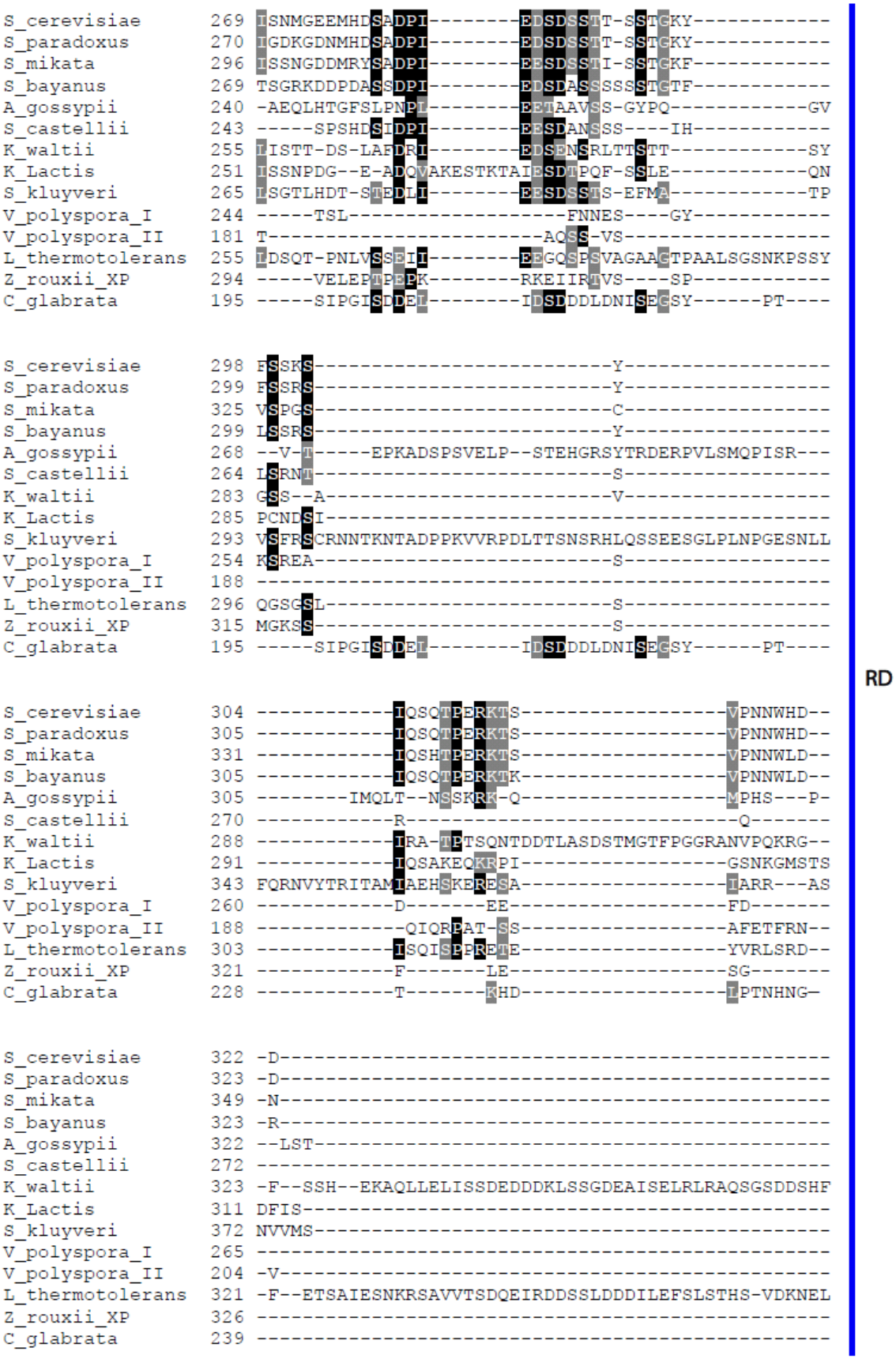
Figure 2.1 Alignments of Cdc13 homologues from *Saccharomyces, Candida and Kluyveromyces* spp.

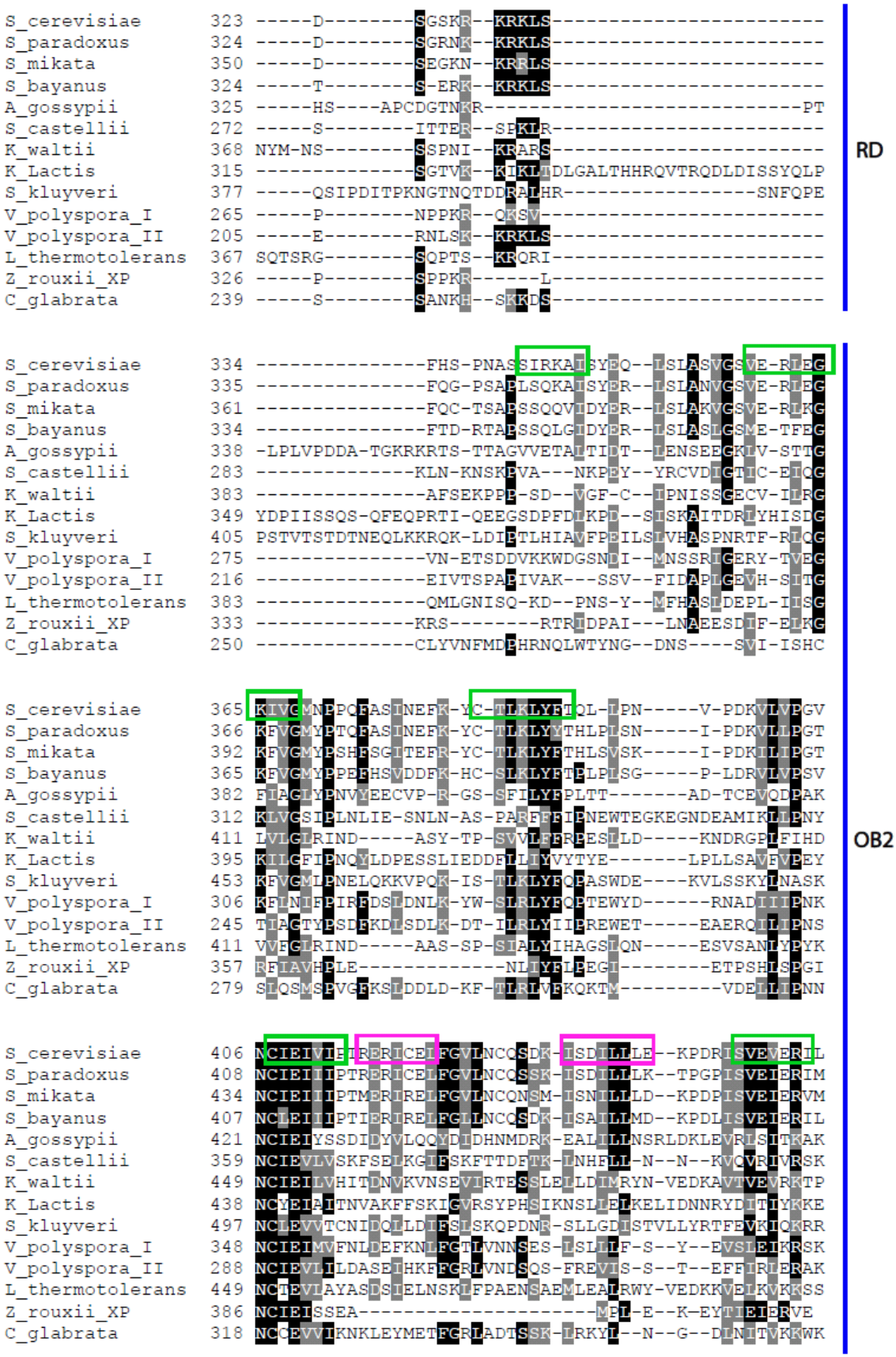
Multiple homologues were identified at the NCBI, SGD and Broad Institute databases using BLAST with default parameters. In all genomes analyzed, a plausible homologue ($E < 0.001$) of each subunit of the CST complex can be identified. Multiple sequence alignments were generated using the T-COFFEE server (http://www.igs.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi) and displayed using Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Structure based assignment of the α helices and β strands within the first, the third, and the last OB folds are indicated by red and blue boxes, respectively, while the predicted α helices and β strands of ScCdc13 are indicated by pink and green boxes. In *Sc*Cdc13$_{OB1}$, red dots denote the *S. cerevisiae* residues important for dimerization, Ile87, Lue91, and tyr95) whereas the pink dot denotes the less important residue Leu84 as shown in the gel filtration, yeast two-hybrid and co-IP assays; yellow dots denote the residues important for dimerization indicated by the crystal structure; and blue dots denote the residues involved in the Cdc13$_{OB1}$-Pol1$_{CBM}$ interaction. In *Cg*Cdc13$_{OB4}$, green dots denote the *C. glabrata* residues important for dimerization. The accession codes (at NCBI or Broad Institute) for the Cdc13 homologues in the alignment are as follows: *D. hansenii*, XP_461188; *C. albicans*, XP_719034; *C. parapsilosis*, CPAG_03609; *L. elongisporus*, XP_001526643; *C. guilliermondii*, XP_001486879, PGUG_00256; *C. lusitaniae*, CLUG_03319; *C. tropicalis*, CTRG_04305.

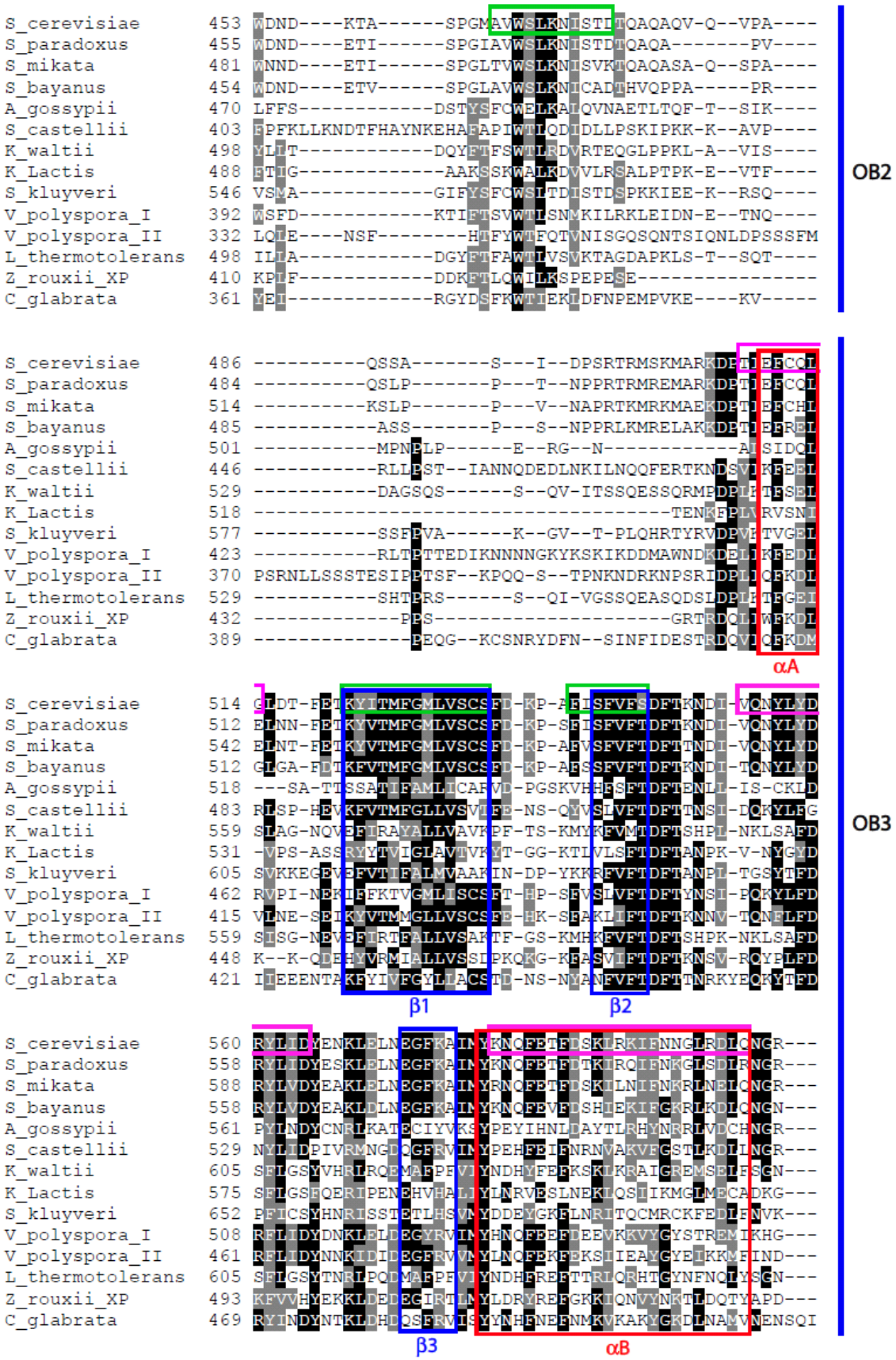
The accession codes for the Stn1 homologues in the alignment are as follows: *S. cerevisiae*, NP_010367; *K. lactis*, XP_452728; *C. albicans*, XP_714522; *D. hansenii*, XP_458626; *C. glabrata*, XP_448655; *L. elongisporus*, XP_001527444; *C. parapsilosis*, CPAG_03600; *C. lusitaniae*, CLUG_02415; *C. tropicalis*, CTRG_01841; *C. guilliermondii*, XP_001485882, PGUG_01553. The accession codes for the Ten1 homologues in the alignment are as follows: *S. cerevisiae*, NP_013110; *K. lactis*, XP_454375; *C. glabrata*, (emb|CR380956.2|); *D. hansenii*, XP_462449; *C. albicans*, XP_717945; *C. guilliermondii*, XP_001485289, PGUG_03018; *C. lusitaniae*, CLUG_01804; *C. tropicalis*, CTRG_00988; *C. parapsilosis*, CPAG_04435; *L. elongisporus*,XP_001527621.

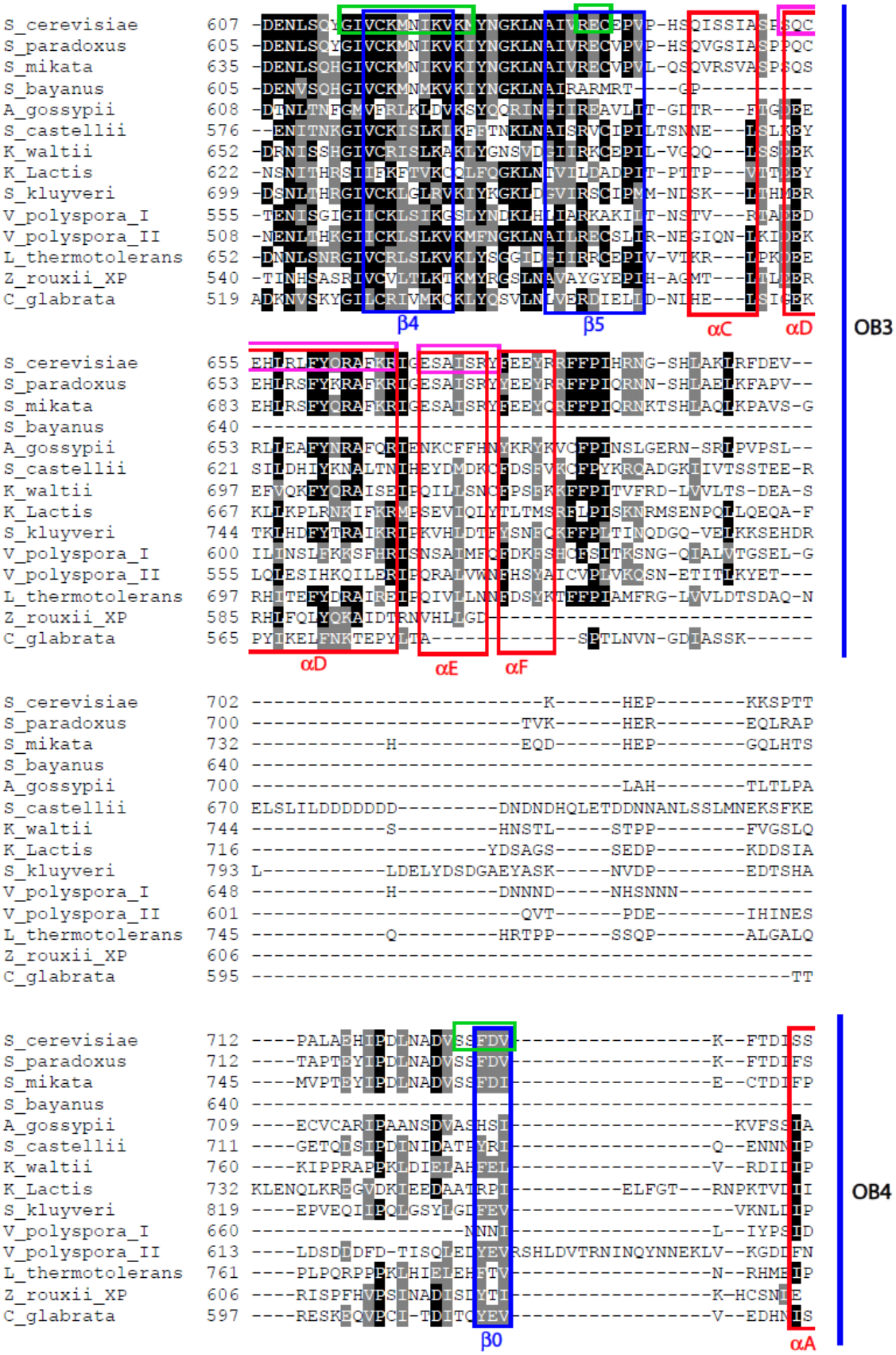

S_cerevisiae
S_paradoxus
S_mikata
S_bayanus
A_gossypii
S_castellii
K_waltii
K_Lactis
S_kluyveri
V_polyspora_I
V_polyspora_II
L_thermotolerans
Z_rouxii_XP
C_glabrata
β0
αA
β1
β2
OB1
β3'
αB
β3
β3
αC
β4
β5

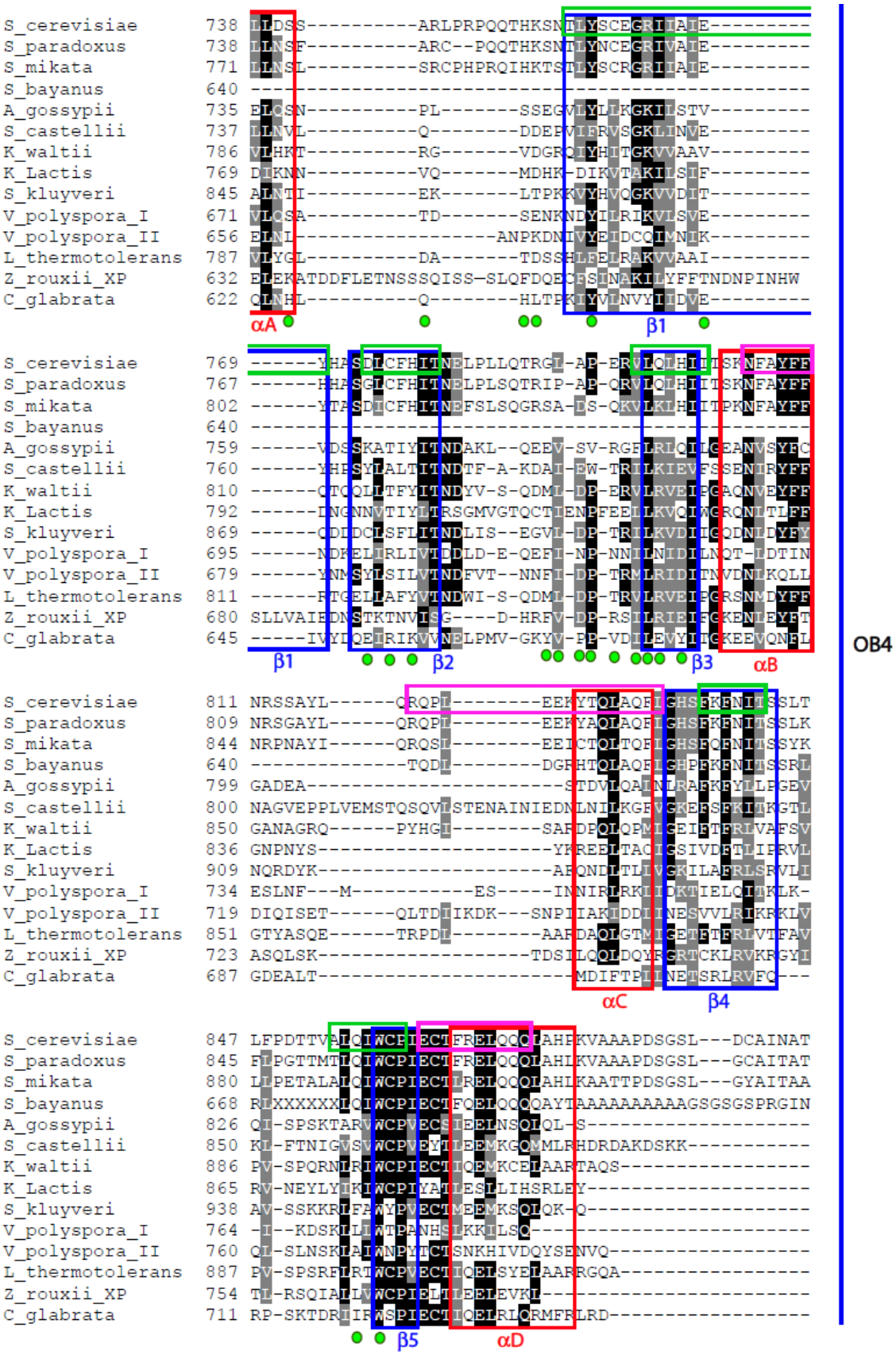

αD
αE
OB1
αF
αG
RD


S_cerevisiae 269 ISNMGEEMHDSADPI--------EDSDSSTT-SSTGKY----------
S_paradoxus 270 IGDKGDNMHDSADPI--------EDSDSSTT-SSTGKY----------
S_mikata 296 ISSNGDDMRYSADPI--------EESDSSTI-SSTGKF----------
S_bayanus 269 TSGRKDDPDASSDPI--------EDSDASSSSSSTGTF----------
A_gossypii 240 -AEQLHTGFSLPNPL--------EETAAVSS-GYPQ----------GV
S_castellii 243 -----SPSHDSIDPI--------EESDANSSS----IH----------
K_waltii 255 LISTT-DS-LAFDRI--------EDSENSRLTTSTT----------SY
K_Lactis 251 ISSNPDG--E-ADQVAKESTKTAIESDTPQF-SSLE----------QN
S_kluyveri 265 LSGTLHDT-STEDLI--------EESDSSTS-EFMA----------TP
V_polyspora_I 244 -----TSL----------FNNES----GY----------
V_polyspora_II 181 T----------AQSS-VS----------
L_thermotolerans 255 LDSQT-PNLVSSEII--------EEGQSPSVAGAAGTPAALSGSNKPSSY
Z_rouxii_XP 294 -----VELEPTPEPK--------RKEIIRTVS----SP----------
C_glabrata 195 -----SIPGISDDEL--------IDSDDDLDNISEGSY------PT----
S_cerevisiae 298 FSSKS----------Y----------
S_paradoxus 299 FSSRS----------Y----------
S_mikata 325 VSPGS----------C----------
S_bayanus 299 LSSRS----------Y----------
A_gossypii 268 --V-T-----EPKADSPSVELP--STEHGRSYTRDERPVLSMQPISR---
S_castellii 264 LSRNT----------S----------
K_waltii 283 GSS--A----------V----------
K_Lactis 285 PCNDSI----------
S_kluyveri 293 VSFRSCRNNTKNTADPPKVVRPDLTTSNSRHLQSSEESGLPLNPGESNLL
V_polyspora_I 254 KSREA----------S----------
V_polyspora_II 188 ----------
L_thermotolerans 296 QGSGSL----------S----------
Z_rouxii_XP 315 MGKSS----------S----------
C_glabrata 195 -----SIPGISDDEL--------IDSDDDLDNISEGSY------PT----
RD
S_cerevisiae 304 ----------IQSQTPERKTS----------VPNNWHD--
S_paradoxus 305 ----------IQSQTPERKTS----------VPNNWHD--
S_mikata 331 ----------IQSHTPERKTS----------VPNNWLD--
S_bayanus 305 ----------IQSQTPERKTK----------VPNNWLD--
A_gossypii 305 --------IMQLT--NSSKRK-Q----------MPHS---P-
S_castellii 270 ----------R----------Q-------
K_waltii 288 ----------IRA-TPTSQNTDDTLASDSTMGTFPGGRANVPQKRG--
K_Lactis 291 ----------IQSAKEQKRPI----------GSNKGMSTS
S_kluyveri 343 FQRNVYTRITAMIAEHSKERESA----------IARR---AS
V_polyspora_I 260 ----------D-------EE----------FD-------
V_polyspora_II 188 ----------QIQRPAT-SS----------AFETFRN--
L_thermotolerans 303 ----------ISQISPPRETE----------YVRLSRD--
Z_rouxii_XP 321 ----------F-------LE----------SG-------
C_glabrata 228 ----------T-------KHD----------LPTNHNG-
S_cerevisiae 322 -D----------
S_paradoxus 323 -D----------
S_mikata 349 -N----------
S_bayanus 323 -R----------
A_gossypii 322 --LST----------
S_castellii 272 ----------
K_waltii 323 -F--SSH--EKAQLLELISSDEDDDKLSSGDEAISELRLRAQSGSDDSHF
K_Lactis 311 DFIS----------
S_kluyveri 372 NVVMS----------
V_polyspora_I 265 ----------
V_polyspora_II 204 -V----------
L_thermotolerans 321 -F--ETSAIESNKRSAVVTSDQEIRDDSSLDDDILEFSLSTHS-VDKNEL
Z_rouxii_XP 326 ----------
C_glabrata 239 ----------


S_cerevisiae 323
S_paradoxus 324
S_mikata 350
S_bayanus 324
A_gossypii 325
S_castellii 272
K_waltii 368
K_Lactis 315
S_kluyveri 377
V_polyspora_I 265
V_polyspora_II 205
L_thermotolerans 367
Z_rouxii_XP 326
C_glabrata 239
RD
S_cerevisiae 334
S_paradoxus 335
S_mikata 361
S_bayanus 334
A_gossypii 338
S_castellii 283
K_waltii 383
K_Lactis 349
S_kluyveri 405
V_polyspora_I 275
V_polyspora_II 216
L_thermotolerans 383
Z_rouxii_XP 333
C_glabrata 250
S_cerevisiae 365
S_paradoxus 366
S_mikata 392
S_bayanus 365
A_gossypii 382
S_castellii 312
K_waltii 411
K_Lactis 395
S_kluyveri 453
V_polyspora_I 306
V_polyspora_II 245
L_thermotolerans 411
Z_rouxii_XP 357
C_glabrata 279
OB2
S_cerevisiae 406
S_paradoxus 408
S_mikata 434
S_bayanus 407
A_gossypii 421
S_castellii 359
K_waltii 449
K_Lactis 438
S_kluyveri 497
V_polyspora_I 348
V_polyspora_II 288
L_thermotolerans 449
Z_rouxii_XP 386
C_glabrata 318


S_cerevisiae 453 WDND----KTA------SPGMAVWSLKNISTITQAQAQV-Q--VPA----
S_paradoxus 455 WDND----ETI------SPGIAVWSLKNISTDTQAQA-------PV----
S_mikata 481 WNND----ETI------SPGLTVWSLKNISVKTQAQASA-Q--SPA----
S_bayanus 454 WDND----ETV------SPGLAVWSLKNICADTHVQPPA-----PR----
A_gossypii 470 LFFS------------DSTYSFCWELKALQVNAETLTQF-T--SIK----
S_castellii 403 FPFKLLKNDTFHAYNKEHAFAPIWTLQDIDLLPSKIPKK-K--AVP----
K_waltii 498 YLLT------------DQYFTFSWTLRDVRTEQGLPPKL-A--VIS----
K_Lactis 488 FTIG------------AAKSSKWALKDVVLRSALPTPK-E--VTF----
S_kluyveri 546 VSVA------------GIFYSFCWSLTDISTDSPKKIEE-K--RSQ----
V_polyspora_I 392 WSFD------------KTIFTSVWTLSNMKILRKLEIDN-E--TNQ----
V_polyspora_II 332 LQIE----NSF--------HTFYWTFQTVNISGQSQNTSIQNLDPSSSFM
L_thermotolerans 498 ILIA------------DGYFTFAWTLVSVKTAGDAPKLS-T--SQT----
Z_rouxii_XP 410 KPLF------------DDKFTLQWILKSPEPESE----------------
C_glabrata 361 YEI-------------RGYDSFKWTIEKIDFNPEMPVKE----KV-----
OB2
S_cerevisiae 486 ----------QSSA-------S---I--DPSRTRMSKMARKDFTIEFCQI
S_paradoxus 484 ----------QSLP-------P---T--NPPRTRMREMARKDFTIEFCQI
S_mikata 514 ----------KSLP-------P---V--NAPRTKMRKMAEKDFTIEFCHI
S_bayanus 485 ----------ASS-------P---S--NPPRLKMRELAKKDFTIEFREI
A_gossypii 501 ----------MPNPLP------E--RG--N----------AISIDQI
S_castellii 446 ----------RLLPST--IANNQDEDLNKILNQQFERTKNDSVIKFEEI
K_waltii 529 ----------DAGSQS------S--QV-ITSSQESSQRMEDPIFTFSEI
K_Lactis 518 ----------TENKHPLVRVSNI
S_kluyveri 577 ----------SSFPVA------K--GV--T-PLQHRTYRVDPVFTVGEI
V_polyspora_I 423 ----------RLTPTTEDIKNNNNGKYKSKIKDDMAWNDKDELIKFEDI
V_polyspora_II 370 PSRNLLSSSTESIPPTSF--KPQQ-S--TPNKNDRKNPSRIDPLIQFKDI
L_thermotolerans 529 ----------SHTPRS------S--QI-VGSSQEASQDSLDPIFTFGEI
Z_rouxii_XP 432 ----------PPS--------------------GRTRDQLIWFKDI
C_glabrata 389 ----------PEQG--KCSNRYDFN--SINFIDESTRDQVIQFKDM
αA
S_cerevisiae 514 GLDT-FETKYITMFGMLVSCSFD-KP-AFISFVFSDFTKNDI-VQNYLYD
S_paradoxus 512 ELNN-FETKYVTMFGMLVSCSFD-KP-SFISFVFTDFTKNDI-VQNYLYD
S_mikata 542 ELNT-FETKYVTMFGMLVSCSFD-KP-AFVSFVFTDFTTNDI-VQNYLYD
S_bayanus 512 GLGA-FDTKFVTMFGMLVSCSFD-KP-AFSSFVFTDFTKNDI-TQNYLYD
A_gossypii 518 ---SA-TISSATIFAMLICAFVD-PGSKVHHFSFTDFTENLI-IS-CKLD
S_castellii 483 RLSP-HEVKFVTMFGLLVSVTFE-NS-QYVSLVFTDFTTNSI-DQKYLFG
K_waltii 559 SLAG-NQVEFIRAYALLVAVKPF-TS-KMYKFVMTDFTSHPL-NKLSAFD
K_Lactis 531 -VPS-ASSRYYTVIGLAVTVKYT-GG-KTLVLSFTDFTANPK-V-NYGYD
S_kluyveri 605 SVKKEGEVEFVTIFALMVAAKIN-DP-YKKKFVFTDFTANPI-TGSYTFD
V_polyspora_I 462 RVPI-NEKIFFKTVGMLISCSFT-HP-SFVSLVFTDFTYNSI-PQKYLFD
V_polyspora_II 415 VLNE-SEIKYVTMMGLLVSCSFE-HK-SFAKLIFTDFTKNNV-TQNFLFD
L_thermotolerans 559 SISG-NEVEFIRTFALLVSAKTF-GS-KMHKFVFTDFTSHPK-NKLSAFD
Z_rouxii_XP 448 K--K-QDEHYVRMIALLVSSIPKQKG-KFASVIFTDFTKNSV-PQYPLFD
C_glabrata 421 IIEEENTAKFYIVFGYLLACSTD-NS-NYANFVFTDFTTNRKYPQKYTFD
β1
β2
S_cerevisiae 560 RYLIDYENKLELNEGFKAINYKNQFETFLSKLRKIFNNGLRDLQNGR---
S_paradoxus 558 RYLIDYESKLELNEGFKAINYKNQFETFLTKIRQIFNKGLSDIRNGR---
S_mikata 588 RYLVDYEAKLELNEGFKAINYRNQFETFLSKILNIFNKRLNEIQNGR---
S_bayanus 558 RYLVDYEAKLDLNEGFKAINYKNQFEVFLSHIEKIFGKRLKDLQNGN---
A_gossypii 561 PYINDYCNRLKATECIYVKSYPEYIHNLDAYTLRHYNRRLVDCHNGR---
S_castellii 529 NYLIDPIVRMNGDQGFRVINYPEHFEIFNRNVAKVFGSTLKDLINGR---
K_waltii 605 SFLGSYVHRLRQEMAFPFVIYNDHYFEFKSKLKRAIGREMSEIFSGN---
K_Lactis 575 SFLGSFQERIPENEHVHALIYLNRVESLNEKLQSTIKMGLMECADKG---
S_kluyveri 652 PFICSYHNRISSTETLHSVNYDDEYGKFLNRITQCMRCKFEDLFNVK---
V_polyspora_I 508 RFLIDYDNKLELDEGYRVINYHNQFEEFTEEVKKVYGYSTREMIKHG---
V_polyspora_II 461 RFLIDYNNKIDIDEGFRVVNYLNQFEKFEKSIIEAYGYEIKKMFIND---
L_thermotolerans 605 SFLGSYTNRLPQDMAFPFVIYNDHFREFTTRLQRHTGYNFNQIYSGN---
Z_rouxii_XP 493 KFVVHYEKKLDEDEGIRTLNYLDRYREFGKKIQNVYNKTLDQTYAPD---
C_glabrata 469 RYINDYNTKLDHDQSFRVISYYNHFNEFNMKVKAKYGKDLNAMVNENSQI
β3
αB
OB3


β4
β5
αC
αD
OB3
αD
αE
αF
β0
αA
OB4


S_cerevisiae 738 LLDSS----------ARLPRPQQTHKSNTLYSCEGRIIAIE---------
S_paradoxus 738 LLNSF----------ARC--PQQTHKSNTLYNCEGRIVAIE---------
S_mikata 771 LLNSL----------SRCPHPRQIHKTSTLYSCRGRIIAIE---------
S_bayanus 640 ----------
A_gossypii 735 ELQSN----------PL-------SSEGVLYLLKGKILSTV---------
S_castellii 737 LLNVL----------Q--------DDEPVIFRVSGKLINVE---------
K_waltii 786 VLHKT----------RG-------VDGRQIYHITGKVVAAV---------
K_Lactis 769 DIKNN----------VQ-------MDHK-DIKVTAKILSIF---------
S_kluyveri 845 ALNTI----------EK-------LTPKKVYHVQGKVVDIT---------
V_polyspora_I 671 VLQSA----------TD-------SENKNDYILRIKVLSVE---------
V_polyspora_II 656 ELNI----------ANPKDNIVYEIDCQIMNIK---------
L_thermotolerans 787 VLYGL----------DA-------TDSSHLFELRAKVVAAI---------
Z_rouxii_XP 632 ELEKATDDFLETNSSSQISS-SLQFDQECFSINAKILYFFTNDNPINHW
C_glabrata 622 QLLNHL----------Q--------HLTPKIYVLNVYIIDVE---------
αA
β1
S_cerevisiae 769 ------LHASDICFHITNELPLLQTRGL-AP-ERVLQLHIITSKNFAYFF
S_paradoxus 767 ------HHASGICFHITNELPLSQTRIP-AP-QRVLQLHIITSKNFAYFF
S_mikata 802 ------VTASDICFHITNEFSLSQGRSA-DS-QKVLKLHIITPKNFAYFF
S_bayanus 640 ----------
A_gossypii 759 ------VDSSKATIYITNDAKL--QEEV-SV-RGFLRLQILGEANVSYFC
S_castellii 760 ------YHESYIALTITNDTF-A-KDAI-EW-TRILKIEVFSSENIRYFF
K_waltii 810 ------QTQQLITFYITNDYV-S-QDML-DP-ERVLRVEIPGAQNVEYFF
K_Lactis 792 ------INGNNVTIYLTRSGMVGTQCTIENPFEELLKVQIWGRQNLTLFF
S_kluyveri 869 ------QDIDCLSFLITNDLIS--EGVL-DP-TRILKVDIIGQDNLIYFY
V_polyspora_I 695 ------NDKELIRLIVTDDLD-E-QEFI-NP-NNILNIDILNQT-LDTIN
V_polyspora_II 679 ------YNMSYLSILVTNDFVT--NNFI-DP-TRMLRIDITNVDNLKQLL
L_thermotolerans 811 ------RTGELLAFYVTNDWI-S-QDML-DP-TRVLRVEIPGRSNMDYFF
Z_rouxii_XP 680 SLLVAIEDNSTKTNVISG----D-HRFV-DP-RSILRIEIFGKENLEYFT
C_glabrata 645 -----IVYIQEIRIKVVNELPMV-GKYV-PP-VDILEVYITGKEEVQNFL
β1
β2
β3
αB
OB4
S_cerevisiae 811 NRSSAYL------QRQPL--------EEKYTQLAQFIGHSFKFNITSSLT
S_paradoxus 809 NRSGAYL------QRQPL--------EEKYAQLAQFLGHSFKFNITSSLK
S_mikata 844 NRPNAYI------QRQSL--------EEICTQLTQFIGHSFQFNITSSYK
S_bayanus 640 ----------TQDL--------DGRHTQLAQFIGHPFKFNITSSRL
A_gossypii 799 GADEA----------STDVLQAINLRAFKFYLLPGEV
S_castellii 800 NAGVEPPLVEMSTQSQVLSTENAINIEDNLNILKGFVGKEFSFKITKGTL
K_waltii 850 GANAGRQ------PYHGI--------SARDPQLQPMLGEIFTFRLVAFSV
K_Lactis 836 GNPNYS----------YKREELTACIGSIVDFTLIPRVL
S_kluyveri 909 NQRDYK----------ARQNDLTLLVGKILAFRLSRVLI
V_polyspora_I 734 ESLNF---M----------ES-----INNIRLRKLIDKTIELQITKLK-
V_polyspora_II 719 DIQISET------QLTDIIKDK---SNPIIAKIDDIINESVVLRIKRKLV
L_thermotolerans 851 GTYASQE------TRPDL--------AARDAQLGTMIGETFTFRLVTFAV
Z_rouxii_XP 723 ASQLSK----------TDSILQQLDQYFGRTCKLRVKRGYI
C_glabrata 687 GDEALT----------MDIFTPLINETSRLRVFQ---
αC
β4
S_cerevisiae 847 LFPDTTVALQIWCPIECTFRELQQQIAHPKVAAAPDSGSL---DCAINAT
S_paradoxus 845 FLPGTTMTLQIWCPIECTFRELQQQLAHLKVAAAPDSGSL---GCAITAT
S_mikata 880 LLPETALALQIWCPIECTLRELQQQLAHLKAATTPDSGSL---GYAITAA
S_bayanus 668 RLXXXXXXIQIWCPIECTFQELQQQQAYTAAAAAAAAAAGSGSGSPRGIN
A_gossypii 826 QI-SPSKTARVWCPVECSIEELNSQLQL-S----------
S_castellii 850 KL-FTNIGVSVWCPVEYTLEEMKGQMMLRHDRDAKDSKK----------
K_waltii 886 PV-SPQRNLRIWCPIECTIQEMKCELAARTAQS----------
K_Lactis 865 RV-NEYLYIKIWCPIYATLESLLIHSRLEY----------
S_kluyveri 938 AV-SSKKRLFAWYPVECTMEEMKSQLQK-Q----------
V_polyspora_I 764 -I--KDSKLLLWTPANHSLKKILSQ----------
V_polyspora_II 760 QL-SLNSKLALWNPYTCTSNKHIVDQYSENVQ----------
L_thermotolerans 887 PV-SPSRFLRTWCPVECTIQELSYELAARRGQA----------
Z_rouxii_XP 754 TL-RSQIALLVWCPIELTLEELEVKL----------
C_glabrata 711 RP-SKTDRIIRWSPIECTIQELRLQRMFRLRD----------
β5
αD

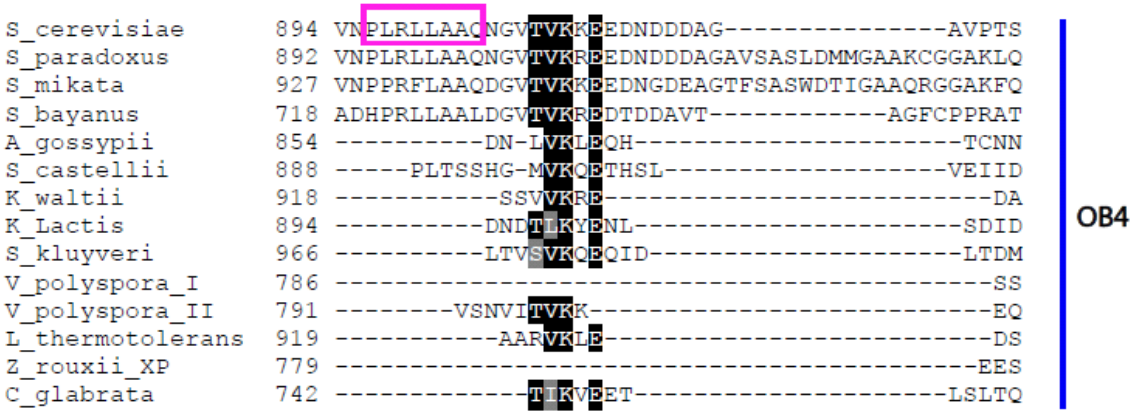

S_cerevisiae 894 VNPLRLLAAQNGVTVKKEEDNDDDAG----------AVPTS
S_paradoxus 892 VNPLRLLAAQNGVTVKREEDNDDDAGAVSASLDMMGAAKCGGAKLQ
S_mikata 927 VNPPRFLAAQDGVTVKKEEDNGDEAGTFSASWDTIGAAQRGGAKFQ
S_bayanus 718 ADHPRLLAALDGVTVKREDTDDAVT----------AGFCPPRAT
A_gossypii 854 ----------DN-LVKLEQH----------TCNN
S_castellii 888 -----PLTSSHG-MVKQETHSL----------VEIID
K_waltii 918 ----------SSVVKRE----------DA
K_Lactis 894 ----------DNDTLKYENL----------SDID
S_kluyveri 966 ----------LTVSVKQEQID----------LTDM
V_polyspora_I 786 ----------SS
V_polyspora_II 791 --------VSNVITVKK----------EQ
L_thermotolerans 919 ----------AARVKLE----------DS
Z_rouxii_XP 779 ----------EES
C_glabrata 742 ----------TIKVEET----------LSLTQ
OB4

Figure 2.2 Phenotypes of the *C. albicans stn1-/-* and *ten1-/-* mutant.
(Experiment and figures prepared by E.Y. Yu)
(A) The slow growing and filamentous morphology of the *stn1-/-* and *ten1-/-* mutants are displayed.
(B) Chromosomal DNAs isolated from the parental *BWP17*, the *stn1-/-* , and the *ten1-/-* mutants were subjected to Southern analysis of the telomere terminal restriction fragments. The mutant samples were from two independently constructed null strains that have undergone ~ 100 cell divisions following construction.
(C, top) Chromosomal DNAs isolated from the parental BWP17, the stn1-/-, and the ten1-/- mutant were subjected to in-gel hybridization analysis of the level of G tails. (Bottom) Subsequently, the DNAs were denatured in the gel and reanalyzed using the same probe. As a positive control, the DNA from a ter1-/- strain, which was demonstrated previously to exhibit an increase in G-tail signal, was analyzed in parallel.
(D) Chromosomal DNAs isolated from the parental *BWP17*, the *stn1-/-*, and the *ten1-/-* mutant were subjected to two-dimensional gel electrophoresis in order to resolve linear and circular telomeric DNA (marked by arrowheads).

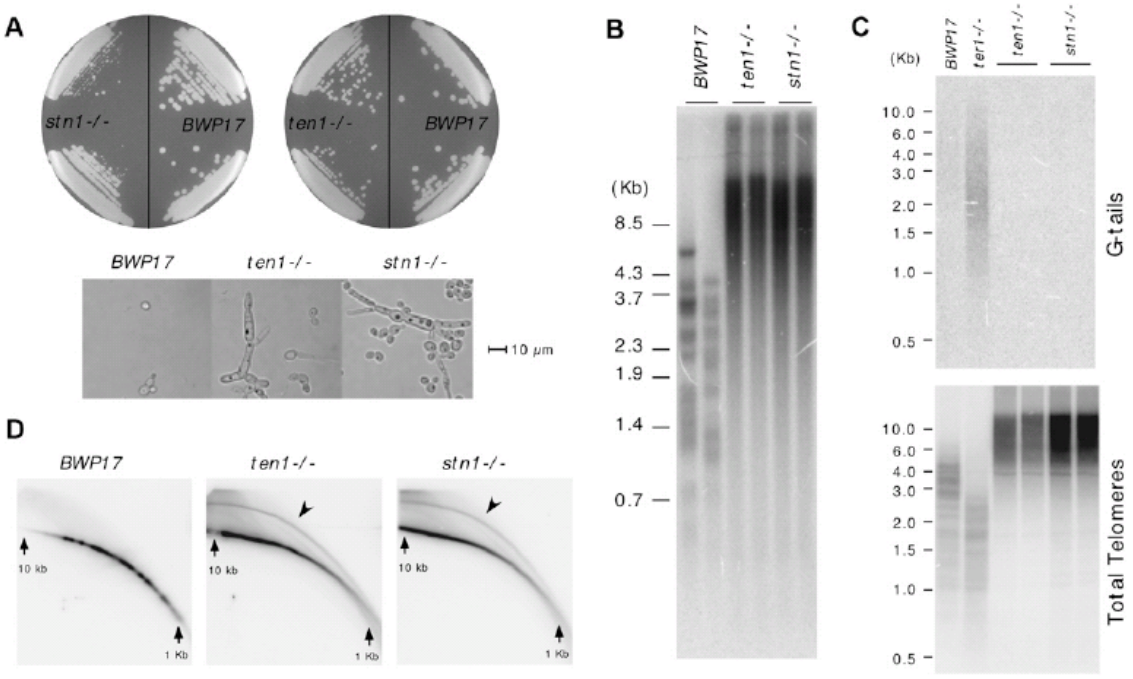


Figure 2.3 *C. tropicalis* Stn1N and Ten1 form a stable complex in solution.
(A) Gel filtration chromatography profile (Hiload Superdex 200) of the Stn1N-Ten1 complex. Elution positions of 40, 63 and 98 kDa protein markers are indicated.
(B) SDS-PAGE of the Stn1N-Ten1 complex corresponding to the peak fraction in the gel filtration profile in A.
(C) Crystals of *C. tropicalis* Stn1N-Ten1

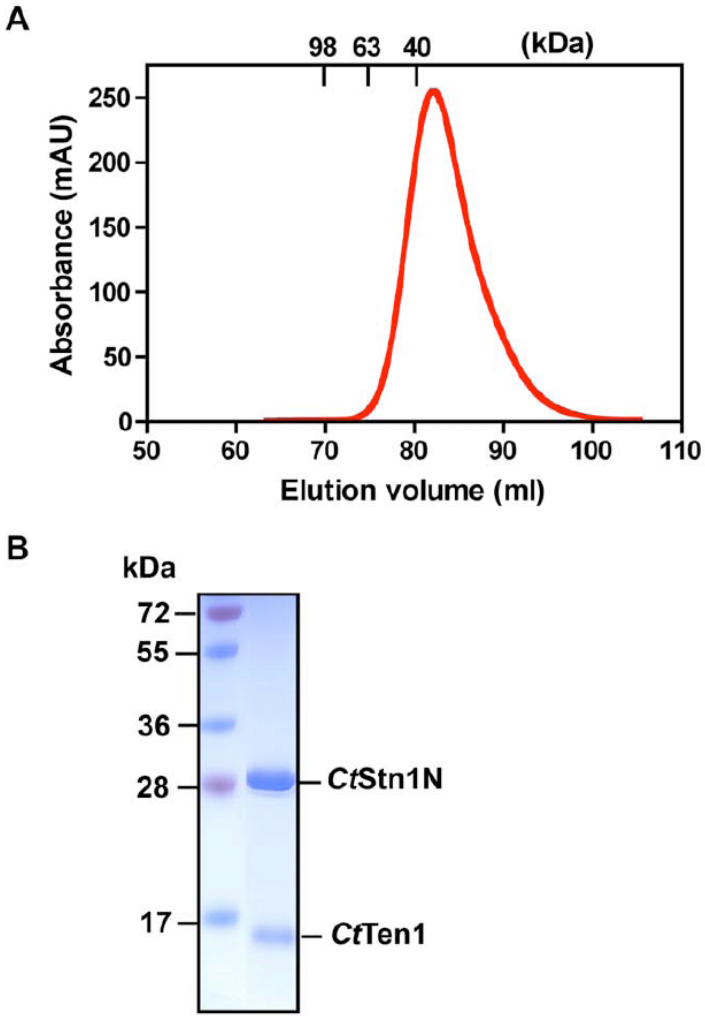


Figure 2.4 Overview of the *C. tropicalis* Stn1N–Ten1 complex structure.
(A) Domain organization of the Stn1 and Ten1 polypeptide chains.
(B) Ribbon diagram of two orthogonal views of the Stn1N–Ten1 complex.
(C) Amino acid sequence alignment of Stn1N and Ten1. (Top panel) Sequence alignment of the N-terminal OB fold regions of the yeast Stn1 family members together with human Rpa32. (Bottom panel) Sequence alignment of the yeast Ten1 family members together with human Rpa14.

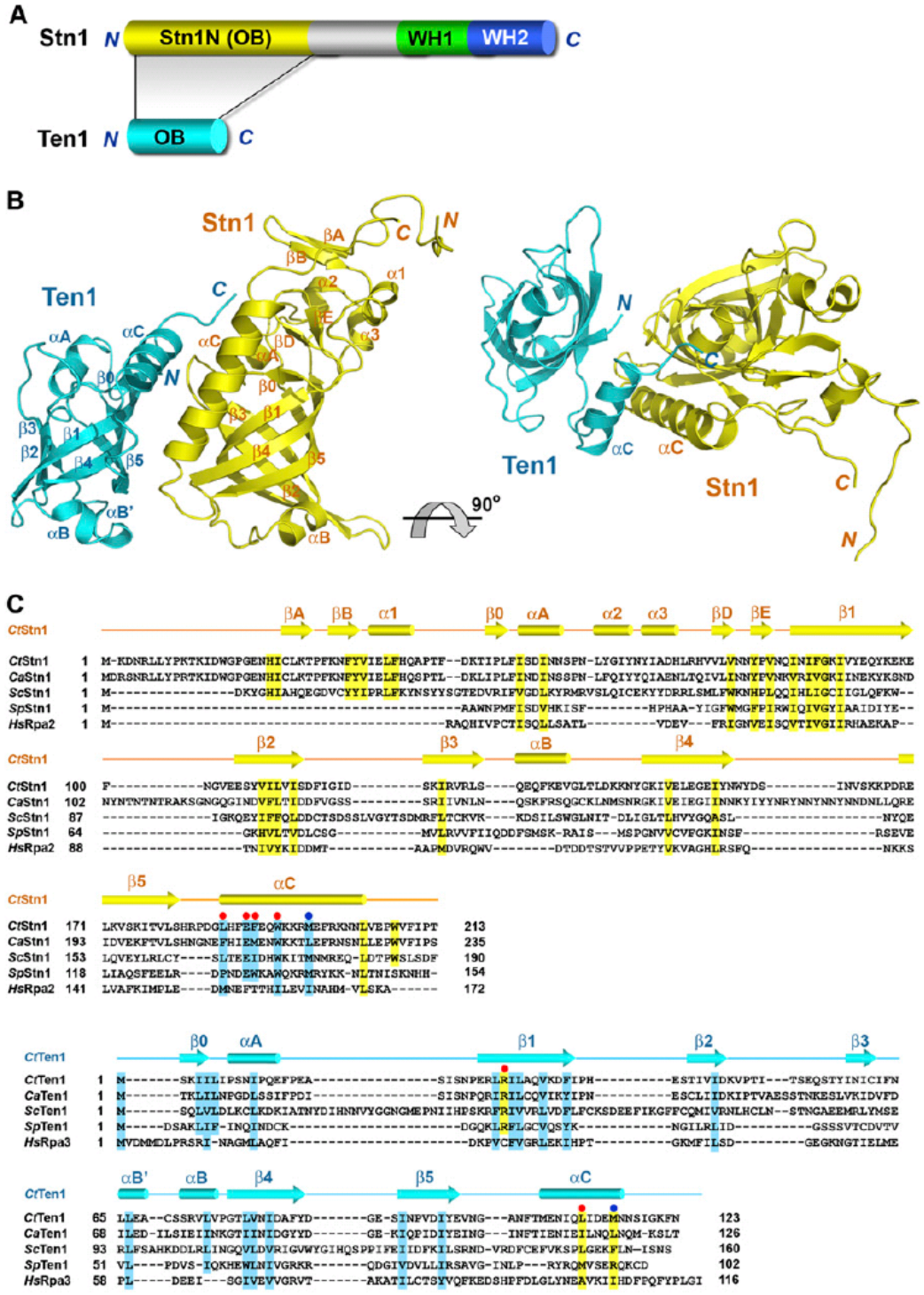


Figure 2.5 The 'Cap' motif of *C. tropicalis* Stn1N is stabilized by the hydrophobic tail C-terminal to helix αC.
The 'Cap' motif and the tail are colored in brown and magenta, respectively. The OB fold core is in yellow. The side chains of the hydrophobic residues are shown in stick.

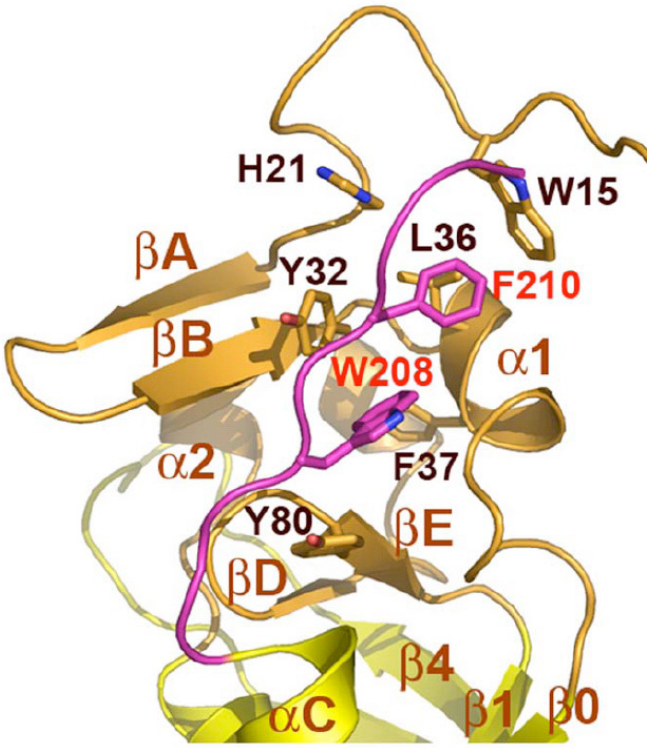


Figure 2.6 The *C. tropicalis* Stn1N–Ten1 complex is structurally similar to Rpa32N–Rpa14.
(A) Superposition of the Stn1N–Ten1 complex on the crystal structure of the human Rpa32N– Rpa14 complex [32]. Stn1N and Ten1 are colored in yellow and cyan and Rpa32N and Rpa14 are shown in blue and magenta. The superposition is based on the structures of Stn1N and Rpa32N. Ten1 and Rpa14 are not aligned well, and Ten1 rotates; 15° relative to the orientation of Rpa14.
(B) Overlay of Ten1 and Rpa14 based on the OB fold β barrels of the proteins.
(C) Superposition of Stn1N and Rpa32N based on the OB fold b barrels shows collisions between the cap motif of Stn1N and the N-terminal b hairpin (βD–βE) of Rpa32N. Residues in Stn1N are drawn as a stick model with dotted surface. Residues in Rpa32N are shown as a space-filling model. Stn1N and Rpa32N are colored as in A. Labels for residues in Rpa32 are in italics, to differentiate them from residues in Stn1.

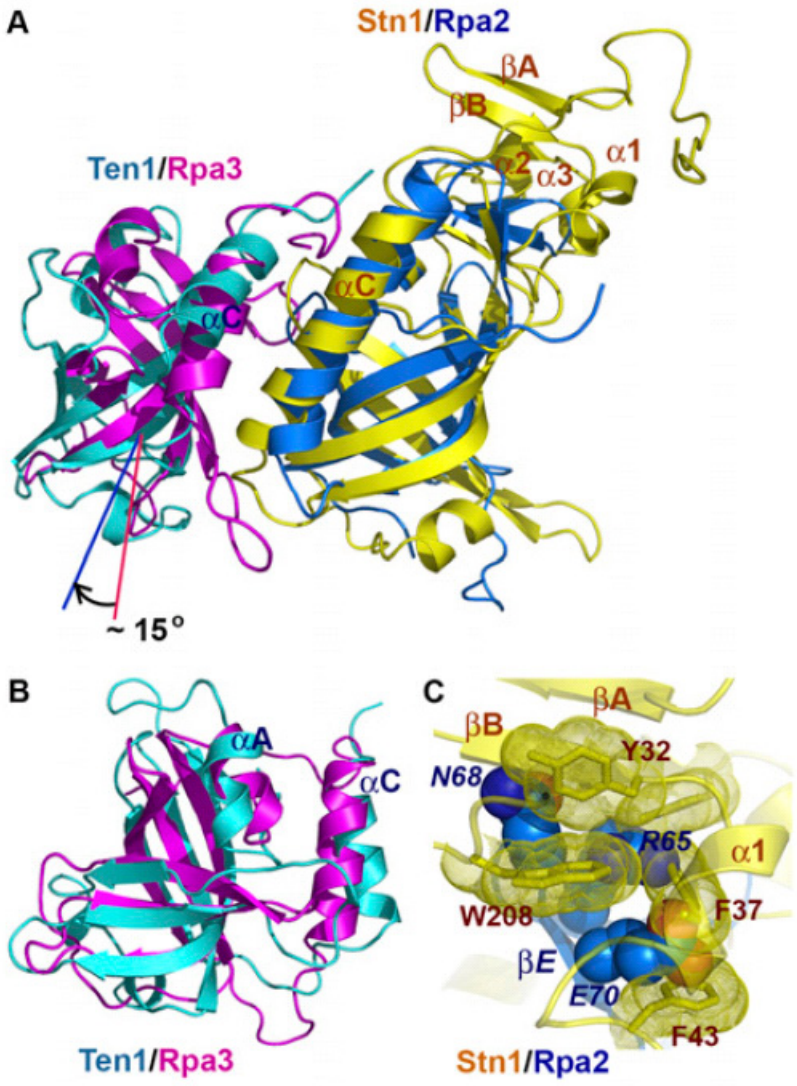


Figure 2.7 The *C. tropicalis* Stn1N–Ten1 interface.
(A) The hydrophobic interface between Stn1N and Ten1. (Left) Stn1N is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red). Ten1 is in ribbon representation. (Right) Ten1 is in electrostatic surface representation, while Stn1N is in ribbon. The orientation of the complex is rotated by 180° about a vertical axis relative to the complex in the left panel.
Hydrophobic interactions (B, C) and electrostatic interactions (D) between Stn1N and Ten1. Side chains of residues important for interaction are shown as stick models and are colored as in A. The intermolecular hydrogen bonds are shown as dashed magenta lines.
(E) Effects of the Stn1 and Ten1 mutations on the Stn1–Ten1 interaction in a yeast two-hybrid assay. Interaction of LexA–Stn1 with GAD–Ten1 was measured as β-galactosidase activity. Data are the average of three independent β-galactosidase measurements normalized to the wild-type Stn1–Ten1 interaction, arbitrarily set to 100.

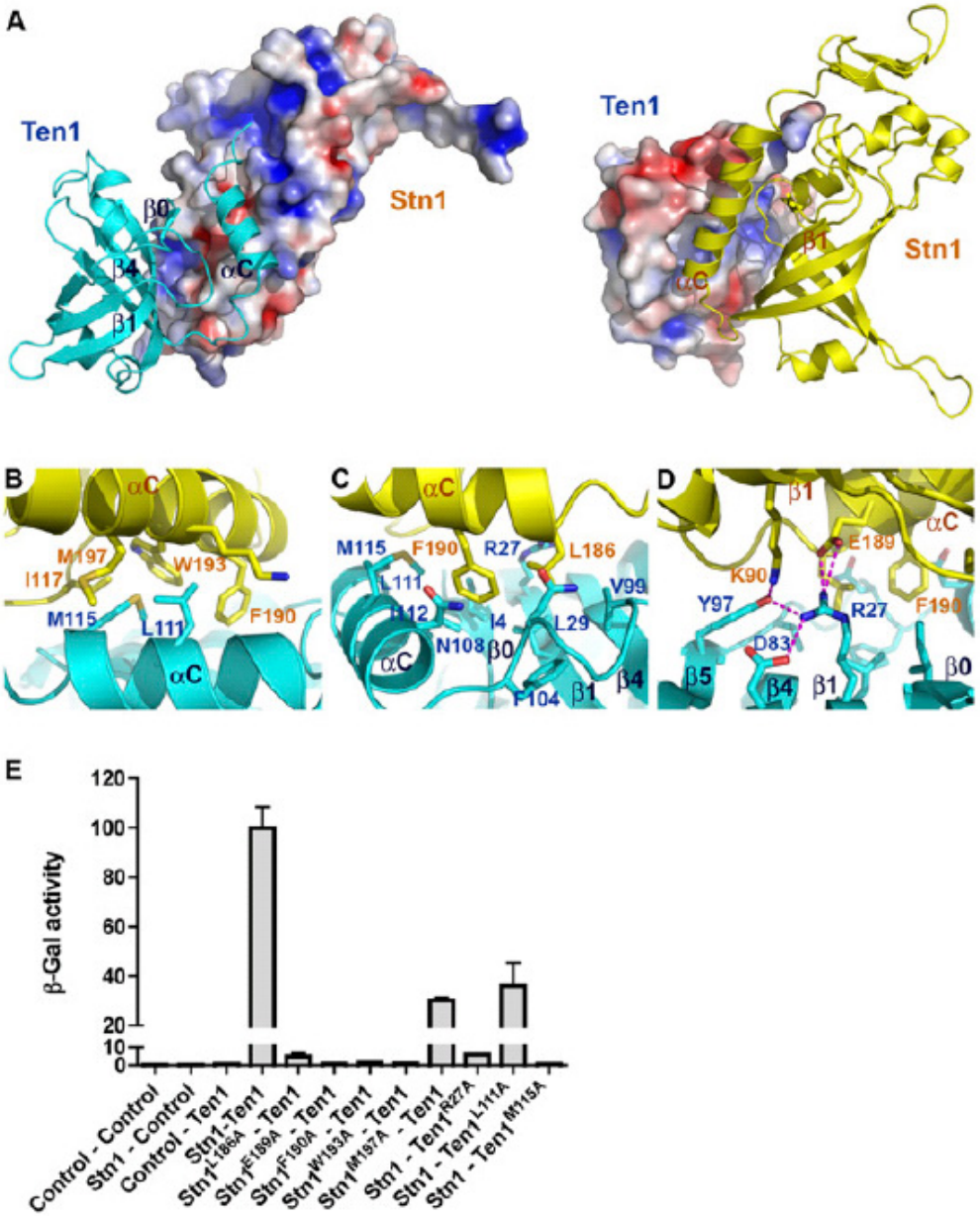


Figure 2.8 The effects of point mutations designed to disrupt the *C. albicans* Stn1–Ten1 interaction on protein levels, telomere length regulation, and protein–telomere association. (Experiment and figures prepared by E.Y. Yu)
(A) Chromosomal DNAs were isolated from the ten1-/- mutant and various reconstituted strains after two to four streaks (~50–100 generations) on plates and were subjected to telomere restriction fragment analysis. (Bottom) As loading controls, the telomere probe was stripped and the blot was rehybridized with a RAD52 fragment.
(B) Chromosomal DNAs were isolated from the stn1-/- mutant and various reconstituted strains after two to four streaks (~50–100 generations) on plates and were subjected to telomere restriction fragment analysis. (Bottom) As loading controls, the telomere probe was stripped and the blot was rehybridized with a RAD52 fragment.
(C) Extracts were prepared from strains containing different GSCP tagged Ten1 mutants and were subjected directly to Western analysis using antibodies against protein A.
(D, top) Extracts from strains bearing different GSCP-tagged Stn1 mutant proteins were subjected to affinity pull-down with IgG-Sepharose, followed by Western analysis using antibodies against protein A. (Bottom) To compare the levels of GSCP-tagged Stn1 and Stn1–E211A, we subjected the wild-type extract to serial dilutions prior to the assays.

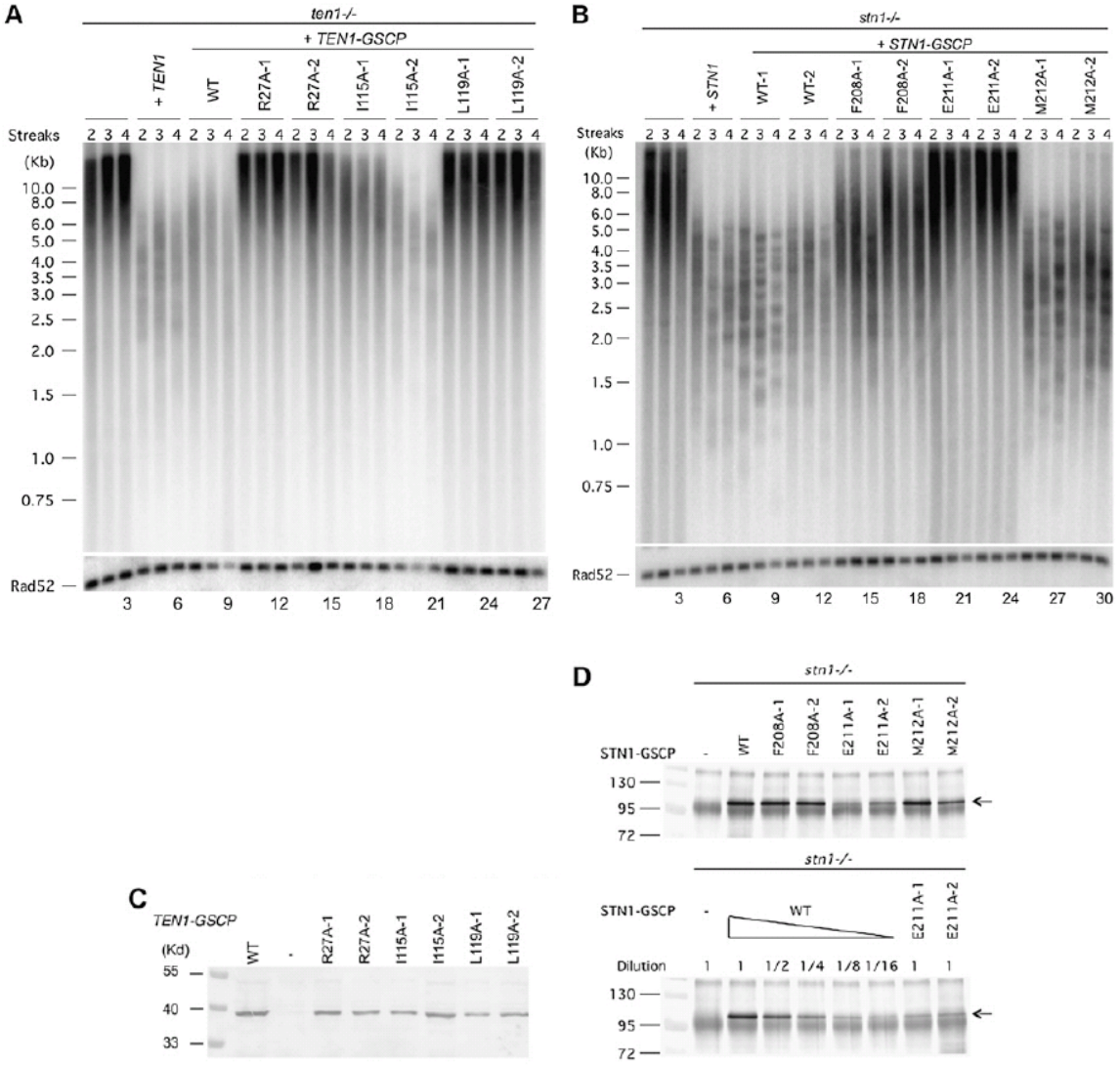
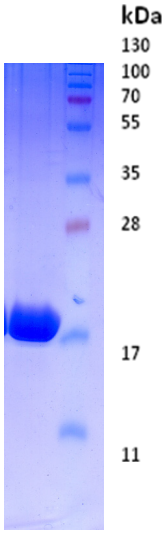


Figure 2.9 C-terminal domain of *S. cerevisiae* Stn1.
(A) SDS-PAGE of purified *Sc*Stn1C
(B) Crystals of *Sc*Stn1C

A



B

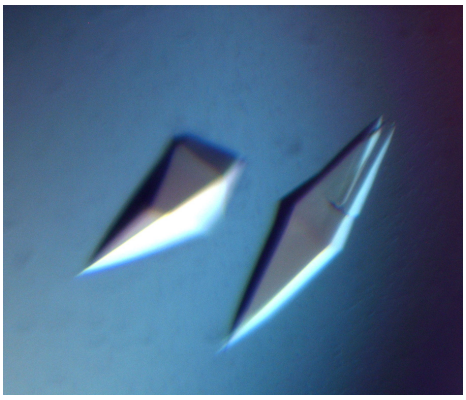
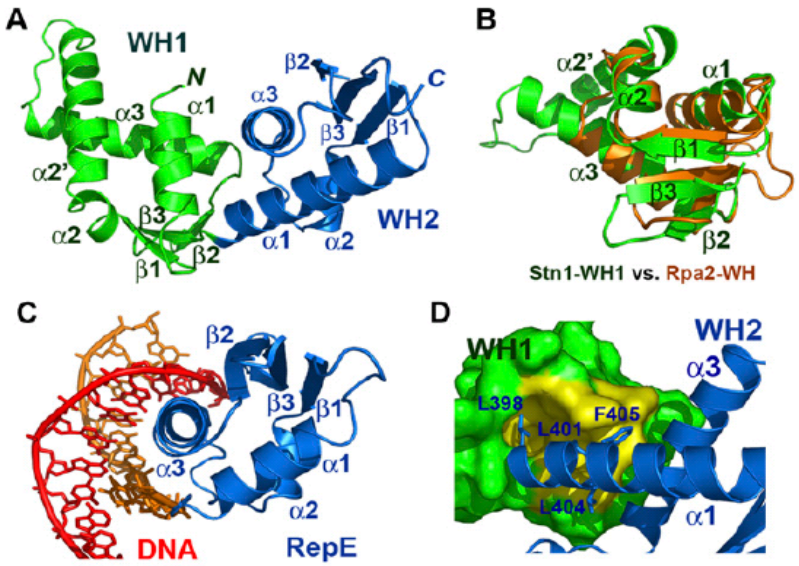


Figure 2.10 Crystal structure of the C-terminal domain of *S. cerevisiae* Stn1.
(A) Ribbon diagram of *Sc*Stn1C. The WH1 and WH2 motifs of *Sc*Stn1N are colored as in Figure 2.4A. The secondary structure elements are labeled. The dotted line represents the disordered loop between strands β2 and β3 of WH2.
(B) Superposition of the WH1 motif of *Sc*Stn1 (in green) on the NMR (nuclear magnetic resonance) structure of the WH motif of Rpa32 (in orange).
(C) Ribbon diagram of the RepE–DNA complex. The orientation of the WH motif of RepE is the same as the WH2 of *Sc*Stn1C in A.
(D) The hydrophobic interactions between the WH1 and WH2 motifs of *Sc*Stn1C.
(E) Amino acid sequence alignment of the C-terminal WH1 and WH2 motifs of budding yeast Stn1 family members together with the WH motifs of *S. pombe* Stn1 and human Rpa32. The alignment with Rpa32 is based on the NMR structure of the Rpa32C–UNG2 complex. Secondary structure assignments from our *Sc*Stn1C crystal structure are shown. Conserved hydrophobic residues in WH1 and WH2 are highlighted in green and blue blocks, respectively. In contrast to the WH motifs in budding yeasts, both WH1 and WH2 of *Sp*Stn1 are similar to Rpa32WH.



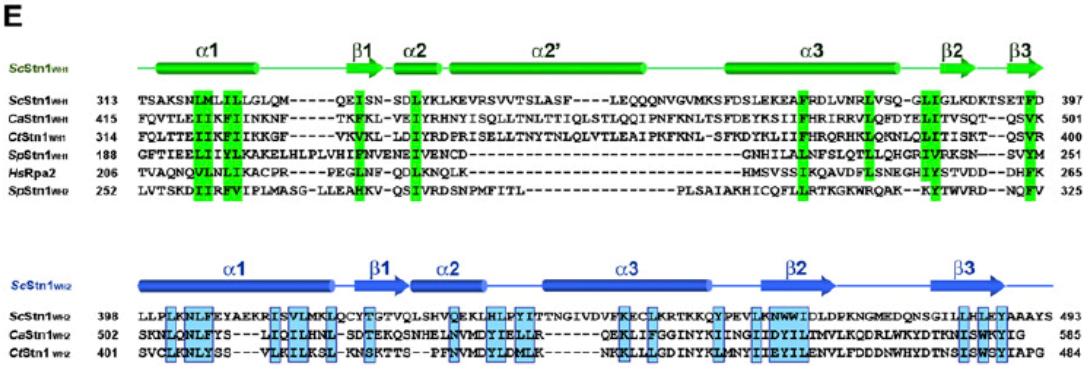
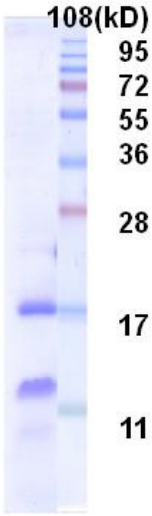


Figure 2.11 *S. pombe* Stn1N and Ten1 form a stable complex.
(A) SDS-PAGE of the *Sp*Stn1N-Ten1 complex
(B) Crystals of *Sp*Stn1N-Ten1

A



B

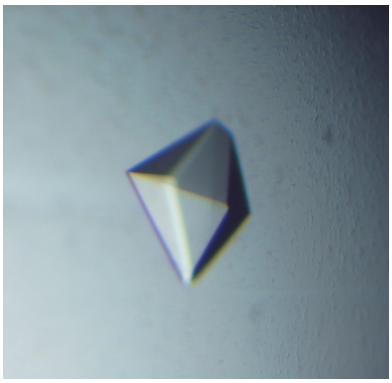


Figure 2.12 Crystal structure of the *S. pombe* Stn1N–Ten1 complex.
(A) Ribbon diagram of the *Sp*Stn1N–Ten1 complex. *Sp*Stn1N and *Sp*Ten1 are colored in pale yellow and sky blue, respectively. The orientation of the complex is the same as that of the left *C. tropicalis* Stn1N–Ten1 complex in Figure 2.4B.
(B) Stereo view of the *Sp*Stn1N–Ten1 interface. *Sp*Stn1N- and *Sp*Ten1-interacting residues are presented as stick models. *Sp*Stn1N and *Sp*Ten1 are colored as in A. The intermolecular hydrogen bonds are shown as dashed magenta lines.

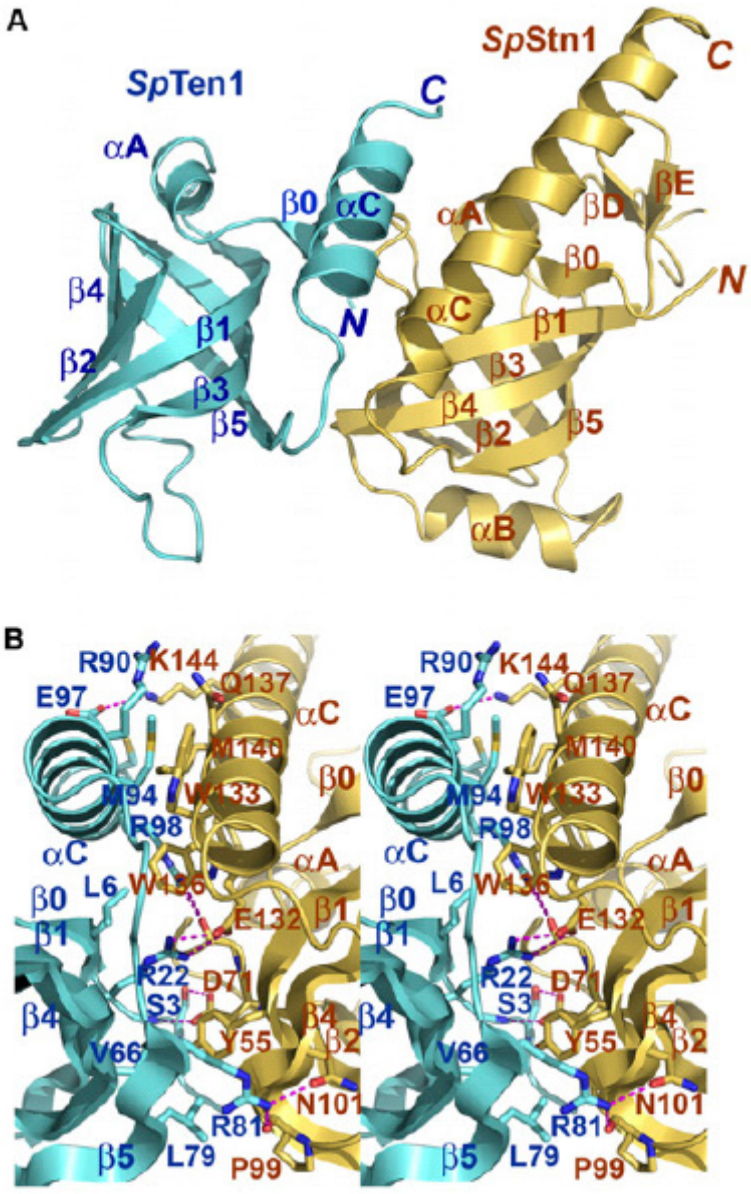


Table 2.1 Data collection, phasing and refinement statistics for *C. tropicalis* Stn1N-Ten1

Sample	***C. tropicalis* Stn1N-Ten1**	
Data collection		
	Hg^{2+} Peak	**Native**
Space group	$P4_12_12$	
Cell dimensions		
a, *b*, *c* (Å)	93.167, 93.167, 200.740	92.072, 92.072, 200.909
α, β, γ (º)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97919	0.97828
Resolution (Å)	50-3.0	100-2.4
R_{merge} (%) (high res. shell)	13.2 (46.0)	7.9 (61.3)
I/σ (high res. shell)	26.7 (5.8)	31.2 (3.1)
Completeness (%) (high res. shell)	91.9 (93.5)	99.8 (99.7)
Redundancy (high res. shell)	10.1 (9.4)	11.0 (8.3)
Phasing		
Acentric Phasing Power	0.832	
Accentric reflections FOM	0.297	
Centric reflections FOM	0.118	
Refinement		
Resolution (Å)		50-2.4
No. reflections		33067
$R_{work/}$ R_{free} (%)		22.8/26.5
B-factors ($Å^2$)		
Protein		48.3
Water		43.1
R.m.s deviations		
Bond lengths (Å)		0.007
Bond angles (º)		1.415

Table 2.2 Data collection, phasing and refinement statistics for *S. cerevisiae* Stn1C

Sample	***S. cerevisiae* Stn1C**	
Data collection		
	Hg^{2+} Peak	**Native**
Space group	$P4_32_12$	
Cell dimensions		
a, *b*, *c* (Å)	52.830, 52.830, 186.683	52.957, 52.957, 186.397
α, β, γ (º)	90, 90, 90	90, 90, 90
Wavelength (Å)	1.008	0.97828
Resolution (Å)	100-2.4	100-2.1
R_{merge} (%) (high res. shell)	7.1 (23.9)	6.3 (24.8)
I/σ (high res. cell)	46.9 (6.7)	34.8(7.7)
Completeness (%) (high res. shell)	98.5 (91.8)	99.3 (96.1)
Redundancy (high res. shell)	14.1 (10.1)	10.4 (8.8)
Phasing		
Acentric Phasing Power	2.09	
Accentric reflections FOM	0.372	
Centric reflections FOM	0.036	
Refinement		
Resolution (Å)		50-2.1
No. reflections		15826
$R_{work/}$ R_{free} (%)		26.3/24.8
B-factors ($Å^2$)		
Protein		43.0
Water		61.7
R.m.s deviations		
Bond lengths (Å)		0.0065
Bond angles (º)		1.024

Table 2.3 Data collection, phasing and refinement statistics for *S. pombe* Stn1N-Ten1

Sample	***S. Pombe* Stn1N-Ten1**	
Data collection		
	Se Peak	**Native**
Space group	$P4_12_12$	
Cell dimensions		
a, *b*, *c* (Å)	93.737, 93.737, 56.385	93.871, 93.871, 56.273
α, β, γ (º)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.9792	0.97828
Resolution (Å)	50-1.53	50-1.63
R_{merge} (%) (high res. shell)	11.7 (66.3)	7.5 (38.8)
I/σ (high res. cell)	48.8 (1.8)	42.3 (3.8)
Completeness (%) (high res. shell)	87.3 (43.7)	97.3 (80.8)
Redundancy (high res. shell)	23.5 (5.2)	13.2 (8.4)
Phasing		
Acentric Phasing Power	5.06	
Accentric reflections FOM	0.624	
Centric reflections FOM	0.182	
Refinement		
Resolution (Å)		50-1.65
No. reflections		29369
$R_{work/}$ R_{free} (%)		24.5/22.9
B-factors (Å^2)		
Protein		28.4
Water		42.1
R.m.s deviations		
Bond lengths (Å)		0.0053
Bond angles (º)		1.2108

Reference:

1. Ferreira, M., K. Miller, and J. Cooper, *Indecent exposure: when telomeres become uncapped.* Mol Cell, 2004. **13**(1): p. p7-18.
2. Bianchi, A. and D. Shore, *How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex.* Mol Cell, 2008. **31**(2): p. 153-65.
3. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
4. Gottschling, D.E. and V.A. Zakian, *Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA.* Cell, 1986. **47**(2): p. 195-205.
5. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**: p. 1171-1175.
6. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.
7. Ye, J.Z., et al., *POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex.* Genes Dev, 2004. **18**(14): p. 1649-54.
8. Xin, H., et al., *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase.* Nature, 2007. **445**(7127): p. 559-62.
9. Garvik, B., M. Carson, and L. Hartwell, *Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint.* Mol. Cell. Biol., 1995. **15**: p. 6128-6138.
10. Grandin, N., C. Damon, and M. Charbonneau, *Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13.* Embo J, 2001. **20**(5): p. p1173-83.
11. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.
12. Petreaca, R.C., et al., *Chromosome end protection plasticity revealed by Stn1p and Ten1p bypass of Cdc13p.* Nat Cell Biol, 2006. **8**: p. 748-755.
13. Bochkarev, A. and E. Bochkareva, *From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold.* Curr Opin Struct Biol, 2004. **14**(1): p. 36-42.
14. Lundblad, V., *Budding yeast telomeres*, in *Telomeres and Telomerase*, T. de Lange, V. Lundblad, and E. Blackburn, Editors. 2006, Cold Spring Harbor Laboratory Press. p. 345-386.
15. Martin, V., et al., *Protection of telomeres by a conserved Stn1-Ten1 complex.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 14038-43.
16. Song, X., et al., *STN1 protects chromosome ends in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19815-20.
17. Grossi, S., et al., *Pol12, the B subunit of DNA polymerase alpha, functions in both telomere capping and length regulation.* Genes Dev, 2004. **18**(9): p. 992-1006.
18. Wold, M.S., *Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism.* Annu Rev Biochem, 1997. **66**: p. 61-92.

19. McEachern, M.J. and E.H. Blackburn, *A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts.* Proc. Natl. Acad. Sci. USA, 1994. **91**: p. 3453-3457.
20. Teixeira, M.T. and E. Gilson, *Telomere maintenance, function and evolution: the yeast paradigm.* Chromosome Res, 2005. **13**(5): p. 535-48.
21. Biswas, K., K.J. Rieger, and J. Morschhauser, *Functional analysis of CaRAP1, encoding the Repressor/activator protein 1 of Candida albicans.* Gene, 2003. **307**: p. 151-8.
22. Enloe, B., A. Diamond, and A. Mitchell, *A single-transformation gene function test in diploid Candida albicans.* J Bacteriol, 2000. **182**(20): p. p5730-6.
23. Fonzi, W. and M. Irwin, *Isogenic strain construction and gene mapping in Candida albicans.* Genetics, 1993. **134**(3): p. p717-28.
24. Legrand, M., et al., *Role of DNA mismatch repair and double-strand break repair in genome stability and antifungal drug resistance in Candida albicans.* Eukaryot Cell, 2007. **6**(12): p. 2194-205.
25. Andaluz, E., et al., *Rad52 depletion in Candida albicans triggers both the DNA-damage checkpoint and filamentation accompanied by but independent of expression of hypha-specific genes.* Mol Microbiol, 2006. **59**(5): p. 1452-72.
26. Iyer, S., A. Chadha, and M. McEachern, *A mutation in the STN1 gene triggers an alternative lengthening of telomere-like runaway recombinational telomere elongation and rapid deletion in yeast.* Mol Cell Biol, 2005. **25**(18): p. p8064-73.
27. Puglisi, A., et al., *Distinct roles for yeast Stn1 in telomere capping and telomerase inhibition.* EMBO J, 2008. **27**(17): p. 2328-39.
28. Neumann, A. and R. Reddel, *Telomerase independent maintenance of mammalian telomeres.*, in *Telomeres and Telomerase*, T. de Lange, V. Lundblad, and E. Blackburn, Editors. 2006, Cold Spring Harbor Laboratory Press. p. 163-198.
29. Nabetani, A. and F. Ishikawa, *Unusual telomeric DNAs in human telomerase-negative immortalized cells.* Mol Cell Biol, 2009. **29**(3): p. 703-13.
30. Bechard, L.H., et al., *Mutant telomeric repeats in yeast can disrupt the negative regulation of recombination-mediated telomere maintenance and create an alternative lengthening of telomeres-like phenotype.* Mol Cell Biol, 2009. **29**(3): p. 626-39.
31. Holm, L. and C. Sander, *Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors.* J Mol Biol, 1991. **218**(1): p. 183-94.
32. Bochkarev, A., et al., *The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding.* EMBO J, 1999. **18**(16): p. 4498-504.
33. Mer, G., et al., *Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA.* Cell, 2000. **103**(3): p. 449-56.
34. Olson, E., et al., *RPA2 is a direct downstream target for ATR to regulate the S-phase checkpoint.* J Biol Chem, 2006. **281**(51): p. 39517-33.
35. Bera, A.K., et al., *Functional dissection of the Bacillus subtilis pur operator site.* J Bacteriol, 2003. **185**(14): p. 4099-109.
36. Komori, H., et al., *Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 A resolution.* EMBO J, 1999. **18**(17): p. 4597-607.

37. Gray, J.T., et al., *Cloning and expression of genes for the Oxytricha telomere-binding protein: specific subunit interactions in the telomeric complex.* Cell, 1991. **67**(4): p. 807-14.
38. Horvath, M.P., et al., *Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA.* Cell, 1998. **95**(7): p. 963-74.
39. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**(5519): p. 1171-5.
40. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.
41. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nat Struct Mol Biol, 2004. **11**(12): p. 1223-9.
42. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
43. Miyake, Y., et al., *RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway.* Mol Cell, 2009. **36**(2): p. 193-206.
44. Surovtseva, Y.V., et al., *Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes.* Mol Cell, 2009. **36**(2): p. 207-18.
45. Wan, M., et al., *OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation.* J Biol Chem, 2009. **284**(39): p. 26725-31.
46. Rost, B., G. Yachdav, and J. Liu, *The PredictProtein server.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W321-6.
47. Casteel, D.E., et al., *A DNA polymerase-{alpha}{middle dot}primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells.* J Biol Chem, 2009. **284**(9): p. 5807-18.
48. Wilson, R.B., D. Davis, and A.P. Mitchell, *Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions.* J Bacteriol, 1999. **181**(6): p. 1868-74.
49. Steinberg-Neifach, O. and N.F. Lue, *Modulation of telomere terminal structure by telomerase components in Candida albicans.* Nucleic Acids Res, 2006. **34**(9): p. 2710-22.
50. Hoffman, C. and F. Winston, *A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli.* Gene, 1987. **57**(2-3): p. p267-72.
51. Yu, E.Y., et al., *A proposed OB-fold with a protein-interaction surface in Candida albicans telomerase protein Est3.* Nat Struct Mol Biol, 2008. **15**(9): p. 985-9.
52. Hsu, M., et al., *Telomerase core components protect Candida telomeres from aberrant overhang accumulation.* Proc Natl Acad Sci U S A, 2007. **104**(28): p. 11682-7.
53. Cohen, S. and S. Lavi, *Induction of circles of heterogeneous sizes in carcinogen-treated cells: two-dimensional gel analysis of circular DNA molecules.* Mol Cell Biol, 1996. **16**(5): p. p2002-14.

54. Brewer, B. and W. Fangman, *The localization of replication origins on ARS plasmids in S. cerevisiae.* Cell, 1987. **51**(3): p. p463-71.
55. Yu, E.Y., et al., *Regulation of telomere structure and functions by subunits of the INO80 chromatin remodeling complex.* Mol Cell Biol, 2007. **27**(16): p. 5639-49.
56. Loayza, D. and T. De Lange, *POT1 as a terminal transducer of TRF1 telomere length control.* Nature, 2003. **423**(6943): p. 1013-8.
57. Moretti, P., et al., *Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1.* Genes Dev, 1994. **8**(19): p. 2257-69.
58. Bricogne, G., *Ab initio macromolecular phasing: blueprint for an expert system based on structure factor statistics with built-in stereochemistry.* Methods Enzymol, 1997. **277**: p. 14-8.
59. Perrakis, A., et al., *ARP/wARP and molecular replacement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 10): p. 1445-50.
60. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
61. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
62. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.
63. Schiltz, M., et al., *Protein Crystallography at Ultra-Short Wavelengths: Feasibility Study of Anomalous-Dispersion Experiments at the Xenon K-edge.* J Synchrotron Radiat, 1997. **4**(Pt 5): p. 287-97.
64. La Fortelle, E.d. and G. Bricogne, *Maximum-likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods*, in *Method in Enzymology*. 1997, Academic Press. p. 472-494.

CHAPTER 3

STRUCTURAL BASES OF DIMERIZATION OF YEAST TELOMERE PROTEIN CDC13 AND ITS INTERACTION WITH THE CATALYTIC SUBUNIT OF DNA POLYMERASE α

3.1 Attributions

This chapter contains the manuscript "Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase α" by J. Sun, Y. Yang, K. Wan, N.H. Mao, T.Y. Yu, Y.C. Lin, D.C. DeZwaan, B.C. Freeman, J.J. Lin, N.F. Lue, and M. Lei published in *Cell Research* (2011) 21: 258–274. Constructs were designed by J. Sun. Mutagenesis was performed by J. Sun. Protein expression, purification and crystallization ware performed by J. Sun and K. Wan. X-ray data collection and structure determination were done by Y. Yang and J. Sun. Sucrose gradient analysis and i*n vivo* yeast genetics experiments were performed by N.H. Mao and T.Y. Yu. The manuscript was written by M. Lei and J. Sun.

3.2 Abstract

Budding yeast Cdc13-Stn1-Ten1 (CST) complex plays an essential role in telomere protection and maintenance and has been proposed to be a telomere specific RPA like complex. Previous genetic and structural studies revealed a close resemblance between

Stn1-Ten and RPA32-RPA14. However, the relationship between Cdc13 and RPA70, the largest subunit of RPA, has remained unclear. Here, we report the crystal structures of multiple OB folds at the N-ends of Cdc13. Although Cdc13 has an RPA70-like domain organization, the structures of Cdc13 OB folds are significantly different from their counterparts in RPA70. Furthermore, our structural and biochemical analyses revealed unexpected dimerization by either the N- or C-terminal OB fold and showed that homodimerization is probably a conserved feature of all Cdc13 proteins. We also uncovered the structural basis of interaction between the Cdc13 N-terminal OB fold and Pol1, and demonstrated a role for N-terminal dimerization in Pol1-binding. Collectively, our findings provide novel insights on the mechanisms and evolution of Cdc13.

3.3 Introduction

Telomeres are specialized nucleoprotein structures that maintain the integrity of eukaryotic chromosomal termini by protecting them from fusion and recombination, and promoting their replication [1, 2]. In most organisms, telomeric DNA consists of short repetitive sequences that terminates in 3' overhangs. Both the double stranded repeats and the 3' overhangs are bound by a multitude of proteins that are crucial for telomere stability. Moreover, because of incomplete end replication, telomeric DNA has to be periodically replenished following rounds of cell division. This task is primarily performed by a ribonucleoprotein (RNP) known as telomerase, which acts as an unusual reverse transcriptase (RT) [1-3]. Both telomere binding proteins and telomerase are critical for the maintenance of telomere integrity through multiple cell divisions, which in turn is pivotal in supporting genome stability and promoting cellular life span.

A key element of the telomere nucleoprotein assembly is the protein complex that binds and protects terminal 3′ overhangs (G-tails). One of the best-studied G-tail binding complex, known as the Cdc13-Stn1-Ten1 (CST) complex, was initially identified and characterized in the budding yeast *S. cerevisiae* [1]. The genes encoding all three components of the complex are essential for cell viability, and hypomorphic alleles of each gene can cause extensive telomere degradation, as well as aberrant telomerase and recombination activities at telomeres. Insights on the mechanisms of this complex have come from analysis of their nucleic acid-binding properties and their interaction partners. Cdc13, the largest subunit, recognizes G-tails with high affinity and sequence specificity through a central OB fold domain [2]. This activity is evidently essential for its capping function [3]. Cdc13 also interacts with the telomerase subunit Est1, thereby promoting the recruitment of the entire telomerase RNP to telomere ends [4, 5]. Another binding partner for Cdc13 is Pol1, the catalytic subunit of pol α-primase complex [5, 6]. Loss of Cdc13-Pol1 interaction is correlated with telomere elongation. The DNA-binding activity of Stn1 and Ten1 are less well characterized [7]. Stn1 also interacts with Pol12, another subunit of the pol α-primase complex, which has likewise been implicated in telomere protection and length regulation [8-10].

Although CST was initially believed to be confined to budding yeast, more recent analyses have revealed broad distribution of the Stn1 and Ten1 components across eukaryotic phyla [7, 11-14]. The discovery of these homologs provided added motivations for ascertaining their mechanisms and the extent of their evolutionary conservation. A particularly provocative notion that emerged was the proposal that CST represents a telomere-specific replication protein A (RPA)-like complex [12]. RPA is a

nonspecific single-stranded DNA-binding complex that contains three subunits (RPA70, RPA32, and RPA14) and mediates critical and diverse DNA transactions throughout the genome [15, 16]. Structural studies provided compelling support for the resemblance between Stn1 and RPA32, and that between Ten1 and RPA14 [17, 18]. The two protein pairs share many structural features and utilize similar motifs for mutual interactions. Stn1 and RPA32, each consists of an N-terminal OB fold and one or two C-terminal WH motifs, whereas Ten1 and RPA14 each consists of a single OB fold. Complex formation in each case is mediated predominantly through α-helices located at the C-termini of OB folds. Thus, the Stn1-Ten1 subcomplex can plausibly be viewed as a telomere specific paralog of the RPA32-RPA14 complex. That Stn1 and Ten1 together act as a close-knit unit is further underscored by their ability to function in the absence of Cdc13. Over-expression of Stn1N (the N-terminal OB fold of Stn1) and Ten1 allows the cells to bypass the essential function of Cdc13 and remain viable [14]. By contrast, even though Cdc13 and RPA70 are both large proteins that have either been shown or proposed to contain multiple OB folds, their evolutionary kinship is less clear [16, 19]. Sequence comparison failed to disclose any convincing similarity between the two families, and the DNA-binding OB fold of Cdc13 does not appear to be closely related to the equivalent OB folds in RPA70 [20, 21].

In this chapter, I will provide structural and biochemical analyses of the N-terminal domain of Cdc13 in detail. The atomic resolution structure confirmed the existence of an OB fold at the N-terminal end of Cdc13. Both structural and biochemical analyses revealed unexpected dimerization by the N-terminal OB fold. I also uncovered the structural basis of interaction between the N-terminal OB fold and Pol1, and

demonstrated a role for N-terminal dimerization in Pol1 binding. Analysis of the phenotypes of mutants defective in Cdc13 dimerization and Cdc13-Pol1 interaction revealed multiple mechanisms by which dimerization regulates telomere lengths *in vivo*. Our findings thus offer novel insights into Cdc13 mechanisms and evolution.

3.4 Prediction of four tandem OB-fold domains in Cdc13

To initiate a comparative analysis of Cdc13 and to uncover possible structural domains in this protein, we systemically searched the NCBI and Broad Institute databases for homologs of Cdc13 using available sequences as queries. This resulted in the identification of many Cdc13 homologs in the *Saccharomyces* and *Kluyveromyces* branches of budding yeast (which also include *Candida glabrata*, but not other *Candida* spp.; Figure 2.1). Multiple sequence alignment of these Cdc13 proteins clearly revealed a pattern of four conserved regions, each of which spans about 150-200 residues (Figure 2.1). These regions in *Saccharomyces cerevisiae* Cdc13 (*Sc*Cdc13) consist of residues 1-231, 323-485, 490-701, and 712-924, respectively (Figure 2.1). Notably, the third conserved region coincides with the DNA-binding domain of *Sc*Cdc13 (*Sc*Cdc13$_{DBD}$) [21]. For simplicity, hereafter, *Sc*Cdc13 is referred to as Cdc13.

I next performed a secondary structural analysis on the four conserved regions of Cdc13 using the program PredictProtein [22]. Supporting the validity of this approach, the program accurately predicted the positions of most of the α-helices and β-strands in Cdc13$_{DBD}$ (Figure 2.1). This analysis also predicted that each of the three remaining regions contains a β-strand-rich core that exhibits a secondary structure pattern of β−β−β−α−β−β (Figure 2.1), which is characteristic of OB folds found in many telomere

proteins including Stn1 and Ten1 [23]. Sequence analyses of several Cdc13 proteins from other yeast species also predicted the existence of four β-strand-rich OB-fold-like domains (Figure 2.1). The less-conserved fragment between the first and the second putative OB folds (~90 resides) exhibited few detectable features of secondary structure (Figure 2.1). Notably, this region, called the recruitment domain (RD), has been reported to play an important role in telomerase recruitment through a direct interaction with Est1 (Figure 3.1) [9, 10].

The largest subunit of the RPA complex, RPA70, also contains four tandem OB-fold domains (Figure 3.1) [20, 23-25]. Furthermore, there is also a ~60-residue unstructured region between the first and second OB folds in RPA70 (Figure 3.1). Both features match well with our bioinformatic analysis of Cdc13 (Figure 3.1 and 2.1). Thus, although there is no primary sequence similarity between Cdc13 and RPA70, the similar domain organization of the two proteins supports the view that Cdc13 is a telomere-specific RPA70-like protein. However, because OB folds are well known for the absence of reliable primary sequence features that can be used for accurate prediction [26-28], decisive confirmation of the existence of four tandem OB folds in Cdc13 and the similarity between Cdc13 and RP70 requires structural characterization of Cdc13.

3.5 Structure of a $Cdc13_{OB1}$ monomer

To address whether Cdc13 contains an OB fold at the N-terminus, recombinant $Cdc13_{OB1}$ (residues 12-243) expressed from *Escherichia coli* was crystallized (Figure 3.2 A and D), and the structure was determined by single anomalous dispersion (SAD) using a mercury compound (MeHgAc) at a resolution of 2.5 Å (Table 3.1). The final atomic model,

refined to an R-value of 21.1% (R_{free} = 26.7%), contains residues 14-225. No electron density is observed corresponding to three loop regions (residues 59-67, 105-111, and 161-170), as well as the C terminal 18 residues, which I presume to be disordered in solution.

The crystal structure demonstrates that the core of $Cdc13_{OB1}$ is indeed made up of an OB fold, consisting of a highly curved five-stranded antiparallel β-barrel with three peripheral α-helices, as expected from our sequence analysis (Figure 3.3A). $Cdc13_{OB1}$ contains a large insertion between helix αB and strand β4 (residues 97-124), part of which forms a short β-strand (β3′) that runs antiparallel to β1 before rejoining to β4. In addition, there is a three-helix bundle at the C-terminus, which packs against the convex side of the β-barrel.

Compared with $Cdc13_{OB1}$, the N-terminal OB fold of RPA70 (RPA70N) only contains a β-barrel core and lacks the C-terminal helix bundle (Figure 3.3B); the size of RPA70N (120 residues) is only about half of that of $Cdc13_{OB1}$ (225 residues). Although the sequences of $Cdc13_{OB1}$ and RPA70N are markedly divergent and share only 8% identity, the β-barrel core of the $Cdc13_{OB1}$ closely resembles that of RPA70N (Figure 3.3B); the two domains can be superimposed with a root-mean-square deviation (r.m.s.d.) of 3.7 Å for 84 equivalent pairs (Figure 3.3B) [23, 29]. Notwithstanding this similarity, there are substantial structural differences evident in the loop and helix regions. Most notably, the αB helix between strands β3 and β4 of $Cdc13_{OB1}$ is much longer and rotates about 45° away from strand β5 relative to the position of αB in RPA70N, resulting in a large hydrophobic groove between αB and β5 (Figure 3.3B). This

displacement of helix αB is essential for the dimeric conformation of $Cdc13_{OB1}$, as described below (Figure 3.3C).

Unexpectedly, the structure of $Cdc13_{OB1}$ closely resembles that of $Cdc13_{OB3}$ (DBD) (Figure 3.4A) [21]. Indeed, an unbiased search for structurally homologous proteins using the Dali server [30] revealed that the structure of $Cdc13_{OB1}$ is most similar to that of $Cdc13_{OB3}$, with a Z-score of 10.3; the two domains can be superimposed with an r.m.s.d. of 3.0 Å for 144 equivalent Cα pairs (Figure 3.4A). However, $Cdc13_{OB3}$ has a very long loop (28 residues), L23, between strands β2 and β3, which packs on one side of the β-barrel and constitutes almost half of the DNA-binding surface (Figure 3.4A) [21, 31]. In contrast, strands β2 and β3 of $Cdc13_{OB1}$ are connected by a much shorter loop (12 residues) that is partially disordered in the current structure (Figure 3.3A).

3.6 Cdc13 is a dimer

In the $Cdc13_{OB1}$ crystals, only one $Cdc13_{OB1}$ molecule is present in each asymmetric unit. However, careful examination of the crystal packing of one protomer against its neighbors revealed that $Cdc13_{OB1}$ makes extensive interactions with one of the crystallographic symmetry related molecules. The two αB helices from both molecules form a tightly packed parallel coiled-coil, whose axis coincides with a crystallographic symmetry dyad (Figure 3.3C). The $Cdc13_{OB1}$ dimer interface buries a total of ~2560 $Å^2$ solvent-accessible surface area, which is substantially larger than other crystal-packing contacts. This strongly implies that the dimeric conformation observed in the crystals is unlikely to be the result of lattice packing. I next asked whether Cdc13 forms a homodimer in solution. Experiments using calibrated gel-filtration chromatography

showed that the elution peak of $Cdc13_{OB1}$ corresponded to a molecular weight of about 45 kDa (Figure 3.3D), as expected if the crystallographic dimer interaction is present in solution. In addition, chemical cross-linking assays with both the OB1 domain and full-length Cdc13 demonstrated that, in both cases, only one higher-molecular-weight band appeared in the presence of cross-linking reagent and the size of this band matched well with a dimer of $Cdc13_{OB1}$ or full-length Cdc13, respectively (Figure 3.4B, 3C and 3D). These results corroborated our crystallographic finding, showing that Cdc13 indeed exists as a dimer in solution. The molecular weight of purified full-length Cdc13 was also estimated by sucrose gradients. Cdc13 expressed and purified from insect cells behaves as an assembly with an apparent molecular weight of ~160-170 kDa (Figure 3.3E). Even though this is less than the expected value of a Cdc13 dimer (210 kDa), it is consistent with our prediction that Cdc13 has a multidomain elongated architecture, which should result in a smaller sedimentation coefficient and thus a reduced apparent molecular mass. To further study the *in vivo* oligomeric state of Cdc13, we tested the dimeric interaction of Cdc13 in yeast cells. Co-immunoprecipitation (Co-IP) experiments with two differently tagged fulllength Cdc13 proteins demonstrated that Cdc13 indeed forms a complex with itself in cells (Figure 3.3F). Finally, we examined the potential role of other Cdc13 domains in dimerization by yeast two-hybrid assays (Figure 3.4E). Self-association was not observed for any other domains, indicating that Cdc13 probably forms a homodimer solely through its N-terminal OB fold.

3.7 The dimer interface of $Cdc13_{OB1}$

The core of the symmetric dimer interface is mediated primarily by helix αB and strand β5 from both $Cdc13_{OB1}$ subunits (Figure 3.3C). Together, αB and β5 from one subunit form a hydrophobic groove that accommodates the αB helix from the other (Figure 3.5A). At one side of the groove, the coiled-coil hydrophobic packing contact between the two αB helices is extensive, consisting of four layers of two-fold symmetry-related interdigitating residues at positions a and d of the heptad repeats from both helices (Ser81, Leu84, Leu91, and Tyr95) (Figure 3.5B). These residues stack closely against each other both within and between adjacent layers. In addition, several hydrophobic residues (Phe142, Leu143, Ile146, and Pro148) of β5 from one monomer make close contacts with helix αB from the opposing monomer so that, except for the two termini, helix αB is almost completely buried into the central core of the dimer (Figure 3.5B and Figure 3.6).

Although the dimeric interface is predominantly hydrophobic, intermolecular electrostatic interactions provide additional specificity and stability to the dimer. In the loop regions before the αB helices in both monomers, two symmetry-related Lys77-Asp78 pairs contribute four salt bridges, sealing one end of the interface (Figure 3.5B). In the center of the coiled-coil, two Thr88 residues form an intermolecular hydrogen bond instead of hydrophobic contacts at position *a* of the heptad (Figure 3.5B). At the side of helix αB, away from the coiled-coil interface, the hydroxyl group of Ser90 mediates an electrostatic interaction with Asp145 from strand β5 of the opposing $Cdc13_{OB1}$ monomer, helping anchor the αB helix into the hydrophobic groove (Figure 3.5B). Besides the helix αB binding groove, we also observed a second smaller interface between the two monomers (Figure 3.5C). Two acidic residues Asp102 and Asp104 in the loop region

between αB and β3 from one monomer form an extensive electrostatic network containing a total of six salt bridges with the side chains of Arg15 and Lys129 from the other monomer (Figure 3.5C).

To confirm the significance of the dimeric contacts observed in the crystal structure, we generated four missense mutations in $Cdc13_{OB1}$. All mutant proteins were purified to homogeneity, and the oligomeric states of these proteins were individually analyzed by gel filtration chromatography (Figure 3.5D). Consistent with the structure, substitution of Ile87, Leu91, or Tyr95 of $Cdc13_{OB1}$ at the hydrophobic interface with a positively charged and bulky arginine residue completely disrupted the dimeric state of the wild-type protein; the elution profiles of these three mutants shifted toward the monomer species on gel filtration (Figure 3.5D). Notably, the L84R mutant had an elution peak between those of the wild-type $Cdc13_{OB1}$ and the monomer mutants, suggesting that this mutant only weakened but did not disrupt the dimeric interface (Figure 3.5D). The effects of these mutants were also confirmed by yeast two-hybrid and Co-IP analyses in yeast cells (Figure 3.5E and 4F). Taken together, we therefore conclude that hydrophobic contact is the major driving force for dimer formation of Cdc13, both *in vitro* and *in vivo*.

3.8 Cdc13 dimerization affects cell growth and telomere length regulation

To determine if dimerization affects the function of Cdc13 *in vivo*, we used a plasmid shuffling system developed previously to study the *in vivo* consequences of Cdc13 mutations [5, 6]. We generated yeast strains that carried nondimeric alleles of CDC13. These alleles contained either a single (L91R) or quadruple (4R:

L84R/I87R/L91R/Y95R) mutations shown earlier to disrupt the OB1 dimer interface. Gel-filtration profile showed that the quadruple mutant protein was well folded and adopted a monomeric conformation in solution (Figure 3.7C). Both proteins were expressed at near wild-type levels in yeast cells, suggesting that residues at the Cdc13 dimeric interface are not required for protein stability. Interestingly, these strains exhibited no apparent growth defects in comparison to the wild-type control at 30 °C, but manifested a moderate reduction in growth at 37 °C (Figure 3.7A). Cdc13 dimerization is thus not essential for cell viability, but appears to promote its function at higher temperatures. Analysis of telomere lengths in both mutant clones revealed a consistent and moderate reduction in average telomere lengths (by ~150 bp) (Figure 3.7B). This reduction was observed about 40 generations following the eviction of plasmids carrying wild type CDC13, and was stable thereafter (data not shown). Collectively, we conclude that Cdc13 dimerization is not essential for cell viability, but is critical for telomere length regulation.

3.9 Characterization of the Cdc13-Pol1 interaction

Although $Cdc13_{OB1}$ is structurally most similar to $Cdc13_{OB3}$ and also contains a basic cleft that corresponds to the canonical nucleic acid-binding pocket of OB folds, $Cdc13_{OB1}$ does not possess DNA-binding activity. Instead, it has been reported to mediate protein-protein interactions at telomeres [5, 6]. One of the $Cdc13_{OB1}$-binding protein is Pol1, the catalytic subunit of DNA polymerase α-primase complex. Disruption of the Cdc13-Pol1 interaction causes cell growth defect and telomere lengthening [5, 6]. An N-terminal region of Pol1 (residues 13-392 reported in one study and residues 47-560 in another)

interacts with $Cdc13_{OB1}$ [5, 6]. To determine the mechanism of Pol1 recognition by Cdc13, we characterized the Cdc13-Pol1 interaction by isothermal titration calorimetry (ITC) (Figure 3.8A). Our data revealed that a short fragment of Pol1 consisting only of residues 215-250 was necessary and sufficient for binding with $Cdc13_{OB1}$ (Figure 3.8A). $Cdc13_{OB1}$ binds to $Pol1_{215-250}$ with an equilibrium dissociation constant (Kd) of 3.8 μM (Figure 3.8B). Hereafter, we will refer to $Pol1_{215-250}$ as $Pol1_{CBM}$ (*C*dc13-*b*inding *m*otif).

3.10 Structural basis for the $Cdc13_{OB1}$-$Pol1_{CBM}$ interaction

To characterize the structural basis of Pol1 recognition by Cdc13, we crystallized the $Cdc13_{OB1}$-$Pol1_{CBM}$ complex and solved its structure by molecular replacement at a resolution of 2.4 Å (Figure 3.2 B and C, Table 3.1). Except for one residue at the N-terminus and five residues at the C-terminus, $Pol1_{CBM}$ is well ordered, as evidenced by good electron density in the crystals and low temperature factors in the final atomic model. The complex structure has been refined to an R-value of 22.4% (R_{free} = 26.4%) with good geometry. The $Cdc13_{OB1}$-$Pol1_{CBM}$ complex structure exhibits a 2:2 stoichiometry between $Cdc13_{OB1}$ and $Pol1_{CBM}$ (Figure 3.8C). Each $Pol1_{CBM}$ peptide is folded into a single amphipathic α-helix that binds into the deep basic groove mostly formed by one $Cdc13_{OB1}$ monomer (Figure 3.8C). The $Cdc13_{OB1}$-$Pol1_{CBM}$ interaction does not interfere with the dimeric interface of $Cdc13_{OB1}$ (Figure 3.8C). The formation of the binary complex causes the burial of ~1997 $Å^2$ of surface area at the interface.

Strikingly, the binding mode of $Pol1_{CBM}$ to $Cdc13_{OB1}$ resembles the interaction between RPA70N and p53 (Figure 3.8D) [23]. In both complexes, a short fragment of

one protein ($Pol1_{CBM}$ and $p53_{38-57}$) adopts a helical conformation and binds into the basic groove of the OB fold of the other component in the complex ($Cdc13_{OB1}$ or RPA70N). Notably, canonical ssDNA-binding OB folds employ exactly the same basic groove for DNA association, as illustrated by the structure of the $Cdc13_{OB3}$-ssDNA complex (Figure 3.8E) [31]. In these structures, both basic and aromatic residues on the ssDNA-binding grooves are required for the interaction; basic residues stabilize the negative phosphate groups of the DNA backbone, whereas aromatic residues are involved in stacking with the bases of the DNA [20, 31-34]. In comparison, although the $Pol1_{CBM}$-binding surface of $Cdc13_{OB1}$ contains many basic residues, there are very few aromatic residues at the expected positions for optimal ssDNA interaction. This is consistent with our data that even at a high protein concentration (~0.5 mM), no $Cdc13_{OB1}$-ssDNA complex was observed in an Electrophoretic Mobility Shift Assay. Thus, we conclude that the N-terminal OB fold of Cdc13 is a protein-protein interaction module.

3.11 The $Cdc13_{OB1}$-$Pol1_{CBM}$ interface

In the $Cdc13_{OB1}$-$Pol1_{CBM}$ complex structure, the two $Pol1_{CBM}$ peptides adopt symmetric conformations and each $Pol1_{CBM}$ interacts with both $Cdc13_{OB1}$ molecules in the dimer (Figure 3.8C). The C-terminal half of $Pol1_{CBM}$ contacts with one $Cdc13_{OB1}$ monomer and this interaction is primarily mediated by a highly positively charged cleft of $Cdc13_{OB1}$ dimer and a negatively charged convex surface of the $Pol1_{CBM}$ helix (Figure 3.9A). The acidic surface of $Pol1_{CBM}$ at the interface contains five negatively charged residues, Asp229, Asp232, Asp235, Asp236, and Glu238 (Figure 3.9A). The more extensive basic groove of $Cdc13_{OB1}$ consists of six lysine residues at positions 30, 50, 73, 75, 77, and 135

(Figure 3.9A). These two surfaces are not only opposite in charge distribution but also complementary in shape. While electrostatic interactions should favor the initial apposition of the two proteins, the interaction specificity between $Cdc13_{OB1}$ and $Pol1_{CBM}$ is mainly provided by van der Waals contacts (Figure 3.9B). The hydrophobic portion of the amphipathic helix of $Pol1_{CBM}$ packs against the hydrophobic floor of the groove formed by strands β1, β4, and β5 of $Cdc13_{OB1}$, accounting for about half of the total buried surface area (Figure 3.9B). The core of this hydrophobic interface consists of the side chains of eight residues, Val230, Leu233, Leu234, and Val237 in $Pol1_{CBM}$, and Ile32, Tyr133, Thr140, and Phe143 in $Cdc13_{OB1}$ (Figure 3.9B). In addition to the helix, the C-terminal tail of $Pol1_{CBM}$ also contributes to the binding to $Cdc13_{OB1}$; it makes a turn at Pro241 and lines the rest of $Pol1_{CBM}$ in an antiparallel direction to strand β5 of $Cdc13_{OB1}$ (Figure 3.9B). The side chains of Val242 and Val243 pack against a hydrophobic patch of $Cdc13_{OB1}$ formed by residues from strands β3 and β5 (Figure 3.9B). This conformation is further stabilized by four hydrogen-bonding interactions between $Pol1_{CBM}$ and $Cdc13_{OB1}$ (Figure 3.9B).

The N-terminal half of the $Pol1_{CBM}$ helix (Pro216-Asp229) protrudes outside the major $Cdc13_{OB1}$-$Pol1_{CBM}$ interface to make direct contacts with the other $Cdc13_{OB1}$ molecule in the dimer (Figure 3.9B and 8C). In this region of the complex, the $Cdc13_{OB1}$-$Pol1_{CBM}$ interface is also dominated by electrostatic interactions; there are a total of seven salt-bridge and hydrogen-bonding interactions between $Pol1_{CBM}$ and $Cdc13_{OB1}$ (Figure 3.9B and 8C). Based on the structure, disruption of the dimeric state of $Cdc13_{OB1}$ would result in a loss of ~596 $Å^2$ of the buried interface area between $Cdc13_{OB1}$ and $Pol1_{CBM}$, suggesting that dimerization of $Cdc13_{OB1}$ might be important for $Pol1_{CBM}$ interaction.

3.12 Both the Cdc13-Pol1 interface and Cdc13 dimerization are required for the Cdc13-Pol1 interaction

Our structural analysis provides plausible explanations for previous mutagenesis data of the Cdc13-Pol1 interaction. Two point mutations of Pol1, D236N and P241T, were reported to abolish the interaction [5]. In the crystal structure, the side chain of Pol1 Asp236 points toward the interface and makes two salt bridges with the amino group of Cdc13 Lys73, whereas the unusual backbone dihedral angles of Pol1 Pro241 allows the C-terminus of $Pol1_{CBM}$ to align with Cdc13 strand β5 for optimal interaction (Figure 3.9B). A third mutation of Pol1, E238K, weakened but did not abolish the interaction [5]. This is also consistent with the structure: the side chain of Glu238, exposed to the solvent, contributes only one hydrogen-bonding interaction (Figure 3.9B).

To further examine the significance of the $Cdc13_{OB1}$-$Pol1_{CBM}$ interface, we assessed the effects of an additional panel of mutations in either $Cdc13_{OB1}$ or $Pol1_{CBM}$ using ITC. In support of the crystal structure, $Cdc13_{OB1}$ mutations of either the hydrophobic residues (Ile32, Val133, Thr140, or Phe142) at the bottom of the groove or the basic residues (K73E/K75E/K77E, R79E, and R83E) at the periphery were sufficient to eliminate the interaction (Figure 3.9D). Similarly, mutations of the hydrophobic or acidic residues of Pol1CBM on the other side of the interface also completely abolished the interaction (Figure 3.9D). Taken together, we conclude that both the electrostatic and hydrophobic interactions observed in the crystal structure are important for the interaction between $Cdc13_{OB1}$ and $Pol1_{CBM}$.

Notably, mutations of residues in both Cdc13 and Pol1 ($Pol1_{CBM}$ D229R, and $Cdc13_{OB1}$ R79E and R83E) at the interface between $Pol1_{CBM}$ and the second $Cdc13_{OB1}$ molecule in the dimer were also able to completely disrupt the $Cdc13_{OB1}$-$Pol1_{CBM}$ interaction (Figure 3.9D). This observation promoted us to examine the role of Cdc13 dimerization in Pol1 binding in solution. As shown in Figure 3.9E, all four monomeric mutants of $Cdc13_{OB1}$ exhibited complete or partial loss of Pol1 association in a manner that is entirely consistent with the severity of the dimerization defects (Figure 3.3D, 2E and 2F). In particular, the L84R mutant, which retained partial function in dimerization, also exhibited the mildest Pol1 association defect (Figure 3.9E). Therefore, we conclude that Cdc13 dimerization is a prerequisite for the stable association between Cdc13 and Pol1.

3.13 Loss of the Cdc13-Pol1 interaction, but not Cdc13 dimerization, results in telomere lengthening

Previous investigations demonstrated that loss of the Cdc13-Pol1 interaction by substitution of wild-type Pol1 with Cdc13-binding-deficient mutants was often correlated with telomere lengthening [5, 6]. The telomere shortening phenotype of the dimerization-deficient CDC13 mutants was thus somewhat surprising, given the mutant's lack of Pol1 binding (Figure 3.7B). One explanation for this apparent discrepancy is that the dimerization of Cdc13 not only disrupts the Cdc13-Pol1 interaction but may also affect the binding of Cdc13 to other partners such as Imp4 and Sir4. We predicted that Cdc13 mutations that only disrupt the Cdc13-Pol1 interface but not the dimerization of Cdc13 would cause telomere lengthening, similar to the phenotype caused by the Pol1 mutants

[5, 6]. To test this idea, we introduced several mutations in Cdc13 to reduce Pol1 binding (I32E, V133E, and K73E/K75E/K77E (3K-3E)) and analyzed the telomere length phenotypes of the resulting mutants. Results from Dr Lue's group showed that strains carrying these cdc13 mutants grew as well as wild-type cells at 25 °C, 30 °C, and 37 °C. Thus, none of the mutant alleles eliminated an essential function of Cdc13. Notably, as we predicted, all three mutants yielded longer telomeres, similar to those caused by the Cdc13-bindingdeficient mutants of Pol1 (Figure 3.9F) [5, 6]. The differences in telomere lengths are unlikely to be caused by differences in the abundance of Cdc13 in cells, as western analysis showed that each of the mutant alleles produced nearly wild-type levels of Cdc13. Clearly, disruption of Cdc13 dimerization caused defects that are distinct from the disruption of Cdc13-Pol1 interface. We therefore suggest that dimerization is likely to affect at least one other function or interaction mediated by Cdc13. Indeed, many other interaction partners for Cdc13 have been identified, and knowing the effect of dimerization on each interaction will be necessary to fully understand the role of dimerization on Cdc13 function.

3.14 Dimerization is a conserved feature of Cdc13 proteins

Multiple sequence alignment revealed a high degree of conservation in most of the residues important for homodimerization of *Sc*Cdc13$_{OB1}$, suggesting that dimerization through the first OB fold is probably conserved for *Saccharomyces* and *Kluyveromyces* Cdc13 proteins (Figure 2.1). To test this idea, we examined the oligomeric state of different OB fold domains of *Kluyveromyces lactis* Cdc13 (*Kl*Cdc13). Even though the putative dimerization interface of *Kl*Cdc13 only shares modest sequence similarity with

*Sc*Cdc13, yeast two-hybrid experiments clearly revealed self-interaction by the N-terminal OB1 domain of *Kl*Cdc13 (Figure 3.10A), strongly supporting the notion that dimerization is a conserved feature of *Saccharomyces* and *Kluyveromyces* Cdc13 proteins.

A notable standout in our sequence alignment was *Candida glabrata* Cdc13 (*Cg*Cdc13), whose OB1 domain has a shorter αB helix and does not contain the conserved residues for dimerization (Figure 2.1). (It should be noted that *Candida glabrata*, despite its name, is evolutionarily closer to *Saccharomyces* and *Kluyveromyces* than other *Candida* spp.) In keeping with the alignment, *Cg*Cdc13$_{OB1}$ failed to self-associate in the yeast two-hybrid assay (Figure 3.10B). Strikingly, the predicted C-terminal OB fold of *Cg*Cdc13, *Cg*Cdc13$_{OB4}$, exhibited a strong self-association activity (Figure 3.10B). By contrast, both *Sc*Cdc13$_{OB4}$ and *Kl*Cdc13$_{OB4}$ behaved as monomers in yeast cells (Figure 3.4E and Figure 3.10A). To further assess the dimerization of the OB folds in different Cdc13 proteins in vitro, recombinant *Kl*Cdc13$_{OB1}$, *Cg*Cdc13$_{OB1}$, and *Cg*Cdc1$_{OB4}$ proteins were purified and individually subjected to gel-filtration chromatography. As shown in Figure 3.10C, the apparent molecular weights of these domains, based on the gel-filtration profiles, are entirely consistent with the yeast two-hybrid results.

Our previous studies showed that Cdc13 homologs in many *Candida* spp. are considerably smaller and lack the N-terminal half of their *S. cerevisiae* counterpart [23]. These *Candida* spp. cluster evolutionarily and form a well-defined clade (Figure 3.11). Sequence alignments suggest that these Cdc13 homologs only contain two OB folds, which correspond to OB3 and OB4 in *Saccharomyces* spp. Cdc13 proteins [23, 40].

Therefore, in keeping with the nomenclature of Cdc13, we refer to the two OB folds of *Candida* Cdc13 proteins as OB3 and OB4, respectively. Given that these smaller Cdc13 proteins lack OB1, we hypothesized that like *Cg*Cdc13, they might form dimeric structures through their OB4 domains. Hence, we examined the oligomeric states of the two OB folds of *Candida albicans* Cdc13 (*Ca*Cdc13). As predicted, *Ca*Cdc13$_{OB4}$, but not the putative DNA-binding domain *Ca*Cdc13$_{OB3}$, associated with itself (Figure 3.10D). Taken together, we propose that homodimerization is likely to be a conserved feature of Cdc13 proteins in all yeast species in the *Saccharomycotina* linage; except for *Cg*Cdc13, *Saccharomyces* like large Cdc13 proteins form dimers through their N-terminal OB1 domains, whereas *Candida*-like small Cdc13 proteins and *Cg*Cdc13 form dimers through their C-terminal OB4 domains.

3.15 Discussion

It has been proposed that CST is a telomere-specific RPA-like complex [7]. Recent structural studies by us and other groups demonstrated a close structural resemblance between Stn1-Ten1 and RPA32-RPA14 [17, 18]. Although the solution structure of the DNA-binding OB fold of Cdc13 is available, the relationship between Cdc13 and RPA70 remains unclear due to the lack of structural information on other regions of Cdc13 and the lack of sequence similarity between Cdc13 and RPA. In this work, our bioinformatic and structural analyses provide the first direct evidence for the existence of multiple OB folds in Cdc13, which is characteristic of RPA70. The similarity between the Cdc13$_{OB1}$-Pol1$_{CBM}$ and the RPA70N-p53 complexes further extends the parallel between Cdc13 and RPA70 (Figure 3.8C and D). However, despite these similarities, there are substantial

differences between Cdc13 and RPA70. First, unlike Stn1-Ten1, none of the two structurally defined OB folds of Cdc13 show similarity to their counterparts in RPA70 outside the central β-barrel cores (Figure 3.3B) [20, 21]. Second, the two central OB folds of RPA70 are required for efficient DNA binding, whereas Cdc13 uses just its OB3 for binding [35]. These marked differences suggest that the resemblance between Cdc13 and RPA70 may be the result of convergent evolution. In other words, Cdc13 may not have evolved from the ancestral RPA70, but were instead recruited by the Stn1-Ten1 complex to provide single-stranded DNA-binding activity. In keeping with this idea, we found that Candida spp. Cdc13 proteins contain only two OB folds that correspond to the C-terminal half of Saccharomyces spp. proteins. In addition, the recently identified CTC1 proteins, the largest components in the human and plant CST complexes, are much larger proteins and show no sequence similarity to either Cdc13 or RPA70, supporting the disparate origins of these proteins [13, 14]. While we cannot rule out the possibility that a common origin for these proteins is obscured by extremely rapid evolutionary divergence, it seems clear that the structural and functional relationships between Cdc13/CTC1 and Stn1-Ten1 are quite distinct from those between RPA70 and RPA32-14.

One striking result of this study is that homodimerization appears to be a conserved feature of Cdc13. Except for *Cg*Cdc13, most *Saccharomyces* and *Kluyveromyces* Cdc13 proteins form dimers through their N-terminal OB1 domains. In contrast, homodimerization of *Candida* Cdc13 proteins and *Cg*Cdc13 is mediated by the C-terminal OB fold. The use of OB4 for dimerization by *Cg*Cdc13 is somewhat surprising, given the closer kinship of this yeast to *Saccharomyces* than to *Candida* spp. Perhaps this represents another case of convergent evolution. For example, an accidental

loss of OB1 dimerization by *Cg*Cdc13 may have provided the selection pressure for the evolution of other dimerization mechanisms, resulting eventually in the utilization of OB4. The prevalence of Cdc13 dimerization suggests that this property may facilitate interaction of Cdc13 with multiple targets. For example, one established function of OB1 dimerization is to facilitate the interaction with Pol1; our mutagenesis data clearly showed that dimerization of *Sc*Cdc13 OB1 domain is required for Pol1 binding. The significance of OB4 dimerization is less clear. A possible function for the dimerization of this domain is suggested by the homodimerization of many telomere binding proteins such as fission yeast Taz1 and human TRF1 and TRF2 [36-39]. Because of the low intrinsic affinity of individual DNA-binding domains, these proteins require dimerization for stable telomere DNA interaction [37, 38]. Thus, even though the *S. cerevisiae* Cdc13 can clearly bind DNA as a monomer, it is possible that dimerization of the smaller Cdc13 proteins in Candida spp. may enhance their DNA-binding activity. Indeed, we found recently that the OB_{DBD} of *Ct*Cdc13 interacts weakly with the cognate telomere repeat and requires the OB4 domain for high-affinity DNA binding [40]. Yet another potential function for Cdc13 dimerization is suggested by the reported multimerization of the telomerase complex. Although the data are somewhat inconclusive, both yeast and human telomerase have been proposed to function as dimers [41, 42]. Because Cdc13 is known to interact with the Est1 component of yeast telomerase, dimerization of Cdc13 could help bring two telomerase complexes into close vicinity for proper function. Further studies are needed to test these possibilities and reveal the full functional significance of Cdc13 dimerization in regulating and maintaining budding yeast telomeres.

3.16 Materials and Methods

Protein expression and purification

S. cerevisiae $Cdc13_{OB1}$ (residues 12-243) and $Pol1_{CBM}$ (residues 215-250) were cloned into a modified pET28b vector with a Sumo protein fused at the N-terminus after the His6 tag [49]. They were expressed in *E. coli* BL21 (DE3). After induction for 16 h with 0.1 mM IPTG at 20 °C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl (pH 8.0), 50 mM NaH_2PO4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol, and homemade protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 2 h at 4 °C before elution with 250 mM imidazole. Then, Ulp1 protease was added to remove the His6-Sumo tag for 12 h at 4 °C. $Cdc13_{OB1}$ was then further purified by passage through Mono-Q ion exchange column and by gel-filtration chromatography on a Hiload Superdex75 (GE Healthcare) equilibrated with 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 5 mM dithiothreitol (DTT). $Pol1_{CBM}$ was further purified by gel-filtration chromatography on Hiload Superdex75 column equilibrated with 100 mM ammonium bicarbonate. The purified $Cdc13_{OB1}$ was concentrated to 20 mg/ml and stored at −80 °C. The purified $Pol1_{CBM}$ peptide was concentrated by SpeedVac and then lyophilized. The lyophilization products were then resuspended in water at a concentration of 50 mg/ml and stored at −80 °C.

Crystallization, data collection, and structure determination

***S. cerevisiae* $Cdc13_{OB1}$:** Crystals were grown at 4 °C by the sitting drop vapor diffusion method. The precipitant/well solution contained 21% PEG3350, 0.2 M NaCl, 0.1 M HEPES (pH 7.0), and 10 mM DTT. Heavy-atom derivatives were obtained by soaking crystals in a solution containing 30% PEG3350, 0.2 M NaCl, 0.1 M HEPES (pH 7.0) and 0.1 mM MeHgAc for 3 h and backsoaking for 2 h in 30% PEG3350, 0.2 M NaCl, and 0.1 M HEPES (pH 7.0). Both native and heavy-atom-derivative crystals were gradually transferred into a harvesting solution (30% PEG3350, 0.2 M NaCl, 0.1 M HEPES (pH 7.0), and 20% glycerol) before being flash-cooling in liquid nitrogen for storage and data collection under cryogenic conditions (100 K). Native and Hg-SAD (at Hg peak wavelength) data sets were collected at APS beamline 21ID-D and processed using HKL2000 [43]. Crystals belong to space group $P2_12_12$ and contain one $Cdc13_{OB1}$ molecule per asymmetric unit. Native crystals diffracted to 2.5 Å resolution with cell parameter a = 62.515 Å, b = 68.641 Å and c = 52.815 Å. Three mercury sites were located and refined, and the SAD phases calculated using SHARP [44]. The initial SAD map was significantly improved by solvent flattening. A model was automatically built into the modified experimental electron density using ARP/WARP [45]. The model was then transferred into the native unit cell by rigid body refinement and further refined using simulated annealing and positional refinement in CNS [46], with manual rebuilding using program O [47]. The majority (86%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

***S. cerevisiae* Cdc13$_{OB1}$-Pol1$_{CBM}$:** Cdc13$_{OB1}$ (20 mg/ml) and Pol1$_{CBM}$ (50 mg/ml) were mixed together in a molecular ratio of 1:1. Crystals were grown at 4 °C by sitting drop vapor diffusion method. The precipitant/well solution contained 23% PEG3350 and 0.2 M magnesium formate, 0.1 M Tris-HCl (pH 8.0), and 5 mM DTT. Crystals were gradually transferred into a harvesting solution (25% PEG3350, 0.2 M magnesium formate, 0.1 M Tris-HCl (pH 8.0), 5 mM DTT, and 25% glycerol) before being flash cooling in liquid nitrogen for storage and data collection under cryogenic conditions. Native data set with a resolution of 2.4 Å was collected at APS beamline 21ID-D and processed using HKL2000 [43]. The crystal belongs to space group $P2_12_12_1$, with unit cell parameters a = 60.393 Å, b = 85.090 Å, and c = 60.376 Å. The structure was determined with the molecular replacement method using Phaser program [48]. Two Pol1 peptides could be identified and modeled unambiguously in the complex. Model building and refinement were carried out following the same procedure as those for Cdc13$_{OB1}$, as described for Cdc13$_{OB1}$. The majority (87%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

Cross-linking assay

Chemical cross-linking experiment was performed with purified Cdc13$_{OB1}$ and full-length Cdc13 in PBS buffer. Cross-linking reagent stock solution was prepared by dissolving 35 mg EDC (3-dimethylaminopropyl carbodiimide hydrochloride, Thermo Scientific) into 532 µl distilled water. Serial two-fold dilutions were made by mixing EDC stock solution with distilled water. A measure of 3 µg of Cdc13$_{OB1}$ or full-length

Cdc13 was mixed with 1 μl EDC solution and incubated at room temperature for 30 min. The reaction was quenched by adding 1 M Tris-HCl (pH 8.0) to a final concentration of 50 mM and incubated at room temperature for 15 min. The reaction mixture was then subjected to SDS-PAGE analysis.

Yeast two-hybrid assay

The yeast two-hybrid assays were performed using L40 strain harboring pBTM116 and pACT2 (Clontech) fusion plasmids. Colonies containing both plasmids were selected on –Leu –Trp plates. β-Galactosidase activities were measured by a liquid assay [49].

Sucrose gradient sedimentation

Sucrose gradient ultracentrifugation of Cdc13 was performed with a 10%-35% (v/v) discontinuous sucrose density gradient. Cdc13 was loaded onto the gradient and then centrifuged at 182,000×g for 16 h at 4°C in a SW41Ti swinging bucket rotor and Optima XL90 ultracentrifuge (Beckman). In all, 300μl each of the fractions were collected from the top. Calibration was done with aldolase, catalase, and ferritin (Amersham).

Co-immunoprecipitation

Yeast cells harboring both HA-tagged and $LexA_{BD}$-tagged Cdc13 proteins were used to analyze the homodimerization of Cdc13. Anti-HA antibody was added to the total yeast extract (~500 μg) in buffer A (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 50 mM NaOAc, 1 mM DTT, 1× protease inhibitor cocktail (Calbiochem), 0.1% Tween 20, and 20% glycerol) and mixed at 4 °C for 1 h. A 50 μl aliquot of protein A-Sepharose 4B beads was

added to the mixture, followed by continued incubation for another 1 h. The beads were then washed three times with buffer A. The immunoprecipitates were eluted with 0.1 M citric acid (pH 3.0) and then subjected to SDS-PAGE analysis. Anti-LexA antibody was used in western blotting analysis to detect the presence of $LexA_{BD}$-Cdc13 in the IP samples.

Complementation of *cdc13Δ* by *CDC13* OB1 mutants

Plasmid loss experiments were carried out to test whether *CDC13* OB1 mutants are sufficient to complement the essential functions of a cdc13Δ mutation. Briefly, the mutations were introduced into the pTHA-NLS-CDC13 plasmid using a QuikChange protocol. The plasmids bearing either the wild-type or mutant *CDC13* genes were transformed into the YJL501 (cdc13Δ::HIS3/YEP24-CDC13) strain, which contains a plasmid carrying CDC13 (YEP24-CDC13) for viability. The resulting transformants were spotted on plates containing 0.5 mg/ml 5-fluoroorotic acid and incubated at different temperatures until colonies formed (~48 h).

Telomere length determination

To determine telomere length, yeast DNA was prepared, digested with either PstI or XhoI, and separated on 1% agarose gels. The DNA fragments were transferred to a Hybond N+ filter (Amersham) for hybridization using either a fragment from the Y' element or poly(dG-dT) • poly(dC-dA) as the probe.

Figure 3.1 Domain organization of the CST and the RPA complexes.
Upper panel: the CST complex; lower panel: the RPA complex. In both Cdc13 and RPA70, the four OB folds from the N- to C-terminus, are colored in yellow, orange, light blue, and green, respectively. The RD domain between the first and second OB folds in Cdc13 is colored in gray. In both Stn1 and RPA32, the OB folds are colored in cyan, the WH1 motif of Stn1 and WH motif of RAP32 in marine, and the WH2 motif of Stn1 in blue. Ten1 and RPA14 are colored in pink. The shaded areas are used to indicate the interdomain interactions among the components within each complex.

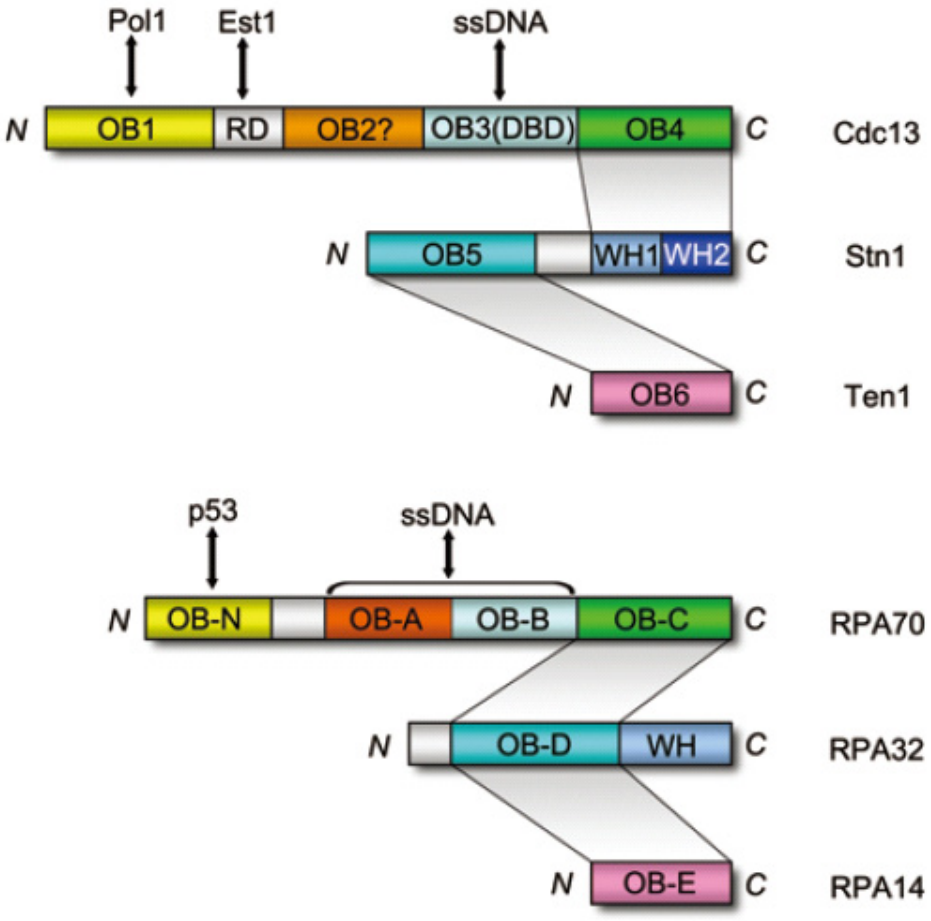


Figure 3.2 Crystals of *Sc*Cdc13$_{OB1}$ and Pol1$_{CBM}$
(A) SDS-PAGE *Sc*Cdc13$_{OB1}$ and Pol1$_{CBM}$ (B)
(C) Crystals of *Sc*Cdc13$_{OB1}$ with Pol1$_{CBM}$ and without Pol1$_{CBM}$ (D)

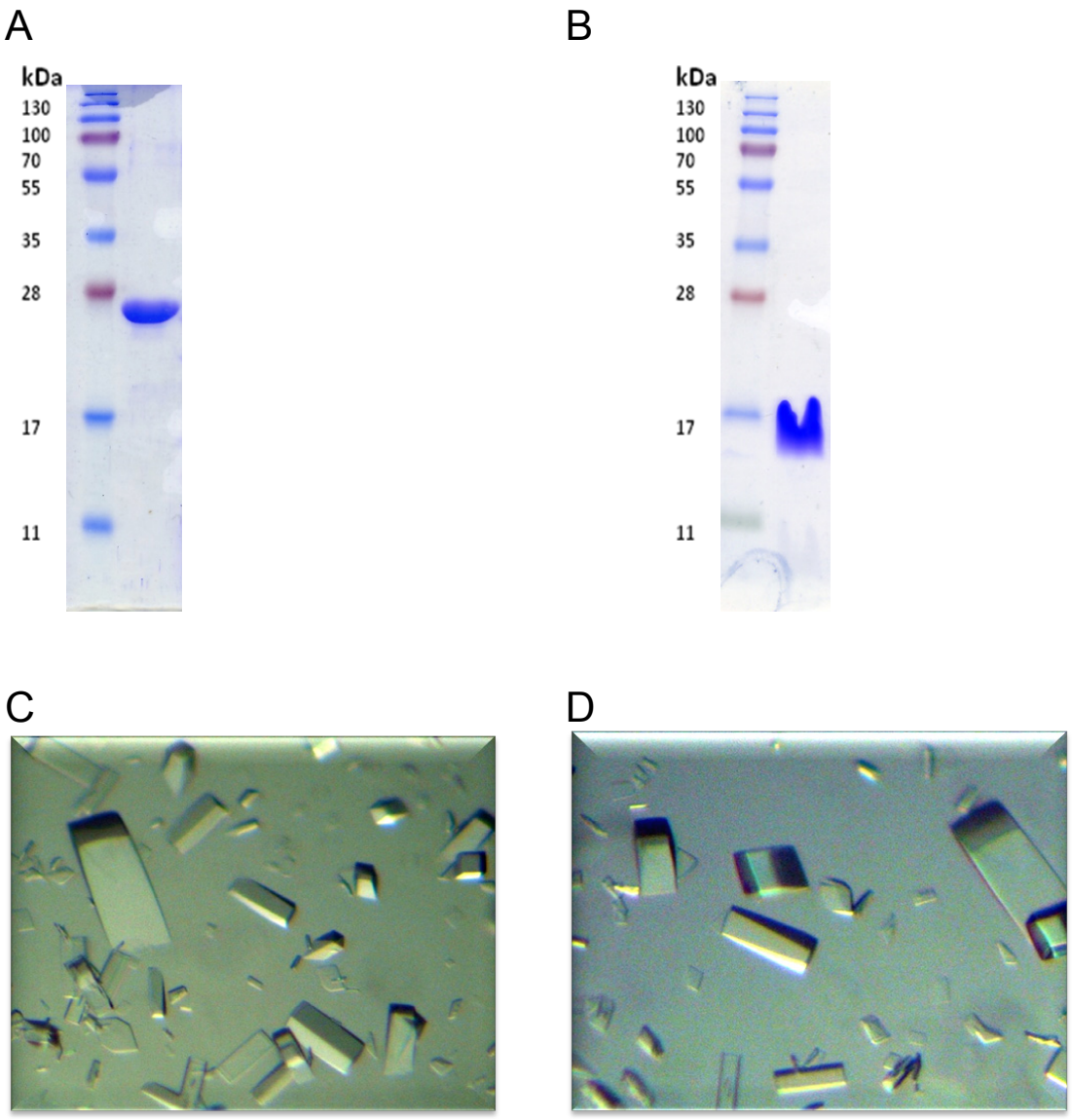


Figure 3.3 $Cdc13_{OB1}$ forms a dimer, both in crystals and in solution.
(A) Ribbon diagram of the monomeric structure of $Cdc13_{OB1}$. (B) Superposition of $Cdc13_{OB1}$ (yellow) on the crystal structure of RPA70N (orange). (C) Ribbon diagram of the $Cdc13_{OB1}$ dimer. The two subunits are colored in yellow and salmon, respectively. (D) Gel-filtration profile revealed that $Cdc13_{OB1}$ behaves as an assembly with an apparent molecular weight of ~45 kDa. (E) Full-length Cdc13 was subjected to a sucrose gradient analysis. The distribution of Cdc13 in the gradient was analyzed by western blot using polyclonal antibodies raised against Cdc13 (upper panel). The band intensities were quantified and plotted (lower panel). Sedimentation positions of three standard proteins are also indicated. (F) Co-IP of Cdc13 fused to different tags in whole cell lysate.

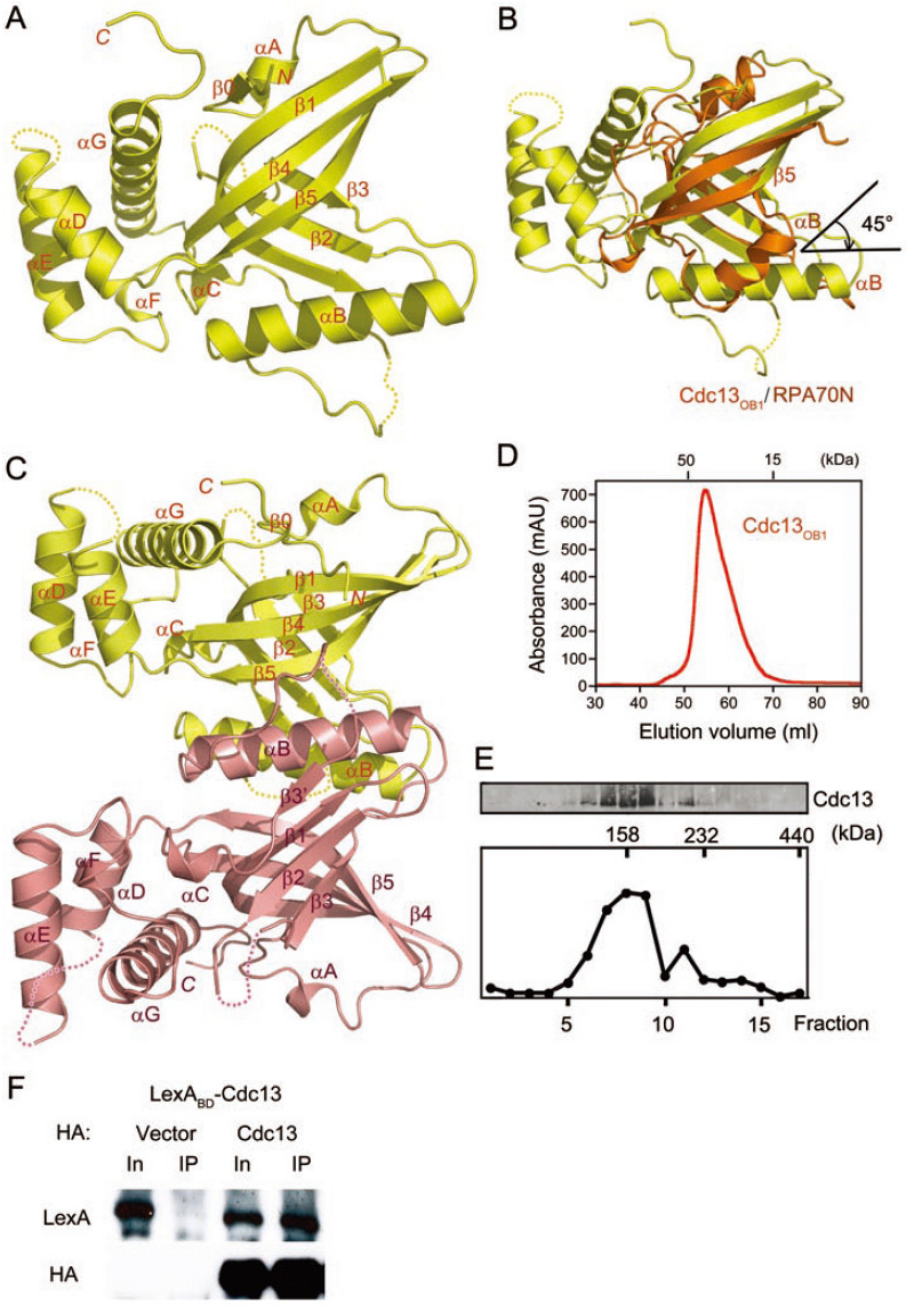


Figure 3.4 Structural and biochemical characterization of *Sc*Cdc13$_{OB1}$ dimer.
(A) Superposition of the structures of *Sc*Cdc13$_{OB1}$ and *Sc*Cdc13$_{OB3}$ (PDB #: 1KXL) [21]. *Sc*Cdc13$_{OB1}$ and *Sc*Cdc13$_{OB3}$ are colored in yellow and cyan. (B) SDS-PAGE of the cross-linked product of Cdc13$_{OB1}$. (C) SDS-PAGE of the cross-linked product of full-length Cdc13. (D) SDS-PAGE of the cross-linked product of the monomeric Cdc13$_{OB1}$ Y95R mutant.(E) Self-association of each OB fold of Cdc13 was examined in yeast two-hybrid assays. The color scheme is the same as in Figure 3.1. Dimeric interaction was measured as β-galactosidase activity. Data are averages of three independent β-galactosidase measurements normalized to the value produced by the dimeric interaction of the OB1 domain, arbitrarily set to 100.

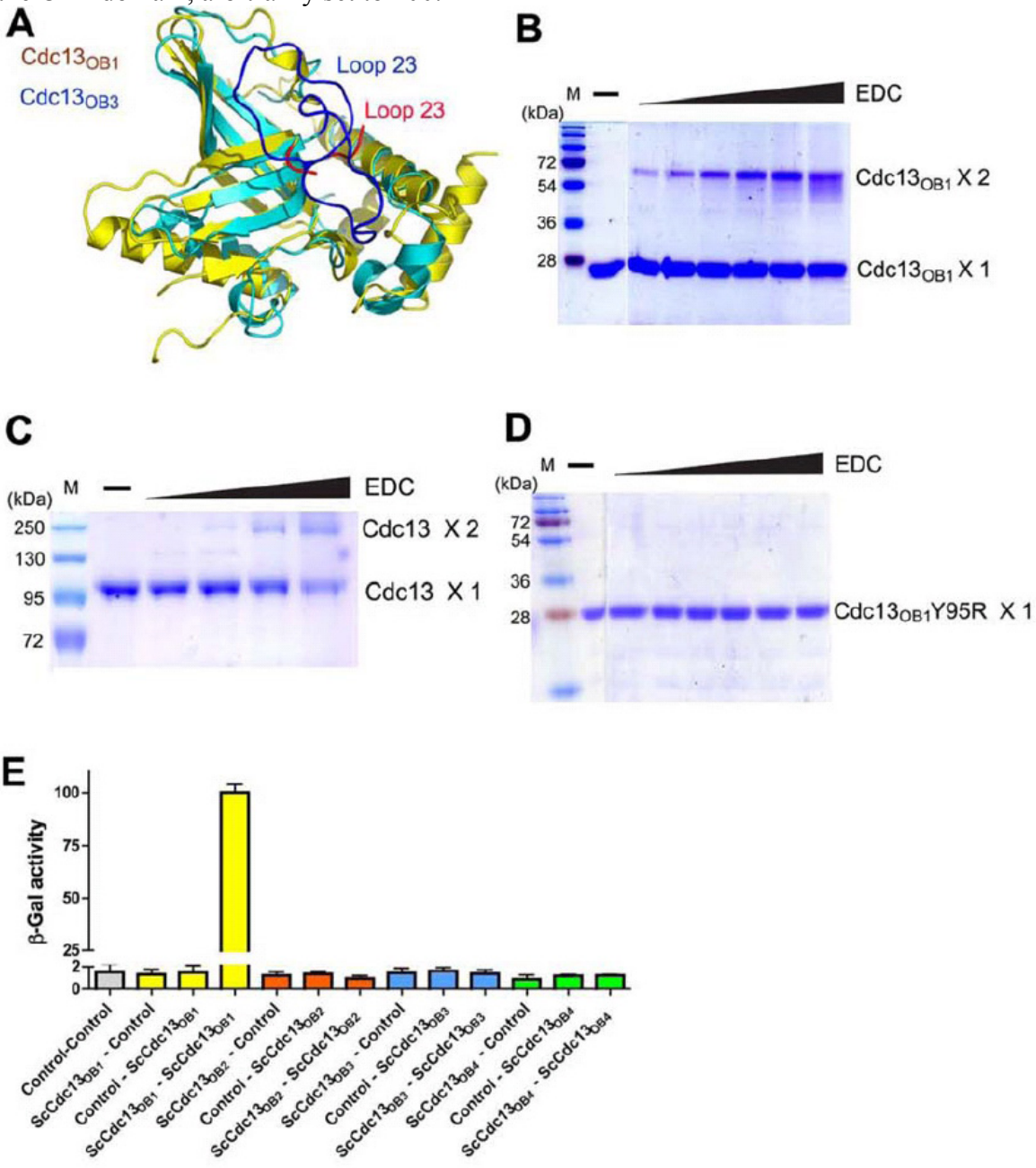


Figure 3.5 The $Cdc13_{OB1}$ dimer interface.
(A) The hydrophobic dimer interface. (B) Helix αB of one $Cdc13_{OB1}$ molecule (in yellow) binds into a hydrophobic groove formed by helix αB and strand β5 of the other (in salmon) in the dimer.(C) The second interface between the two subunits involves two acidic residues (Asp102 and Asp104) from one $Cdc13_{OB1}$, and two basic residues (Arg15 and Lys129) from the other. (D) Superposed chromatographs of wild-type $Cdc13_{OB1}$ and four mutants from gel-filtration columns. (E) Effects of four mutations on dimer formation of $Cdc13_{OB1}$ in yeast two hybrid assays. The color scheme is the same as in D. (F) Co-IP of the same sets of Cdc13 mutants as in panels D and E in whole cell lysate. Conditions are the same as in Figure 3.3F.

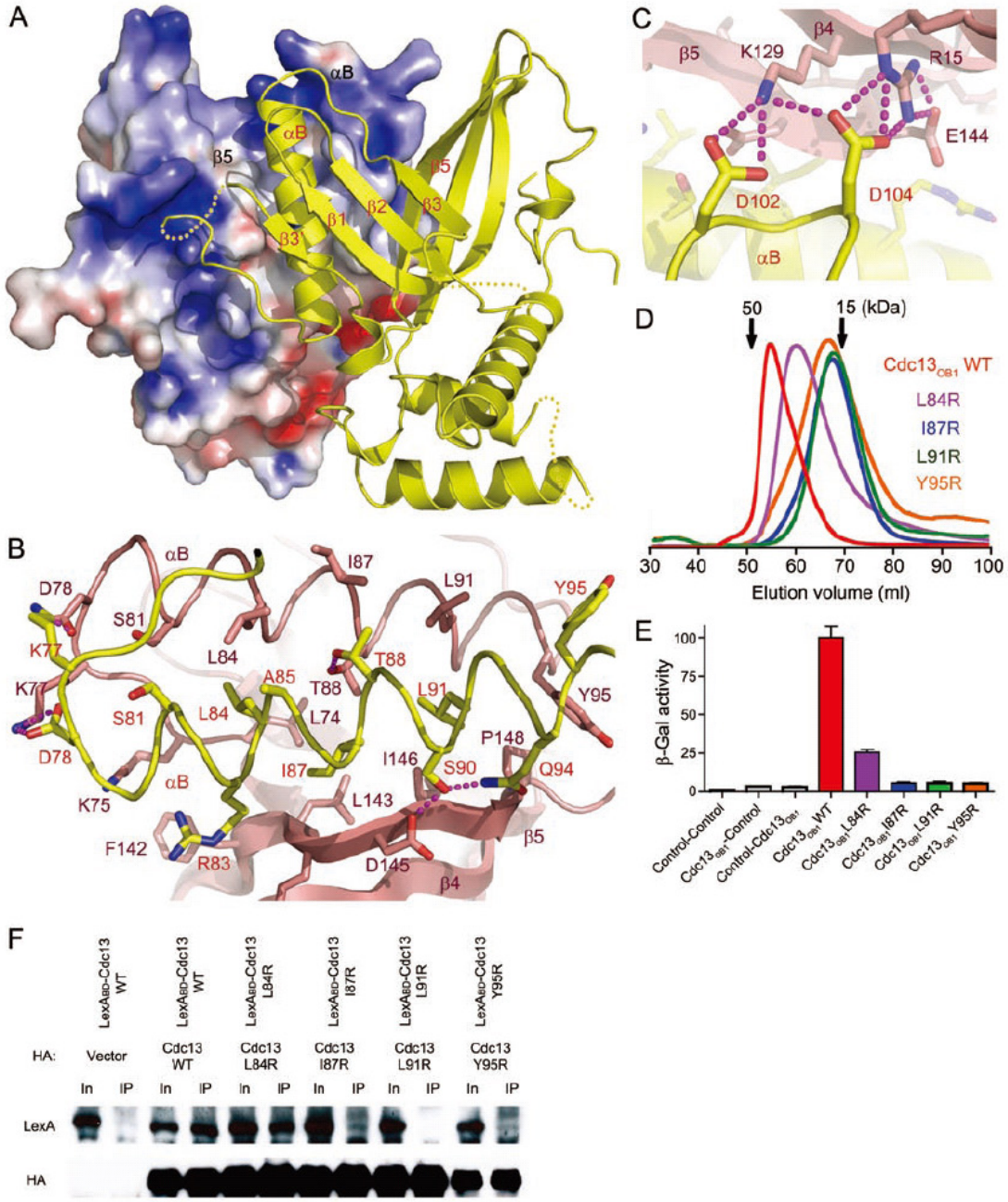


Figure 3.6 Comparison between monomer $Cdc13_{OB1}$ (upper panel) and dimer $Cdc13_{OB1}$ (lower panel) reveals that dimerization of $Cdc13_{OB1}$ almost completely buries the αB helix at the dimeric interface. The two $Cdc13_{OB1}$ molecules are in the surface representation and colored in yellow and salmon, respectively. Helix αB in one $Cdc13_{OB1}$ (in yellow) is colored in red.

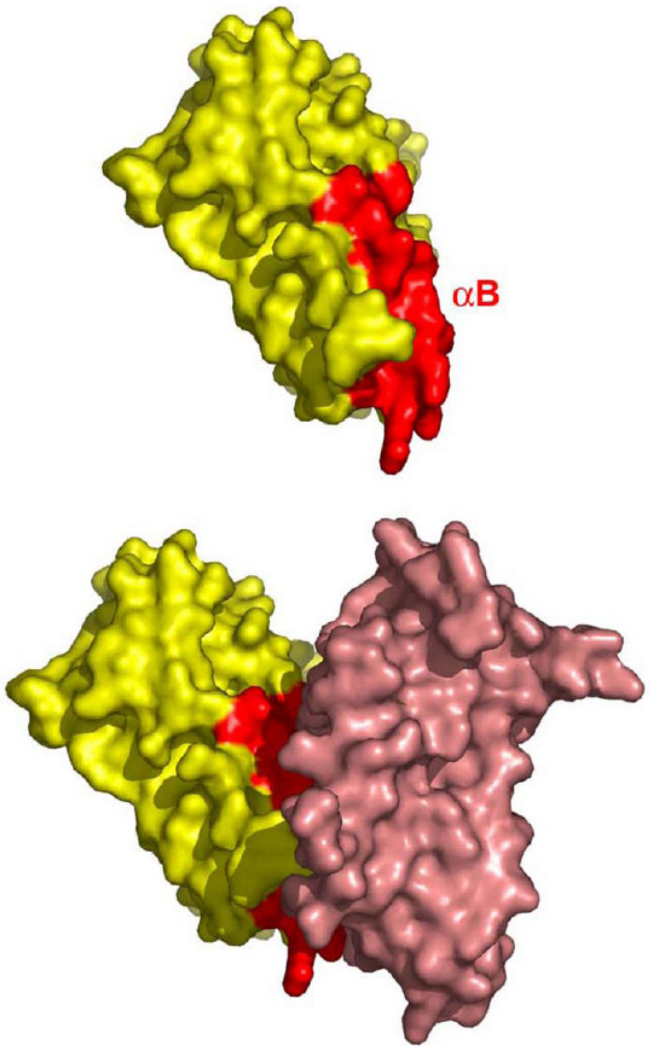
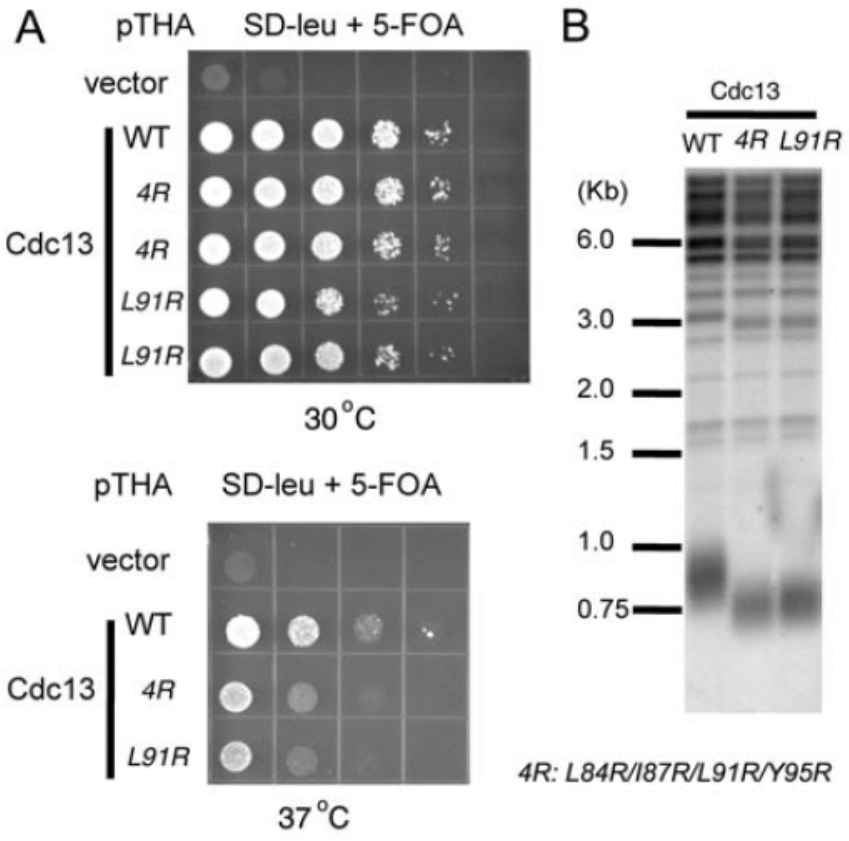


Figure 3.7 Analysis of Cdc13 dimerization mutants *in vivo*.
(Experiments and figures A and B were performed and prepared by E.Y. Yu)
(A) Serial dilutions (10-fold) of strains bearing empty vector or wildtype or mutant CDC13 were spotted on the SD-leu+5-fluoroorotic acid (5-FOA) plates, grown at 30 °C or 37 °C for 2 days, and then photographed.
(B) Chromosomal DNAs were prepared from strains bearing wild-type or the Cdc13 mutants that are deficient in homodimerization, digested with PstI, and subjected to Southern blot analysis using labeled poly(dG-dT) • poly(dC-dA) as the probe.
(C) Gel-filtration profile of $Cdc13_{OB1}$ mutant 4R (L84R/I87R/L91R/Y95R).



C

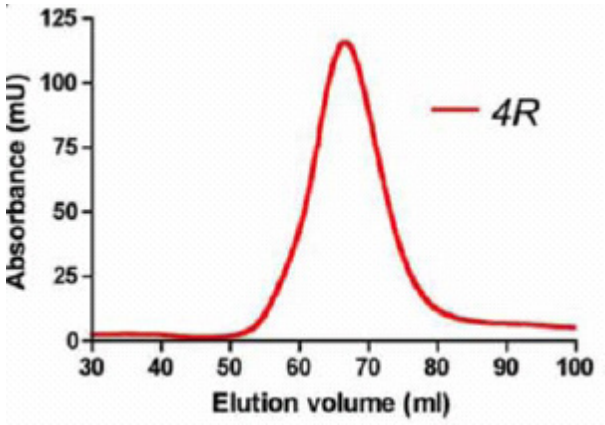


Figure 3.8 The $Cdc13_{OB1}$-$Pol1_{CBM}$ complex structure.
(A) Summary of ITC analysis of the interaction between $Cdc13_{OB1}$ and various Pol1 fragments (nd: not detectable by ITC). A short peptide of Pol1 (residues 215-250) was found to be necessary and sufficient for binding to $Cdc13_{OB1}$.
(B) ITC measurement of the interaction of $Cdc13_{OB1}$ with the $Pol1_{CBM}$ peptide. Insert represents the ITC titration data. The binding curve was fit to a one binding site per $Cdc13_{OB1}$ monomer model.
(C) Overall structure of the dimeric $Cdc13_{OB1}$-$Pol1_{CBM}$ complex. The two $Cdc13_{OB1}$ molecules are colored as in Figure 3.3C. The two $Pol1_{CBM}$ peptides are colored in cyan and blue, respectively. 30 amino acids of the $Pol1_{CBM}$ peptide (residues 216-245) are visible in the electron density map.
(D) The crystal structure of the RPA70N-p53 complex (PDB ID: 2B3G) [23].
(E) The NMR structure of the $Cdc13_{OB3}$-ssDNA complex (PDB ID: 1S40) [31]. In C, D, and E, the OB fold of $Cdc13_{OB1}$, RPA70N, and $Cdc13_{OB3}$ are shown in the same orientation. The interacting partners ($Pol1_{CBM}$, p53, and ssDNA) bind to the same basic grooves of the OB folds.

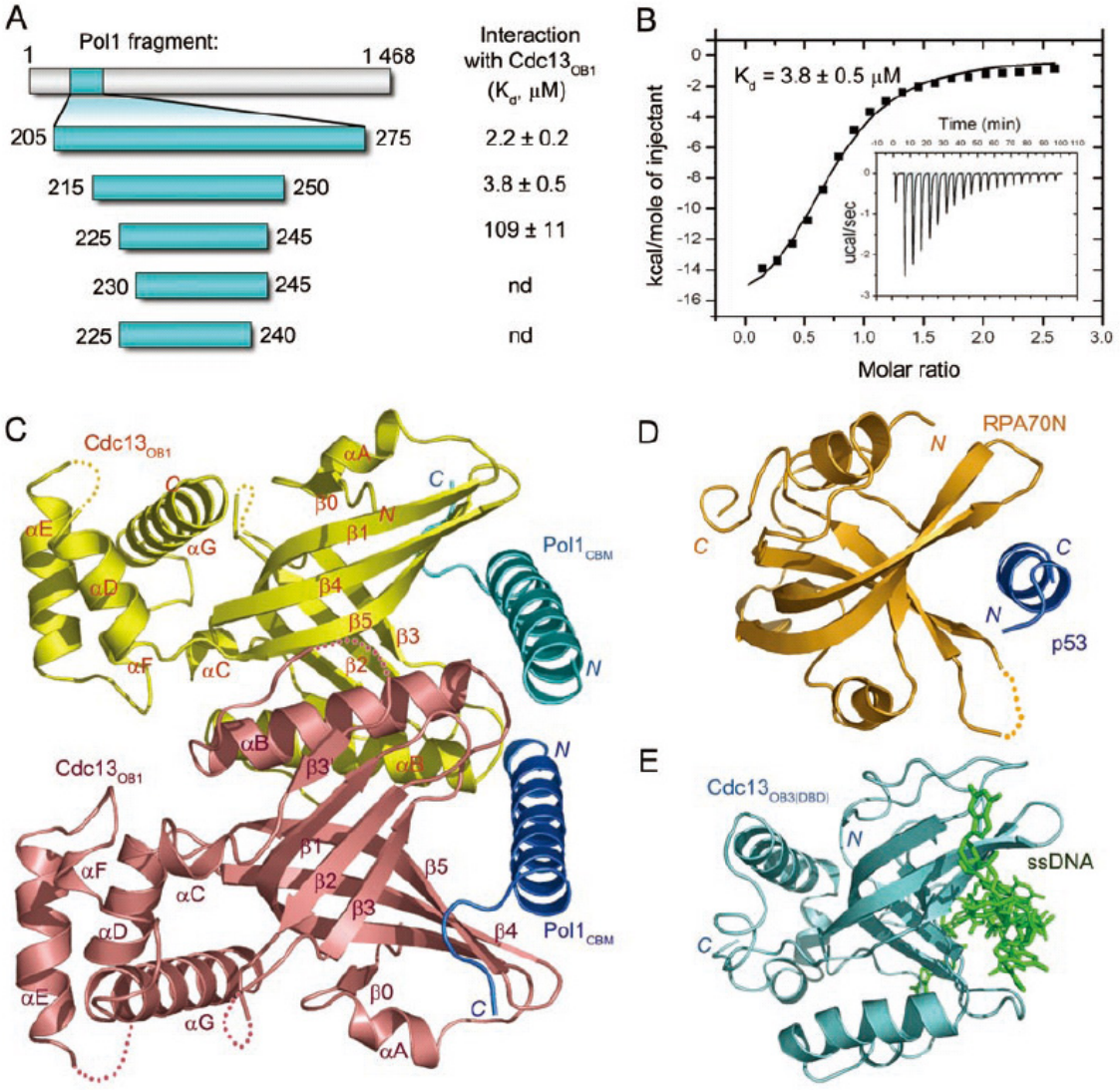
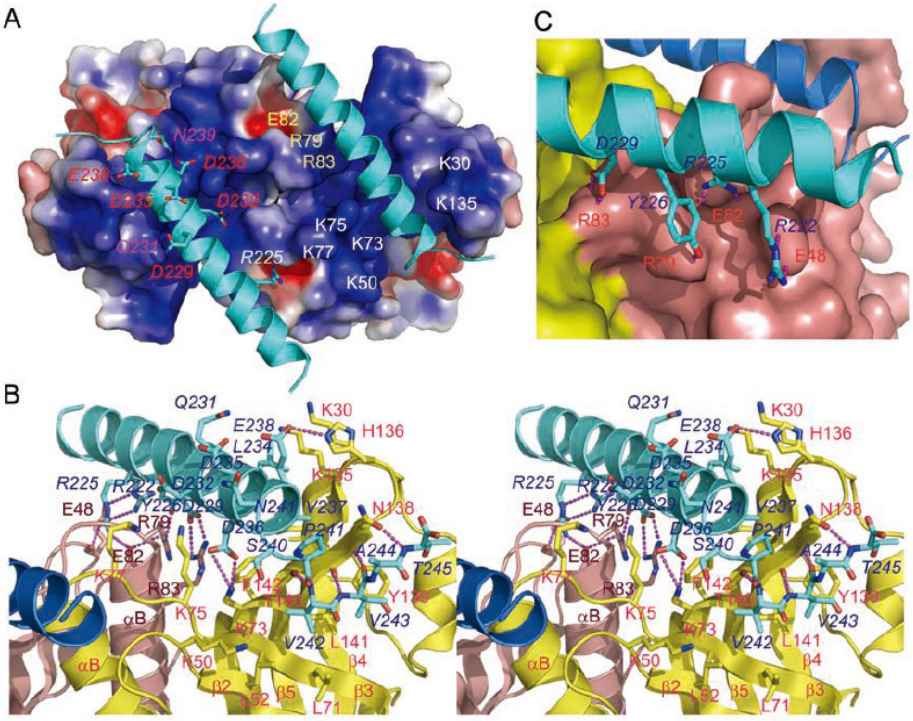


Figure 3.9 The $Cdc13_{OB1}$-$Pol1_{CBM}$ interface.
(A) Electrostatic interaction at the $Cdc13_{OB1}$-$Pol1_{CBM}$ interface. (B) Stereo view of the $Cdc13_{OB1}$-$Pol1_{CBM}$ interface. (C) $Pol1_{CBM}$ interacts with both $Cdc13_{OB1}$ subunits in the dimer. (D) *In vitro* ITC binding of seven $Cdc13_{OB1}$ mutants and six $Pol1_{CBM}$ mutants with wild-type $Pol1_{CBM}$ and $Cdc13_{OB1}$, respectively. (E) *In vitro* ITC binding of four $Cdc13_{OB1}$ mutants that have defects in $Cdc13_{OB1}$ homodimerization with wild-type $Pol1_{CBM}$. (F) Chromosomal DNAs were prepared from strains bearing wild-type or CDC13 mutants that are deficient in Pol1 interaction.



D

$Cdc13_{OB1}$	$Pol1_{CBM}$	K_d (μM)
WT	WT	3.8 ± 0.5
WT	D229R	nd
WT	L233R	nd
WT	D236R	nd
WT	V237A	nd
WT	P241A	nd
WT	V242R	nd
R79E	WT	nd
R83E	WT	nd
3K-3E (K73E/K75E/K77E)	WT	nd
I32E	WT	nd
V133E	WT	nd
T140E	WT	nd
F142A	WT	nd

E

$Cdc13_{OB1}$	$Pol1_{CBM}$	K_d (μM)
WT	WT	3.8 ± 0.5
L84R	WT	12.0 ± 3.5
I87R	WT	nd
L91R	WT	nd
Y95R	WT	nd

F

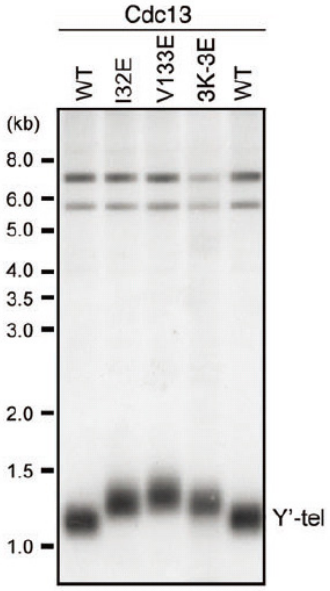


Figure 3.10 Cdc13 proteins employ either the N-terminal OB1 or the C-terminal OB4 domain for dimerization.
Self association of each OB fold of *Kl*Cdc13 (A), *Cg*Cdc13 (B), and *Ca*Cdc13 (D) was examined in yeast two-hybrid assays. The color scheme of the OB folds is the same as in Figure 3.1. Self-association was reflected by the level of β-galactosidase activity produced by the reporter gene. Data are averages of three independent β-galactosidase measurements normalized to the dimeric interaction of the OB1 domain of *Sc*Cdc13 shown in Figure 3.8C, arbitrarily set to 100. (C) Superposed gel-filtration profiles of *Sc*Cdc13$_{OB1}$, *Kl*Cdc13$_{OB1}$, *Cg*Cdc13$_{OB4}$, and *Cg*Cdc13$_{OB1}$.

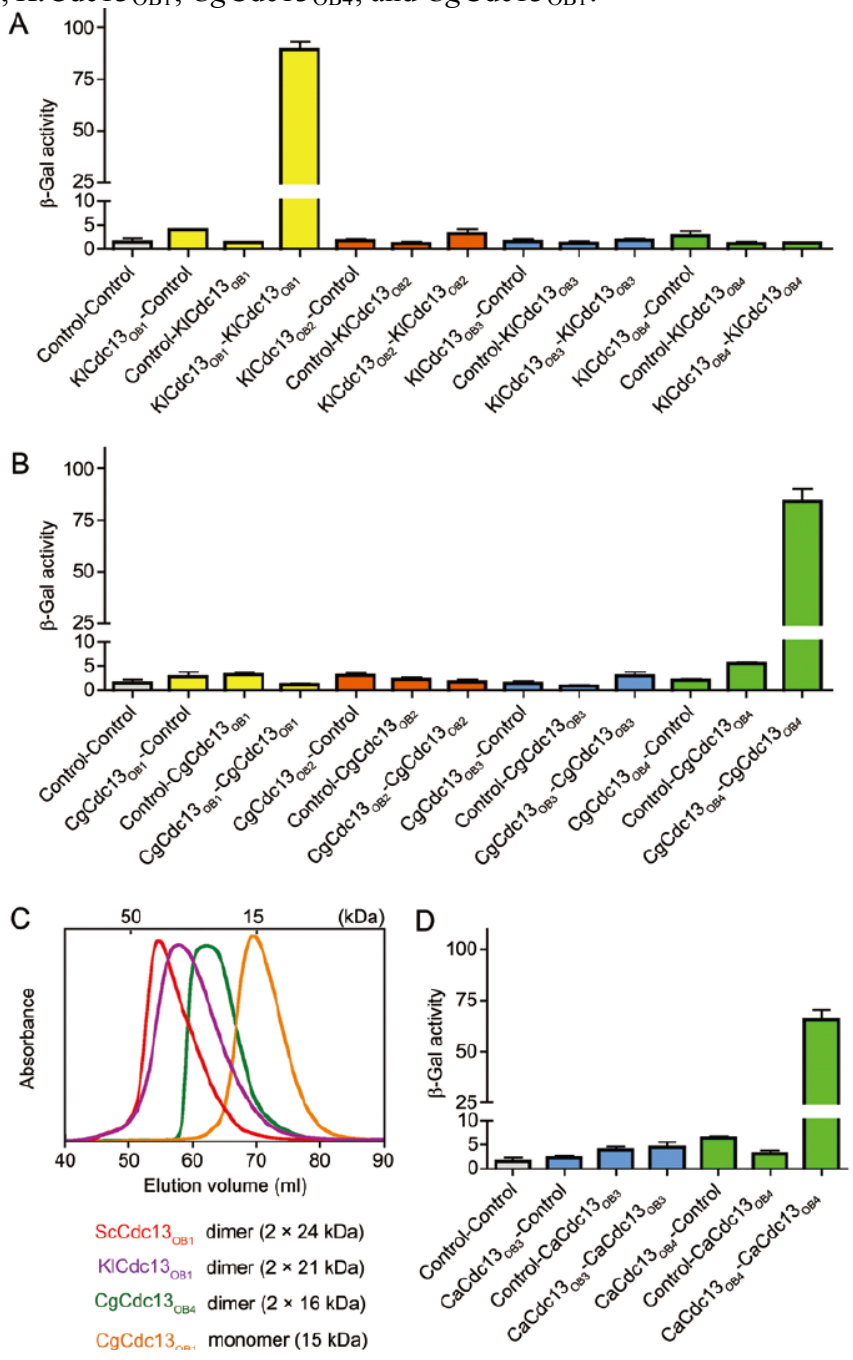


Figure 3.11 The distribution of large and small Cdc13 in the *Saccharomycotina* subphylum of budding yeast.
The evolutionary relationships among the *Saccharomycotina* species and the distribution of large and small Cdc13 homologues in these species are illustrated. The phylogenetic tree is based on comparisons of whole genomes. *C. glabrata* is highlighted because its large Cdc13 protein apparently utilizes an unusual dimerization mechanism in contrast to its close *Saccharomyces* relatives.

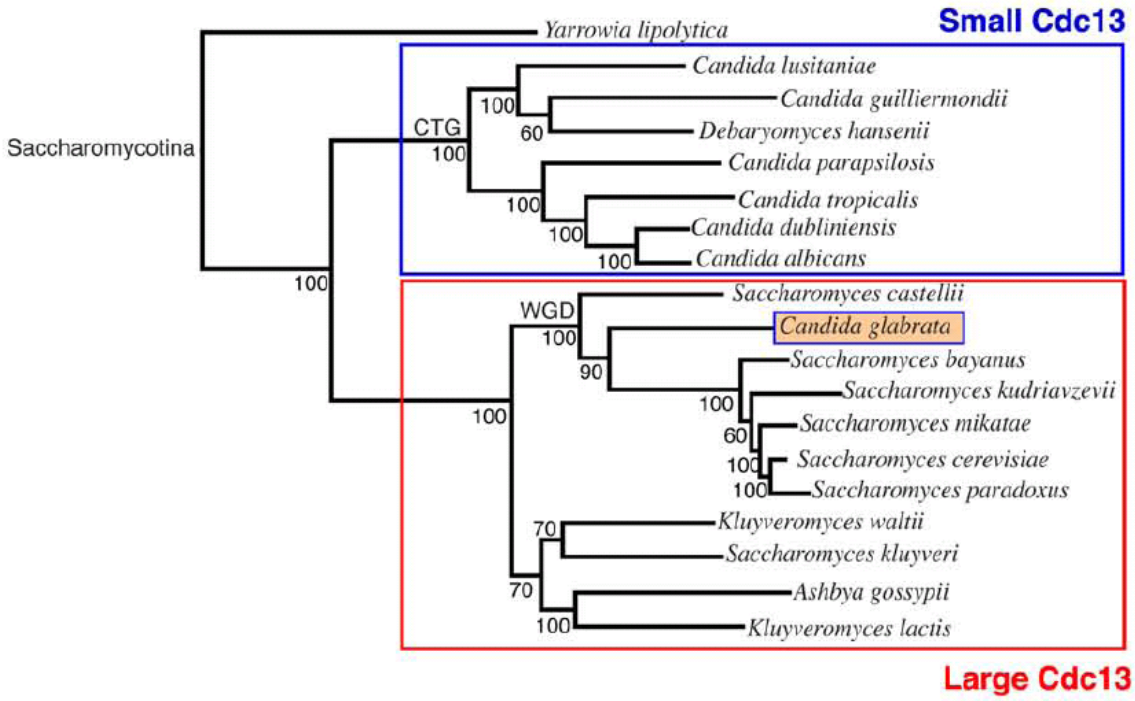


Table 3.1 Data collection, phasing and refinement statistics for *Sc*Cdc13$_{OB1}$ and *Sc*Cdc13$_{OB1}$-Pol1$_{CBM}$

	*Sc*Cdc13$_{OB1}$	*Sc*Cdc13$_{OB1}$	*Sc*Cdc13$_{OB1}$ - Pol1$_{CBM}$
Data collection	*Hg peak*	*Native*	*Native*
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	62.441, 69.024, 52.064	62.515, 68.641, 52.815	60.393, 85.090, 60.376
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.12721	0.97856	0.97872
Resolution (Å)	50-3.3	50-2.5	100-2.4
R_{merge}	0.085(0.263)*	0.063(0.313)*	0.074(0.393)*
$I/\sigma I$	27.3(6.0)*	41.8(3.6)*	42.7(2.9)*
Completeness (%)	93.2(72.0)*	96.4(76.8)*	97.3(83.1)*
Redundancy	8.7(7.1)*	6.6(4.7)*	6.9(3.9)*
Phasing			
Figure of Merit (anomalous)	0.24893(acen) 0.08433(cent)		
Phasing power (anomalous)	0.750		
Refinement			
Resolution (Å)		50-2.5	50-2.4
No. reflections		7498	9706
R_{work}/ R_{free}		0.211/0.267	0.224/0.264
No. atoms			
Protein		1608	1741
Water		35	34
Wilson B-factors (Å^2)		61.9	55.5
B-factors (Å^2)			
Protein		66.293	37.844
Water		61.067	43.306
R.m.s deviations			
Bond lengths (Å)		0.006074	0.004976
Bond angles (°)		1.23652	1.15154
B-factors (Å^2)		5.782	5.821

*The numbers in parentheses represent the highest resolution shell numbers.

REFERENCES

1. Lundblad, V., *Telomeres in the '80s: a few recollections.* Nat Struct Mol Biol, 2006. **13**(12): p. 1036-8.
2. Eldridge, A.M. and D.S. Wuttke, *Probing the mechanism of recognition of ssDNA by the Cdc13-DBD.* Nucleic Acids Res, 2008. **36**(5): p. 1624-33.
3. Hughes, T.R., et al., *Identification of the single-strand telomeric DNA binding domain of the Saccharomyces cerevisiae Cdc13 protein.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6457-62.
4. Pennock, E., K. Buckley, and V. Lundblad, *Cdc13 delivers separate complexes to the telomere for end protection and replication.* Cell, 2001. **104**(3): p. 387-96.
5. Qi, H. and V.A. Zakian, *The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein.* Genes Dev, 2000. **14**(14): p. 1777-88.
6. Hsu, C.L., et al., *Interaction of Saccharomyces Cdc13p with Pol1p, Imp4p, Sir4p and Zds2p is involved in telomere replication, telomere maintenance and cell growth control.* Nucleic Acids Res, 2004. **32**(2): p. 511-21.
7. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.
8. Grossi, S., et al., *Pol12, the B subunit of DNA polymerase alpha, functions in both telomere capping and length regulation.* Genes Dev, 2004. **18**(9): p. 992-1006.
9. Petreaca, R.C., et al., *Chromosome end protection plasticity revealed by Stn1p and Ten1p bypass of Cdc13p.* Nat Cell Biol, 2006. **8**(7): p. 748-55.
10. Puglisi, A., et al., *Distinct roles for yeast Stn1 in telomere capping and telomerase inhibition.* EMBO J, 2008. **27**(17): p. 2328-39.
11. Martin, V., et al., *Protection of telomeres by a conserved Stn1-Ten1 complex.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 14038-43.
12. Surovtseva, Y.V., et al., *Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes.* Mol Cell, 2009. **36**(2): p. 207-18.
13. Miyake, Y., et al., *RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway.* Mol Cell, 2009. **36**(2): p. 193-206.
14. Wan, M., et al., *OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation.* J Biol Chem, 2009. **284**(39): p. 26725-31.
15. Wold, M.S., *Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism.* Annu Rev Biochem, 1997. **66**: p. 61-92.
16. Bochkarev, A. and E. Bochkareva, *From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold.* Curr Opin Struct Biol, 2004. **14**(1): p. 36-42.
17. Sun, J., et al., *Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres.* Genes Dev, 2009. **23**(24): p. 2900-14.

18. Gelinas, A.D., et al., *Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19298-303.
19. Theobald, D.L. and D.S. Wuttke, *Prediction of multiple tandem OB-fold domains in telomere end-binding proteins Pot1 and Cdc13.* Structure, 2004. **12**(10): p. 1877-9.
20. Bochkarev, A., et al., *Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA.* Nature, 1997. **385**(6612): p. 176-81.
21. Mitton-Fry, R.M., et al., *Conserved structure for single-stranded telomeric DNA recognition.* Science, 2002. **296**(5565): p. 145-7.
22. Rost, B., G. Yachdav, and J. Liu, *The PredictProtein server.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W321-6.
23. Bochkareva, E., et al., *Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15412-7.
24. Bochkareva, E., et al., *Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA.* EMBO J, 2002. **21**(7): p. 1855-63.
25. Olson, E., et al., *RPA2 is a direct downstream target for ATR to regulate the S-phase checkpoint.* J Biol Chem, 2006. **281**(51): p. 39517-33.
26. von Bergwelt-Baildon, M.S., et al., *Human primary and memory cytotoxic T lymphocyte responses are efficiently induced by means of CD40-activated B cells as antigen-presenting cells: potential for clinical application.* Blood, 2002. **99**(9): p. 3319-25.
27. Theobald, D.L., R.M. Mitton-Fry, and D.S. Wuttke, *Nucleic acid recognition by OB-fold proteins.* Annu Rev Biophys Biomol Struct, 2003. **32**: p. 115-33.
28. Theobald, D.L. and D.S. Wuttke, *Divergent evolution within protein superfolds inferred from profile-based phylogenetics.* J Mol Biol, 2005. **354**(3): p. 722-37.
29. Jacobs, D.M., et al., *Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker.* J Biomol NMR, 1999. **14**(4): p. 321-31.
30. Holm, L. and C. Sander, *Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors.* J Mol Biol, 1991. **218**(1): p. 183-94.
31. Mitton-Fry, R.M., et al., *Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13.* J Mol Biol, 2004. **338**(2): p. 241-55.
32. Horvath, M.P., et al., *Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA.* Cell, 1998. **95**(7): p. 963-74.
33. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.
34. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nat Struct Mol Biol, 2004. **11**(12): p. 1223-9.
35. Anderson, E.M., W.A. Halsey, and D.S. Wuttke, *Delineation of the high-affinity single-stranded telomeric DNA-binding domain of Saccharomyces cerevisiae Cdc13.* Nucleic Acids Res, 2002. **30**(19): p. 4305-13.

36. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. **17**(2): p. 231-5.
37. Bianchi, A., et al., *TRF1 is a dimer and bends telomeric DNA.* EMBO J, 1997. **16**(7): p. 1785-94.
38. Konishi, A. and T. de Lange, *Cell cycle control of telomere protection and NHEJ revealed by a ts mutation in the DNA-binding domain of TRF2.* Genes Dev, 2008. **22**(9): p. 1221-30.
39. Fairall, L., et al., *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Mol Cell, 2001. **8**(2): p. 351-61.
40. Yu, E.Y., et al., *Analyses of Candida Cdc13 orthologues revealed a novel OB fold dimer arrangement, dimerization-assisted DNA-binding, and substantial structural differences between Cdc13 and RPA70.* Mol Cell Biol, 2011.
41. Prescott, J. and E.H. Blackburn, *Functionally interacting telomerase RNAs in the yeast telomerase complex.* Genes Dev, 1997. **11**(21): p. 2790-800.
42. Wenz, C., et al., *Human telomerase contains two cooperating telomerase RNA molecules.* EMBO J, 2001. **20**(13): p. 3526-34.
43. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.
44. Schiltz, M., et al., *Protein Crystallography at Ultra-Short Wavelengths: Feasibility Study of Anomalous-Dispersion Experiments at the Xenon K-edge.* J Synchrotron Radiat, 1997. **4**(Pt 5): p. 287-97.
45. Perrakis, A., et al., *ARP/wARP and molecular replacement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 10): p. 1445-50.
46. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
47. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
48. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
49. Moretti, P., et al., *Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1.* Genes Dev, 1994. **8**(19): p. 2257-69.

CHAPTER 4

ANALYSES OF *CANDIDA* CDC13 ORTHOLOGUES REVEALED A NOVEL OB FOLD DIMER ARRANGEMENT, SUBSTANTIAL STRUCTURAL DIFFERENCE BETWEEN CDC13 AND RPA70 AND INTERACTION BETWEEN CDC13 AND STN1

4.1 Attributions

This chapter contains the manuscript "Analyses of *Candida* Cdc13 Orthologues Revealed a Novel OB Fold Dimer Arrangement, Dimerization-Assisted DNA Binding, and Substantial Structural Differences between Cdc13 and RPA70" by E.Y. Yu, J. Sun, M. Lei and N.F. Lue published in *Molecular and Cellular Biology* (2011) 23: 186-198. Constructs were designed by J. Sun and E.Y. Yu. Mutagenesis was performed by J. Sun and E.Y. Yu. Protein expression, purification and crystallization ware performed by J. Sun. X-ray data collection and structure determination were done by J. Sun with the help from M. Lei. *In vivo* yeast telomere assays and DNA binding assays were performed by E.Y. Yu. The manuscript was written by J. Sun, M. Lei and N.F. Lue.

4.2 Abstract

The budding yeast Cdc13-Stn1-Ten1 complex is crucial for telomere protection and has been proposed to resemble the RPA complex structurally and functionally. The Cdc13

homologues in *Candida* species are unusually small and lack two conserved domains previously implicated in telomere regulation, thus raising interesting questions concerning the mechanisms and evolution of these proteins. In collaboration with Dr. Lue, we showed that the unusually small Cdc13 homologue in *Candida albicans* is indeed a regulator of telomere lengths and that it associates with telomere DNA *in vivo*. We also determined the crystal structure of the OB4 domain of *C. glabrata* Cdc13, which revealed a novel mechanism of OB fold dimerization. We demonstrated high-affinity telomere DNA binding by *C. tropicalis* Cdc13 (*Ct*Cdc13) and found that dimerization of this protein through its OB4 domain is important for high affinity DNA binding. The structure also exhibits marked differences to the C-terminal OB fold of RPA70, thus arguing against a close evolutionary kinship between these two proteins. Our findings provide new insights on the mechanisms and evolution of a critical telomere end binding protein.

4.3 Introduction

The special structures located at the ends of linear eukaryotic chromosomes, known as telomeres, are critical for chromosome stability; they protect the terminal DNAs from degradation, end-to-end fusion and other abnormal transactions [1-3]. Telomeric DNAs are bound by functionally important proteins through both DNA-protein and protein-protein interactions. In most organisms, telomeres comprise short repetitive G-rich sequences and terminate in 3' overhangs referred to as G-tails. Even though the G-tails represent a shared feature of almost all telomeres, they appear to be bound by divergent protein complexes in different organisms. A widespread dimeric G-tail binding protein

complex was first described in ciliated protozoa and named TEBPα/β in these organisms [4]. Subsequent studies revealed orthologues of these proteins in both fission yeast and mammals (named Pot1-Tpz1 in fission yeast and POT1-TPP1 in mammals) [5, 6]. By contrast, the G-tails of budding yeast are capped by a trimeric complex comprised of Cdc13, Stn1, and Ten1 (CST) [7]. Genetic and structural analyses suggest that CST represents a telomere-specific RPA-like complex [2, 8, 9]. Interestingly, even though CST proteins were initially thought to be confined to budding yeast, recent studies have uncovered Stn1 and Ten1 homologues in *Schizosaccharomyces pombe* as well as CST-like complexes in plants and mammals [10-12]. Thus, in many organisms the CST complex may act as an alternative telomere end protection complex with functions overlapping or parallel to those of the POT1-TPP1 complex.

Among all the CST complexes, the structures and mechanisms of the Saccharomyces cerevisiae subunits are the most extensively characterized. *S. cerevisiae* Cdc13 (*Sc*Cdc13) is a multifunctional protein with a myriad of binding targets (Fig. 1A). It uses a C-terminal OB fold (DNA-binding domain [DBD]) to bind with high affinity and sequence specificity to the irregular, GT-rich repeats of *S. cerevisiae* telomeres [13]. It also employs a recruitment domain (RD) to interact with the telomerase regulatory protein Est1, and this interaction promotes the recruitment of telomerase to chromosome ends and the activation of telomerase [14, 15]. Moreover, I have recently shown that the N-terminal OB fold domain of *Sc*Cdc13 (OB1) mediates *Sc*Cdc13 dimerization and that this dimerization promotes Cdc13-Pol1 (the catalytic subunit of polymerase α [Pol α] interaction and regulates telomere length [16]. Others have reported that dimerization may allow *Sc*Cdc13$_{OB1}$ to bind DNA [17]. In comparison to *Sc*Cdc13, fewer interaction

partners have been identified for *Sc*Stn1 and *Sc*Ten1. Both *Sc*Stn1 and ScTen1 have been reported to bind telomere DNA with moderate to low affinity [8]. *Sc*Stn1 is also known to interact with Pol12, another subunit of the Pol α complex [18]. The multiplicity of interactions between CST and Pol α supports a role for CST in regulating telomere C-strand synthesis, which is thought to be mediated by Pol α [2].

In the budding yeast *Saccharomyces cerevisiae*, Cdc13, Stn1, and Ten1 are thought to form a single ternary complex. Although the interaction between the two smaller subunits, Stn1 and Ten1, has been well studied, the interaction between Cdc13 and Stn1 has remained relatively elusive. Previous studies have identified distinct Stn1 domains that mediate interaction with either Ten1 or Cdc13, allowing analysis of whether the interaction between Cdc13 and Stn1 is indeed essential for telomere capping or length regulation. Consistent with the model that the Stn1 essential function is to promote telomere end protection through Cdc13, *stn1* alleles that truncate the C-terminal 123 residues fail to interact with Cdc13 and do not support viability when expressed at endogenous levels [19]. Also, a region comprising the Stn1-interacting and telomere-binding region of Cdc13 (amino acids 252-924) complemented the growth defects of cdc13 mutants [20].

As alluded to earlier, a provocative recent proposal concerning CST is that it represents a telomere-specific RPA complex [8]. Indeed, we and others have shown a high degree of structural and functional resemblances between Stn1 and RPA32, as well as between Ten1 and RPA14 [8, 9, 21, 22]. By contrast, existing data do not support a paralogous relationship between Cdc13 and RPA70, the largest subunits of the two complexes. Even though both Cdc13 and RPA70 consist of multiple OB fold domains,

neither the first OB fold (OB1) nor the penultimate OB fold (DBD) of Cdc13 displays a strong similarity to the corresponding domain in RPA70 [16]. However, because the structures of other domains of Cdc13 have not been resolved, the possibility remains that additional studies could provide supports for a paralogous relationship between Cdc13 and RPA70.

Our laboratories have employed *Candida* species as alternative model systems for understanding CST structure and mechanisms. The telomere repeat units of *Candida* species are unusual in being long, regular, and non-G-rich [23]. Homologues of the CST proteins can nevertheless be readily identified in most *Candida* genomes [24, 25]. In Chapter 2, I have described the high resolution structure of a complex of *Candida tropicalis* Stn1N and Ten1 and the functions of *C. albicans* Stn1 and Ten1 in telomere regulation [9]. However, our analysis of the *C. albicans* Cdc13 (*Ca*Cdc13) homologue was hampered by the fact that the gene is essential for cell viability. Interestingly, many Cdc13 homologues in *Candida* species are noticeably smaller; they lack the N-terminal half of their *S. cerevisiae* counterpart and contain just two OB fold domains: DBD (responsible for DNA binding of *Sc*Cdc13) and OB4 (implicated in binding Stn1) (Figure 4.1A) [24]. Because the N-terminal half of *Sc*Cdc13 is responsible for dimerization and *Sc*Cdc13-Pol1 and *Sc*Cdc13-Est1 interaction, its absence in *Candida* Cdc13s raises fascinating questions concerning the mechanisms and evolution of these homologues. In this chapter, I provided evidence that the unusually small Cdc13 homologue in *Candida albicans* is indeed a regulator of telomere lengths and structure and that it associates with telomere DNA *in vivo*. We determined the crystal structure of the OB4 domain of *C. glabrata* Cdc13 (*Cg*Cdc13) and uncovered a novel mode of OB fold dimerization. This

dimerization was later discovered to be important for *C. tropicalis* Cdc13 (*Ct*Cdc13) DNA binding, which requires both the DBD and the OB4 domains. Comparative structural analysis revealed marked differences between *Cg*Cdc13$_{OB4}$ and the C-terminal OB fold of RPA70, arguing against a close evolutionary kinship between these two proteins. Our findings provide new insights on the mechanisms and evolution of Cdc13 and underscore the utility of investigating the CST complex in *Candida* species. In addition, I present here detailed analysis concerning the interaction between Cdc13 and Stn1, providing a clear picture of the interaction network within the CST complex.

4.4 *C. albicans* Cdc13 localizes to telomeres and regulates telomere lengths *in vivo*

In Chapter 2, I identified plausible homologues of each CST component in *Candida* and *Saccharomyces* genomes and investigated the functions and mechanisms of *Candida albicans* Stn1 and Ten1 proteins in telomere regulation. Unlike Stn1 and Ten1, *C. albicans* Cdc13 appears to be essential for cell viability, thus hampering analysis of its function [9]. To ascertain a role for the putative *Ca*Cdc13 in telomere regulation, we first attempted to determine if the protein is associated with telomeres *in vivo*. To facilitate chromatin immunoprecipitation (ChIP), a TAP tag was fused to the C terminus of the single *CaCDC13* allele in the heterozygote *CaCDC13* $^{+/-}$ strain background. In collaboration with Dr. Lue's group, we analyzed the telomere association of *Ca*Cdc13-TAP by using ChIP with IgG-Sepharose, which interacts with the protein A epitope of the TAP tag. The *Ca*Cdc13-TAP protein in three independently generated, tagged strains exhibited significant cross-linking to telomeric DNA upon formaldehyde treatment, thus confirming the ability of Cdc13 to localize to telomeres *in vivo* (Figure 4.1B). These

results indicate that *Ca*Cdc13 is indeed a telomere-associated protein and argue that despite the absence of N-terminal domains, *Ca*Cdc13 acts directly at telomeres, possibly forming a CST complex with *Ca*Stn1 and *Ca*Ten1, which are known to be necessary for the maintenance of proper telomere lengths and structure [2, 9, 26, 27].

4.5 Telomere-specific DNA binding activity of *C. tropicalis* Cdc13

The remarkably high degree of telomere sequence divergence in the *Candida* clade raises an interesting question concerning the mechanisms of DNA recognition by Cdc13: how do highly homologous DNA-binding domains (i.e., the DBDs of Cdc13s) recognize such diverse sequence targets? To gain insights into the mechanisms of telomere DNA recognition, we attempted to characterize in detail the DNA-binding properties of small Cdc13s. Initial screening of protein expression and purification indicated that the Cdc13 protein from *C. tropicalis*, but not that from *C. albicans*, can be obtained in large quantities from *E. coli* in an active form. We therefore expressed and purified SUMO-fused *Ct*Cdc13 with a C-terminal FLAG tag in *E. coli*. After removal of the SUMO domain and further purification to near homogeneity, the full-length *Ct*Cdc13 protein was subjected to a series of electrophoretic mobility shift assays (EMSAs) to determine its DNA binding affinity and sequence specificity (Figure 4.2). For comparative purposes, the binding affinity of the putative DBD of *Ct*Cdc13 was also determined. As expected, the full-length *Ct*Cdc13 protein binds to the C. tropicalis telomere repeats with high affinity (K_d [dissociation constant] of ~40 nM) (Figure 4.2D). The formation of the complex was concentration dependent, and all of the probes can be bound when sufficient amounts of proteins were added to the reaction (Figure 4.2D). DNA binding by *Ct*Cdc13

was also highly sequence specific, as revealed by a competition experiment; whereas an unlabeled telomeric competitor at a 2.5-fold molar excess substantially inhibited the formation of the labeled DNA-protein complex, a non-telomeric competitor had no effect even when present at a 200-fold molar excess (Figure 4.2C). In addition, while the telomere repeats from both *C. tropicalis* and *C. albicans* (which differ from each other at 7/23 nucleotide positions) competed well in binding to *Ct*Cdc13, the purely GT repeat of the *S. cerevisiae* telomere sequence did not (Figure 4.2D). These results indicate that *Ct*Cdc13 has a clear sequence preference for the *Candida* telomere repeats but that the recognition is not entirely species specific. Interestingly, the DBD of *Ct*Cdc13 exhibited the same sequence preference as that of the full-length protein but a significantly lower binding affinity (Figure 4.2E) (K_d >>320 nM), suggesting that the OB4 of *Ct*Cdc13 is important for DNA binding affinity but not for sequence specificity. We also analyzed the OB4 domain of *Ct*Cdc13 and found that this domain alone does not possess appreciable DNA-binding activity (data not shown). Thus, unlike the DBD of *Sc*Cdc13, which has an autonomous high-affinity telomere DNA-binding activity, the comparable domain of *Ct*Cdc13 does not, hinting at significant mechanistic differences [28, 29].

Another recent survey revealed low-affinity DNA binding by the DBDs of *C. albicans, C. parapsilosis*, and *C. glabrata* Cdc13 homologues [30]. To determine if other domains of these proteins might contribute to DNA binding (as was observed for *Ct*Cdc13), we attempted to examine the properties of full-length Cdc13s and the DBDs from these species. Thus far, we have been able to isolate only full-length *Cg*Cdc13 and its DBD. Interestingly, full-length *Cg*Cdc13 binds to the cognate telomere repeats with high affinity (K_d=~20 nM), whereas the DBD alone failed to form a stable complex with

the same oligonucleotide (Figure 4.2F). In the DBD assays, broad smears were observed above the free probe, but few distinct bands could be detected, suggesting dissociation of the DBD-DNA complex during native gel electrophoresis. Hence, the *Cg*Cdc13 DBD alone appears to bind telomeric DNA but evidently requires other domains to form a stable complex. Like *Ct*Cdc13, DNA binding by the full-length *Cg*Cdc13 is highly sequence specific: in competition assays, >100-fold-higher concentrations of a non-telomeric oligonucleotide are needed to achieve the same degree of inhibition as with a telomeric oligonucleotide. We conclude that non-DNA-binding domains may modulate the DNA-binding properties of multiple Cdc13 homologues.

4.6 The crystal structure of the Cdc13 OB4 dimer from *C. glabrata*

One way to account for the long DNA binding site (with duplicated consensus motif) and the involvement of the OB4 domain in *Ct*Cdc13-DNA interaction is to invoke OB4 dimerization. The binding of a dimeric Cdc13 complex to an extended and duplicated target site would be expected to enhance substantially the affinity of interaction. In support of this idea, the OB4 domains of *Ca*Cdc13 and *Cg*Cdc13 have been shown to self-associate in two hybrid assays [16]. However, the molecular basis of OB4 dimerization is unknown. In fact, even the notion that the C terminus of Cdc13 comprises an OB fold has not been experimentally verified. We therefore screened several Cdc13 OB4 domains for recombinant expression and crystallization. In the end, we were able to express and purify OB4 of *Cg*Cdc13 (residues 607 to 754) from *E. coli* (Figure 4.3) and solved its crystal structure by single anomalous dispersion (SAD) method using Se-Met-substituted proteins at a resolution of 2.0 Å (Table 1). Indeed as predicted, the structure of *Cg*Cdc13$_{OB4}$ is made of an OB fold with a slightly deformed central β barrel sitting on

a flat surface formed by three peripheral helices, αB, αC, and αD (Figure 4.4A). Between strands β2 and β3, there is a long and extended loop, L23, which is essential for homodimerization of *Cg*Cdc13$_{OB4}$ as described below (Figure 4.4A). Given that the secondary structural elements of *Cg*Cdc13$_{OB4}$ are among the most conserved regions revealed by sequence alignments (Figure 4.5), the crystal structure of *Cg*Cdc13$_{OB4}$ supports the existence of a C-terminal OB fold in all *Saccharomyces* and *Candida* Cdc13 proteins.

Consistent with previous two-hybrid and gel filtration chromatography results, there are two *Cg*Cdc13$_{OB4}$ molecules in each asymmetric unit [16]. The large solvent-accessible surface area buried by the dimer interface (~2,420 Å) implies that *Cg*Cdc13$_{OB4}$ exists as a dimer in solution prior to crystallization. The mode of dimerization is entirely distinct from that observed for *Sc*Cdc13$_{OB1}$; whereas the symmetry dyad is perpendicular to the axis of the β barrel and the two protomers are arranged end to end for *Sc*Cdc13$_{OB1}$, the symmetry dyad is parallel to the axis of the β barrel and the two protomers are arranged side to side for *Cg*Cdc13$_{OB4}$ (Figure 4.4A). The major driving force for dimer formation of *Cg*Cdc13$_{OB4}$ is provided by hydrophobic contacts mediated by three connecting loops (Figure 4.4A). Five resides in loop L23 (^{665}YVPPV^{669}) bind into a hydrophobic cleft formed by two loops, LA1 (between αA and β1) and L45 (between β4 and β5), from the other subunit in the dimer (Figure 4.4B). In particular, Pro$_{667}$ and Pro$_{668}$ of one *Cg*Cdc13$_{OB4}$ fit snugly into a complementary surface of the other molecule (Figure 4.4B). In addition to these hydrophobic contacts, there is another interface involving a cluster of charged and polar residues (Glu$_{644}$, Glu$_{650}$, Arg$_{652}$, Lys$_{654}$, Glu$_{673}$, and Tyr$_{675}$) from strands β1, β2, and β3 of each subunit (Figure 4.4C). Together with

two ordered water molecules, these residues form an extensive and symmetric electrostatic interaction network with a total of 18 salt bridges and hydrogen bonds.

As described in the introduction, even though the Stn1-Ten1 subcomplex is clearly paralogous to RPA32-RPA14, the relationship between Cdc13 and RPA70 has remained unclear. Notably, RPA70 also contains a C-terminal OB fold (RPA70C) (39). Hence, I compared the structures of *Cg*Cdc13$_{OB4}$ and RPA70C in order to glimpse their evolutionary relationship. Three-dimensional superposition analysis revealed several marked differences between the two domains outside the central β-barrel core, arguing against a close evolutionary kinship (Figure 4.4D). RPA70C does not contain a long loop between strands β2 and β3 that is crucial for the dimerization of *Cg*Cdc13$_{OB4}$ (Figure 4.4D). On the other hand, *Cg*Cdc13$_{OB4}$ lacks several features unique to RPA70C. First, RPA70C contains a zinc ribbon motif embedded in the OB fold between strands β1 and β2, which might play a role in single stranded DNA binding (Figure 4.4D). In contrast, strands β1 and β2 in *Cg*Cdc13$_{OB4}$ are connected by a short two-residue loop. Second, the C-terminal helix in RPA70C protrudes away from the β barrel core to interact with the other two components of the RPA complex, RPA32 and RPA14, through an intermolecular three-helix bundle [31]. In contrast, the C-terminal helix of *Cg*Cdc13$_{OB4}$, αD, is short and packs together with helices αB and αC (Figure 4.4A). Hence, it is unlikely that *Cg*Cdc13$_{OB4}$ interacts with Stn1 and Ten1 in the same manner as RPA70 does with its binding partners. Therefore, our comparative structural analysis does not support the idea of a common ancestry for RPA70 and Cdc13.

4.7 The dimerization of the OB4 domain in *Ct*Cdc13

We next attempted to apply the insights derived from the *Cg*Cdc13$_{OB4}$ dimer structure to the analysis of *Ct*Cdc13. First, we investigated the ability of *Ct*Cdc13 to form dimers. A SUMO-fused *Ct*Cdc13 with a His_6 tag (SUMO-*Ct*Cdc13) and a GST-fused *Ct*Cdc13 (GST-*Ct*Cdc13) were coexpressed in *E. coli*. Cell extracts were prepared and subjected to pulldown assays using glutathione-Sepharose beads. As shown in Figure 4.6A, GST-*Ct*Cdc13 but not GST alone can coprecipitate approximately equal amounts of SUMO-*Ct*Cdc13, supporting self-association. Additional pulldown assays using either the *Ct*Cdc13$_{DBD}$ or *Ct*Cdc13$_{OB4}$ domain fusions revealed a much stronger self-association of the OB4 domain, suggesting that this domain is largely responsible for dimerization (Figure 4.6B). Interestingly, the DBD also appears to be capable of self-association, at least when overproduced in *E. coli*. The physiological relevance of this much weaker interaction remains to be determined.

We then attempted to identify dimerization-defective mutants of *Ct*Cdc13$_{OB4}$ by using the structure of *Cg*Cdc13$_{OB4}$ and a multiple sequence alignment of Cdc13 homologues as the guides. As described earlier, three connecting loops in *Cg*Cdc13$_{OB4}$ (named LA1, L23, and L45) are largely responsible for forming the dimer interface. Notably, these loop residues are not well conserved in the *Saccharomyces* and *Candida* Cdc13 homologues (Figure 4.5). Nevertheless, we reasoned that divergent sequences may be compatible with dimerization and proceeded to replace multiple amino acid residues in each corresponding loop in *Ct*Cdc13$_{OB4}$ to generate the LA1 (SISE$_{234-238}$), L23 (TILDDR$_{295-300}$), and L45 (KQKI$_{358-361}$) mutants (Figure 4.5). The abilities of the mutated OB4 domains to self-associate were then tested in pulldown assays (Figure 4.6C). As predicted, each mutant exhibited a significant reduction in self-association,

with the LA1 and L23 mutant manifesting defects more severe (~50–65% reduction) than those of the L45 mutant (~30% reduction). Hence, despite the clear sequence differences between the loops of the *Cg*Cdc13 and *Ct*Cdc13 OB4 domains, these loops appear to mediate a conserved function in protein dimerization.

4.8 The role of dimerization on DNA binding by *Ct*Cdc13

To investigate the role of dimerization on the DNA binding activity of Cdc13, we expressed full-length SUMO-tagged *Ct*Cdc13 proteins carrying the LA1, L23, and L45 mutations in *E. coli*. Notably, all three mutant proteins exhibited reduced affinity for the *C. tropicalis* telomere repeats in comparison to the wild-type protein, suggesting that dimerization contributes to DNA binding (Figure 4.7A). Because the L45 mutant is expressed at a higher level and can be purified in substantial quantities in the untagged form, we performed a more detailed comparison between this mutant and wild-type protein following ULP1 cleavage and further purification (Figure 4.7B). Interestingly, the L45 mutant evidently retained significant DNA-binding activity, as evidenced by decreasing signals for the free probe when substantial amounts of the protein were added to the binding reactions. However, a higher concentration of the mutant was needed to form the same level of complex as the wild type protein. Moreover, a broad smear can be observed below the mutant protein-DNA complex, suggesting a significant dissociation of the complex during native gel electrophoresis (Figure 4.7B). These observations support the notion that the L45 mutant binds telomeric DNA with reduced affinity and stability. Curiously, the presumptive L45 mutant-DNA complex has a reduced mobility in comparison to the wild-type complex, raising questions about its identity (Figure 4.7B,

compare lanes 2 to 4 and lanes 5 to 7). The altered mobility of the DNA-*Ct*Cdc13-L45 complex may be due to an altered conformation of the protein dimer.

4.9 Interaction between *Cg*Cdc13 and *Cg*Stn1

Previously, it has been shown that *stn1* alleles that truncate the C-terminal 123 residues fail to interact with Cdc13 and do not support viability when expressed at endogenous levels [19]. To determine the interaction between Cdc13 and Stn1 in more detail, I characterized the *C. glabrata* Cdc13-Stn1 interaction by yeast two-hybrid assays (Figure 4.8). First, I found that the C-terminal WH motifs, not the N-terminal OB fold, is responsible for the interaction with Cdc13 (Figure 4.8B). Next, I divided full-length Cdc13 into five domains based on existing knowledge about Cdc13 structure and sequence analysis (OB1: 2-165, RD: 161-240, OB2: 240-379, DBD: 403-589, OB4: 607-753). Only one region of *Sc*Cdc13 (amino acids 252-924) has been found to complement the growth defects of *cdc13* mutants [20]. It was surprising to see that *Cg*Cdc13$_{OB4}$ alone can almost fully carry out the Cdc13-Stn1 interaction (Figure 4.8B). The findings here, together with the discovery from Chapter 5, provide new insight into the mechanism of telomerase regulation by Cdc13 and Stn1. More detailed analysis needs to be done to further illustrate the molecular mechanism and physiological importance of this pair of interaction.

4.10 Discussion

We have shown that the unusually small Cdc13 homologues in *Candida* species are indeed regulators of telomere lengths and thus orthologous to the prototypical Cdc13 first

identified and characterized in *S. cerevisiae*. Our determination of the high-resolution structure of *Cg*Cdc13$_{OB4}$ also underscored the remarkable versatility of OB fold domains in mediating protein-protein interactions. We further demonstrated that the small Cdc13s likely form dimers through a homotypic interaction between the OB4 domain and that this dimerization increases the affinity of Cdc13s for the *Candida* telomere repeats and enables the proteins to perform their telomere-dedicated functions. The evolutionary and mechanistic implications of these findings are discussed below.

***Candida* Cdc13s serve telomere-specific functions.** Our detailed analysis of the DNA-binding properties of *Ct*Cdc13 suggests that this protein has sufficient affinity and sequence specificity (data provided by Dr Lue) to interact with Candida telomeres *in vivo* and perform telomere specific functions. This conclusion is supported by ChIP analysis of *Ca*Cdc13, which revealed telomere localization of this small Cdc13 *in vivo*. However, it is at odds with a recent report that posits a more general function for small Cdc13s in chromosome transactions [30]. This alternative proposition was based on analyses of the DNA-binding properties of the DBDs from *C. albicans*, *C. parapsilosis*, and *C. glabrata*. All three DBDs exhibited low affinity (ranging from ~100 to 600 nM) and sequence specificity for short telomeric oligonucleotides, leading the investigators to discount a telomere-specific function. Our results on *Ct*Cdc13 and *Cg*Cdc13 suggest that dimerization-assisted DNA binding may be quite prevalent among Cdc13 homologues and that the DNA binding properties of the DBDs alone do not always reflect those of the full-length proteins.

The propensity of telomere proteins to dimerize. A striking implication of the current report, when juxtaposed against previous findings, is that Cdc13 homologues

have a propensity to dimerize and have evolved different modes of dimerization. As described earlier, whereas *Saccharomyces* and *Kluyveromyces* Cdc13s form dimers through their OB1 domains, *Candida* Cdc13s use the structurally quite distinct OB4 domains to mediate dimerization [16, 17]. How can the distinct modes of dimerization evolve so readily for Cdc13 (and other telomere proteins such as TRF1, TRF2, and Taz1)? An attractive hypothesis invokes the colocalization of multiple molecules of a telomere-binding protein on the iterative telomere sequence [24, 32]. The clustering of a protein greatly increases its local concentration and amplifies the effect of mutations on protein-protein interactions. In this setting, even a low free energy of interaction conferred by a few point mutations may lead to a substantial increase in the fraction of molecules that bind to each other, which may in turn enhance telomere protection sufficiently to allow for selection.

Another notable implication of the combined observations on *Saccharomyces* and *Candida* Cdc13 dimerization is that dimerization can serve different purposes in different organisms. In particular, dimerization of *Sc*Cdc13 is not required for high-affinity DNA binding; the *Sc*Cdc13$_{DBD}$ domain alone interacts with an 11-nt telomere oligonucleotide with a K_d in the picomolar range. Rather, dimerization of *Sc*Cdc13 has been shown to modulate its interactions with Pol1 and to regulate telomere lengths through additional mechanisms [16]. Why then is dimerization of small Cdc13s necessary for high-affinity DNA binding? The answer to this puzzle may reside in the extraordinary telomere sequence divergence exhibited by *Candida* species [23]. This sequence divergence presents a considerable challenge to Cdc13: to evolve suitable affinity and specificity for the different telomere repeats during a short evolutionary time span. However, the OB_{DBD}

domains of *Candida* Cdc13s align well with the corresponding domain in *Sc*Cdc13, and many of the residues implicated in ScCdc13-DNA interactions are conserved in the *Candida* proteins [33]. Furthermore, phylogenetic analysis does not yield evidence of more-rapid evolution of OB_{DBD} relative to OB4 of *Candida* Cdc13s. Thus, instead of evolving unique recognition specificity for each telomere repeat, the *Candida* Cdc13s may have largely retained a universal preference for GT-rich sequence elements within the divergent repeats and used the duplicated binding domains in the dimeric protein complex to enhance binding affinity. Regardless of the potential outcomes, comparative analysis of *Candida* Cdc13-DNA interactions promises to provide a useful paradigm for understanding the coevolution of DNA-binding proteins and their target sequences.

The versatility of OB-fold domains in mediating protein-protein interactions. Even though the OB fold domain was initially defined as an oligonucleotide/oligosaccharide-binding module, more-recent studies have highlighted the remarkable functional diversity of this protein fold and the myriad ways in which this fold can mediate protein-protein interactions [16, 34, 35]. In keeping with this theme, our high-resolution structures of the *Sc*$Cdc13_{OB1}$ dimer and the *Cg*$Cdc13_{OB4}$ dimer revealed dramatically distinct modes of dimerization. In the case of OB1, the two protomers are arranged end to end, and the symmetry dyad is perpendicular to the axis of the β-barrel. By contrast, the *Cg*$Cdc13_{OB4}$ dimer involves a 2-fold symmetry axis that runs parallel to the β-barrel axis and a side-to-side dimerization interface (Figure 4.6A). It is also worth noting that despite our success in identifying dimerization mutants of *Ct*$Cdc13_{OB4}$, the residues implicated in *Cg*$Cdc13_{OB4}$ and *Ct*$Cdc13_{OB4}$ dimerization are in fact not well conserved in other homologues (Figure 4.5). Hence, dramatically different sequences in

the connecting loops of the Cdc13 OB4 domain are compatible with dimerization, making it extremely challenging to infer this biochemical property based on sequence analysis alone. It is tempting to speculate that the repeated utilization of OB fold domains in proteins associated with single-stranded telomeres may be due not only to its nucleic acid binding activity but also to its versatility in binding protein partners.

The evolutionary relationship between CST and RPA. As described before, whereas there are compelling supports for structural and functional similarities between Stn1-Ten1 and RPA32-RPA14, the relationship between Cdc13 and RPA70 has remained unclear. Our results provide additional arguments against a close evolutionary kinship between Cdc13 and RPA70. Specifically, we showed that the last OB fold of Cdc13 does not resemble the corresponding domain in RPA70. Coupled with previous crystallographic and NMR analyses, we now have high-resolution structures of three domains in Cdc13, each of which proved to be quite different from its putative RPA70 counterpart. Thus, CDC13 may not have arisen through a duplication of the RPA70 gene and then undergone functional specialization. Rather, Cdc13 may have originated independently from a different OB fold-containing protein and been recruited later to the Stn1-Ten1 complex to enhance its function. This notion is supported by the apparent absence of a Cdc13 homologue in *S. pombe*, as well as the very disparate sizes and structures of mammalian and plant CTC1s, which are presumed functional equivalents of Cdc13 in these organisms [11, 12]. Further analyses of Cdc13 and other large CST subunits should provide insights on the evolutionary origin and mechanistic diversity of these proteins.

4.11 Materials and Methods

Sequence analysis

Cdc13 homologues from *Candida* and *Saccharomyces* spp. were identified from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Broad Institute (http://www.broad.mit.edu/annotation/genome/candida_group/Blast.html) databases by BLAST or psi-BLAST searches. The multiple sequence alignment was generated using the PROMALS server (http://prodata.swmed.edu/promals/promals.php) and displayed using Boxshade (http://www.ch.embnet.org/software/BOX_form.html).

Telomere analyses

The telomere length analysis and the two dimensional gel analysis of circular and linear telomeric DNA were performed as previously described [36].

Gel electrophoretic mobility shift analysis

Full-length *Ct*CDC13 and individual domains (DBD, amino acids 1 to 195; OB4, amino acids 196 to 369) were cloned into the pSMT3 vector to enable the expression of His6-SUMO-Cdc13 fusion proteins. Because of the atypical translation of the CUG codon in *Candida* species, the CTG triplets encoding amino acids 33 and 132 of *Ct*Cdc13 were mutated to TCG to enable the expression of wild-type proteins in *Escherichia coli* [37]. Following induction, extracts were prepared and the fusion proteins purified with Ni-nitrilotriacetic acid (NTA) chromatography as previously described [36]. The fusion protein was cleaved by the ULP1 protease, and the Cdc13 fragment was purified away from the His_6-SUMO tag by a second round of Ni-NTA affinity chromatography. Some

of the DNA-binding reactions employed *Ct*Cdc13 that had been further purified over a glycerol gradient.

Full-length *Cg*CDC13 and its DBD (amino acids 404 to 594) were cloned into the pSMT3 vector and purified using the same method. Binding reactions contained 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), and 5% glycerol. Following incubation at 25°C for 20 min, the reaction mixtures were electrophoresed through a nondenaturing polyacrylamide gel to resolve the free probe from the DNA-protein complex. Binding activity was analyzed using a Typhoon PhosphorImager and ImageQuant software (GE Healthcare). To examine the effect of dimerization on DNA binding, the following amino acids in three connecting loops in the *Ct*Cdc13 OB4 domain were mutated by QuikChange: SISE234-238 in LA1, TILDDR295-300 in L23, and KQKI358-361in L45. Each His6-SUMOfused Cdc13 mutant protein was expressed in and purified from E. *coli* BL21(DE3). The binding activities of the mutant proteins were analyzed as described above.

Coexpression and GST pulldown assays

The genes encoding full-length *Ct*Cdc13 and individual domains were transferred from the pSMT3 vector into the pGEX4T-2 vector (GE Healthcare). Each His_6-SUMO-Cdc13 fusion protein was coexpressed with either the corresponding glutathione S-transferase (GST)–Cdc13 fusion protein or GST in *E. coli* BL21(DE3). To examine the roles of the connecting loops in *Ct*Cdc13OB4 dimerization, the following three sets of amino acids were mutated by QuikChange: $SISE_{234-238}$ in LA1, $TILDDR_{295-300}$ in L23, and $KQKI_{358-361}$ in L45. Each His_6-SUMO-fused *Ct*$Cdc13_{OB4}$ mutant protein was coexpressed with the

corresponding GST-fused mutant protein in *E.coli* BL21(DE3). Following induction, extracts were prepared and subjected to GST pulldown assays. Briefly, ~ 1 to 3 mg of each extract was incubated with 20 µl of glutathione-Sepharose beads in 300 µl of 1×phosphate-buffered saline (PBS) (10 mM Na_2HPO_4, pH 7.3, 1.8 mM KH_2PO_4, 140 mM NaCl, and 2.7 mM KCl) containing 10% glycerol and 0.1% Triton X-100. Following incubation at 25°C for 1 h, the beads were washed with 1 ml of the same buffer five times. Pulldown samples were analyzed by SDS-polyacrylamide gel electrophoresis, followed by staining with Coomassie brilliant blue R-250 or Western blotting.

ChIP

Chromatin immunoprecipitation (ChIP) of TAP-tagged Cdc13 was carried out using the same procedure as described earlier for tagged *Candida* Rap1 [36].

Expression, purification, and crystallization of *Cg*Cdc13$_{OB4}$

*Cg*Cdc13$_{OB4}$ was cloned into the pMST3 vector (a modified pET28b vector with the SUMO sequence cloned 3' to the His_6 tag [6]), and the resulting expression plasmid was transformed into *E. coli* BL21(DE3). After induction for 16 h with 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) at 20°C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50mM Tris-HCl, pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride [PMSF], 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 2 h at 4°C

before elution with 250 mM imidazole. Then, the Ulp1 protease was added, and the mixture was incubated for 12 h at 4°C to remove the His_6-SUMO tag. *Cg*$Cdc13_{OB4}$ was then further purified by passage through a Mono-Q ion-exchange column and by gel filtration chromatography on a Hiload Superdex75 (GE Healthcare) equilibrated with 25 mM Tris-HCl, pH 8.0, 150mM NaCl, and 5mM dithiothreitol (DTT). The purified *Cg*$Cdc13_{OB4}$ was concentrated to 20 mg/ml and stored at -80°C.

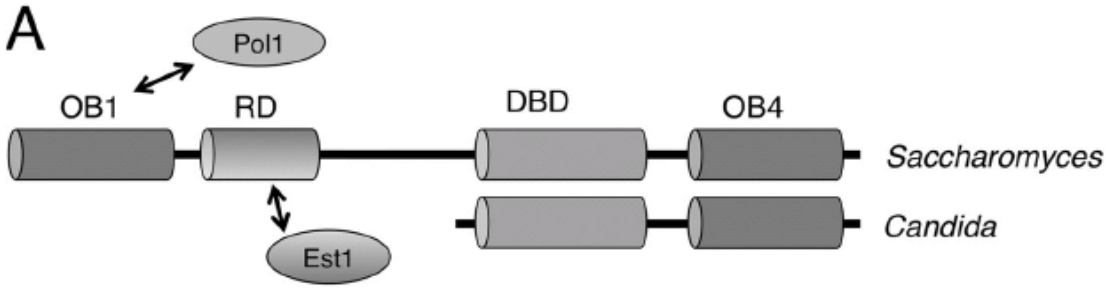
Crystals of the wild-type protein were grown by the sitting-drop vapor diffusion method at 4°C. However, repeated attempts to obtain crystals of Se-Met-substituted wild-type *Cg*$Cdc13_{OB4}$ were unsuccessful. Hence, several single Met-to-Leu point mutations of *Cg*$Cdc13_{OB4}$ were evaluated for crystallization. Eventually, crystals of Se-Met-substituted M661L mutant protein were successfully grown at 4°C by the sitting-drop vapor diffusion method. The precipitant contained 32% PEG4000, 10 mM $CaCl_2$, 0.1 M Tris-HCl, pH 7.4, and 0.2 M ammonium sulfate. Crystals were gradually transferred into a harvesting solution (0.2 M ammonium sulfate, 20% glycerol, 34% PEG 4000, 10 mM $CaCl_2$, 0.1M Tris-HCl, pH 7.4, and 10 mM DTT) before being flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions. A Se-Met single anomalous dispersion (SAD) (at Se peak wavelength) data set with a resolution of 2.0 Å was collected at beam line 21ID-D at APS and processed using HKL2000 [38]. *Cg*$Cdc13_{OB4}$ crystals belong to space group $P2_1$ and contain two *Cg*$Cdc13_{OB4}$ molecules per asymmetric unit. Four selenium atoms were located and refined, and the SAD phases were calculated using SHARP [39]. The initial SAD map was significantly improved by solvent flattening. A model was automatically built into the modified experimental electron density by using ARP/WARP [40]. The model was then transferred into the

native unit cell by rigid-body refinement and further refined using simulated-annealing and positional refinement in CNS [41], with manual rebuilding using program O [42]. The final refined structure shows that Met661, located in the loop region between strands β2 and β3, is solvent exposed and makes no contributions to the dimer interface. Thus, the M661L mutation is unlikely to have any effect on protein folding, stability, or dimerization. The majority (90%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

Yeast two-hybrid assay

The yeast two-hybrid assays were performed using L40 strain harboring pBTM116 and pACT2 (Clontech) fusion plasmids. Colonies containing both plasmids were selected on –Leu –Trp plates. β-Galactosidase activities were measured by a liquid assay [43].

Figure 4.1 Domain organizations of Cdc13s and the role of *Candida* Cdc13 in telomere regulation.
(A) The different domain organizations of Cdc13 homologues from *Saccharomyces* and *Candida* species are illustrated. The OB1 and RD domains of *Saccharomyces* Cdc13 have been shown to interact with Pol1 and Est1, respectively.
(B) (Top) The expression of TAP-tagged Cdc13 protein in extracts derived from the untagged and tagged strains were analyzed by Western blotting using antibodies directed against protein A. The positions of *Ca*Cdc13-TAP and two cross-reacting proteins are indicated by an arrow and two asterisks, respectively. (Bottom) Strains with or without TAP-tagged Cdc13 were subjected to ChIP analysis using IgG-Sepharose. The input (0.13, 0.64, and 3.2%) and precipitated DNAs (100%) were spotted on nylon filters and probed with labeled *C. albicans* telomere repeats.



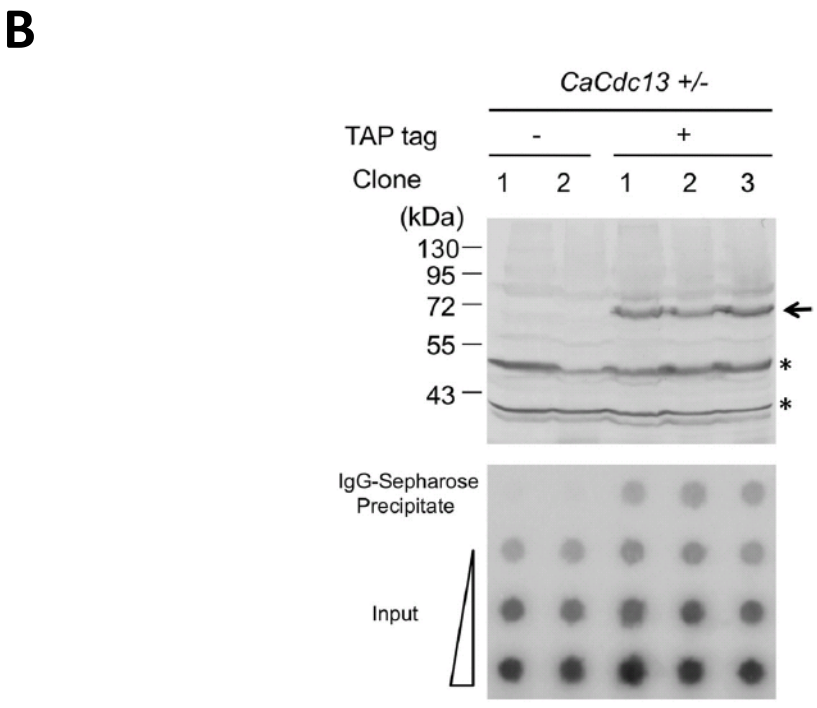


Figure 4.2 Specific binding of *Candida* telomeric DNA by *C. tropicalis* [*C.tro*] Cdc13 (Experiment and figures prepared by E.Y. Yu)
(A) The *C. tropicalis* Cdc13 protein and the domains tested for DNA binding are illustrated.
(B) Purified full-length (FL) *Ct*Cdc13 and the DBD were analyzed by SDS-PAGE and Coomassie staining.
(C) *Ct*Cdc13 was incubated with 7.5nM labeled *C. tropicalis* TEL-GX1.5B and the indicated competitor oligonucleotides (*C. tro* telomere, same as the probe; nontelomeric competitor,AATTGTCGACTTATGGAGCAATTCTTGTTAAACA). The resulting DNA-protein complexes were analyzed by native gel electrophoresis. The concentrations of *Ct*Cdc13 and the levels of the competitors relative to the probe for the reactions are listed at the top.
(D) The indicated concentrations of full-length *Ct*Cdc13 were incubated with 7.5 nM probe consisting of two copies of the *C. tropicalis* telomere repeat (*C.tro* TEL-GX2). The resulting DNA-protein complexes were analyzed by native gel electrophoresis. The K_d for this DNA-protein interaction was estimated to be ~40 nM based on the concentration of protein needed to reduce the free probe by 50% (four left lanes). Some assays also included excess unlabeled oligonucleotides consisting of various telomere repeat sequences. These competitor oligonucleotides were added at 2.5-fold or 10-fold molar excess.
(E) The indicated concentrations of full-length *Ct*Cdc13 or DBD were incubated with 7.5nM probe consisting of two copies of the *C. tropicalis* telomere repeat (*C. tro* TEL-GX2). The indicated competitor oligonucleotides were added at 10-fold molar excess. *C.alb, C. albicans; S.cer, S. cerevisiae.*
(F) The indicated concentrations of full length *Cg*Cdc13 or DBD were incubated with 7.5nM probe consisting of three copies of the *C. glabrata* telomere repeat. The resulting DNA-protein complexes were analyzed by native gel electrophoresis. The K_d for the DNA-*Cg*Cdc13 interaction was estimated to be ~20 nM based on the concentration of protein needed to reduce the free probe by 50%.

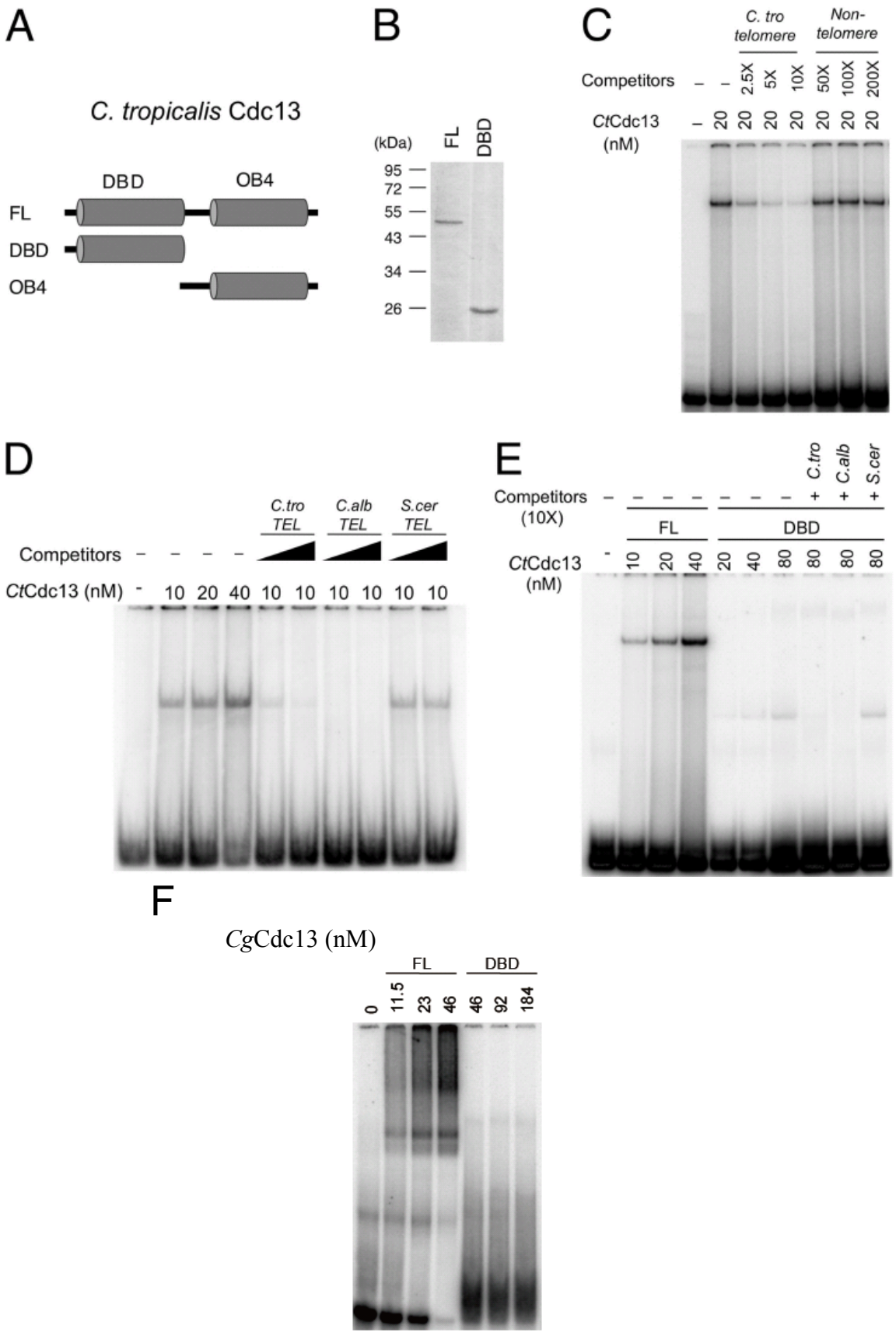
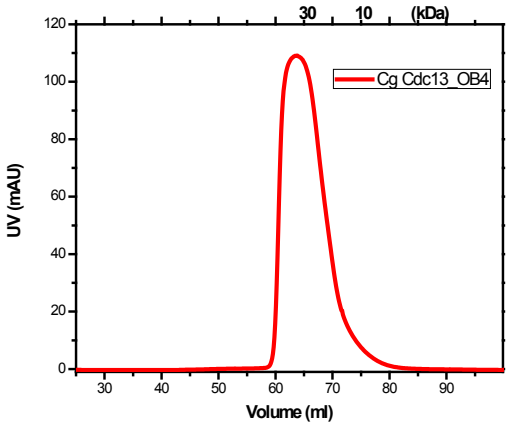
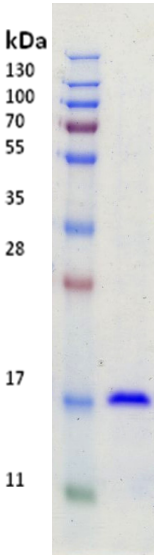

A
C. tropicalis Cdc13
DBD
OB4
FL
DBD
OB4
B
(kDa)
FL
DBD
95
72
55
43
34
26
C
C. tro telomere
Non-telomere
Competitors
2.5X
5X
10X
50X
100X
200X
CtCdc13 (nM)
20
D
C.tro TEL
C.alb TEL
S.cer TEL
Competitors
CtCdc13 (nM)
10 20 40 10 10 10 10 10 10
E
Competitors (10X)
C.tro
C.alb
S.cer
FL
DBD
CtCdc13 (nM)
10 20 40 20 40 80 80 80 80
F
CgCdc13 (nM)
FL
DBD
0 11.5 23 46 46 92 184

Figure 4.3 C-terminal domain of *C. glabrata* $Cdc13_{OB4}$.
(A) Gel filtration chromatography profile (Hiload Superdex 750) of *Cg*$Cdc13_{OB4}$
(B) SDS-PAGE *Cg*$Cdc13_{OB4}$ corresponding to the peak fraction in the gel filtration profile in A
(C) Crystals of *Cg*$Cdc13_{OB4}$

A



B



C

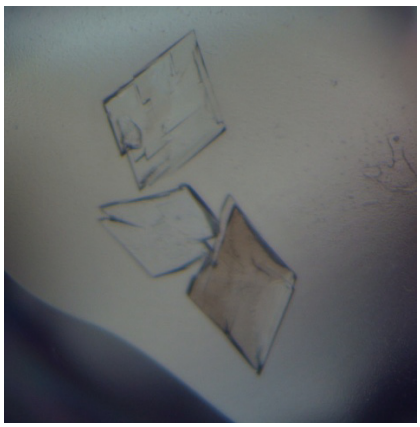


Figure 4.4 Structure of the C-terminal OB fold of *C. glabrata* Cdc13

(A) Ribbon diagram of two views of the *Cg*Cdc13$_{OB4}$ dimer. The two subunits are colored in green and cyan, respectively. The secondary structural elements are labeled. The *Cg*Cdc13$_{OB4}$ dimer at right is rotated by 70° about a horizontal axis relative to the dimer at left.

(B) The hydrophobic dimer interface of *Cg*Cdc13$_{OB4}$. One *Cg*Cdc13$_{OB4}$ molecule is in surface representation and colored according to its electrostatic potential. The other molecule is in ribbon representation and colored in green. Side chains of residues in loops LA1, L23, and L45 important for dimerization are shown as stick models.

(C) An extensive electrostatic interaction network is formed by a cluster of symmetry-related charged and polar residues on the β1-β2-β3 side of the barrel. The intermolecular hydrogen bonds are shown as dashed magenta lines.

(D) Superposition of *Cg*Cdc13$_{OB4}$ on the crystal structure of human RPA70C reveals that *Cg*Cdc13$_{OB4}$ is not structurally similar to RPA70C. *Cg*Cdc13$_{OB4}$ and RPA70C are colored in green and light blue, respectively. The superposition is based on the OB fold β-barrels of the proteins.

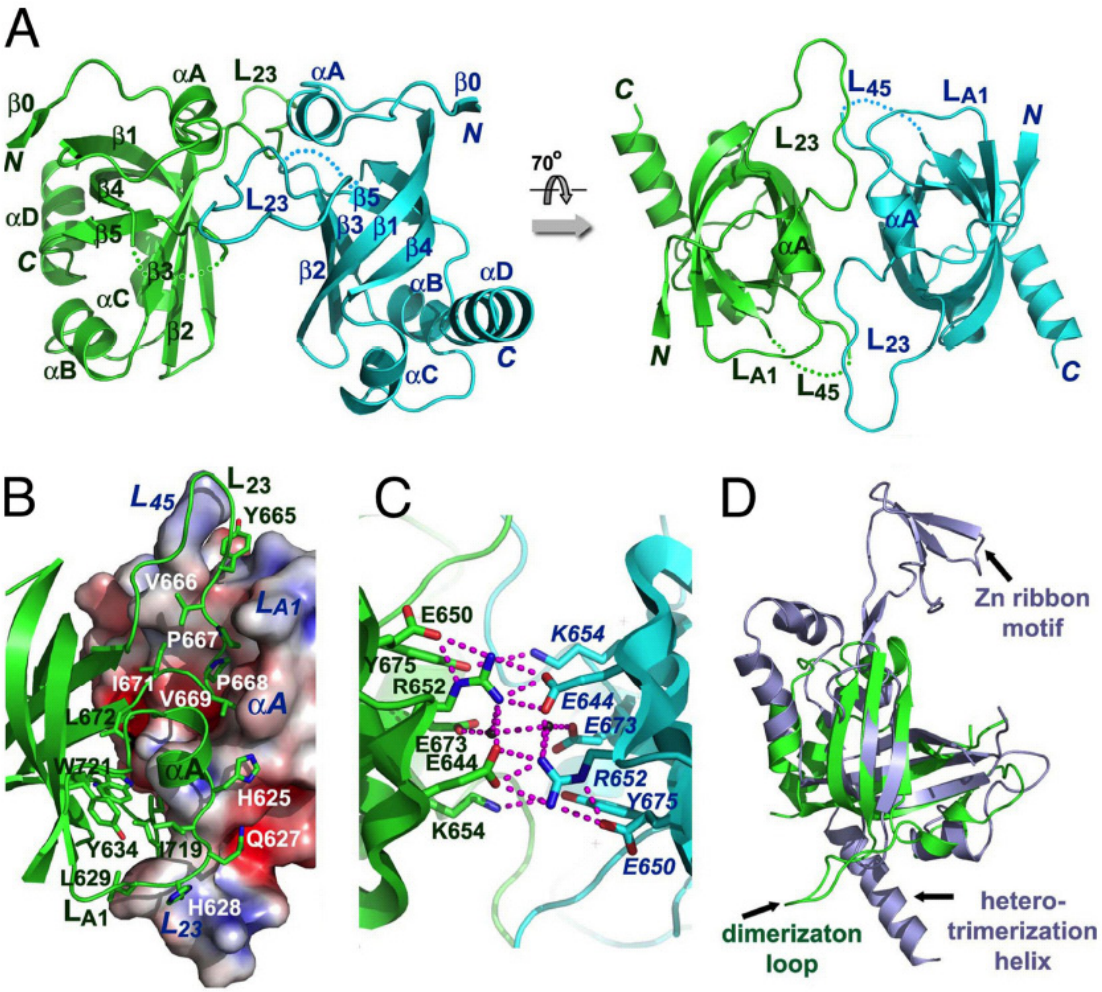


Figure 4.5 Alignment of the OB4 domains of Cdc13 homologues from *Saccharomycotina* yeast.

The OB4 domains of Cdc13 homologues were aligned using the PROMALS server and displayed using Boxshade. The abbreviations are as follows: *S_cer, S. cerevisiae; A_gos, A. gossypii; K_lac, K. lactis; C_gla, C. glabrata; C_tro, C. tropicalis; C_alb, C. albicans; D_han, D. hansenii; C_par, C. parapsilosis; L_elo, L. elongisporus; C_lus, C. lusitaniae; C_gui, C. guilliermondi.*

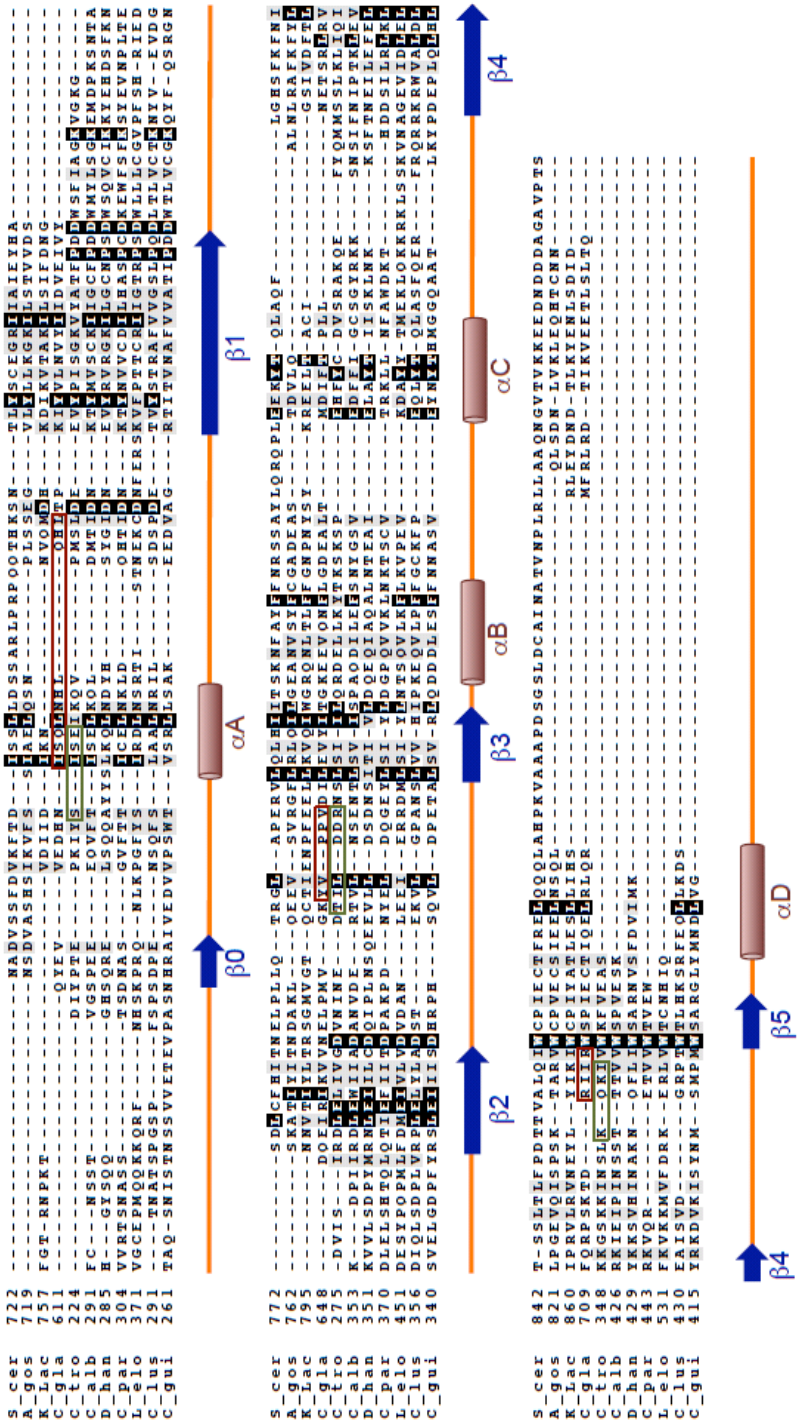


Figure 4.6 Self-association of *Ct*Cdc13 (Experiment and figures prepared by E.Y. Yu)
(A) (Top) The indicated proteins were coexpressed and subjected to GST pulldown analysis. The bound proteins were analyzed by SDS-PAGE and either Coomassie staining or Western blotting (WB) using anti-His tag antibodies. (Bottom) The levels of His-tagged SUMO-Cdc13 protein in the input extracts were analyzed by Western blotting.
(B) (Top) The indicated proteins were coexpressed and subjected to GST pulldown analysis. The glutathione-Sepharose-bound proteins were analyzed by SDS-PAGE and Coomassie staining. (Bottom) The levels of His-tagged SUMO fusion proteins (SUMO-DBD or SUMO-OB4) in the input extracts were analyzed by Western blotting.
(C) GST pulldown assays were performed using either wild-type or mutated OB4 domains fused to the GST and SUMO tags. The ratio of the SUMO fusion to GST fusion protein in each precipitated sample was quantified, normalized to the wild-type sample, and then plotted. The results are from three independent experiments.

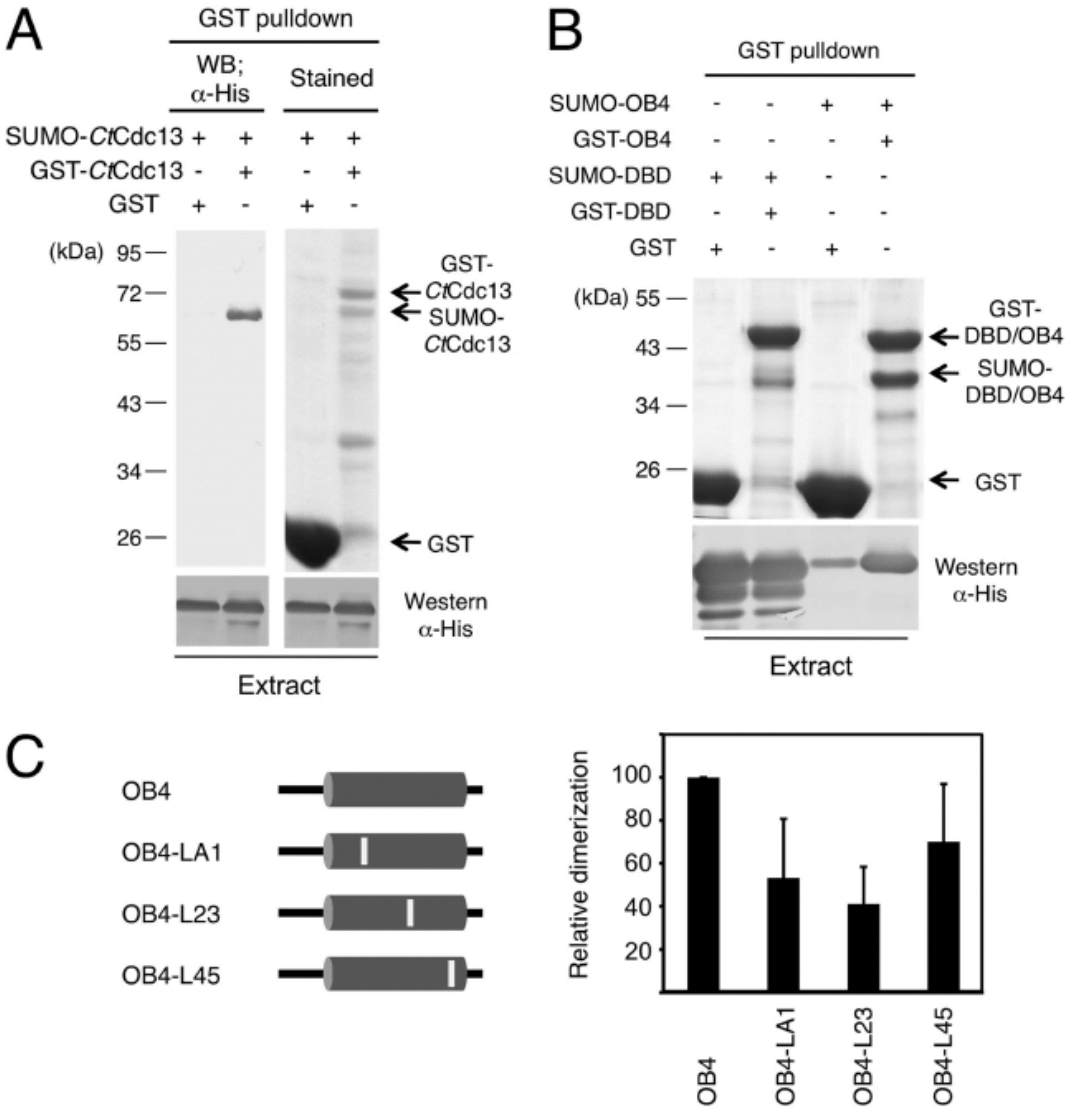


Figure 4.7 The effects of OB4 mutations on DNA binding by *Ct*Cdc13 (Experiment and figures prepared by E.Y. Yu)
(A) EMSAs were performed using increasing concentrations (116, 232, and 464 nM) of SUMO-fused wild-type and mutated *Ct*Cdc13s bearing amino acid replacements in the OB4 domain and the TEL-GX1.5B probe (7.5 nM).
(B) The indicated concentrations of wild-type and L45 mutant proteins were tested for binding to the TEL-GX2 probe.

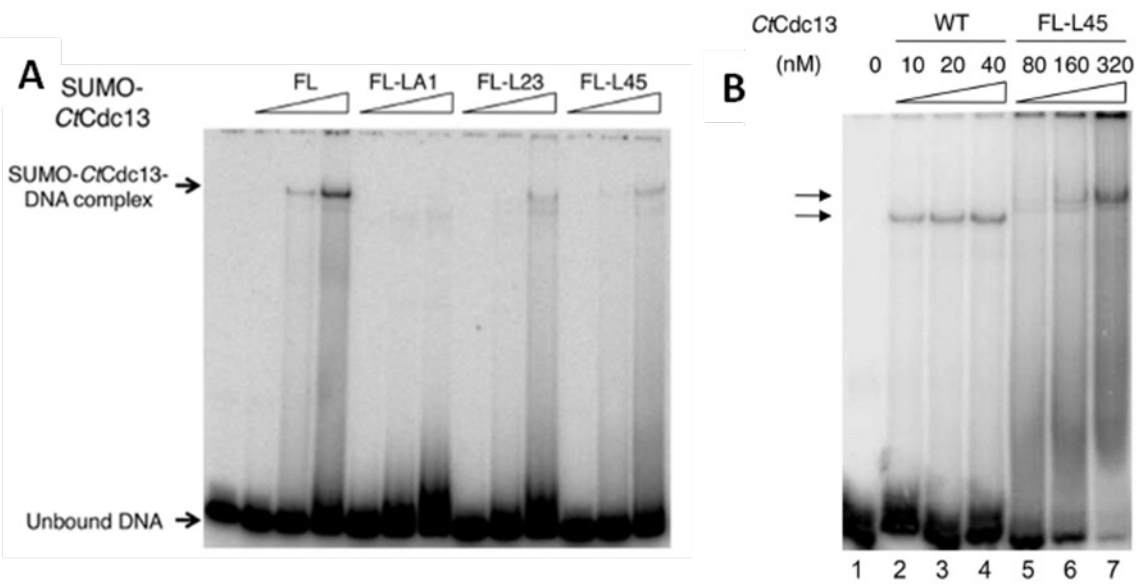


Figure 4.8 *Candida glabrata* Cdc13 uses OB4 to interact with the C-terminus of Stn1
(A) Domain organizations of Cdc13 and Stn1
(B) *Cg*Cdc13 interaction with different domains of *Cg*Stn1 in a yeast two-hybrid assay. Interaction of LexA–Stn1 with GAD–Cdc13 was measured as β-galactosidase activity. Data are the average of three independent β-galactosidase measurements normalized to the wild-type Stn1–Ten1 interaction, arbitrarily set to 100.

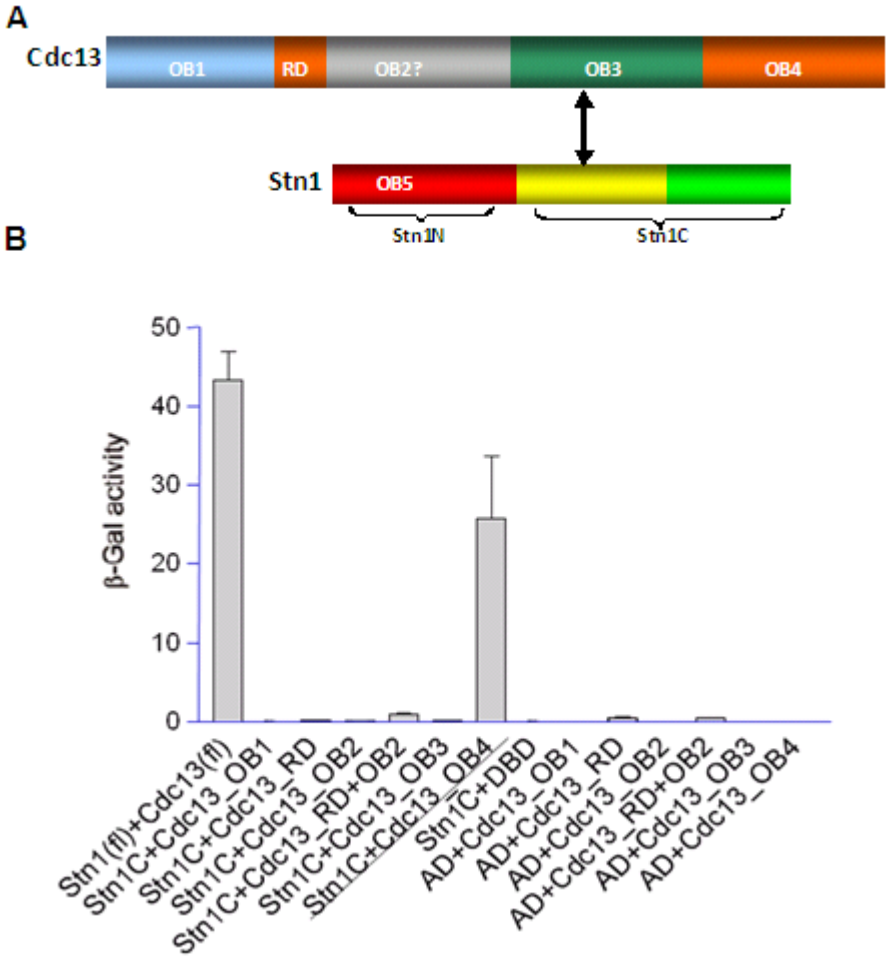


Table 4.1 Data collection, phasing and refinement statistics and model validation with Ramachandran plot for *Cg*Cdc13$_{OB4}$

Data collection	
Parameter[a]	**Value**[b] **for *Cg*Cdc13$_{OB4}$**
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	61.815, 38.745, 62.709
α, β, γ (º)	90, 109.021, 90
Wavelength (Å)	0.97949
Resolution (Å)	100-2.0
R_{merge} (%) (high res. shell)	6.4 (16.2)
I/σ (high res. cell)	53.7 (11.9)
Completeness (%) (high res. shell)	98.7 (92.3)
Redundancy (high res. shell)	7.2 (6.2)
Phasing	
Acentric Phasing Power	1.781
Accentric reflections FOM	0.26324
Centric reflections FOM	0.10262
Refinement	
Resolution (Å)	30-1.90
No. reflections	22461
$R_{work/}$ R_{free} (%)	20.59/23.82
B-factors (Å^2)	
Protein	42.655
Water	46.244
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (º)	1.075

Reference:

1. de Lange, T., *How telomeres solve the end-protection problem.* Science, 2009. **326**(5955): p. 948-52.
2. Giraud-Panis, M.J., et al., *CST meets shelterin to keep telomeres in check.* Mol Cell. **39**(5): p. 665-76.
3. Karlseder, J., et al., *p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2.* Science, 1999. **283**(5406): p. 1321-5.
4. Price, C.M. and T.R. Cech, *Telomeric DNA-protein interactions of Oxytricha macronuclear DNA.* Genes Dev, 1987. **1**(8): p. 783-93.
5. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.
6. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
7. De Lange, T., V. Lundblad, and E.H. Blackburn, *Telomeres*. 2nd ed. Cold Spring Harbor monograph series. 2006, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. viii, 576 p.
8. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.
9. Sun, J., et al., *Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres.* Genes Dev, 2009. **23**(24): p. 2900-14.
10. Martin, V., et al., *Protection of telomeres by a conserved Stn1-Ten1 complex.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 14038-43.
11. Miyake, Y., et al., *RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway.* Mol Cell, 2009. **36**(2): p. 193-206.
12. Surovtseva, Y.V., et al., *Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes.* Mol Cell, 2009. **36**(2): p. 207-18.
13. Mitton-Fry, R.M., et al., *Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13.* J Mol Biol, 2004. **338**(2): p. 241-55.
14. Chan, A., J.B. Boule, and V.A. Zakian, *Two pathways recruit telomerase to Saccharomyces cerevisiae telomeres.* PLoS Genet, 2008. **4**(10): p. e1000236.
15. Pennock, E., K. Buckley, and V. Lundblad, *Cdc13 delivers separate complexes to the telomere for end protection and replication.* Cell, 2001. **104**(3): p. 387-96.
16. Sun, J., et al., *Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase alpha.* Cell Res, 2010. **21**(2): p. 258-74.
17. Mitchell, M.T., et al., *Cdc13 N-terminal dimerization, DNA binding, and telomere length regulation.* Mol Cell Biol. **30**(22): p. 5325-34.
18. Grossi, S., et al., *Pol12, the B subunit of DNA polymerase alpha, functions in both telomere capping and length regulation.* Genes Dev, 2004. **18**(9): p. 992-1006.
19. Petreaca, R.C., H.C. Chiu, and C.I. Nugent, *The role of Stn1p in Saccharomyces cerevisiae telomere capping can be separated from its interaction with Cdc13p.* Genetics, 2007. **177**(3): p. 1459-74.

20. Wang, M.J., et al., *Telomere-binding and Stn1p-interacting activities are required for the essential function of Saccharomyces cerevisiae Cdc13p.* Nucleic Acids Res, 2000. **28**(23): p. 4733-41.
21. Gelinas, A.D., et al., *Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19298-303.
22. Paschini, M., E.K. Mandell, and V. Lundblad, *Structure prediction-driven genetics in Saccharomyces cerevisiae identifies an interface between the t-RPA proteins Stn1 and Ten1.* Genetics, 2010. **185**(1): p. 11-21.
23. McEachern, M.J. and E.H. Blackburn, *A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3453-7.
24. Lue, N.F., *Plasticity of telomere maintenance mechanisms in yeast.* Trends Biochem Sci, 2010. **35**(1): p. 8-17.
25. Teixeira, M.T. and E. Gilson, *Telomere maintenance, function and evolution: the yeast paradigm.* Chromosome Res, 2005. **13**(5): p. 535-48.
26. Lin, J.J. and V.A. Zakian, *The Saccharomyces CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13760-5.
27. Nugent, C.I., et al., *Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance.* Science, 1996. **274**(5285): p. 249-52.
28. Hughes, T.R., et al., *Identification of the single-strand telomeric DNA binding domain of the Saccharomyces cerevisiae Cdc13 protein.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6457-62.
29. Lin, Y.C., et al., *Specific binding of single-stranded telomeric DNA by Cdc13p of Saccharomyces cerevisiae.* J Biol Chem, 2001. **276**(27): p. 24588-93.
30. Mandell, E.K., et al., *Sequence-specific binding to telomeric DNA is not a conserved property of the Cdc13 DNA binding domain.* Biochemistry, 2011. **50**(29): p. 6289-91.
31. Bochkareva, E., et al., *Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA.* EMBO J, 2002. **21**(7): p. 1855-63.
32. Kuriyan, J. and D. Eisenberg, *The origin of protein interactions and allostery in colocalization.* Nature, 2007. **450**(7172): p. 983-90.
33. Anderson, E.M., W.A. Halsey, and D.S. Wuttke, *Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13.* Biochemistry, 2003. **42**(13): p. 3751-8.
34. Bochkarev, A. and E. Bochkareva, *From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold.* Curr Opin Struct Biol, 2004. **14**(1): p. 36-42.
35. Wang, F., et al., *Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex.* Structure, 2010. **18**(9): p. 1159-70.
36. Yu, E.Y., et al., *Rap1 in Candida albicans: an unusual structural organization and a critical function in suppressing telomere recombination.* Mol Cell Biol, 2010. **30**(5): p. 1254-68.

37. Santos, M.A., G. Keith, and M.F. Tuite, *Non-standard translational events in Candida albicans mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon.* EMBO J, 1993. **12**(2): p. 607-16.
38. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.
39. Bricogne, G., *Ab initio macromolecular phasing: blueprint for an expert system based on structure factor statistics with built-in stereochemistry.* Methods Enzymol, 1997. **277**: p. 14-8.
40. Perrakis, A., et al., *ARP/wARP and molecular replacement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 10): p. 1445-50.
41. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
42. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
43. Moretti, P., et al., *Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1.* Genes Dev, 1994. **8**(19): p. 2257-69.

CHAPTER 5

ANALYSES OF THE *KLUYVEROMYCES* CDC13-EST1 INTERACTION

5.1 Introduction

In budding yeast, *Saccharomyces cerevisiae*, five genes are required for the telomerase pathway: an RNA (TLC1), a reverse transcriptase (Est2) and at least three regulatory proteins (Est1, Est3 and Cdc13) [1-4]. The gene encoding the RNA subunit, TLC1, was first identified by Singer and Gottschling [1]. The protein components were identified from genetic screening based on the EST phenotype in budding yeast *S. cerevisiae* [2, 3]. *TLC1* and *EST2* encode the RNA and reverse transcriptase subunits of telomerase, respectively. The two encoded subunits are essential for catalysis and telomerase activity is absent in extracts from strains defective in *EST2* and *TLC1* [4]. In contrast, mutations in *EST1*, *EST3*, and *CDC13* do not diminish enzyme activity *in vitro*, although they result in similar severe telomere replication defects as Δ*est2* or Δ*tlc1*. The highly basic 82-kDa Est1 protein possesses three distinct biochemical functionalities. First, it associates with the telomerase RNP through TLC1. Second, it also interacts with single-stranded telomeric DNA. Finally, Est1 can be recruited to the telomere region by Cdc13, thus localize the telomerase catalytic core to the chromosome ends [5, 6].

Est1 interacts with telomerase by binding to a bulged stem loop in the TLC1 RNA [7]. Sequence alignment of the telomerase RNAs from *Saccaromyces cerevisiae* and six

Kluyveromyces species revealed a conserved region that is essential for telomere maintenance. The conservation is not only observed in primary nucleotide sequence but also in secondary structure: a bulged-stem structure. Overexpression of Est1p has been shown to compensate the phenotype of bulged-stem mutant RNAs [7]. Co-immunoprecipitation also indicated the co-localization of Est1 and a small RNA containing the bulged stem, suggesting a direct interaction. Notably, this interaction is only dependent on the bulged stem and independent of other regions of the RNA. It was proposed that this bulge links the enzymatic core of telomerase with Est1, allowing Est1p to recruit and thus activate telomerase at the telomere in subsequent steps [8].

Both yeast and human Est1 proteins bind to single-stranded G-rich DNA [9-14], but the functional importance of this interaction is not clear. It has been shown that Est1p not only binds to telomeric G-rich ssDNA, but also possesses a biochemical activity of converting telomeric G-rich ssDNA into G-quadruplex structures [14]. It has also been proposed that its DNA binding activity is secondary to the Cdc13-ssDNA interaction [13]. Though both Cdc13 and Est1 bind single-stranded DNA, they make separate contributions to telomere replication and stability. Est1 only participates in the telomerase pathway whereas Cdc13 has an additional essential function in protecting the end of the chromosome.

Although Est1 displayed specific DNA and RNA binding, neither activity contributed significantly to telomerase stimulation. Rather, Est1 mediated telomerase up-regulation through direct contacts with the reverse transcriptase subunit. Regulation of telomerase could take place at three levels: at the level of recruitment to the telomere terminus, at the initiation of elongation, or at the rate and processivity of the elongation

cycles. Most notably, the Est1 and Cdc13 interaction at the recruitment stage has been the focus of multiple research groups. The primary evidence for the recruitment model stems from a number of gene fusion experiments in which Cdc13 or its DNA-binding domain were fused to Est2, Est1, or Est3. The chimeric proteins could mitigate or even completely rescue the telomere maintenance defects of *cdc13–2* and *est1Δ* strains [5, 15]. For example, a $Cdc13_{DBD}$-Est2 fusion can bypass Est1 in telomere maintenance. Consequently, these experiments suggest that the recruitment step is essential for telomere maintenance and the Cdc13-Est1 interaction is central to recruit telomerase to the very end of the chromosomes. Also, a Cdc13-Est1 fusion introduced into yeast resulted in substantial telomere elongation, suggesting the recruitment function of Cdc13 can be enhanced by fusing it to a telomerase component [5]. Furthermore, a "charge swap" mutant of Cdc13, *cdc13-2* ($Cdc13^{E252K}$), a mutation within the RD, confers a telomerase-null phenotype on its own [2, 16] but is suppressed by a charge-swap allele of Est1, *est1-60* ($Est1^{K444E}$) [15]. These results suggest that interaction between Cdc13 and Est1 is supported by the electrostatic attraction of a specific Lys-Glu pair [15]. The requirement of the salt-bridge was confirmed with another mutant series (*cdc13-9* $Cdc13^{E252R}$ and *est1-62* ($Est1^{K444D}$) [15]). Additionally, another "activation" model has been proposed by the Zakian group [17]. Contrary to the expectations of a recruitment model, the *cdc13-2* protein can interact with Est1 normally by both *in vitro* and *in vivo* criteria, indicating that the functional interaction between Cdc13 and Est1 lost in a *cdc13-2* strain occurs at a step other than recruitment [13, 17, 18]. If Est1 was expressed in conjunction with the Cdc13-Est2 fusion then the telomeric DNA was hyperelongated, which suggests Est1 upregulated telomerase DNA extension activity. Their findings

suggest a model in which Est1 binds telomere late in S phase and interacts with Cdc13 to convert inactive, telomere-bound Est2 to an active form [17].

Thus, I aim to understand the Est1 regulatory mechanism by investigating: 1) the structure of Est1p conserved core; 2) the interaction between of Est1 and Cdc13; and 3) combined telomerase regulatory function of Est1 and Cdc13. Here I present some preliminary *in vitro* result on the first two goals.

5.2 Identification of the Conserved Core of Est1

To initiate a comparative analysis of Est1 and to uncover possible structural domains in this protein, I systemically searched the NCBI and Broad Institute databases for homologs of Est1 using available sequences as queries. This resulted in the identification of many Est1 homologs in the *Saccharomyces* and *Kluyveromyces* branches of budding yeast (which also include *Candida* spp.; Figure 5.1). Multiple sequence alignment of these Est1 proteins clearly revealed a conserved N terminal region, containing about 470-600 residues in different species (selective examples are shown in Figure 5.1).

I next performed a secondary structural analysis on *Kluyveromyces lactis* Est1 using the program PredictProtein [19]. Supporting the validity of this approach, the program predicted a single N terminal domain composed of predominately α helices (data not shown). Sequence analyses of several Est1 proteins from other yeast species also predicted the existence of this α-helices rich domain. Therefore, my following efforts have focused on characterizing this conserved core and establishing the interaction between Est1 and Cdc13.

5.3 Purification and Characterization of Recombinant $Cdc13_{RD}$ and Est1

Yeast telomerase subunit Est1 has been widely studied *in vivo* and most of our current knowledge about its interaction with Cdc13 came from yeast two-hybrid assays and Co-IP. Based on the results of yeast two-hybrid assays, the interaction between Est1 and Cdc13 is very weak. Many groups reported difficulties in obtaining positive result and can only observe the interaction by over-expressing Est1 in Co-IP experiments [12, 18, 20]. The difficulties lie in the fact that it has been nearly impossible to obtain large quantity of full-length Est1 in any species to conduct *in vitro* biochemical assays. This is the first problem I need to solve before I can continue with my structural studies on Est1 and Cdc13.

Due to the fact that full-length Est1 proteins are hard to come by and the region outside of the conserved N-terminus seems largely unstructured and varied, I have focused my efforts on trying to express the N terminal core only, in both *E. coli* and baculovirus protein expression systems. Est1 from multiple origins, including *Saccharomyces, Candida* spp. and *Kluyveromyces* branches of budding yeast, has been investigated (Figure 5.2 A and B, and data not shown). Notably, out of the five yeast species, the only construct that can yield a good amount of soluble protein is *K. lactis* Est1 (residues 2-600), hereafter referred to as *Kl*Est1N. Recombinant *Kl*Est1N has been subjected to crystallization screening but hasn't generated any crystals yet. Notably, repeated freezing and thaw would break the domain into two parts. To further improve the construct, the recombinant protein has been subjected to limited proteolysis analysis (Figure 5.2C).

I also expressed and purified *Sc*Cdc13$_{RD}$ in *E. coli* in the hope of crystallizing the domain as well. Despite the high purity and quality of the fragment, it failed to yield any crystals in the screening. In order to investigate whether this domain contains any structural elements, it was subjected to different proteases. Limited proteolysis (Figure 5.3 A, B and C) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Figure 5.3D) identified a protease-resistant core domain of *Sc*Cdc13$_{RD}$ containing residues 233-289. My collaborator, Dr. Hongyu Hu, in Chinese Academy of Science, Shanghai, solved the solution structure of RD by Nuclear magnetic resonance (NMR). The result revealed an α helix core.

5.4 Cdc13 and Est1 Interact Directly to Form a 1:1 Complex *in vitro*

In vivo studies suggest that Cdc13 and Est1 interact and this interaction is important for the recruitment of telomerase to telomeres [5, 15, 18, 21]. The interaction between Cdc13 and Est1 appears to be either a very weak or transient one as it has been unsuccessful for multiple groups to use exogenously expressed subunits to study the interaction. Previously, the interaction has been studied predominantly using the yeast two-hybrid system with full-length Cdc13p and Est1p that allows others to observe this interaction. For example, two mutations, *cdc13-2* and *est1-60*, have been shown to abolish this interaction [4]. To determine whether this interaction is direct, I tested purified untagged Est1N and N-terminally GST-tagged Cdc13 fragments for their ability to interact *in vitro*, using glutathione sepharose beads pulldown experiments (Figure 5.4A). The result clearly indicated that Cdc13 and Est1 interact, by biochemical standard, in about 1:1 ratio. Sequence alignment and secondary sequence analysis of Cdc13 revealed an

unstructured and variable linker region between the RD and OB2 domain (Figure 5.1 and Figure 2.1). This linker was removed and its binding with *Kl*Est1N was also tested following the same method. Noticeably, the interaction was retained (Figure 5.4B).

5.5 Cdc13 RD and OB2 are Both Necessary to Maintain Interaction with Est1

The N terminus of Cdc13 contains two additional OB folds in addition to the genetically defined RD [22]. The first OB fold (OB1) is required for Cdc13 dimerization (41, 42) as well as Pol1 (the catalytic subunit of DNA polymerase α) interaction [18, 22]. Previously, only $Cdc13_{RD}$ has been indicated to be important for this pair of interactions [5, 15, 18, 21]. I therefore attempted to apply the method I developed above to further map the region on Cdc13 that is responsible for interacting with Est1. A series of constructs of GST-*Kl*Cdc13N containing different domains in the presence or absence of *Kl*Est1N (Figure 5.5A, lane 7) were mixed with glutathione Sepharose beads that capture the N-terminal affinity tag of Cdc13. *Kl*Est1N was not pulled down by the mix of GST protein and beads (Fig 5.5A, lane 6). Surprisingly, GST-$Cdc13_{RD}$ seems to interact with Est1N very weakly, if they interact as all. However, in the presence of both $Cdc13_{RD}$ and $Cdc13_{OB2}$, *Kl*Est1N was bead-associated, and the amount of bead-associated *Kl*Est1N with both $Cdc13_{RD}$ and $Cdc13_{OB2}$ is greater than when only $Cdc13_{RD}$ is present (Figure 5.5A, lanes 1–5). This, in part, explains the weak positive result in the yeast two-hybrid assays performed using *Sc*$Cdc13_{RD}$ and *Sc*Est1N by other research groups. Using this method, I was able to define the minimal *Kl*Est1N binding sequence of Cdc13 as *Kl*$Cdc13_{213\text{-}507\Delta300\text{-}374}$. To further validate my findings, I used yeast two-hybrid to test the same pairs of interactions (Figure 5.5C). Consistent with previous discovery, inclusion of

$Cdc13_{OB2}$ greatly enhanced the positive readout of the interaction, thus providing a plausible explanation for the difficulties of previous experiments. There appears to be additional interaction between $Cdc13_{OB2}$ and Est1, besides the recruitment domain and Est1, providing either greater stability for the interaction or another site of action.

This represents a big step towards understanding the interaction between Cdc13 and Est1. To further investigate the molecular mechanism of this interaction, it would require the use of x-ray crystallography under the condition that a large amount of high quality *Kl*Est1N-$Cdc13_{RD+OB2}$ complex could be obtained. Aforementioned coexpression strategy has been used to achieve this goal (Figure 5.6A-C). After a series of affinity and size-exclusion chromatography, a stable complex with high homogeneity was achieved. Initial crystallization screening hasn't yielded any crystals so far.

5.6 Discussion

Previously, it has been shown genetically that the Cdc13-Est1 interaction is critical to recruit telomerase to telomeres *in vivo* [5, 15, 21]. The weak interaction was identified by yeast two-hybrid and *in vivo* Co-IP experiments [18, 23, 24]. What is missing is the quantitative information about the strength of the interaction. Here I used the purified N-terminal core domain of *Kl*Est1 and different subdomains of *Kl*Cdc13 to provide unique *in vitro* evidence that Est1 and Cdc13 do interact directly and form a 1:1 complex (Figures 5.5A and 5.6A). This interaction, which is essential for Est1 recruitment to telomeric ssDNA *in vitro*, mimics the *in vivo* role of both proteins as comediators for telomerase recruitment [5]. Unexpectedly, this interaction involves both the genetically defined RD and the second OB fold of Cdc13. At the same time this project was being

carried out, Dr. Zakian's laboratory determined that the apparent K_d for the Cdc13-Est1 interaction was ~250nM [13], which falls within the range of other transient interaction between yeast nuclear proteins, for example the replication machinery components PCNA and Polη (~100nM) [25], but is stronger than Cdc13 and Pol1 interaction (~3.8μM) [22]. There is significant discrepancy with my results as they claimed the RD is solely responsible for the interaction.

Combined with my discoveries in Chapters 3 and 4, the findings provided new insight into the replication of telomeres. Self-associated Cdc13 can interact with DNA polymerase and telomerase using different subdomains, probably at the same time. As telomerase and DNA polymerase each synthesizes leading- and lagging- strands of the chromosome, Cdc13's coordination of DNA synthesis is critical. The fundamental mechanism for telomere replication seems to be highly evolutionarily conserved. Cdc13, TEBPα and Pot1 are structural homologs responsible for G-rich ssDNA binding in *S. cerevisiae*, *S. lemnae* and *Homo sapiens*, respectively [26]. They recruit their partners, Est1, TEBPβ and TPP1, respectively, which further recruit telomerase to the location where it functions [15, 27-30]. Insights drawn from Cdc13's action in *Saccharomyces cerevisiae* would be instrumental in understanding the synthesis of telomeres in humans.

5.7 Materials and Methods

Sequence analysis

Est1 homologues from *Candida, Kluyveromyces* and *Saccharomyces* spp. were identified from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Broad Institute (http://www.broad.mit.edu/annotation/genome/candida_group/Blast.html) databases by

BLAST or psi-BLAST searches. The multiple sequence alignment was generated using the PROMALS server (http://prodata.swmed.edu/promals/promals.php) and displayed using Boxshade (http://www.ch.embnet.org/software/BOX_form.html).

Coexpression and GST pulldown assays

The numerous N-terminal domains of *K. lactis* Est1 (residues 2-560, 2-580, 2-600, 2-614, 2-635, 2-660, 2-685), *S. cerevisiae* Est1 (residues 2–590) and full-length Est1 from *C. albicans* and *C. tropicalis* were cloned into a GST fusion protein expression vector, pGEX6p-1 (GE healthcare). *S. cerevisiae* Cdc13 (residues 233-289) and *K. lactis* Cdc13 (residues 2-357, 213-357, 213-507Δ300-374, 213-305, and 357-507) were cloned into a modified pET28b vector with a sumo protein fused at the N-terminus after the His_6 tag [29].

The *K. lactis* Est1N-Cdc13 complexes was coexpressed in *E. coli* BL21(DE3). After induction for 16 hours with 0.1 mM IPTG at 25°C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 6 hours at 4˚C before elution with 250 mM imidazole. Then Ulp1 protease was added to remove the His_6-Sumo tag. The complex was then mixed with glutathione Sepharose beads (GE Healthcare) and rocked for 8 hours at 4˚C before elution with 15 mM glutathione. Protease 3C was added to remove the GST-tag. Finally, the Est1N-

Cdc13 complex was further purified by passage through Mono-Q ion-exchange column and by gel-filtration chromatography on Hiload Superdex200 equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM dithiothreitol (DTT). The purified Est1N-Cdc13 complex was concentrated to 30 mg/ml and stored at -80˚C.

For pull-down assays, different GST-Cdc13 fragments (10 μg), Est1N (10 μg), or GST (5 μg) were used. The indicated proteins were incubated in 30 μl of buffer 25 mM Tris-HCl, pH 7.5, 0.005% Triton, 150 mM NaCl, and 1 mM DTT) for 30 min at 4°C. The reactions were mixed with 10 μl of glutathione sepharose beads (which recognize the GST-tag at the N-terminus of GST-Cdc13 and GST) at 4°C for 30 min. After washing the beads twice with 200 μl of the same buffer, 20 μl SDS loading dye were added to each sample and analyzed by 15% SDS-PAGE and Coomassie Blue staining.

Expression and purification of *K. lactis* Est1N and GSTCdc13 fragments

The numerous N-terminal domains of *K. lactis* Est1 (residues 2-560, 2-580, 2-600, 2-614, 2-635, 2-660, 2-685), *S. cerevisiae* Est1 (residues 2–590), full-length Est1 from *C. albicans* and *C. tropicalis* and *S. cerevisiae* Cdc13 (residues 233-289) were cloned into a modified pET28b vector with a Sumo protein fused at the N-terminus after the His_6 tag [29]. *K. lactis* Cdc13 (residues 2-357, 213-357, 213-507Δ300-374, 213-305, and 357-507) were cloned into a GST fusion protein expression vector, pGEX6p-1 (GE healthcare).

The resulting expression plasmids of *Kl*Est1N and *S. cerevisiae* $Cdc13_{RD}$ were transformed into *E. coli* BL21(DE3). After induction for 16 h with 0.1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) at 20°C, the cells were harvested by centrifugation

and the pellets were resuspended in lysis buffer (50mM Tris-HCl, pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride [PMSF], 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 2 h at 4°C before elution with 250 mM imidazole. Then, the Ulp1 protease was added, and the mixture was incubated for 12 h at 4°C to remove the His_6-SUMO tag. *Kl*Est1N was then further purified by passage through a Mono-Q ion-exchange column and by gel filtration chromatography on a Hiload Superdex200 (GE Healthcare) equilibrated with 25 mM Tris-HCl, pH 8.0, 150mM NaCl, and 5mM dithiothreitol (DTT). The purified *Kl*Est1N and *S. cerevisiae* $Cdc13_{RD}$ was concentrated to 30 mg/ml and 10mg/ml, respectively, and stored at -80°C.

The *K. lactis* GSTCdc13 fragments were expressed in *E. coli* BL21(DE3). After induction for 16 hours with 0.1 mM IPTG at 25°C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with glutathione Sepharose beads (GE Healthcare) and rocked for 4 hours at 4˚C before elution with 15 mM glutathione. Finally, the GSTCdc13 fusion protein was further purified by gel-filtration chromatography on Hiload Superdex200 (GE Healthcare) equilibrated with 25 mM Tris-

HCl pH 8.0, 150 mM NaCl and 5 mM dithiothreitol (DTT). The purified GSTCdc13 protein was concentrated to 30 mg/ml and stored at -80˚C.

Limited proteolysis (Subtilisin) of *S. cerevisiae* $Cdc13_{RD}$ and the analysis of *K. lactis* Est1N

The protein of *S. cerevisiae* Cdc13 (residues 233-289) was incubated with 0.2% w/w subtilisin (Roche) at 25 °C in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 5 mM DTT. At various time points, 10 µl aliquots of the reaction mixture were withdrawn, diluted with 10 µl of water and 5 µl of SDS loading dye, and run on 15% SDS-PAGE visualized with Coomassie brilliant blue stain.

The protein of *K. lactis* Est1 (residues 2-600) was incubated with 0.2% w/w subtilisin, trypsin, papain, pepsin, Glu-C and elastase at 25 °C in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 5 mM DTT. After 30 min, 10 µl aliquots of the reaction mixture were withdrawn, diluted with 10 µl of water and 5 µl of 5xSDS loading dye, and run on 15% SDS-PAGE visualized with Coomassie brilliant blue stain.

MALDI mass spectrometry of the limited protease (subtilisin) cleavage products

For MALDI mass spectrometry analysis, *S. cerevisiae* Cdc13 (residues 211-331) was incubated with 0.2 % w/w subtilisin (Roche) at 25 °C in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 5 mM DTT. Aliquots were withdrawn as described above for SDS-PAGE analysis. At the 90 min time point, 2 µl of the reaction mixture was co-crystallized with 2 µl sinapinic acid matrix. The samples were analyzed by MALDI-TOF-MS in linear mode. The major product by MALDI had an MH(+1) of 6109 Da. Examination of the

map of predicted trypsin sites revealed that this fragment corresponds to the predicted fragments: Cdc13 [MH(+1) 6109.7 Da].

Yeast two-hybrid assay

The yeast two-hybrid assays were performed using L40 strain harboring pBTM116 and pACT2 (Clontech) fusion plasmids. Colonies containing both plasmids were selected on –Leu –Trp plates. β-Galactosidase activities were measured by a liquid assay [31].

Figure 5.1 Alignments of Est1 homologues from *Saccharomyces, Candida and Kluyveromyces* spp.

Multiple homologues were identified at the NCBI, SGD and Broad Institute databases using BLAST with default parameters. In all genomes analyzed, a plausible homologue (E < 0.001) of each known subunit of the telomerase holoenzyme can be identified. Multiple sequence alignments were generated using the PROMALS server (http://prodata.swmed.edu/promals/promals.php) and displayed using Boxshade (http://www.ch.embnet.org/software/BOX_form.html). The abbreviations are as follows: *Sc, S. cerevisiae; Klac, K. lactis; Cglab, C. glabrata; Calb, C. albicans.*

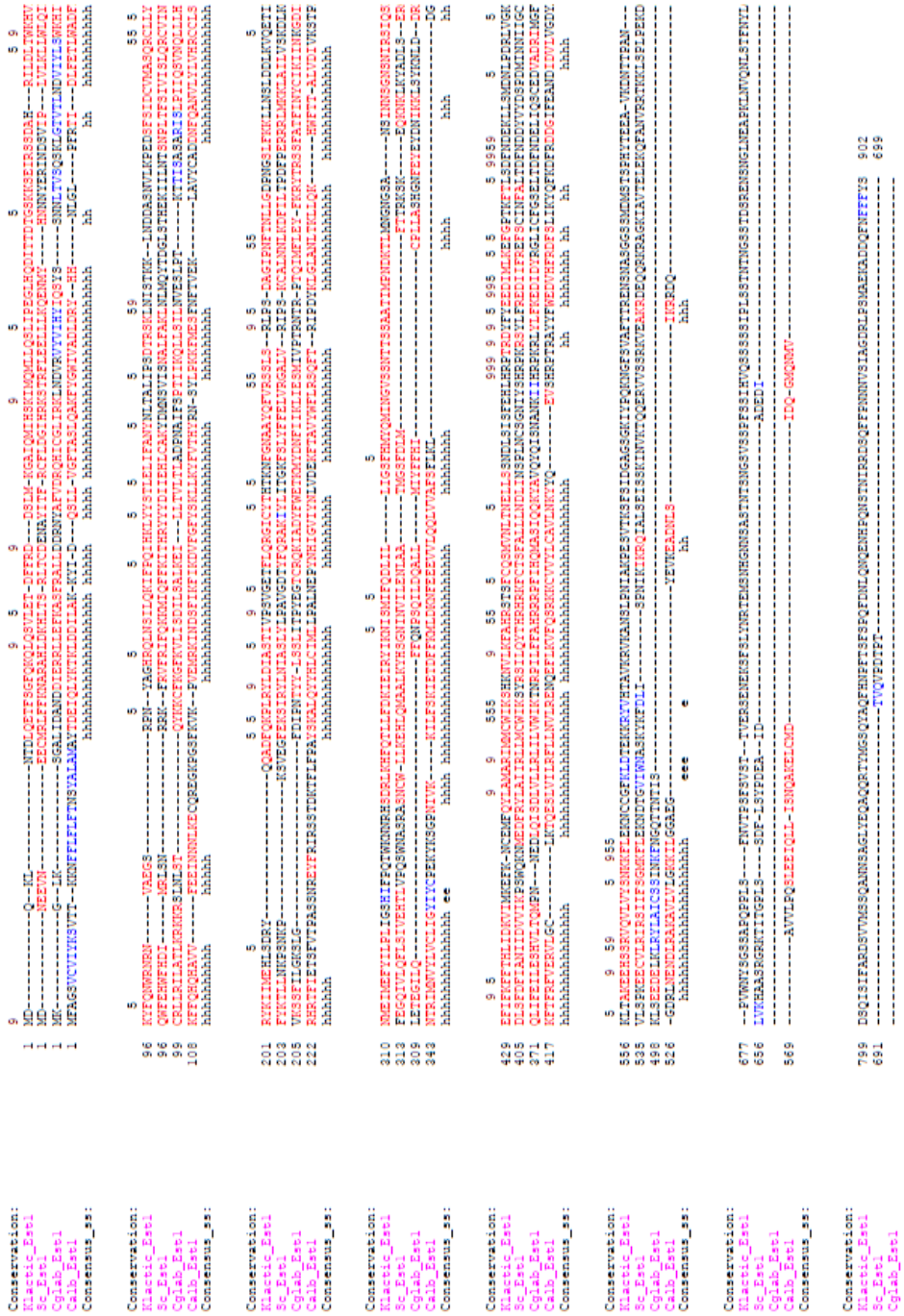


Figure 5.2 Purification and limited proteolysis of *Kl*Est1N
(A) SDS-PAGE of purified protein of *Kl*Est1N for crystallization
(B) Gel filtration chromatography profile (Hiload Superdex 200) of *Kl*Est1N
(C) Limited proteolysis of *Kl*Est1N by six different proteases: lane 1, control; lane 2, 1/10 elastase, lane 3, 1/100 elastase; lane 4, 1/10 Glu-C; lane 5, 1/100 Glu-C; lane 6, 1/10 papain; lane 7, 1/100 papain; lane 8, 1/10 pepsin; lane 9, 1/100 pepsin; lane 10, 1/100 substilisin; lane 11, 1/1000 subtilisin; lane 12, 1/10 trypsin; and lane 13, 1/100 trypsin.

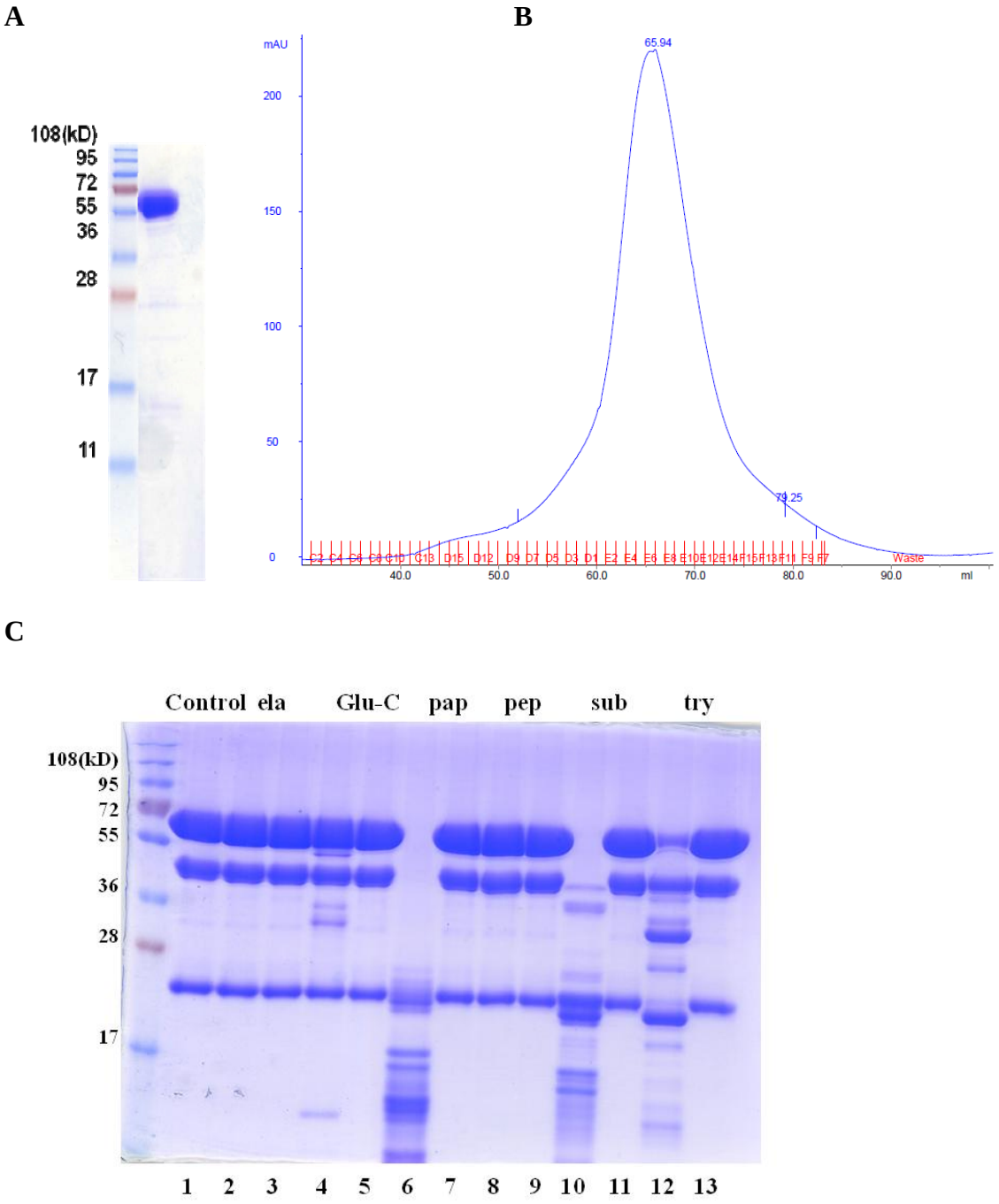
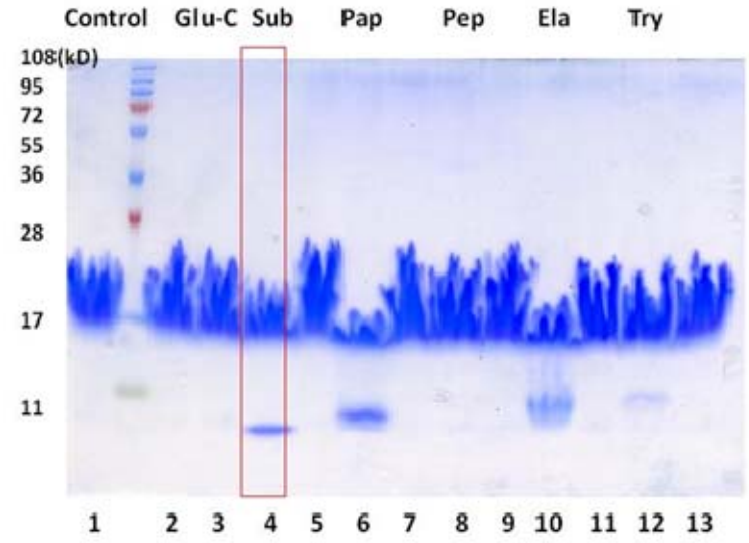
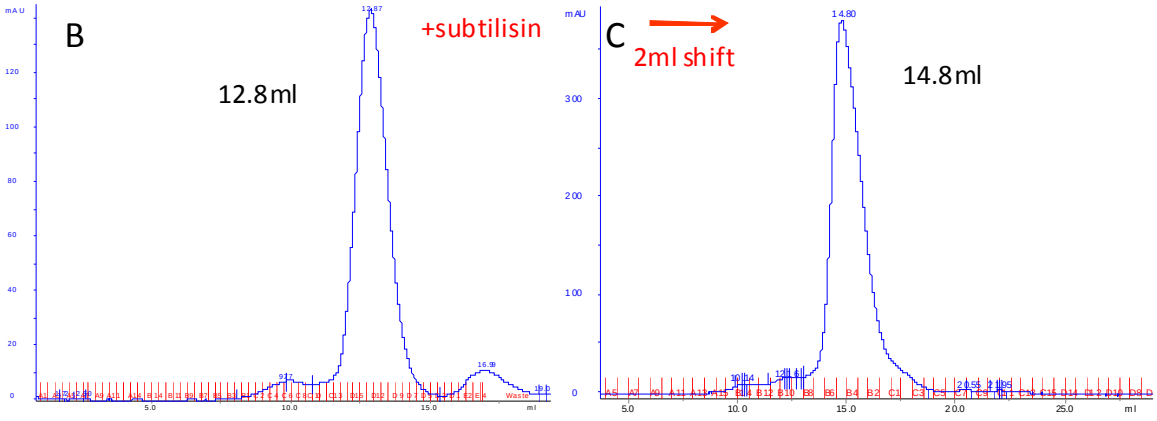


Figure 5.3 Limited proteolysis of *Sc*Cdc13$_{RD}$ and MALDI mass spectrometry result

(A) Limited proteolysis of *Sc*Cdc13$_{RD}$ by six different proteases: lane 1: control; lane 2, 1/10 Glu-C, lane 3, 1/100 Glu-C; lane 4, 1/100 subtilisin; lane 5, 1/1000 subtilisin; lane 6, 1/10 papain; lane 7, 1/100 papain; lane 8, 1/10 pepsin; lane 9, 1/100 pepsin; lane 10, 1/10 elastase; lane 11, 1/100 elastase; lane 12, 1/10 trypsin; and lane 13, 1/100 trypsin.
(B) Gel filtration chromatography profile (Superdex 75) before *Sc*Cdc13$_{RD}$ subjected to subtilisin treatment
(C) Gel filtration chromatography profile (Superdex 75) after *Sc*Cdc13$_{RD}$ subjected to subtilisin treatment
(D) MALDI mass spectrometry result of *Sc*Cdc13$_{211-331}$ digested by subtilisin

A





D

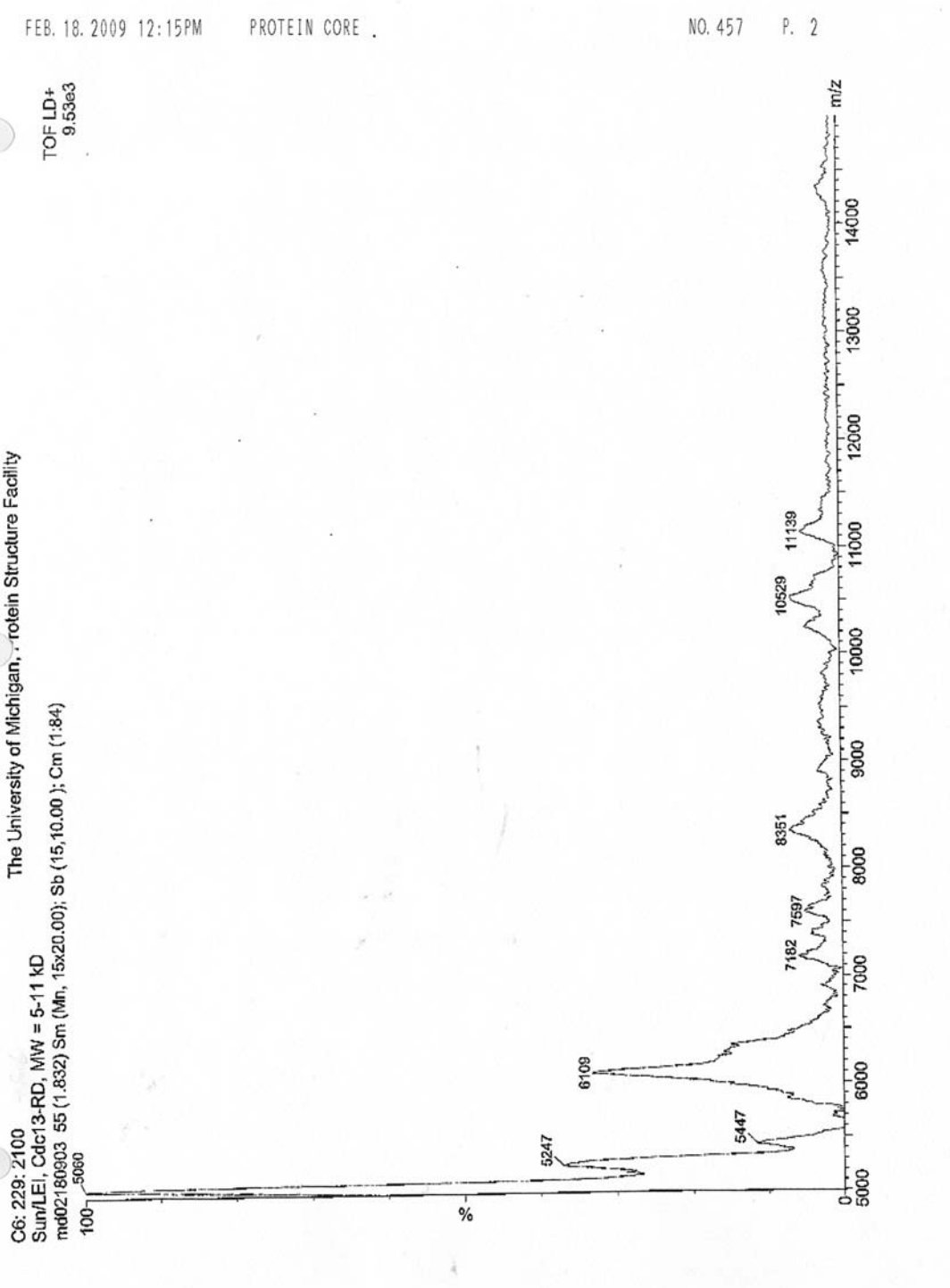
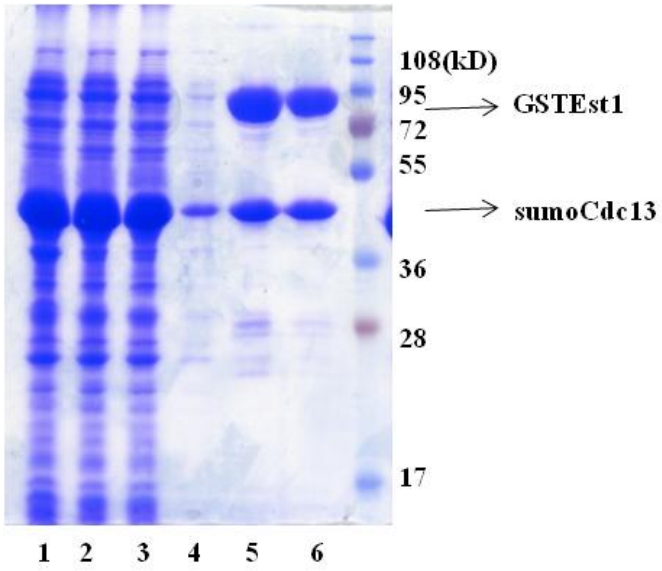

FEB. 18. 2009 12:15PM PROTEIN CORE . NO. 457 P. 2
TOF LD+
9.53e3
The University of Michigan, . rotein Structure Facility
C6: 229: 2100
Sun/LEI, Cdc13-RD, MW = 5-11 kD
md02180903 55 (1.832) Sm (Mn, 15x20.00); Sb (15,10.00); Cm (1:84)
5060
100
%
0
5247
6109
5447
7182
7597
8351
10529
11139
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
m/z

Figure 5.4 Coexpression of *Kl*Est1N and Cdc13 and GST pulldown

(A) Coexpression of *Kl*Est1N and $Cdc13_{213-507}$, lane 1: crude cell lysate, lane 2: supernatant, lane 3: flow-through, lane 4: wash with 10mM imidazole, lane 5: 1st elution of 300mM imidazole, lane 6: 3rd elution of 300mM imidazole
(B) Coexpression of *Kl*Est1N and $Cdc13_{213-507(\Delta 263-374)}$, lane 1: crude cell lysate, lane 2: supernatant, lane 3: flow-through, lane 4: wash with 10mM imidazole, lane 5: 1st elution of 300mM imidazole, lane 6: 3rd elution of 300mM imidazole

A



B

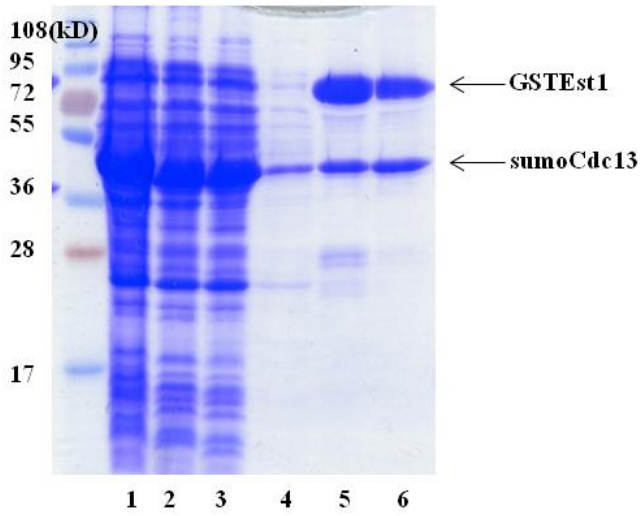
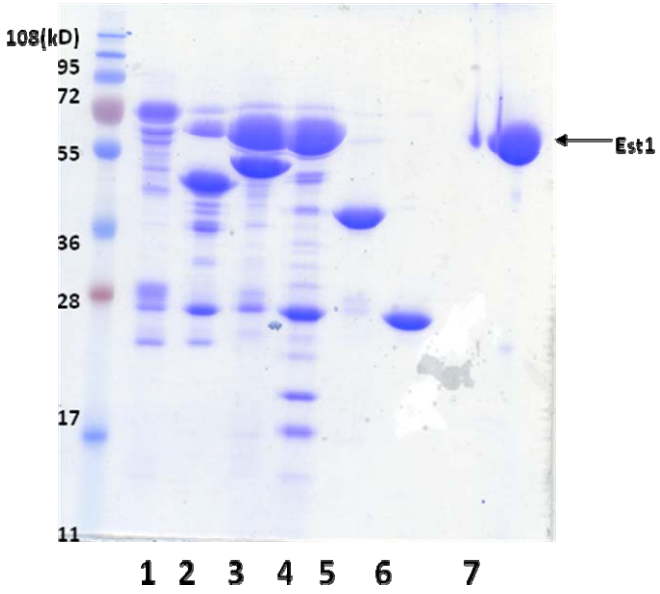


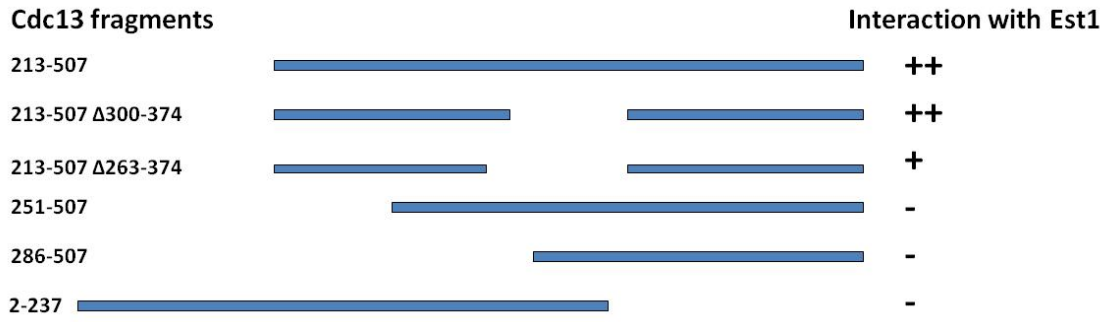
Figure 5.5 GST-*Kl*Cdc13 pulldown of *Kl*Est1N

(A) Bait: Lane 1: GST-*Kl*Cdc13$_{2\text{-}357}$; lane 2: GST-*Kl*Cdc13$_{213\text{-}357}$; lane 3: GST-*Kl*Cdc13 $_{213\text{-}357\Delta300\text{-}374}$; lane 4: GST-*Kl*Cdc13$_{213\text{-}507}$; lane 5: GST-*Kl*Cdc13$_{357\text{-}507}$; lane 6: GST alone; lane 7: *Kl*Est1N input
(B) Schematic map of the *Kl*Est1N and *Kl*Cdc13 interaction
(C) Yeast two-hybrid of *Kl*Est1N and *Kl*Cdc13 interaction

A



B



C

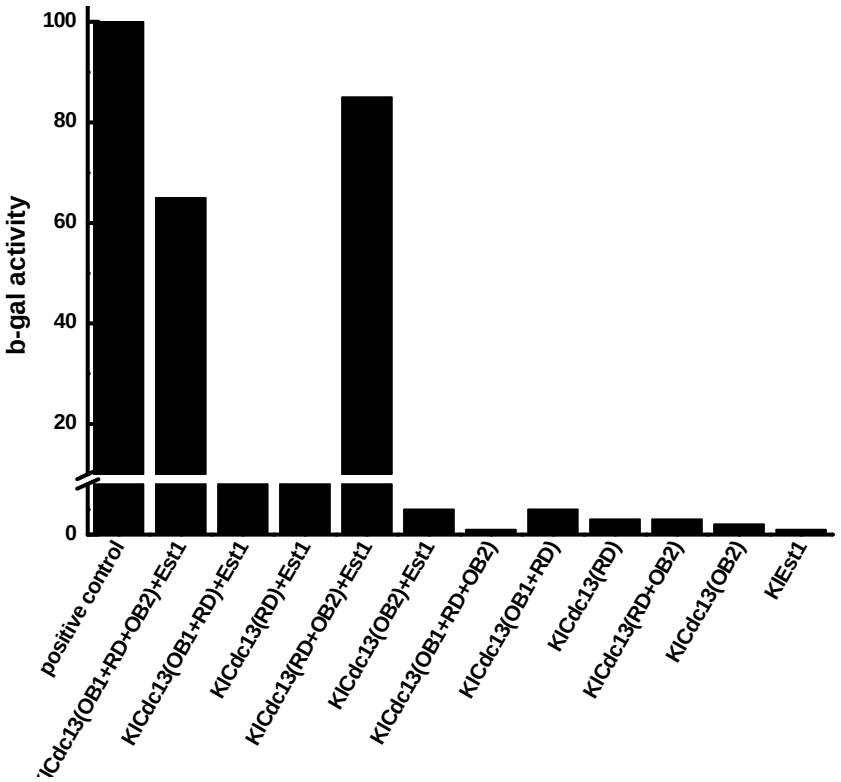
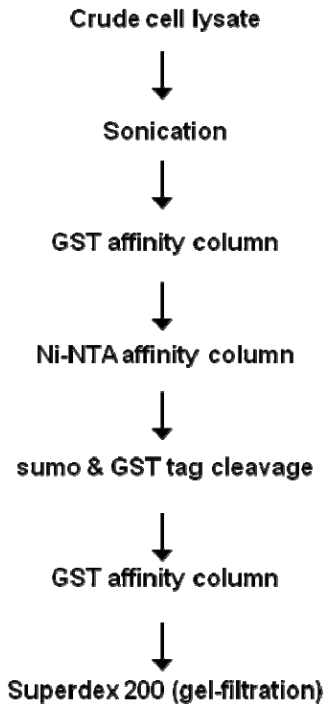

b-gal activity
100
80
60
40
20
0
positive control
KlCdc13(OB1+RD+OB2)+Est1
KlCdc13(OB1+RD)+Est1
KlCdc13(RD)+Est1
KlCdc13(RD+OB2)+Est1
KlCdc13(OB2)+Est1
KlCdc13(OB1+RD+OB2)
KlCdc13(OB1+RD)
KlCdc13(RD)
KlCdc13(RD+OB2)
KlCdc13(OB2)
KlEst1

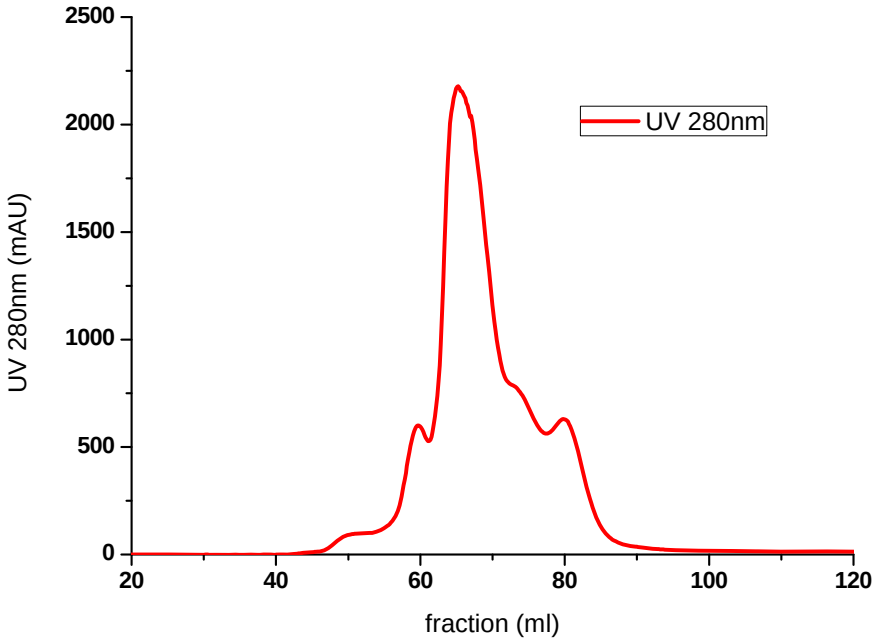
Figure 5.6 Co-purification of *Kl*Est1N and Cdc13

(A) Schematic diagram of purification procedure of *Kl*Est1N and Cdc13 complex
(B) Gel filtration chromatography profile (Hiload Superdex 200) of *Kl*Est1N and Cdc13 complex (C) SDS-PAGE of purified *Kl*Est1N and Cdc13 complex

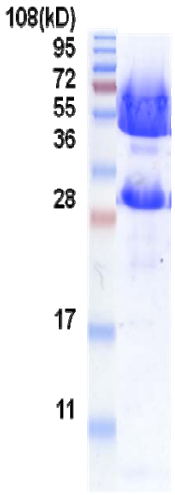
A



B



C



Reference:

1. Singer, M.S. and D.E. Gottschling, *TLC1: template RNA component of Saccharomyces cerevisiae telomerase.* Science, 1994. **266**(5184): p. 404-9.
2. Lendvay, T.S., et al., *Senescence mutants of Saccharomyces cerevisiae with a defect in telomere replication identify three additional EST genes.* Genetics, 1996. **144**(4): p. 1399-412.
3. Lundblad, V. and J.W. Szostak, *A mutant with a defect in telomere elongation leads to senescence in yeast.* Cell, 1989. **57**(4): p. 633-43.
4. Lingner, J., et al., *Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity.* Proc Natl Acad Sci U S A, 1997. **94**(21): p. 11190-5.
5. Evans, S.K. and V. Lundblad, *Est1 and Cdc13 as comediators of telomerase access.* Science, 1999. **286**(5437): p. 117-20.
6. Blackburn, E.H., *Switching and signaling at the telomere.* Cell, 2001. **106**(6): p. 661-73.
7. Seto, A.G., et al., *A bulged stem tethers Est1p to telomerase RNA in budding yeast.* Genes Dev, 2002. **16**(21): p. 2800-12.
8. Webb, C.J. and V.A. Zakian, *Schizosaccharomyces pombe Ccq1 and TER1 bind the 14-3-3-like domain of Est1, which promotes and stabilizes telomerase-telomere association.* Genes Dev, 2012. **26**(1): p. 82-91.
9. Lin, J.J. and V.A. Zakian, *An in vitro assay for Saccharomyces telomerase requires EST1.* Cell, 1995. **81**(7): p. 1127-35.
10. Virta-Pearlman, V., D.K. Morris, and V. Lundblad, *Est1 has the properties of a single-stranded telomere end-binding protein.* Genes Dev, 1996. **10**(24): p. 3094-104.
11. Snow, B.E., et al., *Functional conservation of the telomerase protein Est1p in humans.* Curr Biol, 2003. **13**(8): p. 698-704.
12. DeZwaan, D.C. and B.C. Freeman, *The conserved Est1 protein stimulates telomerase DNA extension activity.* Proc Natl Acad Sci U S A, 2009. **106**(41): p. 17337-42.
13. Wu, Y. and V.A. Zakian, *The telomeric Cdc13 protein interacts directly with the telomerase subunit Est1 to bring it to telomeric DNA ends in vitro.* Proc Natl Acad Sci U S A, 2011.
14. Zhang, M.L., et al., *Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation.* Nat Struct Mol Biol, 2010. **17**(2): p. 202-9.
15. Pennock, E., K. Buckley, and V. Lundblad, *Cdc13 delivers separate complexes to the telomere for end protection and replication.* Cell, 2001. **104**(3): p. 387-96.
16. Nugent, C.I., et al., *Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance.* Science, 1996. **274**(5285): p. 249-52.
17. Taggart, A.K., S.C. Teng, and V.A. Zakian, *Est1p as a cell cycle-regulated activator of telomere-bound telomerase.* Science, 2002. **297**(5583): p. 1023-6.
18. Qi, H. and V.A. Zakian, *The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein.* Genes Dev, 2000. **14**(14): p. 1777-88.

19. Rost, B., G. Yachdav, and J. Liu, *The PredictProtein server.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W321-6.
20. Chan, A., J.B. Boule, and V.A. Zakian, *Two pathways recruit telomerase to Saccharomyces cerevisiae telomeres.* PLoS Genet, 2008. **4**(10): p. e1000236.
21. Bianchi, A., S. Negrini, and D. Shore, *Delivery of yeast telomerase to a DNA break depends on the recruitment functions of Cdc13 and Est1.* Mol Cell, 2004. **16**(1): p. 139-46.
22. Sun, J., et al., *Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase alpha.* Cell Res, 2010. **21**(2): p. 258-74.
23. Li, S., et al., *Cdk1-dependent phosphorylation of Cdc13 coordinates telomere elongation during cell-cycle progression.* Cell, 2009. **136**(1): p. 50-61.
24. Gao, H., et al., *Telomerase recruitment in Saccharomyces cerevisiae is not dependent on Tel1-mediated phosphorylation of Cdc13.* Genetics, 2010. **186**(4): p. 1147-59.
25. Zhuang, Z., et al., *Regulation of polymerase exchange between Poleta and Poldelta by monoubiquitination of PCNA and the movement of DNA polymerase holoenzyme.* Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5361-6.
26. Luke, B., et al., *The Rat1p 5′ to 3′ exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae.* Mol Cell, 2008. **32**(4): p. 465-77.
27. Paeschke, K., et al., *Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates.* Nat Struct Mol Biol, 2008. **15**(6): p. 598-604.
28. Xin, H., et al., *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase.* Nature, 2007. **445**(7127): p. 559-62.
29. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
30. Fang, G. and T.R. Cech, *Oxytricha telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits.* Proc Natl Acad Sci U S A, 1993. **90**(13): p. 6056-60.
31. Moretti, P., et al., *Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1.* Genes Dev, 1994. **8**(19): p. 2257-69.

CHAPTER 6

CONCLUSIONS AND PERSPECTIVES

Telomere is the specialized protein-DNA complex localizing at the end of the linear chromosomes. Telomeres are essential for genomic stability and long-term cellular proliferation. The two major roles that telomeres play are: protection and replication of chromosomal ends [1]. The protecting function prevents chromosome ends from being recognized as the DNA breaks so that they won't undergo inappropriate DNA repair pathways (reviewed in [1-3]). Telomeres are synthesized and extended by a special reverse transcriptase named telomerase. TERT (telomerase reverse transcriptase) and TR (telomerase RNA component) are the major components of telomerase. Using TR as the internal template, TERT adds the repetitive telomeric sequence (telomerase repeats) to the end of chromosomes in a processive manner [2]. Besides telomeric DNA, telomeric proteins, the permanent residents at the telomere region also play important roles in regulating telomerase activity and protecting the telomere (reviewed in [1, 3, 4]). The focus of my research, the CST (Cdc13-Stn1-Ten1) complex, recently sparked a lot of interests in the telomere community because of the discovery of its homologues in a wide range of organisms, mammals included [5-9]. During my pursuit of a doctoral degree in chemical biology, I have been focusing on solving the following problems: (1) characterizing the structure and function of CST complex; (2) elucidating the difference

and similarity between CST and RPA; and (3) probing the interaction between CST and other telomere-associated proteins. In this chapter, I will summarize the findings with regard to the above questions and propose new directions going forward.

6.1 Stn1-Ten1 is an Rpa32-Rpa14-like Complex at Telomere while there is Substantial Structural Differences between Cdc13 and Rpa70

An emerging theme in chromosome biology has been the discovery of protein complexes that resemble striking structural similarities to complexes required for canonical semiconservative DNA replication. It has been proposed that Cdc13, Stn1 and Ten1 proteins form an RPA-like complex that is specifically dedicated to binding chromosome termini [6]. This proposal is largely based on the examination of their DNA binding abilities and bioinformatic prediction of the structure of CST subunits [6]. By solving the crystal structure of Stn1-Ten1 complex structure from both budding and fission yeast, I showed that they share the same three-dimensional architecture as the Rpa32-Rpa14 complex despite minimal sequence similarity, thus providing the first direct confirmation of structural similarity between components of the CST and the RPA complexes. The reliability of my structures was further corroborated by mutational analyses of Stn1 and Ten1, which underscored the importance of functional heterodimerization between Stn1 and Ten1 for telomere localization of Ten1 and telomere length regulation. Besides Stn1 and Ten1, Cdc13 has indeed been shown to be composed of multiple OB folds as well. However, my results provided arguments against a close evolutionary kinship between Cdc13 and Rpa70. Coupled with previous crystallographic and NMR analyses, we now

have high-resolution structures of three domains in Cdc13, each of which proved to be quite different from its putative Rpa70 counterpart. With this caveat in mind, my findings still provide a foundation for leveraging insights from the analysis of RPA to study of the CST complex.

Additionally, the proposal that Cdc13, Stn1 and Ten1 form a telomere-dedicated RPA-like complex also leads to a stoichiometry comparable to that of the canonical RPA complex [6]. However, recent advancement indicated that the stoichiometry among Cdc13, Stn1 and Ten1 is probably 1:3:1, different from that of RPA (personal communications with Dr. Neal Lue).

6.2 The Versatility of OB Fold Domains in Mediating Protein-Protein Interaction

Even though the OB fold domain was initially defined as an oligonucleotide/oligosaccharide-binding module, it is tempting to speculate that the repeated utilization of OB fold domains in proteins associated with single-stranded telomeres may be due not only to its nucleic acid binding activity but also to its versatility in binding protein partners. More-recent studies have highlighted the remarkable functional diversity of this protein fold and the myriad ways in which this fold can mediate protein-protein interactions [10-12]. In keeping with this theme, my high-resolution structures of the *Sc*Cdc13$_{OB1}$ dimer and the *Cg*Cdc13$_{OB4}$ dimer revealed dramatically distinct modes of dimerization. In the case of OB1, the two protomers are arranged end to end, and the symmetry dyad is perpendicular to the axis of the β-barrel

(see Chapter 3 and Chapter 4 for details). While *Sc*Cdc13$_{OB1}$ dimerization is involved in interaction with DNA polymerase catalytic subunit Pol1, the centrally located recruitment domain (RD) and the putative second OB fold (*Sc*Cdc13$_{OB2}$) have been found to mediate the interaction with Est1. This represents one big step forward towards understanding the regulation of telomerase activity because this interaction, which is essential for Est1 recruitment to telomeric ssDNA *in vitro*, mimics the *in vivo* role of both as comediators for telomerase recruitment [13]. All in all, by utilizing different subdomains, Cdc13 functions as a large platform that harbors different functionalities, such as high affinity and specificity telomere binding, DNA polymerase α and telomerase recruitment.

6.3 The Evolution of CST

Budding yeast was believed to have evolved a very different set of telomeric proteins to protect and maintain chromosome ends. Hence, the budding yeast CST complex has been considered to serve as the functional equivalent of the POT1–TPP1 complex in fission yeast and other POT1-containing organisms. However, putative homologs of the CST proteins have been identified recently in both plants and humans [7, 8, 14], suggesting that this telomere regulatory complex is probably more widespread in nature than previously believed, even in organisms that use POT1 for telomere protection. On the other hand, the almost complete lack of sequence similarity between the CST components from budding yeast and POT1-containing organisms raised serious doubts concerning the structural and functional conservation of these proteins in these two groups of organisms. These doubts are now substantially alleviated by my structural data

showing that the budding and fission yeast Stn1–Ten1 complexes share similar three-dimensional structures. Nevertheless, it would be premature to extrapolate from the current findings to other features of the CST complexes. In particular, whether the remaining components of the CST complexes in different organisms (i.e., Cdc13 in yeast and Ctc1 in plants and humans) [7, 8] resemble one another is largely unresolved. Clarifying these and other key issues in CST structure, assembly, and mechanisms will require detailed structural and functional analyses of the entire complex. Again, insights from my structural studies are expected to provide a platform for functional studies of the CST complexes in a wide range of organisms, including humans.

A possible evolutionary scenario has been proposed (Figure 6.1) [15]: POT1 might have been lost from telomeres as a result of mutations in the telomere repeat sequence during budding yeast evolution. Its dissociation might have resulted in TPP1's dissociation from telomeres as well [16]. Significant selection pressures would force the yeast mutant to develop alternative mechanisms of telomere protection and telomerase stimulation. The function of telomere protection was apparently assumed by the Cdc13–Stn1–Ten1 complex. On the other hand, the telomerase-recruitment/activation function was complemented by new interactions between Cdc13 and Est1, as well as Est3 and Est1, which stabilized the association between Est3 and TERT [17, 18].

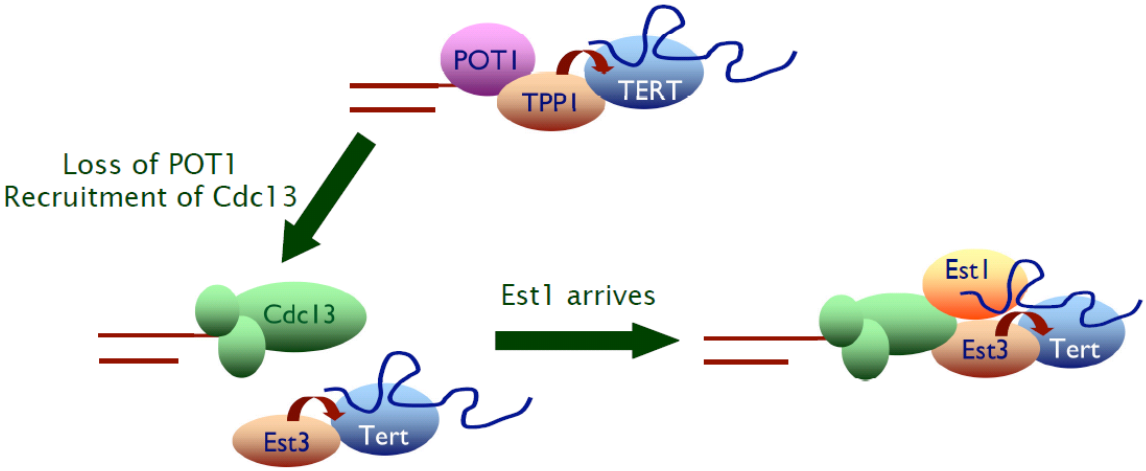
6.4 Future Directions

Moving forward, there are still a number of questions waiting to be answered. To name a few: (1) what is the architecture of the CST ternary complex? Does it indeed follow a one

Cdc13: three Stn1: one Ten1 stoichiomitry? (2) What's the mechanism of telomerase recruitment to telomeres? It would almost solely rely on the determination of Cdc13 and Est1 complex structure to answer this question. (3) How does Cdc13 coordinate the action of DNA polymerase and telomerase? Answering these questions would surely provide valuable insights on telomere function and regulation. In the long term, combining all the information from structural studies, genetics, cell biology and etc, we aim to paint a holistic picture showing all telomere-associated proteins, thus further our understanding of the mysterious chromosomal ends.

Figure 6.1 A possible evolutionary scenario for yeast telomere binding proteins and telomerase (adopted from [15])

During budding yeast evolution, TPP1 may be lost from the telomeres and its function in telomere protection and telomerase activation taken up by the CST complex and telomerase-bound Est3, respectively.



References

1. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
2. Cech, T.R., *Beginning to understand the end of the chromosome.* Cell, 2004. **116**(2): p. 273-9.
3. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. **19**(18): p. 2100-10.
4. Smogorzewska, A. and T. de Lange, *Regulation of telomerase by telomeric proteins.* Annu Rev Biochem, 2004. **73**: p. 177-208.
5. Martin, V., et al., *Protection of telomeres by a conserved Stn1-Ten1 complex.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 14038-43.
6. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.
7. Miyake, Y., et al., *RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway.* Mol Cell, 2009. **36**(2): p. 193-206.
8. Surovtseva, Y.V., et al., *Conserved telomere maintenance component 1 interacts with STN1 and maintains chromosome ends in higher eukaryotes.* Mol Cell, 2009. **36**(2): p. 207-18.
9. Wellinger, R.J., *The CST complex and telomere maintenance: the exception becomes the rule.* Mol Cell, 2009. **36**(2): p. 168-9.
10. Bochkarev, A. and E. Bochkareva, *From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold.* Curr Opin Struct Biol, 2004. **14**(1): p. 36-42.
11. Sun, J., et al., *Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase alpha.* Cell Res, 2010. **21**(2): p. 258-74.
12. Wang, F., et al., *Crystal structures of RMI1 and RMI2, two OB-fold regulatory subunits of the BLM complex.* Structure, 2010. **18**(9): p. 1159-70.
13. Evans, S.K. and V. Lundblad, *Est1 and Cdc13 as comediators of telomerase access.* Science, 1999. **286**(5437): p. 117-20.
14. Casteel, D.E., et al., *A DNA polymerase-{alpha}{middle dot}primase cofactor with homology to replication protein A-32 regulates DNA replication in mammalian cells.* J Biol Chem, 2009. **284**(9): p. 5807-18.
15. Yu, E.Y., et al., *A proposed OB-fold with a protein-interaction surface in Candida albicans telomerase protein Est3.* Nat Struct Mol Biol, 2008. **15**(9): p. 985-9.
16. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-83.
17. Hsu, M., et al., *Mutual dependence of Candida albicans Est1p and Est3p in telomerase assembly and activation.* Eukaryot Cell, 2007. **6**(8): p. 1330-8.
18. Osterhage, J.L., J.M. Talley, and K.L. Friedman, *Proteasome-dependent degradation of Est1p regulates the cell cycle-restricted assembly of telomerase in Saccharomyces cerevisiae.* Nat Struct Mol Biol, 2006. **13**(8): p. 720-8.