The forkhead-associated (FHA) domain of Fkh1 as a key organizer of chromosome biology in *Saccharomyces cerevisiae*

by

Allison J. Hollatz

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 12/11/24

The dissertation is approved by the following members of the Final Oral Committee:

Catherine A. Fox, Professor, Biomolecular Chemistry
Joshua J. Coon, Professor, Biomolecular Chemistry
John M. Denu, Professor, Biomolecular Chemistry
Jason M. Peters, Assistant Professor, Pharmaceutical Sciences
Michael D. Sheets, Professor, Biomolecular Chemistry

Abstract

The yeast *Saccharomyces cerevisiae* (a.k.a. yeast) is a single-celled model organism used for examining the mechanisms underlying the duplication and stability of eukaryotic chromosomes. This thesis focused on the role that a sequence-specific DNA binding protein, forkhead homolog 1 (Fkh1), plays in these mechanisms. Fkh1 was initially defined as a gene-specific transcription factor important for the normal yeast cell cycle, but when I started my research, it was becoming clear that Fkh1 also functioned directly in 1) promoting the activity of a subset of DNA replication origins and 2) in a locus-specific form of homologous recombination (HR). Fkh1 is named for its forkhead DNA binding domain (DBD) that defines a conserved family of DNA binding proteins known as FOX (forkhead box) proteins. Although many important questions remain about the specific Fkh1-DNA interactions required for Fkh1's various roles, my research has focused on Fkh1's interaction with partner proteins. In particular, Fkh1 contains a protein-protein interaction domain named the forkhead associated (FHA) domain. FHA domains are enriched in proteins that function in DNA metabolism and are distinct for their ability to recognize and bind phosphothreonine- (pT-) containing peptide regions within their partner proteins. The pT-recognition ability of the Fkh1 FHA domain is critical to Fkh1's roles at DNA replication origins and in HR. Therefore, the goal of my thesis research was to identify and characterize Fkh1-FHA binding partners required for Fkh1's roles in promoting DNA replication origin function and HR. Chapter 1 includes an overview of the yeast forkhead proteins and their role in cell cycle-linked transcription, as well as summaries of Fkh1's roles in DNA replication control and HR. Chapter 1 concludes with a brief discussion of protein-protein interaction assays that have or could be applied to the Fkh1-FHA domain. In Chapter 2, I addressed whether a distinct pT-containing peptide region within the essential S-phase kinase

Cdc7 could explain Fkh1's role at DNA replication origins. *In vivo* and *in vitro* FHA-protein-protein interaction assays, as well as *in silico* and genetic experiments, supported the conclusion that this region of Cdc7 was more likely a relevant partner of a distinct FHA domain belonging to the checkpoint kinase Rad53, than of the FHA domain in Fkh1. This work also revealed an unanticipated ability of these two similar FHA domains to distinguish between two different but highly similar peptides in a manner relevant to their biological functions. In Chapter 3, I address another FHA-dependent Fkh1-partner interaction that occurs between Fkh1 and the non-conserved protein Fdo1 (Forkhead 1 interacting protein involved in donor preference). The experiments supported a model wherein this interaction is important for promoting HR. Analyses of the Fdo1 sequence and the predicted AlphaFold structural model as well as analyses of mutant proteins led me to propose that Fdo1 might interact with Fkh1 through a non-classic protein-protein interaction mechanism, involving liquid-liquid phase separation (LLPS) or prion-related conformational states. Limited sequence analyses in *Saccharomycotina* species, literature searches, and direct experiments led me to speculate that the Fkh1-Fdo1 interaction contributes to the high efficiency of HR that has helped make *S. cerevisiae* an important model eukaryote. In Chapter 4, I discuss modifications of some of the protein-protein interaction methods I used that could be helpful in defining Fkh1-FHA partners using non-targeted discovery-based proteomics.

Acknowledgements

I am grateful to everyone who supported me throughout my graduate journey, in both personal and professional capacities. Here, I will acknowledge just a few people for their professional contributions.

First, I would like to thank my advisor, Dr. Catherine Fox, for her dedication toward helping me grow as a scientist.

Thanks to past and present members of the Fox lab, especially Dr. Tim Hoggard, who has been an immense supporter and collaborator.

I would also like to thank my thesis committee – Drs. Joshua Coon, John Denu, Jason Peters, and Mike Sheets – for their thoughtful insights on my projects.

I would like to acknowledge support from the MBTG training grant during my earlier years of graduate school. Dr. Christina Hull has remained a constant guiding presence throughout my graduate career.

Lastly, thanks to others for their expertise on various aspects of my work: Drs. Melissa Harrison, Peter Lewis, Xiaolan Zhao, Jim Haber, and Xiaoxue Zhou.

Table of Contents

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Appendix

Chapter 1.

1.1. Forkhead family of transcription factors

Members of the FOX (forkhead box) family of proteins are defined by the presence of a forkhead box DNA binding domain. The first FOX member to be identified was the *Drosophila melanogaster* gene fork head (*fkh*). *fkh* mutant flies show ectopic head structure development in the gut [1]. A subsequent co-crystal structure of the human FOX protein HNF-3γ in complex with DNA revealed the presence of a unique helix-turn-helix DNA-binding motif called a winged helix motif [2]. Thereafter, many FOX proteins were identified in other organisms, including at least 50 FOX family members present in humans, which also show tissue-specific expression patterns.

FOX family members frequently serve as transcription factors for genes that act at specific points in the cell cycle. For example, *FOXM1* promotes the G1/S transition and G2/M transition. *FOXM1* is highly expressed in cells that actively proliferate, with lowered expression in differentiated, non-proliferating tissues. Unsurprisingly, *FOXM1* is overexpressed in many cancer types [3]. FoxA and certain FoxE proteins can function as pioneer factors: they can bind to sites within compacted chromatin and facilitate chromatin opening, allowing other factors to interact with the now-open chromatin [4,5].

The FOX family of proteins are of great interest due to the roles they play in human development and disease. Given their roles in cell proliferation and differentiation, it follows that when FOX proteins are dysregulated, they can lead to the unregulated cell proliferation and tissue invasion (leading to metastasis) that are hallmarks of cancer [6].

FOX proteins play pivotal roles in controlling cell proliferation and differentiation in complex multicellular organisms. Interestingly, in contrast to many other tissue-specific

transcription factors, FOX proteins have orthologs dating back much earlier in evolutionary history. The single-celled yeast *Saccharomyces cerevisiae* (budding yeast) has identifiable FOX protein orthologs that regulate cell proliferation and differentiation-like processes in this single-celled eukaryotic microbe.

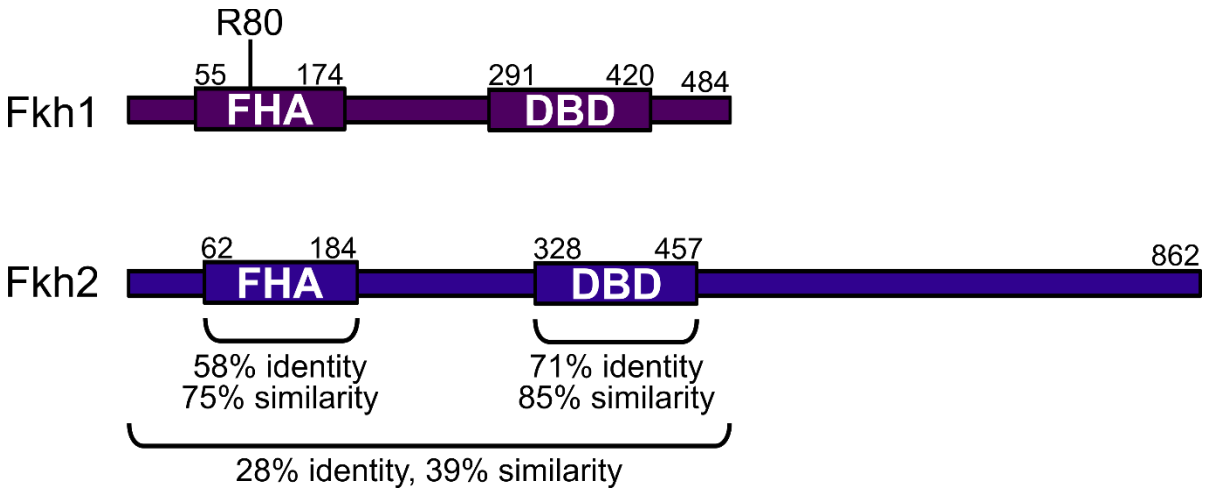
1.2. Forkhead proteins in *Saccharomyces cerevisiae*

Fkh1 and Fkh2 are paralogs resulting from an ancient whole genome duplication in *Saccharomyces cerevisiae*. The two proteins share the same basic structural features and organization, with an N-terminal forkhead-associated (FHA) domain and a more C-terminal forkhead box DNA binding domain. Save for a unique C-terminal region in Fkh2 that lacks strong conservation to any known domains, Fkh1 and Fkh2 share substantial identity, especially across their FHA and DNA binding domains (**Figure 1.1**).

As FOX proteins, Fkh1 and Fkh2 each contain a winged helix DNA binding domain. While both proteins bind to some of the same genomic loci, other genomic targets are uniquely bound by either Fkh1 or Fkh2, with more loci overall uniquely bound by Fkh1 than loci uniquely bound by Fkh2. Fkh2 cooperates with Mcm1 and Ndd1 to bind to DNA, which contributes to Fkh2's unique binding profile. Fkh1 and Fkh2 play some overlapping roles: both Fkh1 and Fkh2 are known to act as transcription factors for certain genes, and indeed both Fkh1 and Fkh2 bind near the transcription start sites of these genes. Overall, the DNA binding domains of Fkh1 and Fkh2 are capable of binding to the same target sites, which helps explain their overlapping, partially redundant roles in the yeast cell cycle. However, due to auxiliary factors (some defined and some yet unknown), under wild-type conditions, Fkh1 and Fkh2 each bind to some unique target sites as well.

Figure 1.1. Comparison of Fkh1 and Fkh2 domains.

Except for Fkh2's unique C-terminal region, Fkh1 and Fkh2 share substantial similarity. They share the highest similarity across their FHA (forkhead-associated) and DBD (DNA binding) domains.



What is clear is that the DNA binding domain is critical for each of these proteins' functions: mutants with a deleted DBD (*fkh1-dbdΔ* or *fkh2-dbdΔ*) behave equivalently to *fkh1Δ* or *fkh2Δ* cells respectively [7]. Models for Fkh1 and Fkh2 functions propose that each protein's direct role occurs at the genomic loci to which these proteins bind.

1.3. Forkhead-associated (FHA) domains

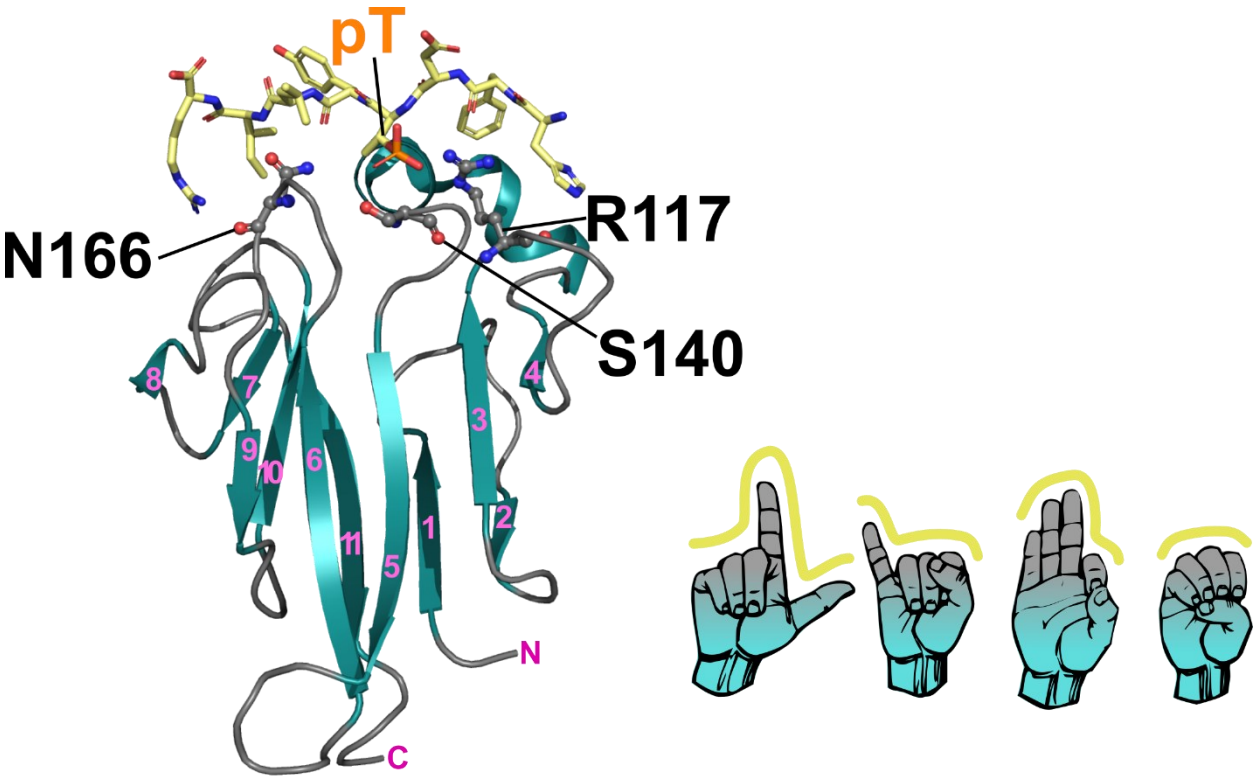
In addition to the winged helix DNA binding domain that defines all FOX family members, Fkh1 and Fkh2 each also contain an FHA domain. FHA domains are protein-protein interaction domains unique for their specific recognition of phosphorylated threonine (pT) residues on their protein partners.

Minimally, FHA domains contain 11 β sheets connected by loops of variable sequences and lengths (**Figure 1.2**). Within one of these loops is an important arginine residue that recognizes pT residues on FHA-containing proteins' partners – for *S. cerevisiae* Fkh1, for example, the critical arginine residue is residue 80 (R80). Sequence variations and structural flexibility in the loops give rise to distinct ligand-binding preferences. Therefore, aside from recognizing a phosphorylated threonine residue, FHA domains interact with residues adjacent to the pT residue. However, beyond the requirement for a pT residue, a general rule defining FHA-ligand interactions does not yet exist, so it is not yet possible to confidently predict which adjacent residues are most important for a given FHA domain. Thus, a major challenge in the FHA domain field is defining the biologically relevant protein partners of FHA-containing proteins. It is also important to identify the specific pT-containing epitopes at the protein-protein interaction interface, and to assign these FHA-pT interactions to specific biochemical step(s) in the cell.

Figure 1.2. Example of general FHA domain structure.

Crystal structure of Chk2-FHA and a phosphothreonine-containing peptide (PDB: 1GXC) [8]. The 11 β sheets common to all FHA domains are depicted in teal. Loops, shown in gray, have sequence variability and structural flexibility that give rise to their substrate specificity. An FHA-bound peptide is shown in yellow. FHA domains recognize epitopes containing a key phosphothreonine residue (pT).

Right: FHA domains can be compared to hands, with a palm that stays the same shape connected to fingers that can form various shapes.



1.4. Fkh1 and Fkh2 regulate the yeast cell cycle through transcriptional regulation of the *CLB2* gene cluster

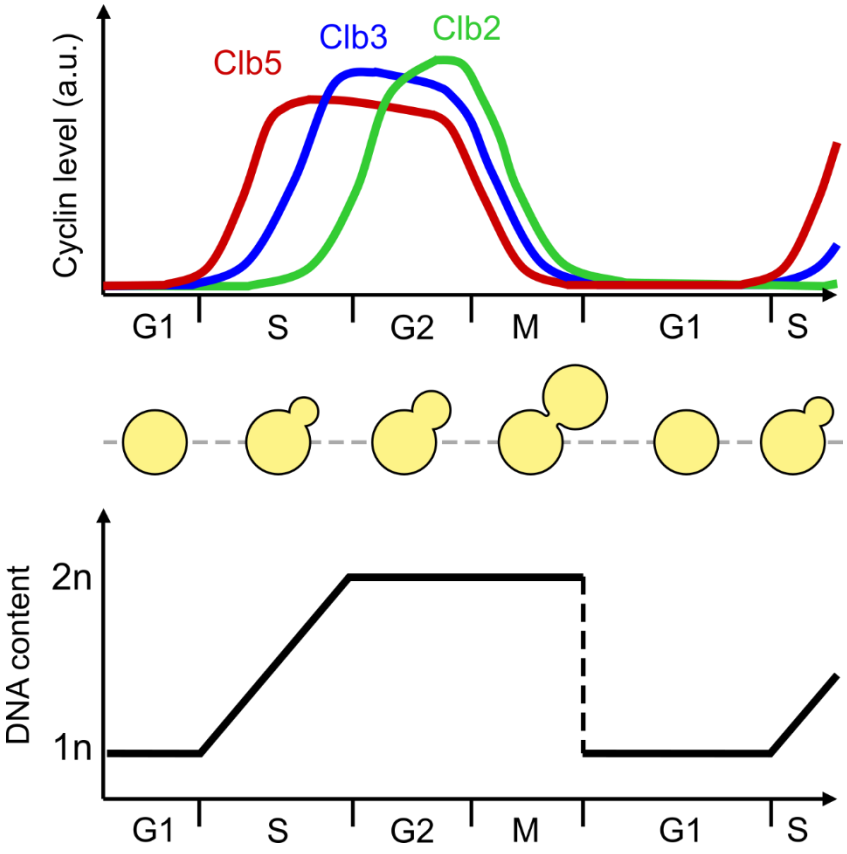
Fkh1 and Fkh2 were discovered as the transcription factors that participate in regulation of the cell cycle, and both their DBDs and FHA domains are important for their transcriptional roles. Specifically, Fkh1 and Fkh2 were found to drive the G2/M phase of the cell cycle by regulating the transcription of the *CLB2* gene cluster (**Figure 1.3**), a group of about 33 genes important for mitosis that are transcribed in late G2 and early mitosis [9]. In yeast strains where Fkh1 and Fkh2 are both absent (*fkh1Δ fkh2Δ*), *CLB2* genes largely stop oscillating in tandem with the cell cycle, and the peak expression levels of these genes are reduced. *fkh1Δ fkh2Δ* cells also enter a pseudohyphal growth state: they grow in filaments and penetrate a solid agar growth medium, a phenotype that is linked to cell cycle alterations but can be suppressed by overexpressing a target of Fkh1 and Fkh2 regulation, the G2/M phase cyclin *CLB2* [10]. The pseudohyphal morphological state provides a simple assay for the functionality of *FKH1* and *FKH2* that I use in Chapter 2 to test whether my Fkh1-TurboID construct compensates for Fkh1 function, at least in terms of cell cycle transcription.

While the genetic phenotype shows that Fkh1 and Fkh2 proteins can partially compensate for each other's role in controlling cell cycle progression, more detailed studies of *fkh1Δ* and *fkh2Δ* mutants provide evidence that each protein normally plays a distinct transcriptional role in wild-type cells [10]. The current view is that Fkh2 is normally the primary positive regulator of *CLB2* transcription, with Fkh1 being able to substitute only when Fkh2 is absent [11,12]. In fact, Fkh1 may actually play an "anti-activation" role, reducing the peak level of *CLB2* mRNA levels even as mRNA levels cycle normally [10,13]. This attenuation role is consistent with the observation that Fkh1 interacts with the Sin3 repressor [14].

Figure 1.3. Cell cycle diagram including waves of Clb gene expression.

Fkh1 and Fkh2 regulate transcription of *CLB2* cluster of G2/M-specific genes. Clb gene expression is linked to the cell cycle; Clb2 and associated *CLB2* cluster genes peak in transcription early in mitosis.

Other processes closely linked to the cell cycle include cell division and DNA replication.



One example of Fkh1's distinct function compared to Fkh2 is that a deletion of Fkh2 decreases *CLB2* mRNA levels while a deletion of Fkh1 causes elevated *CLB2* mRNA levels, stemming from a release of Fkh1-mediated silencing of the *CLB2* gene cluster. These opposite phenotypes provide evidence for the divergent roles of Fkh1 and Fkh2 in transcription.

In summary, while Fkh1 and Fkh2 can at least partially substitute for one another in terms of controlling *CLB2* cluster transcription, they have diverged to play distinct roles in transcriptional regulation in wild-type cells, with Fkh2 being the primary cell cycle transcriptional activator of the *CLB2* gene cluster.

1.5. The Fkh1 protein also has transcription-independent roles in chromosomal biology

Fkh1 is of particular interest to our lab because it has direct roles in chromosomal processes that are mechanistically separable and distinct from Fkh1's roles in gene-specific transcription. Specifically, Fkh1 contributes directly to promoting the process of homologous recombination (HR) and the activity of DNA replication origins that act in the first half of S-phase. These origins are termed early origins [15–21]. While Fkh2 can (at least partially) substitute for loss of Fkh1 in this function, Fkh1 is the primary forkhead protein that regulates HR and DNA replication origins through its binding sites within the recombinational enhancer (RE) and adjacent to early origins. These non-transcription roles for Fkh1, with a particular emphasis on Fkh1's FHA domain, are the focus of my thesis research (Chapter 2).

1.5.1. The Fkh1-FHA domain promotes homologous recombination (HR)

Fkh1 promotes mating-type switching in yeast, a process by which the genes that control mating type are exchanged (**Figure 1.4.a**). Mating-type switching is a locus-specific form of HR involving homologous loci on chromosome III. Fkh1 performs this function by binding to a

discrete DNA element found on the left end of chromosome III named the recombination enhancer, or RE. The purpose of the RE is to promote a genetic switch of *MAT*a cells to the opposite mating type in yeast, *MAT*α. Specifically, after the *MAT*a locus is cut by a specialized endonuclease called HO, *MAT*a can recombine with either of the two silent mating-type loci that exist on the ends of chromosome III, *HML*α or *HMR*a [22]. *HMR*a would normally be favored as a recombination target for *MAT*a because these two loci are closer together; however, such an event does not result in a productive change in mating type. To prevent nonproductive recombination events, the RE is active in *MAT*a cells; thanks to Fkh1 binding, *MAT*a cells favor the RE-proximal *HML*α as the donor 90% of the time. In summary, the directionality of mating-type switching is regulated by Fkh1 bound to the RE.

Arginine 80 (R80) of the Fkh1 protein is the residue predicted to be critical for Fkh1's FHA domain to recognize phosphorylated threonines (pTs). Thus, the *fkh1-R80A* mutant allele should disable Fkh1's ability to bind pT-containing protein partners. The Haber lab showed that the *fkh1-R80A* mutant allele was equivalent to a *fkh1Δ* mutation or a RE deletion in terms of promoting the directional recombination between *MAT*a and *HML*α [17]. These and other genetic experiments supported a model in which Fkh1's FHA domain interacts with proteins present at the double-strand break at *MAT*a to promote the a-to-α mating-type switch. However, no specific phosphothreonine-containing partner proteins required for mating-type switching directionality were known until our lab identified Mph1 and Fdo1 [20].

For Mph1, co-immunoprecipitation and *in vitro* fluorescence polarization experiments with purified Fkh1 and synthesized Mph1 peptides identified the discrete region containing two different pTs, each serving as an individual Fkh1-FHA binding site.

Mutations in *MPH1* that substitute the relevant threonines to alanine residues, hereafter known as the *mph1-2TA* allele, reduce the directionality of mating-type switching (though not to the same degree as the reduction in a *fkh1-R80A* strain). The Mph1 protein is the best characterized direct protein partner of Fkh1's FHA domain, where the relevant pTs and peptide region have been identified. Precisely defining the relevant interaction surface is important: while the *mph1-2TA* allele acts equivalently to *mph1Δ* in terms of mating-type switching, it does not phenocopy other *mph1Δ* phenotypes, suggesting that the Fkh1-Mph1 interaction is dedicated to Mph1's role in HR. Defining the precise Fkh1-FHA partners as well as the relevant pT-dependent interaction surface is critical to performing definitive *in vivo* experiments, especially because many of these interactions could involve partner proteins that have multiple nuclear functions and/or are essential so null mutations cannot be made.

Mph1 acts additively alongside Fdo1 to promote the directionality of mating-type switching. However, deletions of both proteins (*mph1Δ fdo1Δ*) do not reduce directionality to the same degree that *fkh1-R80A* does (**Figure 1.4.b**). To date, no single Fkh1-FHA partner has been defined that is as critical as the Fkh1-FHA domain itself. Thus, our model is that the Fkh1-FHA domain transiently contacts multiple distinct pT-containing partners bound to the DSB at *MAT*a, localizing *HML*α near *MAT*a and enhancing the probability that the double-strand break at *MAT*a can perform the strand invasion into *HML*α required for HR.

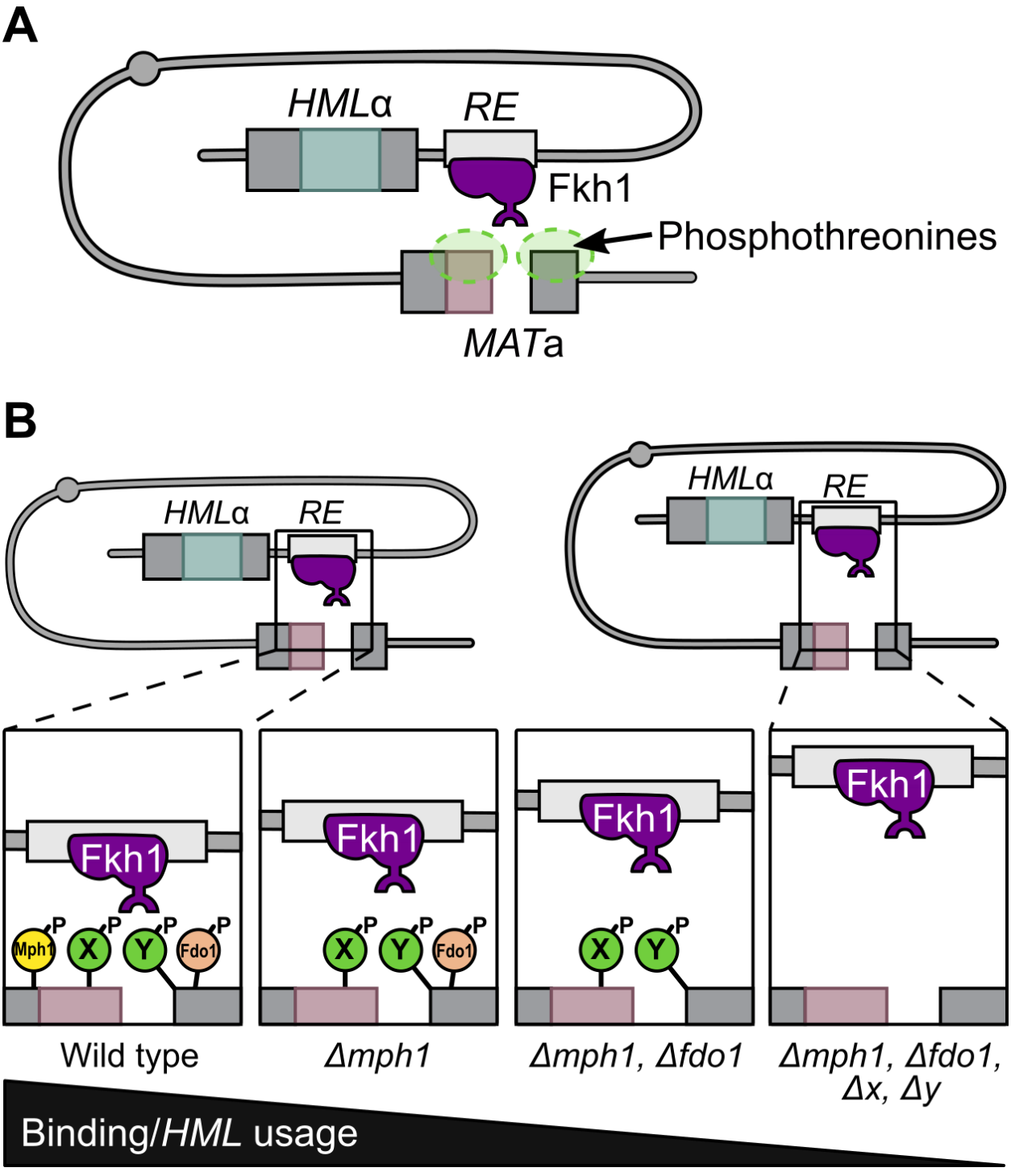
Mating-type switching is a specialized, locus-specific form of HR; the current thinking is that, aside from the endonucleolytic cut made by the HO endonuclease that forms the DSB at *MAT*, the proteins and steps involved are the same as those used for general HR events occurring between any two loci in yeast, including between a double-strand break and an exogenously added homologous DNA fragment intended to generate a specific mutant by CRISPR.

Figure 1.4. Model of Fkh1's regulation of donor preference during mating-type switching.

A. Fkh1 binds near the recombination enhancer (RE) and interacts with phosphothreonine-containing proteins present at the double-strand break.

B. Donor preference in mating-type switching depends partially on both Mph1 and Fdo1: *HML*α usage decreases if either is deleted. However, a *fkh1-R80A* mutation decreases *HML*α usage even further, suggesting the involvement of additional unknown Fkh1 interactors (X and Y).

Figure 1.4.


A
HMLα
RE
Fkh1
Phosphothreonines
MATa
B
HMLα
RE
HMLα
RE
Fkh1
Mph1
X
Y
Fdo1
Fkh1
X
Y
Fdo1
Fkh1
X
Y
Fkh1
Wild type
Δmph1
Δmph1, Δfdo1
Δmph1, Δfdo1, Δx, Δy
Binding/HML usage

Consistent with this idea, when Fkh1 (as a LexA-Fkh1 fusion) is tethered to a donor by a LexA binding site, this donor is used far more frequently to repair a double-strand break generated by CRISPR-Cas9 [15]. Thus, more general cases of HR initiated by Cas9 can also be enhanced by the localized activity of Fkh1.

1.5.2. The Fkh1-FHA domain promotes the activity of most DNA replication origins acting in early S-phase

This thesis mainly focuses on how Fkh1's FHA domain promotes origin activity. To appreciate this issue, this section of the Introduction will present some of the basic issues in the field of yeast DNA replication origin function and regulation.

1.5.2.1. Overview of DNA replication origins

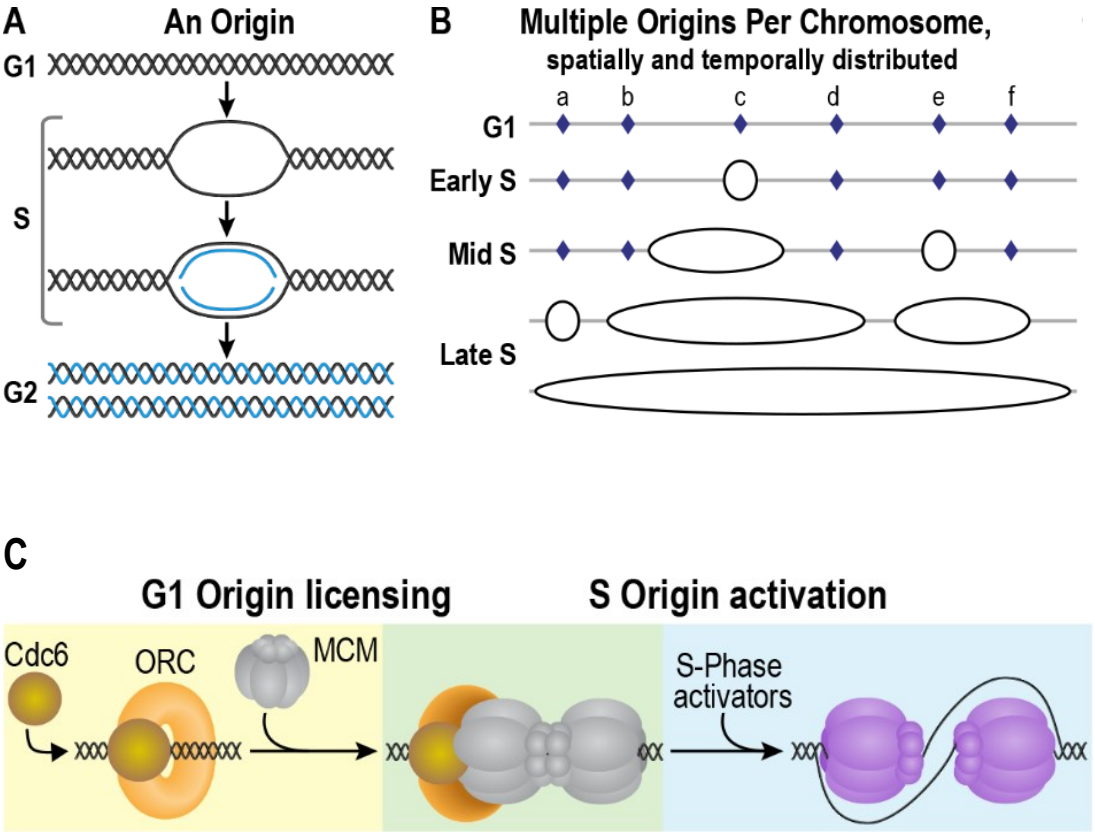
DNA replication origins (origins) are the regions on chromosomes where the process of DNA replication begins or initiates (**Figure 1.5**). This initiation step is essential for the complete duplication of the cell's genomic DNA, a defining requirement of cellular proliferation. Before one cell divides to form two new identical daughter cells, each base pair of its genome must be copied exactly. However, this process is made more complicated in eukaryotic cells, whose genomes consist of multiple linear chromosomes that each must be duplicated independently. Another complication is that each chromosome contains multiple independent DNA replication origins. In addition, it is critical that these origins are used in a sequential manner throughout S-phase to adequately distribute the replication machinery, prevent depletion of dNTPs, and distribute higher-risk replication intermediates that exacerbate DNA fragility, such as termination zones [23–25]. The temporal and spatial pattern of origin use shows cell type specificity, likely reflecting cell type-specific chromatin structures and/or transcriptional states.

Figure 1.5. General aspects of eukaryotic chromosome replication.

A. Origins are sites on the genome where DNA unwinds and bidirectional replication begins.

B. In eukaryotes, each chromosome's replication relies on multiple origins that are activated at different times during S-phase.

C. Origins proceed through a series of defined steps intimately linked to the cell cycle. During G1 phase, origins are "licensed": ORC-Cdc6 load an inactive version of the MCM replicative helicase. During S phase, various factors act to convert MCM to an active form. MCM unwinds DNA and replication begins.



Even slight perturbations in origin control can promote developmental defects or contribute to cancers [26,27].

1.5.2.2. Origins are licensed in G1-phase

Active DNA synthesis is confined to S-phase. However, the eukaryotic "origin cycle" is intimately linked to the cell cycle, and important steps occur even outside of S-phase [28,29].

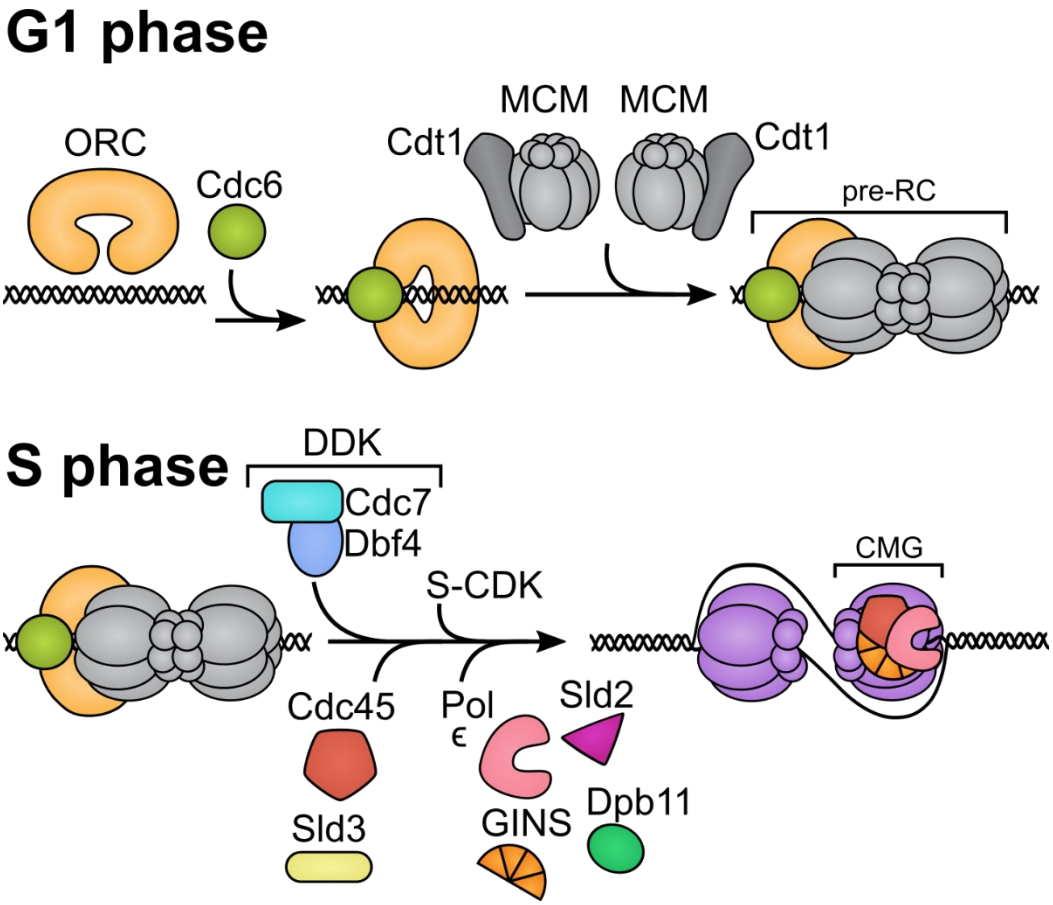
During the M→G1-phase transition, origins are "licensed", or made competent for initiation in the subsequent S-phase. In *S. cerevisiae*, specific origin locations are identifiable by a conserved sequence element called the ACS, short for ARS (autonomously replicating sequence) consensus sequence, so named because an ACS is found in each yeast chromosomal DNA fragment that allows plasmids to autonomously replicate in yeast [29].

In yeast, the origin recognition complex, or ORC, is a hexamer composed of subunits Orc1-Orc6. ORC binds to ACSs throughout the cell cycle, but only in G1-phase can ORC play its active role in licensing origins. Licensing refers to the ORC-directed macromolecular reaction that leads to loading of the replicative helicase complex onto chromosomal DNA.

First, ORC recruits the protein Cdc6, which is expressed in G1-phase and degraded outside of G1 (**Figure 1.6**). Both Orc1 and Cdc6 are ATPases that can drive protein conformational changes *in vitro*. The ORC-Cdc6 complex recruits a Cdt1-MCM complex. Cdt1 is a cell cycle-regulated chaperone and MCM is the catalytic core of the replicative helicase, though MCM is present in an inactive form in G1-phase. Each of the six MCM subunits, Mcm2-7, are ATPases. Through a series of protein-protein recruitments and major protein conformational changes that include the exit and re-entry of Cdc6 and Cdt1 into the macromolecular complex, two MCMs are loaded in a head-to-head arrangement [30,31].

Figure 1.6. Key players in origin initiation.

A more detailed view of the essential machinery required for origin licensing and activation.



This double-hexamer version of MCM (dhMCM) encircles double-stranded DNA and is extremely stable [32]. dhMCM is the substrate that will be converted into two active helicases when the origin initiates, or "fires" [33].

When S-phase begins, the reactions that promote licensing in G1-phase are actively inhibited and origin activation is promoted by the cyclin dependent kinase bound by S-phase cyclins (S-CDK). The inhibition of licensing occurs via multiple mechanisms and is important to prevent re-replication in a single cell cycle. That is, it is this type of inhibition that enforces the eukaryotic rule that each base pair of the genome is replicated once and only once per cell cycle. Mutations that bypass these mechanisms lead to DNA breakage and are likely mutagenic [34–37].

1.5.2.3. The MCM complex is converted to two active helicases in S-phase, causing origin unwinding (a.k.a. firing, activation)

In S-phase, the inactive MCM complex is converted into two active helicases that will bi-directionally unwind the parental DNA [29]. This activation causes the origin DNA to unwind, or "fire." An early step in this process is phosphorylation of the MCM complex by the Dbf4-dependent kinase (DDK), an S-phase kinase that consists of two different proteins: Cdc7, a Ser/Thr kinase, and Dbf4, a substrate-targeting subunit. A collection of transiently associated pre-initiation factors and the Cdc45 protein are then recruited to the MCM complex, followed by recruitment of another multiprotein complex named GINS. Once Cdc45 and GINS associate with the MCM helicase complex, the active helicase holoenzyme, which contains one copy each of Cdc45, Mcm2-7 and GINS, can form. This active form of the MCM helicase is called the CMG helicase. The individual components of the CMG were discovered in yeast, but the biochemical advance that showed that the active helicase required Cdc45, Mcm2-7 and GINS was made using extracts of *Drosophila melanogaster* embryos [38].

1.5.2.4. Stochastic control of origins and the spatial and temporal pattern of chromosome duplication

A eukaryotic chromosome is not replicated all at once during S-phase. Instead, certain regions of the chromosome are replicated early in S-phase, other regions are replicated in mid-S-phase, while still other regions are replicated in late S-phase after most of the genome has already been duplicated. The temporal activation of origins is the primary determinant of the spatial and temporal pattern of chromosome replication [39,40]. Each origin in yeast is typically responsible for replicating about 30-50 kb of chromosomal DNA, called a replicon.

To act in S-phase, a given origin must be both licensed and activated. However, not every origin will be licensed in a given cell cycle, and not every licensed origin will be activated: both the G1-phase licensing and S-phase activation reactions are stochastic [23]. The probability that any given origin will serve as an initiation site is a product of the probability that an MCM complex will be loaded at that origin and the probability that the MCM complex will be activated [40]. The origin-probability landscape is what leads to the spatiotemporal pattern of chromosome replication that can be observed at a population scale (tens of millions of cells). Even within a clonal population of dividing cells, each individual cell uses a distinct collection of origins to achieve full replication of its genome. This stochasticity itself has been proposed to contribute to genome stability across a clonal population of cells [41].

However, at a population level (tens of millions of cells), reasonably predictable spatial and temporal patterns of chromosome duplication emerge. Origins that have a high probability of being licensed and activated are among the earliest, most efficient origins in the genome at the cell population level and are responsible for chromosomal regions that are replicated early in S-phase. In contrast, origins that are less efficiently licensed and activated are generally concentrated in late-replicating regions. The precise mechanisms that govern an origin's

probability of being licensed or its probability of being activated are not known. However, a general consensus is that multiple chromatin-associated proteins, including histone modifiers, sequence-specific DNA binding proteins, and even higher-order chromosome structures such as loops (e.g. topologically associated domains) collaborate to establish distinct chromatin environments that directly influence the efficiency of the G1-phase licensing and/or S-phase activation steps discussed above. Our lab has focused on identifying chromatin- and DNA sequence-mediated mechanisms acting in *cis* (e.g. the ORC site) or in *trans* (e.g. non-DNA accessory elements) to origins.

1.5.2.5. Origin-adjacent chromatin context can enhance or diminish origin activity and can work at either the G1-phase licensing or S-phase activation steps

Chromosomal context can have a substantial effect on origin activity in yeast [42]. For example, origins within silenced chromatin regions, the yeast version of heterochromatin, are relatively inactive, but become more active in mutants that disrupt the silent chromatin state [43,44]. Engineered recruitment of a histone acetyltransferase near a late origin leads to that origin acting earlier in S-phase [45]. However, while there are many examples of chromatin context affecting origin activity, the molecular mechanisms are largely undefined. In many cases, even assigning a chromatin-mediated form of origin regulation to either the G1-phase licensing step or the S-phase activation step can be challenging and controversial, perhaps in part because these two reaction phases are not as strictly separated as models present them to be (see Chapter 2 Discussion).

ORC itself can affect local chromatin structure. For example, ORC binding to origins creates a positioned nucleosome array at these loci composed of ~3 precisely-positioned nucleosomes on either side of ORC [46]. In the absence of ORC, this positioned array collapses toward the origin sequence, and the nucleosomes become more randomly positioned. There are

~400 yeast origins, but over 10,000 matches to the ORC site, so using information about both nucleosome positioning and the ORC site sequence is a stronger predictor of yeast origins than using only the ORC site match alone.

The Orc1 subunit contains a nucleosome binding module, a bromo-adjacent homology (BAH) domain. The Orc1 BAH domain is not essential for yeast viability, but it does enhance ORC's binding to chromatin and ORC-origin complex formation at several specific origins. In addition, an *orc1-bahΔ* mutation is synthetically lethal in combination with the viable but temperature-sensitive *orc2-1* allele, providing genetic evidence that the Orc1 BAH domain contributes to origin activity *in vivo* [47].

More recent studies demonstrate that ORC can remodel nucleosomes, evicting H2A-H2B histone dimers, an activity for which the Orc1 BAH domain is required [48]. Another recent study indicates that ORC can recruit several distinct chromatin remodeler ATPases, and that these two components, ORC and any one of four remodelers, are sufficient to create the characteristic positioned nucleosome array on most origins *in vitro* [49]. Some genetic data suggest that this nucleosome array is essential for origin activity *in vivo*, but interpretations of the *in vivo* data are complicated because the relevant mutant alleles of ORC cause cells to be inviable, which can be caused by multiple molecular defects. In addition, it is difficult to generate definitive evidence that these mutant alleles of ORC do not alter the efficiency of the essential biochemical role for ORC in loading the MCM complex.

These observations suggest a way for origin-adjacent chromatin to affect the G1 origin licensing step by influencing ORC-origin binding; however, other experiments point to the recruitment of limiting S-phase activation factors as a culprit in the control of origin activation timing. For example, overexpression of multiple distinct S-phase origin regulatory proteins leads

to virtually every yeast origin firing simultaneously in S-phase. This condition is terminal – yeast cells do not recover from this catastrophic version of S-phase – supporting the idea that the spatiotemporal control of replication is important [50,51]. While these experiments create an artificial experimental situation, they suggest that an early origin remodeling step in S-phase, prior to origin unwinding, is the replication timing control point. Consistent with this idea, experiments focused on understanding why centromere-associated origins in yeast act in early S-phase indicate that recruitment of Dbf4 may be the limiting step for the activation of these origins [52].

The experimental observations that S-phase MCM activation factors can be limiting for origin activation have supported a view in the field that this activation step is the major regulatory "decision point" that controls when and how frequently an origin acts in a population of clonal yeast cells. However, there is another argument in the field that the efficiency of licensing is equally important in determining the efficiency and replication time of chromosomal regions [53,54]. It is likely that both reactions, origin licensing and activation, can be limiting for successful firing at a given origin locus, and thus relevant to regulating origin activity and replication timing of chromosomal regions.

Sir2 is a histone deacetylase required for formation of silent chromatin at telomeres, which also contain "dormant" origins that show almost no detectable activity in their chromosome contexts. However, in mutant yeast that fail to form silent chromatin at telomeres, these origins act more efficiently and earlier in the cell cycle [44]. More recent genome-scale studies from our lab show that both the replication time of telomeres and MCM binding to telomeric origins (according to MCM ChIP-seq) is enhanced in *sir2Δ* mutants that abolish telomeric chromatin [55]. These data suggest that telomeric chromatin inhibits the G1 licensing

step, such that the normally very late replication time of telomeres results largely from these regions being under-licensed relative to other regions of the genome. Thus, telomeres are normally "passively" replicated very late in S-phase, after most of the genome has been fully duplicated, by replication forks that have traveled from distant origins.

In contrast to telomeres that replicate very late in S-phase, centromere regions, positioned centrally in yeast chromosomes, are replicated very early in S-phase. This early replication is caused by the sequences that define the centromeres themselves [56]. Early S-phase replication of centromeres promotes accurate chromosome segregation at mitosis by allowing for efficient assembly of the kinetochore on the duplicated sister chromosomes. A specific component of the kinetochore, the Ctf19 complex, is required for the DDK to accumulate at centromeres. This observation supports a model in which the early, efficient activation of centromere-adjacent origins is due to their increased efficiency of activation in S-phase [52].

As discussed above, the Fkh1 protein promotes the activity of many early origins in yeast. When I began my thesis work this fact was widely accepted, but the mechanism(s) by which Fkh1 performed this role were unclear. Observations from a few labs, including ours, provided evidence that Fkh1 promoted the activity of early origins by binding to origin-adjacent FKH binding sites. For example, FKH sites are enriched near early origins compared to late origins [18]. Fkh1 binding *in vivo*, as measured by ChIP-seq, shows that Fkh1 binding coincides with origins, and is substantially greater at early origins relative to late ones [57]. A normally early origin placed within a late-replicating region will cause the region to replicate early only when an intact FKH site is positioned near the origin [58].

Our lab defined a small subset of origins where Fkh1 acted directly to promote ORC-origin binding [21]. This discovery builds upon an earlier study from our lab where we defined a

subset of ~20 origins, termed positive chromatin (PC) origins where ORC-origin binding affinity measured *in vivo* could not be fully explained by the comparatively weaker ORC-origin DNA binding affinity measured *in vitro* (**Figure 1.7.a**) [59]. In our more recent study, we showed that a specific FKH site positioned 5' of the ORC site was important for these origins' activity and for ORC-origin binding (**Figure 1.7.c-d**) [21]. More critically for my project, we also showed that the pT-binding activity of Fkh1's FHA domain, or the lack thereof in *fkh1-R80A* mutant strains, could explain the requirement for Fkh1 and this specific 5' FKH site for these origins' full activity (**Figure 1.7.b**). In our most recent study on this topic, we showed that the Fkh1-FHA domain's pT-binding activity, as defined by the *fkh1-R80A* mutant allele, was important for both the activity and normal ORC-origin interactions at most early origins [60]. This outcome was significant for my thesis project because it focused my mechanistic questioning toward defining the pT-containing protein partner(s) of the Fkh1-FHA domain that are relevant to early origin activity.

To address this question, I have worked to establish protein-protein interaction assays *in vivo* and *in vitro* that will enable us to rigorously define partner proteins interacting with Fkh1's FHA domain, as well as defining the relevant pT-dependent interaction surfaces on these proteins. Because exploring a specific protein-protein interaction network was my central thesis goal, in the next section I will discuss protein-protein interaction assays that have been used by our lab and others to provide insights into Fkh1 protein partners, and briefly mention the strengths and weaknesses of each type of approach.

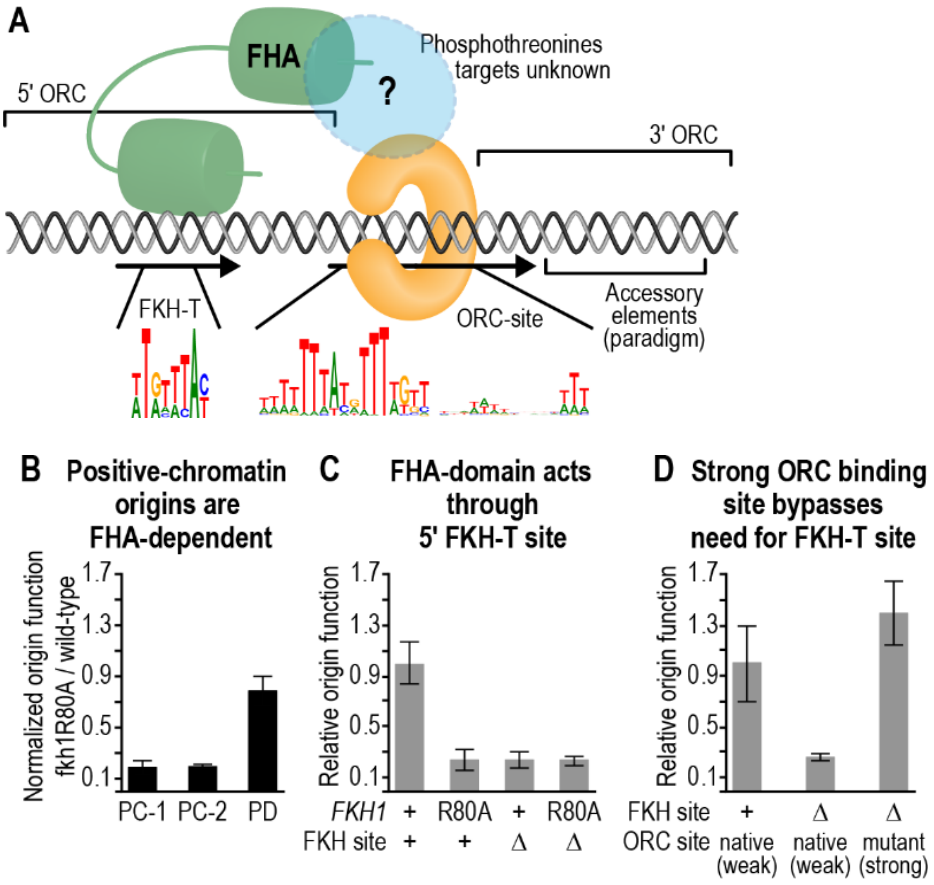
Figure 1.7. Fkh1's effect at positive-chromatin origins.

A. Model showing Fkh1 behavior at a subset of origins known as positive-chromatin origins. Fkh1 (green) binds to the conserved T-rich FKH site 5' of the ORC binding site to which ORC (orange) is bound. Fkh1's FHA domain interacts with phosphothreonine-containing epitopes on its protein partners to promote origin activity through a yet-unknown mechanism.

B. Positive-DNA (PD) origins, or those with strong ORC binding sites whose function does not require additional non-DNA factors, retain their activity when a *fkh1-R80A* mutation is present. Positive-chromatin (PC) origins, on the other hand, lose activity in *fkh1-R80A* cells.

C. Both Fkh1's FHA domain and a 5' FKH binding site are required for full PC origin function.

D. Substituting the weaker ORC binding site of a PC origin with the strong ORC binding site from a PD origin relieves the requirement for a 5' FKH-T site.



1.6. Approaches for defining partners of the Fkh1-FHA domain: advantages, challenges, and limitations

The canonical FHA-protein interaction occurs through non-covalent contacts between the loop region of the FHA domain and a pT-containing peptide region on the partner protein [61]. *In vitro*, the interaction affinities thus far described are in the single to low double-digit micromolar range. *In vivo*, these interactions may occur within multiprotein complexes in which other protein-protein interactions also occur. Because these interactions are regulated by phosphorylation, they may occur transiently *in vivo* in response to cell cycle or metabolic signals. These facets of FHA domains make it challenging to define direct interaction surfaces and then use that information to test the relevance of that particular FHA-dependent interaction to a biological process.

In the interest of using precise terminology, I need to clarify that by "direct" interactions, I mean that two proteins' physical interaction occurs through non-covalent contacts between amino acids within defined epitopes on the two proteins under consideration. To prove such directness, *in vitro* biochemical interaction assays with purified proteins or peptide substrates must be used. However, proteins can also interact physically though indirectly, such as those interactions that occur in multiprotein complexes. Thus, some biologically important physical protein-protein interactions can be indirect. *In vivo* protein interaction assays cannot definitively prove directness, but they are useful for defining biologically relevant physical interactions that can then be followed up with biochemical assays to address whether the interactions are direct. Ultimately, my goal was to define the direct physical partners of the Fkh1-FHA domain; to

identify relevant candidates I have exploited or established *in vivo* protein-protein interaction assays.

1.6.1. 2-hybrid assay

The 2-hybrid assay is a molecular genetic assay suitable for identifying physical interactions between two proteins, though these interactions may not necessarily be direct [62,63]. In this assay, a fusion protein is created in which a protein of interest is fused to one domain of a transcription factor, creating the "bait" protein. The "prey" protein is a fusion of another protein to the other domain of the transcription factor. If the bait and prey proteins come together in close proximity, the complete transcription factor is reconstituted, and it can transcribe a reporter gene whose expression can be measured *in vivo*. This assay was developed for use in yeast, a system in which protein-protein interactions from any eukaryotic organism could be examined. Variations on the assay have been developed using fluorescent or bioluminescent proteins as reporters [64].

A two-hybrid screen is performed by creating a library of prey proteins in which each protein from the organism of interest is fused to one half of the transcription factor; the bait protein is the protein whose interactions are being probed.

Our lab performed a 2-hybrid screen using Fkh1 as bait and identified five potential Fkh1 interactors, including Mph1 and Fdo1 [20]. We chose to pursue both Mph1 and Fdo1 further because each of these interactors was also found in a proteomic screen in which purification of FLAG-tagged Fkh1 also co-immunoprecipitated Mph1 and Fdo1, affirming the accuracy of our 2-hybrid screen [65]. Because two independent assays identified the same Fkh1-interacting proteins, we were more confident that Mph1 and Fdo1 were relevant factors worth investigating [20]. For the Fkh1-Mph1 interaction, we also pursued *in vitro* biochemical assays to show that

this interaction is direct and occurs between the FHA domain of Fkh1 and a specific pT region of Mph1.

Another lab performed a 2-hybrid screen using Dbf4 as a bait protein and identified Fkh1 as a physical interactor [66]. They also performed co-immunoprecipitation experiments that generated data in support of this interaction. There is no independent evidence from another lab that Fkh1 and Dbf4 interact physically, nor is there evidence of a direct interaction. For example, Dbf4 was not identified in the affinity purification screen mentioned above [65]. The relevant interaction regions in Dbf4 and Fkh1 were only partially investigated, and the role of the Fkh1-FHA domain was unclear. In Chapter 2, I address the possibility that the Fkh1-Dbf4 interaction may be supported in part by a direct physical interaction between Fkh1's FHA domain and a pT-containing peptide in Cdc7, the kinase partner of Dbf4.

The advantage of the 2-hybrid screen is that it is technically simple and cheap. In addition, it can be used to rapidly identify putative interaction regions on partner proteins. For example, for the Fkh1-Mph1 interaction, the 2-hybrid assay was used to identify regions within both Fkh1 and Mph1 required for the interaction, which in turn allowed the generation of a precisely defined allele of *MPH1*, *mph1-2TA*, for genetic studies [20]. However, the artificial nature of the interaction assay (e.g. fusion proteins that must come together to stimulate transcription or some other event) and/or the challenge of generating the most appropriate protein fusions and/or comprehensive libraries can limit the effectiveness of this approach. In addition, not all regulated phosphorylation will be preserved in this context, which also limits the 2-hybrid's potential usefulness for examining FHA-dependent protein-protein interactions.

1.6.2. Affinity purification

In an affinity purification assay, a protein of interest (referred to here as the "bait" protein) is pulled down using an antibody or other high-affinity reagent against the bait protein. To define interacting proteins, this enrichment step is performed under non-denaturing conditions to preserve non-covalent interactions. "Prey" proteins are those that co-purify with the bait proteins, and can be identified with targeted approaches, such as protein immunoblots, or through a non-biased approach, such as mass spectrometry. If this approach is applied to a crude mixture of prey proteins, such as a crude cellular extract, the "prey" proteins identified are not necessarily interacting directly with the bait protein.

Our lab was able to enrich for (co-immunoprecipitate) Mph1 when Fkh1 served as the bait in this approach and conversely found that Fkh1 could be enriched when Mph1 was used as bait [20]. I have used a version of this assay in which I use an immobilized Mph1 peptide to bind to recombinant purified Fkh1 in Chapter 2. In this version of the assay, I can conclude more confidently that I am examining direct interactions.

Affinity purification methods for identifying new interactions require that interactions between bait and prey proteins are stable enough such that complexes survive the often-harsh processing methods required to extract lysate from yeast cells and the subsequent affinity purification and wash steps. While Fkh1 and Mph1 appear to interact stably enough for their interactions to be detected with both affinity purification and 2-hybrid methods, I was interested in discovering and examining Fkh1-FHA partner proteins that might have been missed by other approaches and that might occur transiently, for example, only during particular stages of the cell cycle. I expected these types of interactions to be more difficult to detect by the methods discussed above. Therefore, I invested substantial effort in establishing a Fkh1-FHA-dependent

proximity labeling assay using the engineered TurboID biotin ligase. I describe my work in establishing this assay system in Chapter 2.

1.6.3. Biotin proximity labeling: "tagging" with biotin

Proximity labeling is an *in vivo* method in which a protein of interest creates a label on nearby proteins, providing a sensitive method for examining possible physical interactors of the protein of interest [67]. Essentially, proximity labeling allows a protein of interest to create a persistent mark on other proteins that have encountered the protein of interest, creating a "history" of the proteins that have been near the protein of interest during the labeling time.

Biotin proximity labeling methods use a modified biotin ligase. The *E. coli* biotin ligase BirA converts the substrate, biotin, to a reactive biotinoyl-AMP molecule that can subsequently be attached to a partner protein. The innovation of the BioID system was to create a promiscuous version of the BirA protein, in which biotinoyl-AMP is released from the active site and diffuses outward in a narrow radius (10 Å) to nonspecifically label lysine residues on other proteins [68]. The BioID protein has undergone directed evolution in yeast to improve its enzymatic efficiency, resulting in the TurboID biotin ligase [69].

The TurboID protein was selected because of its high efficiency, the short labeling time it requires, and its activity at lower temperatures compatible with yeast growth (30°C as opposed to 37°C). Unlike the 2-hybrid and affinity purification methods previously discussed, TurboID proximity labeling does not require interactions between Fkh1 and its partners to stably persist for long periods of time, and its short labeling time also makes TurboID compatible with the cell cycle arrest methods that may be helpful in examining the regulation of cell cycle-linked processes.

To carry out biotin proximity labeling, I have created fusion proteins consisting of Fkh1 fused to TurboID. Much of my thesis work was devoted to establishing this method in the lab so it worked reproducibly to identify Fkh1 FHA-specific partners. This work will be discussed in Chapters 2 and 3.

In theory, the TurboID proximity labeling method has the potential to detect biologically relevant interactions *in vivo* in an unbiased manner when coupled with mass spectrometry. However, I encountered several challenges, primarily that yeast contain an abundance of endogenous biotinylated proteins, and that substantial purification of nuclear proteins will likely be essential to generate samples that are useful for detecting biologically relevant interactions by mass spectrometry.

Chapter 2

Examining specific FHA-dependent protein-protein interactions to gain insights into how the forkhead associated (FHA) domain of *Saccharomyces cerevisiae* Fkh1 promotes the activity of early DNA replication origins

This chapter is in preparation for publication in 2025, on which I will be first author. I (with the help of my advisor, Catherine Fox) conceived and designed all experiments. Timothy Hoggard conducted S-phase flow cytometry and Sort-seq experiments and developed R analysis pipelines to analyze the data. Michael Sheets helped design Fkh1-TurboID constructs.

2.1. Abstract

Hundreds of independent DNA replication origins are both spatially and temporally distributed across the 16 nuclear chromosomes that make up the *Saccharomyces cerevisiae* genome. Each origin acts in its own local chromatin context that can regulate its efficiency (i.e. the percent of cell cycles in which the origin will function) and timing (i.e. the point in S-phase the origin will function). The Fkh1 protein is a transcription factor that binds near early origins, defined as the ~100 origins that function in the first third of S-phase, and enhances their function using its conserved FHA domain. FHA domains are protein-binding modules with specificity for phosphothreonine (pT)-containing peptides. The relevant pT-containing protein partner(s) of the Fkh1-FHA domain required for early origin function are unknown. Here, we showed that the one known FHA-dependent protein partner of Fkh1, the Mph1 DNA helicase, could not explain Fkh1's role at early origins. Next, we used our knowledge of the FHA-binding peptide on Mph1 to identify putative FHA-binding peptides within the known essential origin-regulatory proteins. A C-terminal peptide on Cdc7, centered on Cdc7's T484 residue, emerged as a strong candidate for an FHA-dependent Fkh1 partner relevant to early origins. A Fkh1-directed protein proximity labeling assay provided evidence of a Fkh1-Cdc7 interaction *in vivo*. To better address mechanism, an *in vitro* peptide binding assay was used to examine interactions between two FHA-binding peptides: the Mph1-T776 peptide, considered as the canonical substrate of the Fkh1-FHA domain, and the Cdc7-T484 peptide, considered as the canonical substrate of the Rad53-FHA1 domain. Like Fkh1, Rad53 also functions in S-phase, except its major role is to function at the S-phase checkpoint, slowing DNA replication progression when problems arise. There is strong biological rationale for each of these domains needing to interact with Cdc7, an S-phase kinase essential for origin function. Therefore, we posited that the Rad53-FHA1 and Fkh1-FHA domains might share one pT peptide binding partner. In support of this idea, the

Mph1-T776 and Cdc7-T484 peptides each bound to the Fkh1-FHA and Rad53-FHA1 domains in a pT-dependent manner. However, experiments using mutant versions of these FHA domains, in which the conserved pT-anchoring arginine residue characteristic of FHA domains is altered, challenged this interpretation. While the interactions of Fkh1 with Mph1-pT776 and Rad53 with Cdc7-pT484 required the key pT-anchoring residue within the corresponding FHA domain, the swapped substrate partner interactions, i.e. Fkh1 with Cdc7-pT484 and Rad53 with Mph1-pT776, did not. Therefore, despite the substantial similarity between both the peptide substrates and the FHA domains, the *in vitro* data argued against the Fkh1-FHA and Rad53-FHA1 proteins sharing substantial cross-specificity for peptide partners. Consistent with this revised idea, S-phase progression of *fkh1-R80A* and *cdc7-T484A* mutant cells monitored by flow cytometry revealed that while *fkh1-R80A* cells delayed S-phase progression, c*dc7-T484* cells did the opposite, accelerating S-phase progression. Thus, the Cdc7-T484 peptide is not the Fkh1-FHA target required for early origin function.

2.2. Introduction

Genome replication initiates at discrete regions on chromosomes called origins of DNA replication. At origins, the parental double strand DNA is unwound to expose the two single-strand DNA templates for new DNA synthesis. Eukaryotic cells rely on many independent origins that are both spatially distributed across each chromosome and temporally distributed throughout the duration of S-phase. In multicellular eukaryotes, this spatiotemporal pattern of origin use shows cell type specificity, indicating that features beyond chromosomal DNA sequence (i.e. chromatin, in its broadest sense) have a large impact on the loci that are permissive to forming origins [70]. *Saccharomyces cerevisiae* (hereafter referred to as yeast) is a single-celled model eukaryote. As a single-celled organism, yeast lacks the chromatin-dictated cell type

specialization of multicellular organisms. Yet, yeast has been an important model for understanding how chromatin features can shape the spatial and temporal pattern of origin use [29,42].

S. cerevisiae is the only eukaryote for which a discrete, small DNA element can be defined as essential for origin function. Many individual origins in yeast function as small ~150 bp elements that can be easily engineered onto a plasmid or into different locations in the genome. These types of experiments reveal that different chromatin environments in yeast can influence the efficiency and temporal activity of an origin. For example, an origin within heterochromatin normally acts in late S-phase, but in mutants incapable of forming heterochromatin, these origins initiate replication earlier in S-phase [43,44,55,71]. In addition, engineered recruitment of a histone acetyltransferase near an origin that normally fires late in S-phase can advance the point in S-phase that this origin functions [45]. Generally, features that define active euchromatin are associated with origins that act in early S-phase while features that define inactive/repressive heterochromatin house dormant or late origins. However, the molecular identity of the relevant chromatin factors and how they control the essential origin-regulatory proteins are not well understood.

A chromatin environment capable of permitting or inhibiting an origin must influence one or more function(s) performed by the core origin-regulatory proteins. These proteins are essential for origin function *in vivo* and *in vitro* and, not surprisingly, given their biochemical function, they are also essential for yeast viability [29]. Yeast have been a key model for defining these essential origin proteins and the molecular steps required to form an origin.

The formation of a eukaryotic origin is inextricably linked to the cell cycle and essential steps occur in two different phases. In G1-phase, the origin is “licensed”. Licensing is a multistep

reaction that results in the loading of an inactive version of the DNA replicative helicase. First, a six-subunit protein complex called the origin recognition complex (ORC) binds to DNA. In yeast, ORC binds to a specific ~33 bp sequence element called the ORC site. A protein named Cdc6 binds to ORC, forming a scaffold on DNA. The ORC-Cdc6-DNA complex recruits a complex of Cdt1 and the hexameric catalytic core of the replicative helicase, the MCM complex. Through a series of incompletely understood conformational changes and Cdc6/Cdt1 exits and re-entries, two MCM complexes are loaded in a head-to-head conformation, forming a semi-stable double hexamer (dhMCM) [32].

In S-phase, the origin is activated, meaning that the origin DNA is unwound. This is sometimes referred to as activation of the dhMCM or origin firing. First, an S-phase kinase named the DDK (Dbf4-dependent kinase), consisting of the Ser/Thr kinase Cdc7 and the regulatory subunit Dbf4, binds to dhMCM and phosphorylates it. This step stimulates recruitment of several transitory pre-initiation proteins and another kinase, the CDK (cyclin dependent kinase) in partnership with an S-phase cyclin, that ultimately results in the formation of two activated replicative helicases. The conversion from the inactive dhMCM state to the active helicase state drives DNA unwinding and requires that each MCM complex associate with two additional factors, the Cdc45 protein and the GINS complex, forming the active helicase CMG (short for Cdc45-MCM-GINS).

The assembly of multiple essential factors during each step presents multiple opportunities for regulation. During G1-phase origin licensing, ORC-DNA binding and Cdc6 recruitment can be limiting for origin formation [72]. During S-phase activation, multiple steps may be limiting, including recruitment of the DDK, several transient accessory pre-initiation factors, and Cdc45 [50,51]. Chromatin has been shown to impact ORC-DNA binding and

recruitment of the DDK [21,52,59]. Furthermore, it is possible that chromatin factors could influence many steps other than these.

In this study, we explore how a non-histone chromatin-associated factor, the Fkh1 DNA binding protein, promotes the activity of early origins, the ~100 (out of ~400 total) yeast origins that act in the first third of S-phase. Several independent experiments have shown that Fkh1 can act directly through Fkh1 binding sites (FKH sites) located near early origins, though for only a small subset of these origins has a precise native FKH site been definitively defined [21]. Until recently, it was also unclear which domains or epitopes on Fkh1, aside from the Fkh1 DNA binding domain, were required for Fkh1's ability to promote early origin activity.

In two recent studies, our lab has observed that the canonical function of the conserved FHA domain within Fkh1 can explain much, if not all, of Fkh1's ability to promote early origin activity [21,60]. This observation was unexpected because several other mechanisms had been proposed, and the FHA domain had even been dismissed as a key issue [73].

FHA domains are protein binding modules with unique specificity for epitopes that contain a phosphorylated threonine residue (pT). A key arginine residue present in each FHA domain is critical for pT specificity. The *fkh1-R80A* allele produces a mutant Fkh1^{R80A} protein in which the key arginine, R80, is substituted with alanine. This mutant protein is expressed at normal levels and maintains Fkh1-specific DNA binding ability [20,60]. However, several lines of evidence show that *fkh1-R80A* leads to a reduction in ORC's binding at a subset of origins, specifically those that depend on a specific FKH site positioned 5' of the origin [21]. These data and other experiments support a role for Fkh1's FHA domain in promoting normal levels of ORC-origin binding. However, in another independent study, the Fkh1-FHA domain was

implicated in the recruitment of the DDK to origins via interactions with Dbf4, though its role in origin activity was not addressed [66].

Recently, our lab used Sort-seq to examine cells' progression through S-phase and observed that *fkh1-R80A* cells showed reduced activity of early origins [60] (**Figure 2.1**). This outcome supported a simple model in which Fkh1's FHA domain interacted with a specific protein(s) to promote the full activity of early origins. In this report, we used our knowledge of the known FHA-dependent protein interaction occurring between Fkh1 and Mph1 that we had defined in a previous study to explore how Fkh1's FHA domain promotes early origin activity.

First, we used Sort-seq to ask whether an *mph1-2TA* mutant abolishing Mph1's interaction with Fkh1 could explain the requirement for Fkh1's FHA domain in early origin activity, but *mph1-2TA* could not explain this. We then used *in silico* approaches to identify putative Fkh1-FHA protein partners among the collection of essential origin regulatory proteins. This work and additional information from the literature pointed to the Cdc7 kinase as a putative direct partner of the Fkh1-FHA domain.

I further explored this possibility by establishing and using an *in vivo* proximity labeling assay and an *in vitro* protein-peptide interaction assay. While the *in vivo* assay was consistent with Fkh1-FHA-dependent interactions between Fkh1 and Cdc7, the data were not definitive, nor can they address the issue of whether a Fkh1-Cdc7 interaction is direct. The *in vitro* assay revealed that two similar FHA domains, Fkh1's FHA domain and Rad53's FHA1 domain, could bind to the relevant Mph1- and Cdc7-derived peptides in a pT-specific manner, consistent with the canonical FHA-peptide interface. However, we were surprised that while the *in vitro* Fkh1-Mph1 interaction was dependent on Fkh1's key arginine R80, the Fkh1-Cdc7 interaction was not. Conversely, the canonical Rad53-Cdc7 interaction was dependent on the key arginine in

Rad53, R70, but the Rad53-Mph1 interaction was not. Thus, despite the structural and biological similarities of these FHA domains and the two pT-containing peptides, these two different FHA domains each have distinct specificities that are biologically meaningful. Consistent with this conclusion, a relevant yeast strain with a T→A substitution in Cdc7 exhibits an S-phase progression phenotype consistent with a defect in checkpoint surveillance. We discuss the implications of our work for defining biologically relevant substrates for Fkh1's FHA domain.

2.3. Results

2.3.1. The defined Fkh1-Mph1 interaction could not account for the requirement of the Fkh1-FHA domain in promoting early origin function.

Fkh1's FHA domain and the multifunctional DNA helicase Mph1 interact directly via a specific phosphothreonine (pT)-containing peptide [20]. This Mph1 peptide region is the only molecularly defined and biologically validated Fkh1-FHA partner interface defined thus far. The Mph1 helicase is not essential for yeast cell viability and has no known role in DNA replication initiation. Nevertheless, it was plausible that, given that Mph1 is a conserved DNA helicase, the Fkh1-FHA-Mph1 interaction might play a role in enhancing the activity of early origins and thus account for the requirement for the Fkh1-FHA domain by early origins [60] (see **Figure 2.1.b**). For example, because early origins are located near highly transcribed genes, they could require additional helicase activity to protect the licensed origin and/or the early replication fork.

To test this hypothesis, we examined the effect that the *mph1-2TA* mutant allele had on yeast genome replication by Sort-seq, as described in **Figure 2.1.a**. The *mph1-2TA* allele substitutes two threonine residues in Mph1, T776 and T785, with alanine residues [20] (**Figure 2.2.a**). *In vitro*, at least one of these two threonine residues must be phosphorylated to promote a Fkh1-Mph1-peptide interaction.

Figure 2.1. Schematic of Sort-seq experiment.

A. Sort-seq schematic in which flow cytometry is used to sort cells by cell cycle phase. DNA from cells in S-phase and G2-phase is sequenced. Reads from the G2-phase population are used for normalization, helping to calculate the S-phase copy number, or the per-nucleotide copy number of the S-phase population, indicating the extent of replication of that locus among the population.

B. Sort-seq data comparing the mean value for S-phase copy number for three independent *FKH1* experiments (black line) and the mean for two independent *fkh1-R80A* experiments (red line). Shaded regions indicate the 95% confidence value for the mean, essentially indicating the variation observed between independent experiments.

Arrows mark loci where the S-phase copy number varies between *FKH1* and *fkh1-R80A* genotypes; downward arrows indicate origins whose activity has been reduced by the *fkh1-R80A* allele. Based on several experiments discussed in the text and published by our lab and others, Fkh1 acts directly as a positive regulator at origins that act early in S-phase (early origins). A consequence of this regulation is that this same mutant allele acts as an indirect negative regulator of less efficient late origins (see upward pointing arrows).

The asterisk indicates a termination zone (TZ), i.e. where two forks traveling in opposite directions meet and resolve to complete DNA replication. This region completes replication earlier in S-phase in *fkh1-R80A*, probably as an indirect byproduct of origin dysregulation in the population.

Figure 2.1.

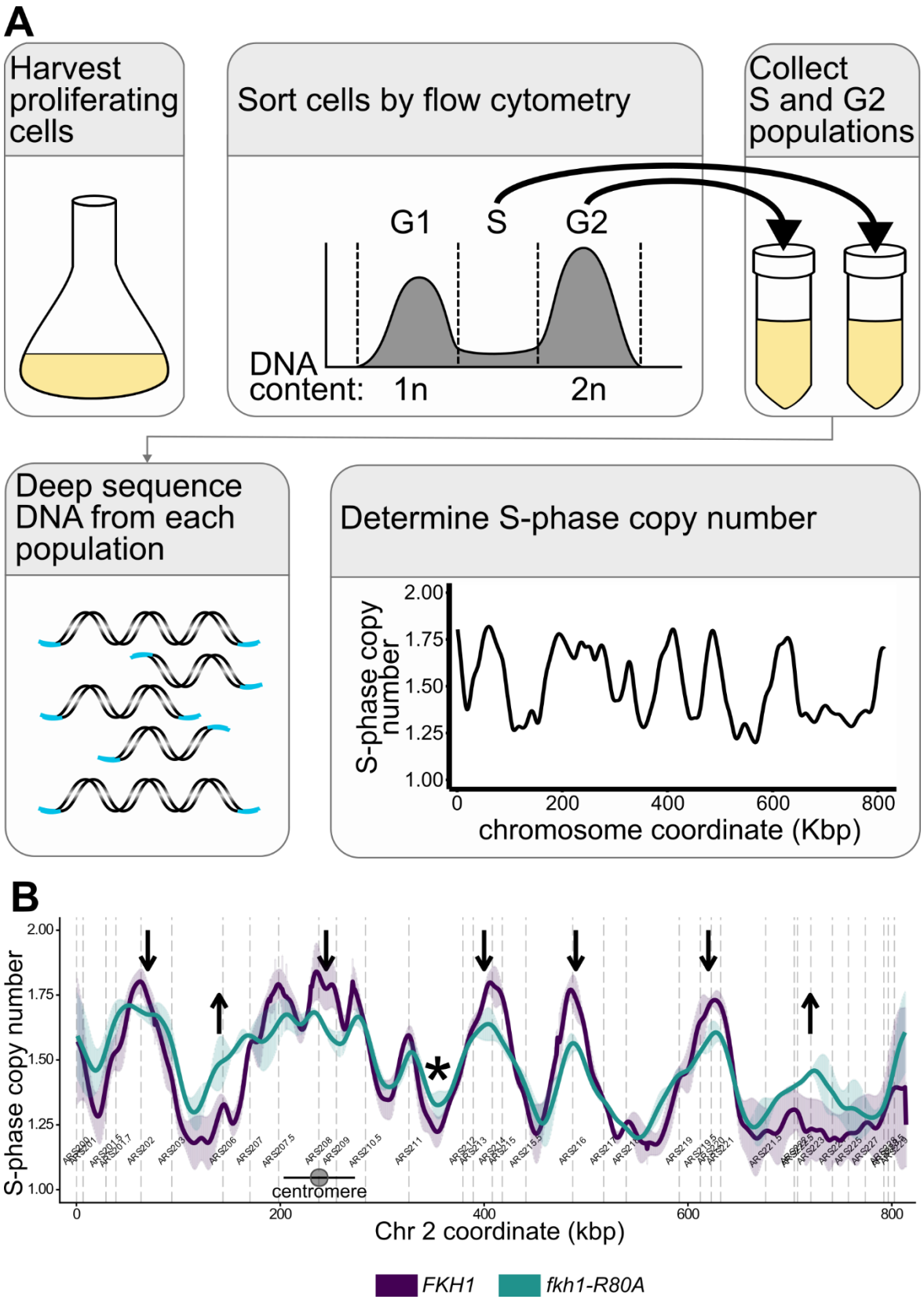

A
Harvest proliferating cells
Sort cells by flow cytometry
G1
S
G2
DNA content:
1n
2n
Collect S and G2 populations
Deep sequence DNA from each population
Determine S-phase copy number
S-phase copy number
2.00
1.75
1.50
1.25
1.00
0
200
400
600
800
chromosome coordinate (Kbp)
B
S-phase copy number
centromere
Chr 2 coordinate (kbp)
FKH1
fkh1-R80A

Figure 2.2. Summary of interactions between Fkh1-Mph1 and Dbf4-Cdc7 (DDK).

A. Domains of Fkh1 and Mph1. Arrows mark specific residues required for Fkh1-Mph1 interaction. Inset: zoom in on Fkh1-FHA-interacting epitope of Mph1 containing phosphorylated threonine residues (highlighted). Residues are marked with asterisks indicating their level of contribution to the Fkh1-Mph1 interaction.

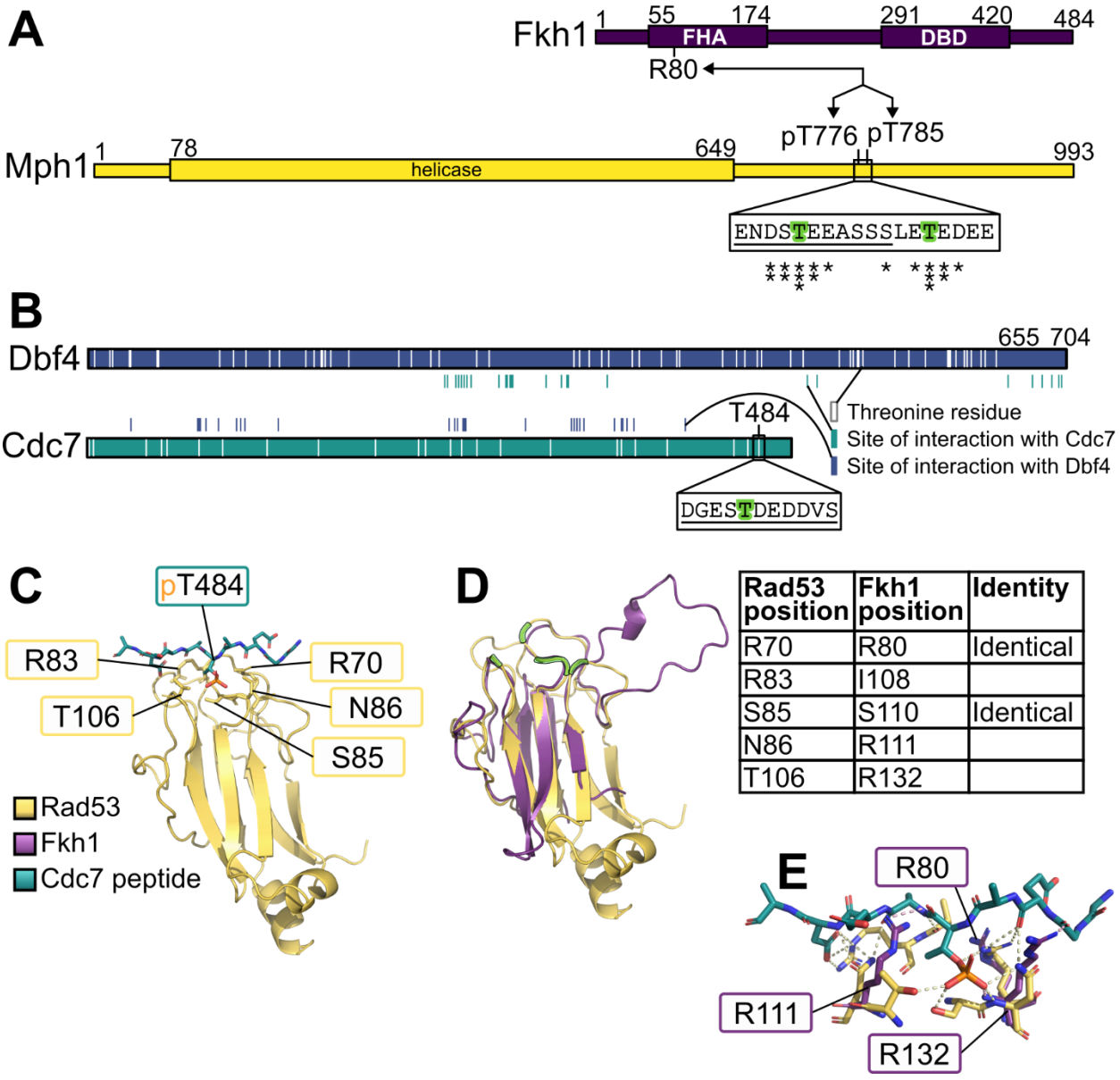
B. Cartoon of Dbf4 and Cdc7. White lines mark threonine residues within each protein. Colorful lines mark residues involved in Dbf4-Cdc7 interaction. Inset: zoom in on Cdc7 epitope containing threonine 484.

C. Structure of Rad53's FHA1 domain interacting with a Cdc7-derived peptide [74]. Amino acids involved in the pT-dependent interaction between Rad53 (in yellow) and a Cdc7-derived peptide (in teal).

D. Rigid-body alignment between Rad53-FHA1 (PDB: 5T2S) and AlphaFold model of Fkh1-FHA (in purple). Cdc7 amino acids from part C are marked in green. Table lists these amino acids and the amino acid in the corresponding position on Fkh1.

E. Detail of the polar contact network between Cdc7-derived peptide and the surrounding residues from Rad53 (yellow) and the aligned Fkh1 model (purple). Hydrogen bonds are marked with dashed lines.

Figure 2.2.



Rad53 position	Fkh1 position	Identity
R70	R80	Identical
R83	I108	
S85	S110	Identical
N86	R111	
T106	R132	

The *mph1-2TA* allele also abolishes Mph1's interaction with Fkh1's FHA domain *in vivo*, and abrogates *MPH1*'s contribution to mating-type switching [20]. Thus, it is clear that this allele causes a defect in the Fkh1-Mph1 interaction within yeast cells. However, based on Sort-seq, the *mph1-2TA* allele did not produce a replication origin phenotype equivalent to that of the *fkh1-R80A* allele (**Figure 2.3**). This failure of *mph1-2TA* to phenocopy *fkh1-R80A* led us to conclude that the previously characterized Fkh1 FHA-Mph1 interaction did not explain the Fkh1 FHA domain's role in promoting early origin function in yeast.

2.3.2. Using knowledge of the Fkh1-FHA-Mph1 interface to guide the identification of putative Fkh1-FHA-protein partners required for promoting early origin function.

Several essential origin-regulatory proteins and steps are potentially rate-limiting for origin function at any specific origin locus *in vivo* [18,21,23,50,51,54,59,72,75]. Thus, individual origins within yeast compete to complete the reactions that will transform them into origins, and this competition is what creates the spatial and temporal pattern of genome duplication. For example, early origins recruit limiting factors and/or more efficiently complete limiting steps required for origins in late G1-phase and/or in early S-phase, which results in late origins going unlicensed or, even if they are licensed, being passively replicated by a neighboring replication fork passing through before they have the chance to fire.

With this framework in mind, we postulated that the Fkh1 FHA domain, when positioned in *cis* and proximal to core origin sequences of early origins, enhanced the recruitment of essential origin-regulatory proteins. Therefore, we considered the essential origin-binding proteins that act in the early stages of origin function as potential Fkh1-FHA binding partners (**Table 2.1**).

Figure 2.3. The Fkh1-FHA-dependent interaction between Fkh1 and Mph1 could not explain the requirement for Fkh1's FHA domain in early origin activity.

Sort-seq was performed on wild-type, *fkh1-R80A*, and *mph1-2TA* cells (n=2). S-phase copy number (line) and 95% confidence interval (shaded area) are plotted across chromosome II. (The *fkh1-R80A* and wild-type (*FKH1*) data were previously shown in Figure 2.1.) The *mph1-2TA* allele reduces the Fkh1-Mph1 interaction below detectable levels *in vivo*. Arrows indicate regions where *fkh1-R80A* copy number uniquely differs from both wild-type and *mph1-2TA* strains.

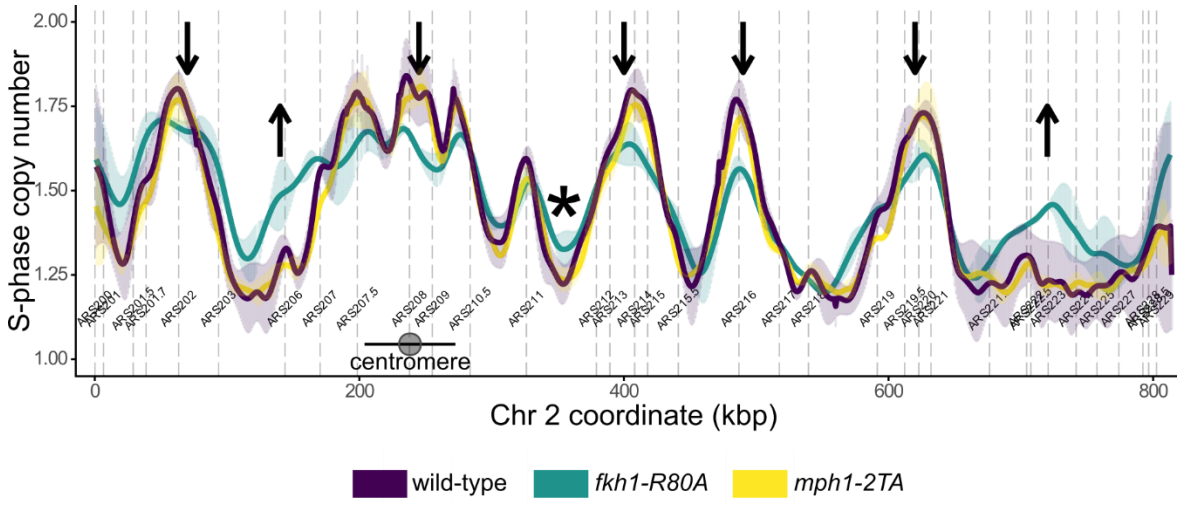


Table 2.1. Phosphopeptide candidates among replication proteins.

Table assessing similarity to peptides containing Mph1-pT776 and Mph1-pT785, phosphorylated threonine residues important for Mph1's interaction with Fkh1's FHA domain. Proteins making up the core replication machinery were searched for matches meeting the following criteria:

1. Experimental evidence exists for a phosphorylated threonine residue at this position.
2. Phosphorylation probability is above 0.5 according to either NetPhos 2.0 or DisPhos 1.3.
3. The threonine residue is predicted to be phosphorylated by a kinase in the CMGC group of cell cycle kinases.
4. The surrounding residues share sequence similarity to the Mph1 peptide (cutoff 75%).
5. The peptide is solvent-exposed and accessible for interactions with other proteins (experimental structures were evaluated if available, otherwise AlphaFold models were used).

Gray highlight indicates highest-ranking mutants.

Protein, position	Sequence	Experimental evidence	pT probability	Predicted kinase?	Alignment to Mph1	In solvent-exposed region?	Score
Mph1 T776	ENDSTEEAS	[20]	0.078 | 0.719	no prediction	100%	Yes	★★★★★
Mph1 T785	SSLETEDEE	[20]	0.845 | 0.623	CMGC (4/5)	100%	Yes	★★★★★
Cdc7 T484	DGESTDEDD	[74,76]	0.883 | 0.552	CMGC (2/5)	79.8%	Yes	★★★★★
Cdc45 T453	NNDDTDGEE	[74,77–80]	0.573 | 0.470	CMGC (3/5)	78.8%	Yes	★★★★★
Orc6 T227	ESDPTSEEE	[81]	0.842 | 0.783	CMGC (2/5)	76.8%	Yes	★★★★★
Mcm7 T811	GTMDTDQED	[77,82–88]	0.531 | 0.326	CMGC (1/5)	81.8%	Yes	★★★★★
Cdc6 T368	QVPLTPTTS	[81,89–91]	0.138 | 0.741	CMGC (4/5)	67.7%	Yes	★★★★☆
Cdc28 T169	LRAYTHEIV	[79,81,82,92–97]	0.357 | 1.000	CMGC (1/5)	64.6%	Yes	★★★★☆
Dbf4 T413	HADQTDEKN	[84]	0.506 | 0.789	AGC (1/5)	77.8%	Yes	★★★★☆
Mcm3 T868	TEVGTPRLP	[82,83]	0.958 | 0.893	CMGC (2/5)	64.6%	Yes	★★★★☆

Mcm5 T350	VTMFTEEEE		0.744 | 0.180	CMGC (3/5)	81.8%	Yes	★★★★☆
Mcm6 T9	FPADTPSSN	[83]	0.906 | 0.925	CMGC (3/5)	64.6%	Yes	★★★★☆
Orc2 T70	LKPKTPSKA	[81,98,99]	0.688 | 0.986	CMGC (4/5)	57.6%	Yes	★★★★☆
Orc2 T174	PEPATPSKK	[81,99–101]	0.850 | 0.960	CMGC (5/5)	60.6%	Yes	★★★★☆
Sld2 T168	SSTRTPTSS	[102]	0.516 | 0.860	CMGC (3/5)	72.7%	Yes	★★★★☆
Sld2 T241	SLQKTPSKD	[102]	0.703 | 0.789	CMGC (3/5)	69.7%	Yes	★★★★☆
Sld3 T442	SSPNTVETY	[103]	0.100 | 0.573	CMGC (1/5)	71.7%	Yes	★★★★☆
Sld3 T609	KRTVTPNKK	[104]	0.913 | 0.895	CMGC (4/5)	60.6%	Yes	★★★★☆
Cdc6 T432	NSDATIDES		0.830 | 0.180	CMGC (1/5)	75.8%	No	★★★☆☆
Cdc45 T189	DDEATDADE		0.339 | 0.920	CMGC (3/5)	69.7%	Yes	★★★☆☆
Cdc45 T195	ADEVTDEDE		0.124 | 0.904	CMGC (3/5)	73.7%	Yes	★★★☆☆
Clb5 T107	YNDRTAAEQ		0.056 | 0.632	CMGC (2/5)	66.7%	Yes	★★★☆☆
Dbf4 T424	LLNETETKE		0.025 | 0.514	CMGC (1/5)	74.7%	Yes	★★★☆☆
Mcm4 T9	SSSPTKEDN		0.172 | 0.861	no prediction	75.8%	Yes	★★★☆☆
Orc1 T247	ATDITDNED		0.092 | 0.691	CMGC (2/5)	73.7%	Yes	★★★☆☆
Orc4 T194	STTKTRNED		0.641 | 0.614	no prediction	80.8%	Yes	★★★☆☆
Pol2 T1222	LFEPTVEED		0.736 | 0.406	CMGC (2/5)	64.6%	Yes	★★★☆☆
Pol2 T1250	RNQLTNEED		0.462 | 0.523	CMGC (2/5)	74.7%	Yes	★★★☆☆
Dpb2 T567	EDVYTENDN		0.061 | 0.675	CMGC (1/5)	74.7%	No	★★☆☆☆

We postulated that a relevant origin binding protein that interacted with the Fkh1 FHA domain would contain a peptide(s) with sequence similarity to the known Fkh1 FHA-binding peptide in Mph1, and that this peptide would exist in a solvent-accessible and unstructured region (**Figure 2.2.a** and **Table 2.1**). The Mph1 peptide contains two threonine residues; each threonine residue, when phosphorylated, contributes to the Fkh1 FHA-Mph1 interaction *in vitro* [20]. A 2-hybrid assay identified additional amino acids adjacent to these T residues that also contribute to the Fkh1-FHA-Mph1 interaction. These additional residues are indicated by asterisks (*) in **Figure 2.2.a,** with the number of *'s indicating the extent of their contribution to the 2-hybrid interaction. We used this information, and additional criteria as described in the legend for **Table 2.1**, to identify similar peptides within the known essential origin-binding proteins, compiling a list of putative protein partners of the Fkh1 FHA domain. The four highest-ranking peptides are indicated in shaded gray and exist within the Cdc7, Cdc45, Orc6 and Mcm7 protein subunits. Given what is currently known about the mechanism of origin licensing and activation, the ability of Fkh1's FHA domain to enhance the concentration of any (or all) of these proteins near origins could enhance early origin activity. For reasons described below, this study focused on examining a potential interaction between the Fkh1 FHA domain and the Cdc7 peptide, the highest-ranking peptide by our criteria in Table 2.1.

2.3.3. A C-terminal pT peptide on Cdc7 is a strong candidate for a Fkh1-FHA binding interface that promotes the activity of early origins.

We were intrigued by the highest ranked putative Fkh1-FHA partner peptide in **Table 2.1**, the most C-terminal T-containing peptide within Cdc7, because there was other evidence in the literature that a Fkh1-DDK interaction was important (**Figure 2.2.b**) [74]. First, Cdc7 is the catalytic kinase subunit of the S-phase Dbf4-dependent kinase (DDK) that directly

phosphorylates dhMCM in the earliest S-phase activation step in origin function. Second, the DDK is one of the limiting S-phase origin activation factors, and its recruitment via a chromatin-mediated mechanism is implicated in enhancing the activity of one type of early origin, those associated with centromeres [52]. Third, biochemical and structural evidence indicates that this Cdc7 peptide binds directly to the FHA1 domain of the Rad53 checkpoint kinase, which monitors S-phase progression. The Rad53 FHA1 domain shares substantial similarity to the Fkh1 FHA domain: its structure guided our earlier homology model for the Fkh1 FHA domain [20]. A rigid superposition of the Rad53 FHA1 domain structure with the AlphaFold model of the Fkh1 FHA domain (**Figure 2.2.d**) shows substantial similarity even when the Fkh1 model is restricted from flexing to align better with the Rad53 structure. Fourth, Fkh1 was identified as a Dbf4-interacting protein in a 2-hybrid screen in which Dbf4 was used as bait [66]. Co-immunoprecipitation experiments also supported a Fkh1-Dbf4 interaction. However, relevant interaction regions were not mapped. Because Cdc7 forms a stable interaction with Dbf4 in the DDK, it seemed plausible that an interaction between this Cdc7 pT-containing peptide could be serving as a bridge that stabilized an interaction between Fkh1 and Dbf4. Therefore, we postulated that Fkh1 was interacting with the DDK, and an FHA-relevant interface in this interaction might be the Cdc7 C-terminal pT peptide shown in **Figure 2.2.b**.

2.3.4. Establishing a proximity labeling assay to examine Fkh1-FHA-dependent interactions *in vivo*.

To examine protein-protein interactions, I established a Fkh1-FHA-specific TurboID protein labeling system (**Figure 2.4.a**).

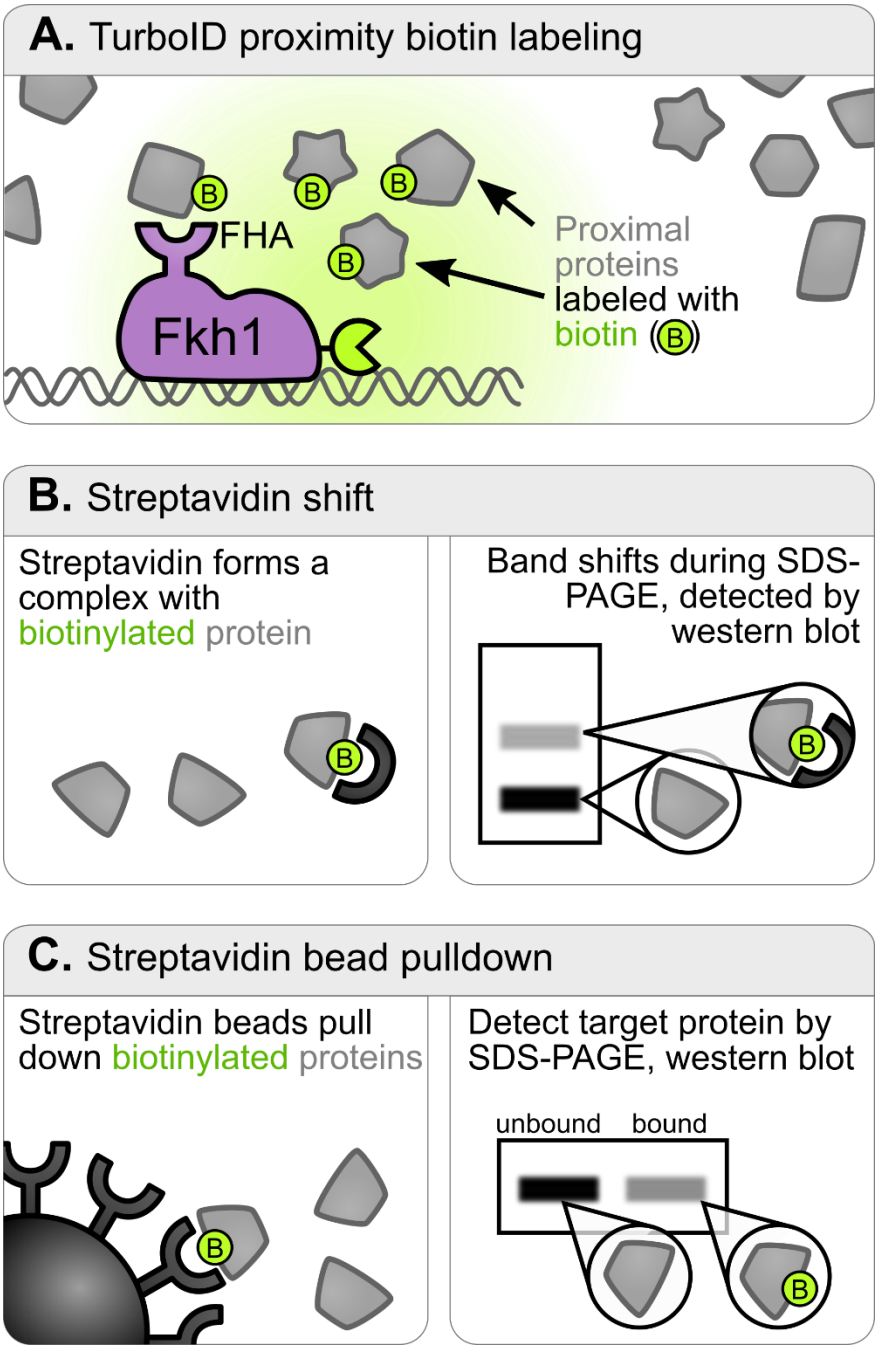
Figure 2.4. Schematic of the TurboID system.

A. Cartoon showing Fkh1-TurboID fusion protein labeling proximal proteins with biotin.

B. Schematic of streptavidin shift experiment in which streptavidin protein binds directly to biotinylated proteins, slowing their progress through an SDS-PAGE gel and resulting in a supershifted band that is detected by protein immunoblot.

C. Schematic of streptavidin bead pulldown in which biotinylated proteins are pulled down by streptavidin beads, and bead-bound and unbound proteins are detected by protein immunoblot.

Figure 2.4.


A. TurboID proximity biotin labeling
FHA
Fkh1
Proximal proteins labeled with biotin (B)
B. Streptavidin shift
Streptavidin forms a complex with biotinylated protein
Band shifts during SDS-PAGE, detected by western blot
C. Streptavidin bead pulldown
Streptavidin beads pull down biotinylated proteins
Detect target protein by SDS-PAGE, western blot
unbound
bound

I posited that this system would offer advantages over the 2-hybrid and co-IP methods in that TurboID would allow me to capture protein-protein interaction information in as close to a native state as possible, in contrast to the 2-hybrid assay which uses a non-native configuration. I also reasoned that TurboID might help with the detection of transient, phosphorylation-dependent interactions that might be harder to detect by other approaches, and more easily adapted to using cell cycle arrest protocols for yeast, which could help examine the cell cycle-regulated phases of the origin cycle.

In this system, a protein fusion of Fkh1 fused to TurboID is expressed in yeast cells. Labeling is triggered by the addition of exogenous biotin to the proliferating culture for a pulse of 5-60 minutes. A sample of yeast cells is then isolated, and proteins are extracted for analyses. Several approaches to detect specific proteins were used in this study. In one approach, termed the "streptavidin shift", a molar excess of streptavidin is added to a denatured protein extract. When bound to a biotinylated protein, the streptavidin slows the target protein's migration through a denaturing SDS-PAGE gel. The specific protein of interest is detected in a protein immunoblot, and I look for a band that appears at a larger molecular weight, indicative of biotinylation (**Figure 2.4.b**). This method works because of the high-affinity interaction between streptavidin and biotin. In a second approach, chromatin-bound proteins are enriched using either a crude chromatin pelleting step or a cleaner but more labor-intensive nuclear isolation step. The unbiotinylated proteins (unbound) versus biotinylated (bound) protein fractions are then analyzed for the target protein by protein immunoblotting (**Figure 2.4.c**). Alternatives include first enriching the protein of interest with an antibody and then probing a protein blot with streptavidin-HRP. I found that the most useful approach varied with the protein target under consideration. For example, the streptavidin shift method worked with FLAG-tagged Mph1 and

another protein, Fdo1 (see Chapter 3). However, this method was only minimally useful with Cdc7.

To adapt this system for assessing Fkh1-FHA-dependent interactions *in vivo*, two different TurboID fusion proteins were expressed in yeast. One expressed the wild-type Fkh1 protein, and the other, the Fkh1^{R80A} mutant protein. Each protein was fused to the TurboID enzyme at its C-terminus (**Figure 2.5.a**). I wanted the fusions to be expressed at near native levels, and so the native *FKH1* promoter and terminator regions were used to drive the expression of these fusion genes. A protein immunoblot using an anti-Fkh1 protein to detect the level of fusion protein in yeast showed that the wild-type and mutant Fkh1-TurboID proteins were expressed at levels similar to each other and to the native Fkh1 protein, fulfilling a key requirement for this system (**Figure 2.5.b**). Another goal for this system was to enhance the ability to define biologically relevant partners of the Fkh1-FHA domain. Therefore, the Fkh1-TurboID fusions were expressed in a *fkh1Δ* cell so that the only source of Fkh1 would be the expressed fusion proteins.

The proteins also needed to recapitulate a known biological function of Fkh1. For this goal, we used a simple colony-based assay to assess Fkh1 function. Specifically, in the absence of both Fkh1 and Fkh2 (i.e. in *fkh1Δ fkh2Δ* double mutant cells), the yeast cell cycle is so defective that the cells grow in a pseudohyphal-like state that results in "rough" colonies that "scar" the agar substrate on which they grow (**Figure 2.5.c**). The scarring is detected by simply washing the colonies off the agar plate with water. Expression of Fkh1 from a plasmid fully rescues the defect and Fkh1^{R80A} partially rescues it [20]. We observed that plasmid expression of Fkh1-TurboID rescued the colony morphology phenotypes of roughness and agar penetration similarly to the plasmid-expressed Fkh1.

Figure 2.5. Construction and validation of Fkh1-TurboID system to identify putative FHA-specific interacting proteins *in vivo*.

A. Schematic of TurboID expression constructs. TurboID biotin ligase and Fkh1-TurboID fusion proteins are each flanked by the untranslated region (UTR) upstream and downstream of the *FKH1* gene. This construction means that Fkh1-TurboID construct expression is driven by the native *FKH1* promoter and terminator and will be expressed more similarly to endogenous *FKH1*.

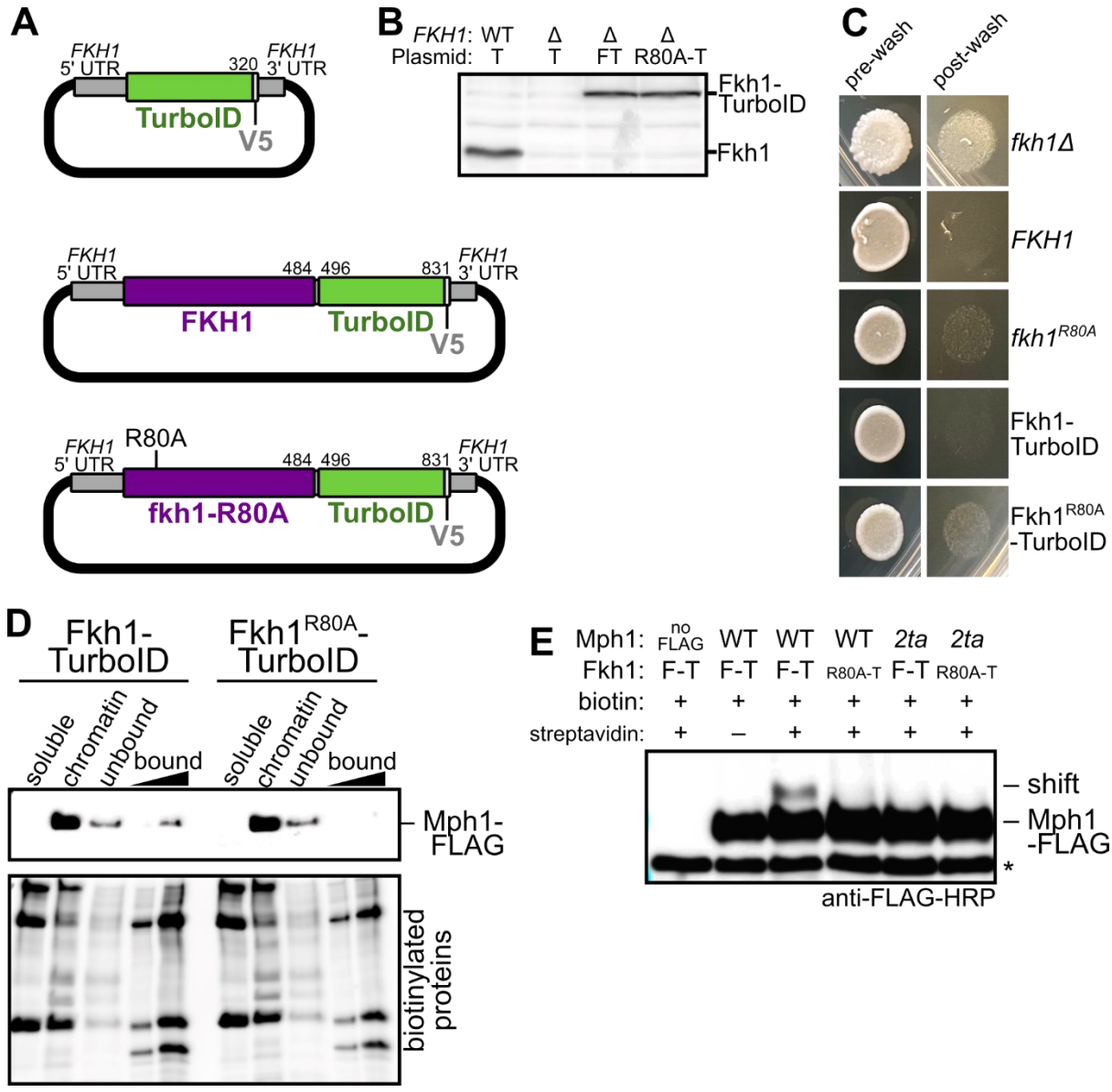
B. Protein immunoblot of whole protein extracts showing expression of endogenous Fkh1 protein in a wild-type yeast strain compared to expression of Fkh1-TurboID fusion proteins in a *fkh1Δ* yeast strain.

C. Morphology assay in which each *S. cerevisiae* strain is spotted on a plate to observe rough colony morphology (pre-wash). Yeast are then removed to observe scarring of the agar beneath (post-wash).

D. Protein immunoblots showing streptavidin bead pulldown of chromatin-enriched samples. Top: anti-FLAG antibody detects Mph1-FLAG. Bottom: streptavidin-HRP detects biotinylated proteins.

E. Protein immunoblot showing streptavidin shift experiment performed on whole cell extracts. Mph1-FLAG and shifted Mph1-FLAG band are detected using anti-FLAG-HRP antibody. Asterisk marks a nonspecific anti-FLAG cross-reactive band present in all strains.

Figure 2.5.


A
FKH1 5' UTR
320
FKH1 3' UTR
TurboID
V5
FKH1 5' UTR
484 496
831
FKH1 3' UTR
FKH1
TurboID
V5
FKH1 5' UTR
R80A
484 496
831
FKH1 3' UTR
fkh1-R80A
TurboID
V5
B
FKH1: WT Δ Δ Δ
Plasmid: T T FT R80A-T
Fkh1-TurboID
Fkh1
C
pre-wash
post-wash
fkh1Δ
FKH1
fkh1R80A
Fkh1-TurboID
Fkh1R80A-TurboID
D
Fkh1-TurboID
Fkh1R80A-TurboID
soluble
chromatin
unbound
bound
soluble
chromatin
unbound
bound
Mph1-FLAG
biotinylated proteins
E
Mph1: no FLAG WT WT WT 2ta 2ta
Fkh1: F-T F-T F-T R80A-T F-T R80A-T
biotin: + + + + + +
streptavidin: + – + + + +
– shift
– Mph1-FLAG
*
anti-FLAG-HRP

The plasmid-expressed Fkh1^{R80A} and Fkh1^{R80A}-TurboID only partially rescued these same morphological phenotypes, and to similar extents. Together, these observations provided evidence that the TurboID expression constructs were recapitulating at least this aspect of native Fkh1 behavior and therefore were more likely to be interacting with biologically relevant FKH binding sites and protein partners.

Next, I examined the ability of the system to label Mph1, the characterized protein partner of Fkh1, in the expected FHA-dependent manner. First, after enriching for chromatin and chromatin-associated proteins, I used a version of the streptavidin pulldown assay (see **Figure 2.4.c**). While this chromatin fractionation approach produced proof-of-principle data, the amount of Mph1 protein signal recovered was low, making this approach less robust than ideal (**Figure 2.5.d**). Therefore, I also tried the streptavidin shift approach that exploited crude cell extracts and avoided the sometimes-finicky chromatin fractionation step (**Figure 2.5.e**). This approach was more robust in terms of reproducibility and allowed me to demonstrate that I could detect the known Fkh1-Mph1 interaction in proliferating cells and observe its dependence on both the Fkh1-FHA domain and the two threonine residues within the relevant Mph1 peptide shown in **Figure 2.2.a**.

2.3.5. Assessing Fkh1-Cdc7 interactions *in vivo* using the Fkh1-TurboID system.

The first indication that Fkh1 interacted with Cdc7 *in vivo* was generated by a Fkh1-TurboID experiment using an extract of protein generated from a proliferating yeast culture and the streptavidin shift method described in **Figure 2.4.b**. Specifically, I prepared yeast chromatin extracts from relevant yeast expressing a Cdc7-FLAG fusion protein, incubated denatured proteins with excess streptavidin, separated the protein mixture by SDS-PAGE and

transferred the proteins to a membrane. The membrane was then probed with anti-FLAG-HRP, a reagent in which the horseradish peroxidase used for chemiluminescent imaging is directly coupled to an anti-FLAG antibody (**Figure 2.6.a**). I observed a weak streptavidin- and Fkh1-FHA-dependent supershift of Cdc7-FLAG, indicating that a minute fraction of Cdc7-FLAG was in close proximity (≤10 Å) to Fkh1. Unfortunately, the reagent that allowed for this detection became unavailable. In addition, the Cdc7-FLAG epitope on Cdc7 caused a growth defect, raising additional concerns.

Next, I assessed a Fkh1-Cdc7 interaction using an anti-Cdc7 polyclonal against native Cdc7 in relevant yeast strains. For these experiments, I had to enrich for the Cdc7 protein by making a nuclear extract (**Figure 2.6.b**). I then separated the nuclear extract with streptavidin magnetic beads into a bound fraction (representing biotinylated proteins) and an unbound fraction (representing non-biotinylated proteins) and analyzed these fractions by SDS-PAGE and protein immunoblotting. The blot was cut in half so that I could examine Mph1-FLAG (MW=114 kDa) and Cdc7 (MW=58.3 kDa) with different antibodies. Recovery of biotinylated proteins and total proteins on the protein blot were assessed with streptavidin-HRP and Ponceau S staining, respectively (**Figure 2.6.d-f**). I observed that Fkh1-TurboID, but not TurboID alone, resulted in a weak signal of biotinylated Cdc7, while the signal for Mph1, my control protein, was more robust (**Figure 2.6.c**). A limitation to this interpretation was that the biotinylated proteins in the extract had not been quantitatively enriched – some biotinylated proteins remained in the unbound fraction. The experiment was repeated with Fkh1^{R80A}-TurboID. In this experiment, a robust biotinylated Cdc7 signal was observed with Fkh1-TurboID, but not with Fkh1^{R80A}-TurboID (**Figure 2.6.d**).

Figure 2.6. Evidence that Fkh1 and Cdc7 interact *in vivo*.

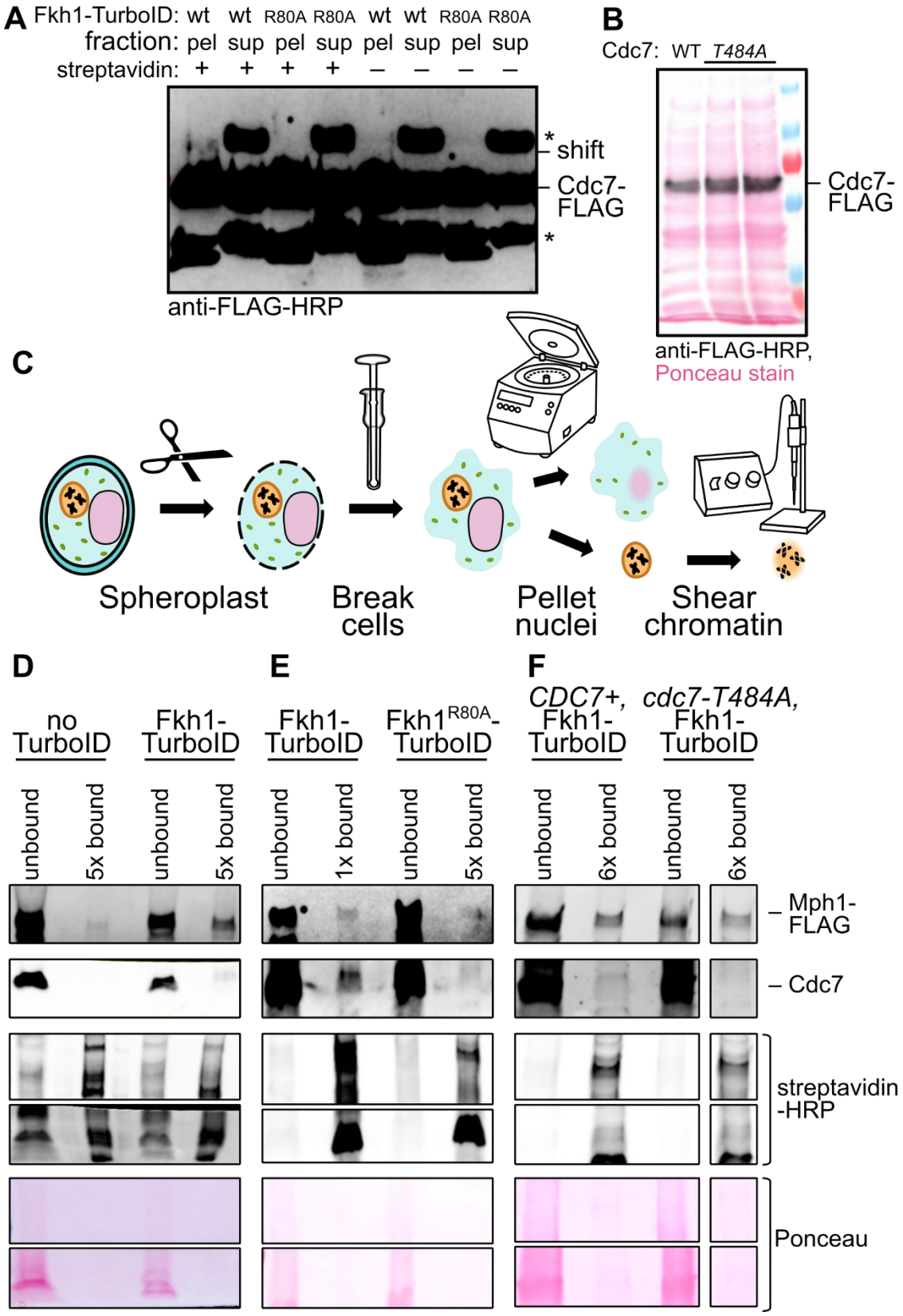
A. Protein immunoblot showing streptavidin shift experiment performed on chromatin-enriched samples. Cdc7-FLAG and faint shifted Cdc7-FLAG band are detected using anti-FLAG-HRP antibody. Asterisks mark nonspecific anti-FLAG cross-reactive bands.

B. Schematic showing general steps to extract nuclei from *S. cerevisiae* cells and release chromatin-bound protein.

C. Protein immunoblot comparing expression of Cdc7-FLAG and mutant Cdc7^{T484A}-FLAG protein. Cdc7-FLAG bands are detected by anti-FLAG-HRP (black). As a loading control, total protein is detected with Ponceau S stain (pink).

D, E, F. Protein immunoblots showing streptavidin bead pulldown of nuclear extracted samples. Generally, unbound samples containing 1x cell equivalents were loaded, while bound sample lanes contain more concentrated samples (5x-6x cell equivalents compared to their corresponding unbound sample). Top to bottom: Mph1-FLAG is detected by anti-FLAG antibody, Cdc7 protein is detected by anti-Cdc7 antibody, biotinylated proteins are detected by streptavidin-HRP, total protein is stained with Ponceau S stain.

Figure 2.6.


A
Fkh1-TurboID: wt wt R80A R80A wt wt R80A R80A
fraction: pel sup pel sup pel sup pel sup
streptavidin: + + + + – – – –
* shift
Cdc7-FLAG
*
anti-FLAG-HRP
B
Cdc7: WT T484A
Cdc7-FLAG
anti-FLAG-HRP, Ponceau stain
C
Spheroplast
Break cells
Pellet nuclei
Shear chromatin
D
no TurboID
Fkh1-TurboID
unbound
5x bound
unbound
5x bound
E
Fkh1-TurboID
Fkh1R80A-TurboID
unbound
1x bound
unbound
5x bound
F
CDC7+, Fkh1-TurboID
cdc7-T484A, Fkh1-TurboID
unbound
6x bound
unbound
6x bound
– Mph1-FLAG
– Cdc7
streptavidin-HRP
Ponceau

However, a qualification to the interpretation that the Fkh1-Cdc7 interaction is FHA-dependent was that a reduced amount of biotinylated protein in general was recovered in the Fkh1^{R80A}-TurboID experiment, which necessitated loading 5x more bound protein onto the gels. Nevertheless, I noted that the ratio of Cdc7 signal to Mph1 signal was higher for the Fkh1-TurboID experiment than for the Fkh1^{R80A}-TurboID experiment, which was consistent with the Fkh1-Cdc7 interaction's dependence on the Fkh1-FHA domain *in vivo*. Finally, I performed the experiment with Fkh1-TurboID in *CDC7* or *cdc7-T484A* mutant cells (**Figure 2.6.e**). In this experiment, the low signals made interpreting these data challenging. No difference in expression level was observed between wild-type Cdc7 and the Cdc7^{T484A} mutant protein, so variations in expression level could not explain the low signal (**Figure 2.6.f**).

2.3.6. The pT residues on the Mph1 and Cdc7 peptides promoted their binding to both the Fkh1 FHA and Rad53 FHA1 domains *in vitro*.

Because it was so challenging to glean robust information about the Fkh1-Cdc7 interaction using the *in vivo* TurboID system, and because that system could not address the directness of this interaction regardless, I turned to an *in vitro* FHA-peptide binding assay. For these experiments, the Mph1 and Cdc7 peptides indicated in **Figure 2.2** were commercially synthesized in either the phosphorylated threonine (pT) or non-phosphorylated forms as shown in **Figure 2.7.b**. Each of these peptides included an N-terminal biotin moiety. Rad53-FHA1 and Fkh1-FHA are similar domains with a decent amount of alignment (see **Figure 2.2.c**). Both of these FHA domains are involved in DNA replication, with Fkh1's FHA domain promoting early origin use (**Figure 2.1**) and Rad53's FHA1 domain monitoring S-phase progression as a major S-phase checkpoint kinase. Thus, it seemed plausible that these two FHA domains might share some protein targets. Indeed, the DDK interacts with Rad53 through multiple protein-protein

contacts [74,105–107]. In fact, the pT-containing Cdc7 peptide we are examining was shown to be a direct partner target of Rad53's FHA1 domain, with a co-crystal structure of Rad53's FHA1 domain in complex with this peptide [74] (see **Figure 2.2.c**). Nevertheless, the Mph1-pT and the Cdc7-pT peptides in **Figure 2.7.a** are currently understood as canonical peptide partners of Fkh1-FHA and Rad53-FHA1 respectively.

The peptide-binding assay is outlined in **Figure 2.7.b**. The phosphorylated forms of both Mph1 and Cdc7 peptides bound ~4x more efficiently to either the Fkh1-FHA or Rad53-FHA1 domain compared to the non-phosphorylated versions of these peptides. Specifically, each of these peptides interacted in a pT-dependent manner with purified Fkh1, consistent with their recognition by the Fkh1-FHA domain (**Figure 2.7.c**). When the analogous peptide-binding experiments were performed with purified Rad53-FHA1 domain, I also observed pT-enhanced binding of both the Mph1 and Cdc7 peptides to the Rad53-FHA1 domain (**Figure 2.7.d**).

2.3.7. Rad53's FHA1 and Fkh1's FHA domains bind similar pT peptides differently.

The pT dependence of the interactions between the Mph1 and Cdc7 peptides with both Fkh1-FHA and Rad53-FHA1 proteins supported the idea that these FHA domains shared protein partners. However, another defining criterion of a canonical FHA-pT peptide interaction is the dependence on a conserved arginine residue located within a flexible peptide-binding loop that binds the pT residue. Other residues contribute to stabilizing this pT residue, either by contacting it directly or contacting adjacent residues in the peptide. We noted that although the Mph1 and Cdc7 peptides were similar, some contacts between the Rad53-FHA1 domain and the relevant Cdc7 peptide are formed by residues within Rad53-FHA1 that were not conserved with Fkh1-FHA (see **Figure 2.2.c-e**).

Figure 2.7. Both Fkh1 and Rad53 FHA domains can bind Mph1- and Cdc7-derived peptides in a pT-dependent manner.

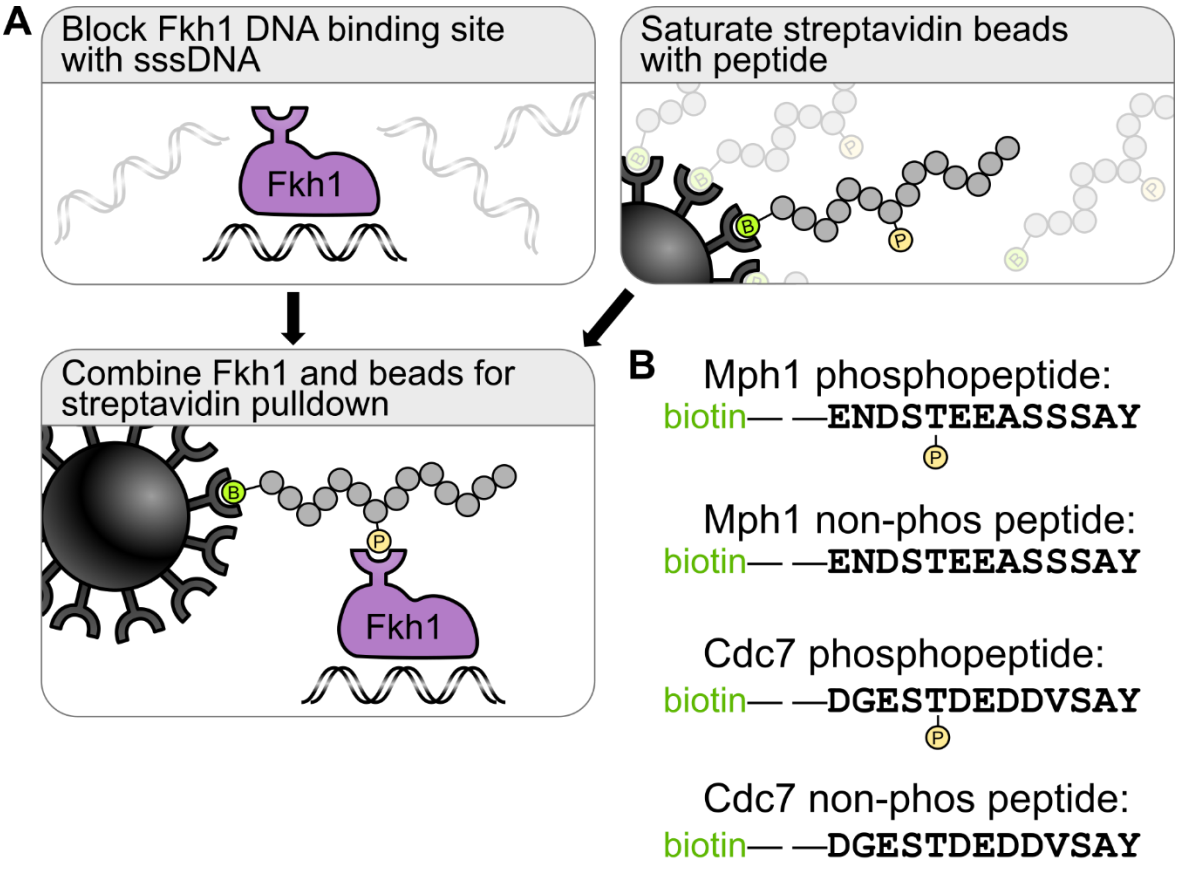
A. Schematic of *in vitro* peptide pulldown experiment. Purified recombinant Fkh1 or Rad53-FHA1 proteins are incubated in binding buffer plus sssDNA with peptide-saturated streptavidin beads. Both the bound protein and unbound protein fractions are detected by protein immunoblot.

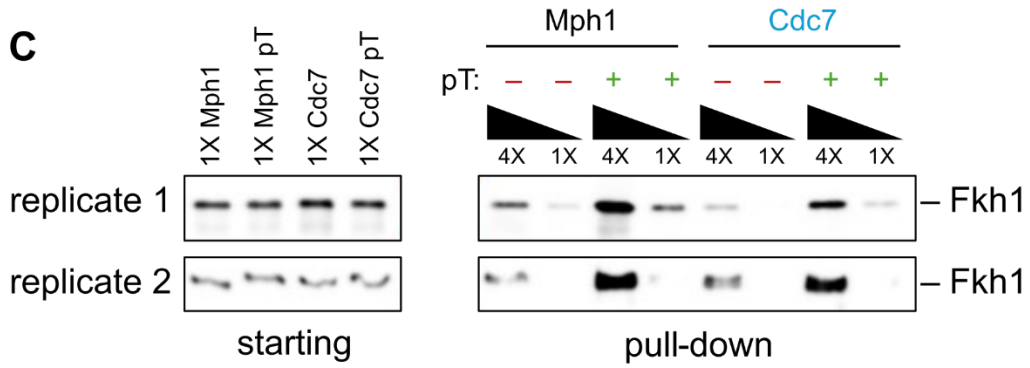
B. Sequences of Mph1- and Cdc7-derived phosphopeptides.

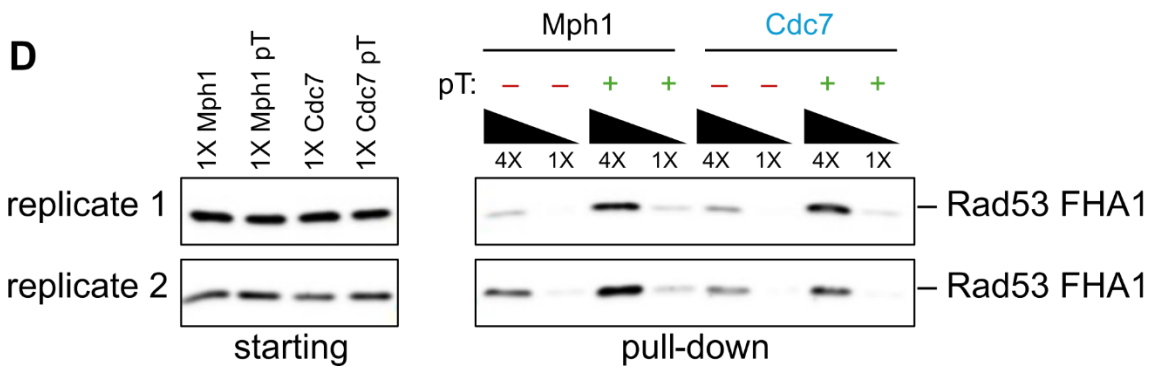
C. Protein immunoblots showing Fkh1 pulled down by either Mph1- or Cdc7-derived peptides. Left: a sample of Fkh1 protein that served as the starting material for each pulldown experiment is shown. Right: peptides either contain the phosphorylated threonine shown in part B (+) or have a non-phosphorylated threonine residue in that position (–). Two replicate experiments are shown.

D. As in part C, except the protein assessed is Rad53's FHA1 domain.

Figure 2.7.


A
Block Fkh1 DNA binding site with sssDNA
Fkh1
Saturate streptavidin beads with peptide
Combine Fkh1 and beads for streptavidin pulldown
Fkh1
B
Mph1 phosphopeptide:
biotin— —ENDSTEEASSSAY
Mph1 non-phos peptide:
biotin— —ENDSTEEASSSAY
Cdc7 phosphopeptide:
biotin— —DGESTDEDDVSAY
Cdc7 non-phos peptide:
biotin— —DGESTDEDDVSAY


C
1X Mph1
1X Mph1 pT
1X Cdc7
1X Cdc7 pT
Mph1
Cdc7
pT: – – + + – – + +
4X 1X 4X 1X 4X 1X 4X 1X
replicate 1
replicate 2
– Fkh1
– Fkh1
starting
pull-down


D
1X Mph1
1X Mph1 pT
1X Cdc7
1X Cdc7 pT
Mph1
Cdc7
pT: – – + + – – + +
4X 1X 4X 1X 4X 1X 4X 1X
replicate 1
replicate 2
– Rad53 FHA1
– Rad53 FHA1
starting
pull-down

Therefore, I used the peptide-binding approach to test whether the conserved pT-contacting arginine residues, R80 in Fkh1-FHA and R70 in Rad53-FHA1, were required for these FHA domains to bind pT-containing versions of the relevant Mph1 and Cdc7 peptides (**Figure 2.8**). Fkh1's ability to bind to the pT-Mph1 peptide required that the FHA domain contain R80, as a Fkh1^{R80A} mutant showed 4-fold reduced binding to a pT-Mph1 peptide (**Figure 2.8.a**), mirroring the pT-dependence of this interaction observed in **Figure 2.7.c**. However, I did not observe an analogous outcome with pT-Cdc7 peptide binding to Fkh1 (**Figure 2.8.b**). Specifically, the Fkh1^{R80A} mutant protein bound the pT-Cdc7 peptide as well as, if not better than, Fkh1. Thus, while Fkh1's interaction with the pT-Cdc7 peptide was pT-enhanced, this interaction did not depend on the key R80 residue that mediates the verified Fkh1-Mph1 interaction. To further challenge this outcome, I performed the analogous experiments with a purified Rad53-FHA1 domain. I observed that the Rad53-FHA1 interaction with the pT-Cdc7 peptide was enhanced by the conserved R70, as expected based on previous structural and biochemical work on this canonical FHA-pT peptide interaction (**Figure 2.8.c**) [74]. However, the mutant Rad53^{R70A}-FHA bound the pT-Mph1 peptide slightly better than the wild-type Rad53-FHA (**Figure 2.8.d**). Thus, while the interaction of Rad53-FHA1 with pT-Mph1 was pT-enhanced, this interaction did not depend on the key R70 residue that mediates the verified Rad53-Cdc7 interaction.

2.3.8. *cdc7-T484A* mutant yeast show delayed progression through S-phase.

The lack of cross-substrate specificity observed for binding of Fkh1-FHA to the pT-Mph1 peptide and Rad53-FHA1 to the pT-Cdc7 peptide, at least in terms of the requirement for the pT-binding arginine residue within each of the FHA domains here, suggested that the Cdc7 epitope containing pT484 might be most critical *in vivo* for regulating interactions with Rad53, the checkpoint kinase that monitors S-phase progression.

Figure 2.8. Only the "canonical" FHA-pT peptide interactions show a requirement for the conserved anchoring arginine in the FHA domain.

A. Peptide pulldown experiments testing an Mph1-derived peptide's ability to pull down Fkh1 vs. Fkh1^{R80A}. Fkh1 protein is detected using anti-Fkh1 polyclonal antibody.

B. As in part A, except using Cdc7-derived peptide.

C. Peptide pulldown experiments testing an Mph1-derived peptide's ability to pull down a truncated protein containing Rad53's FHA1 domain, or the same protein containing a R70A mutation. These Rad53-derived proteins are FLAG tagged and detected using anti-FLAG-HRP.

D. As in part C, except using Cdc7-derived peptide.

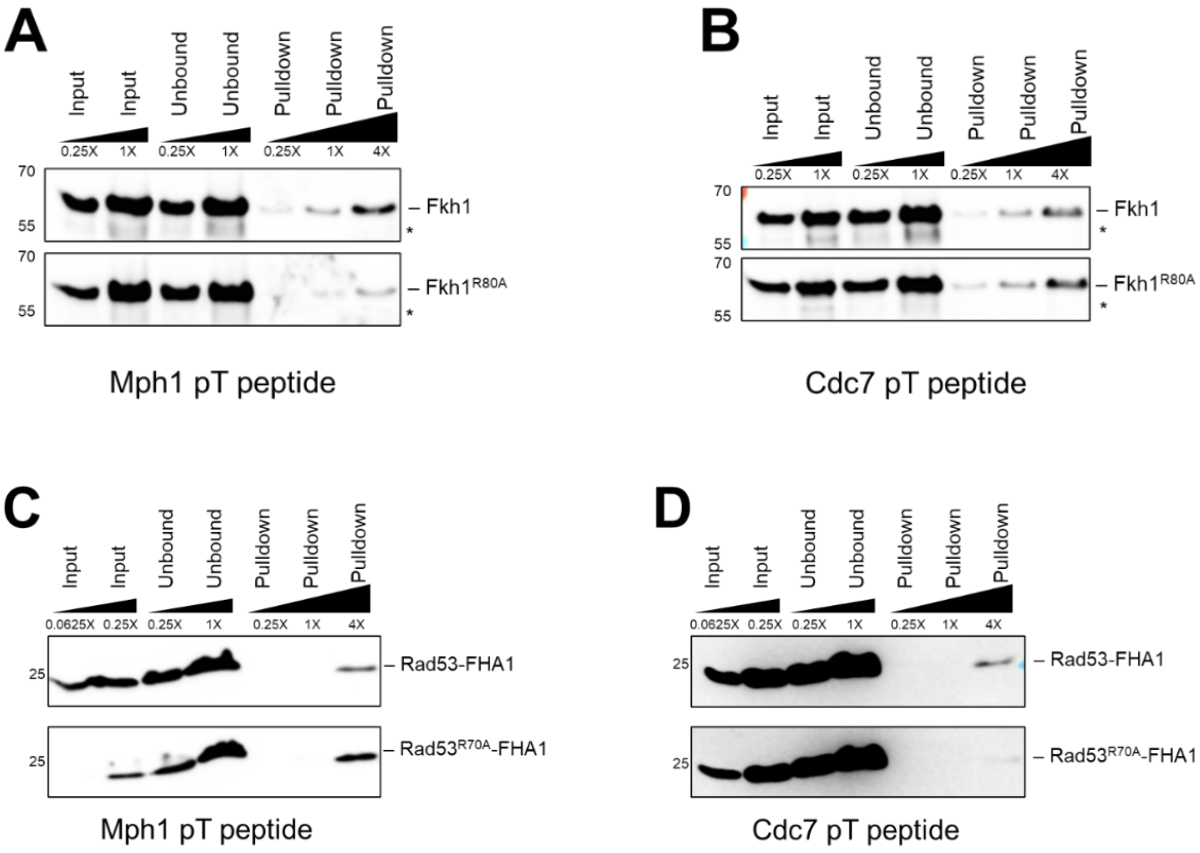
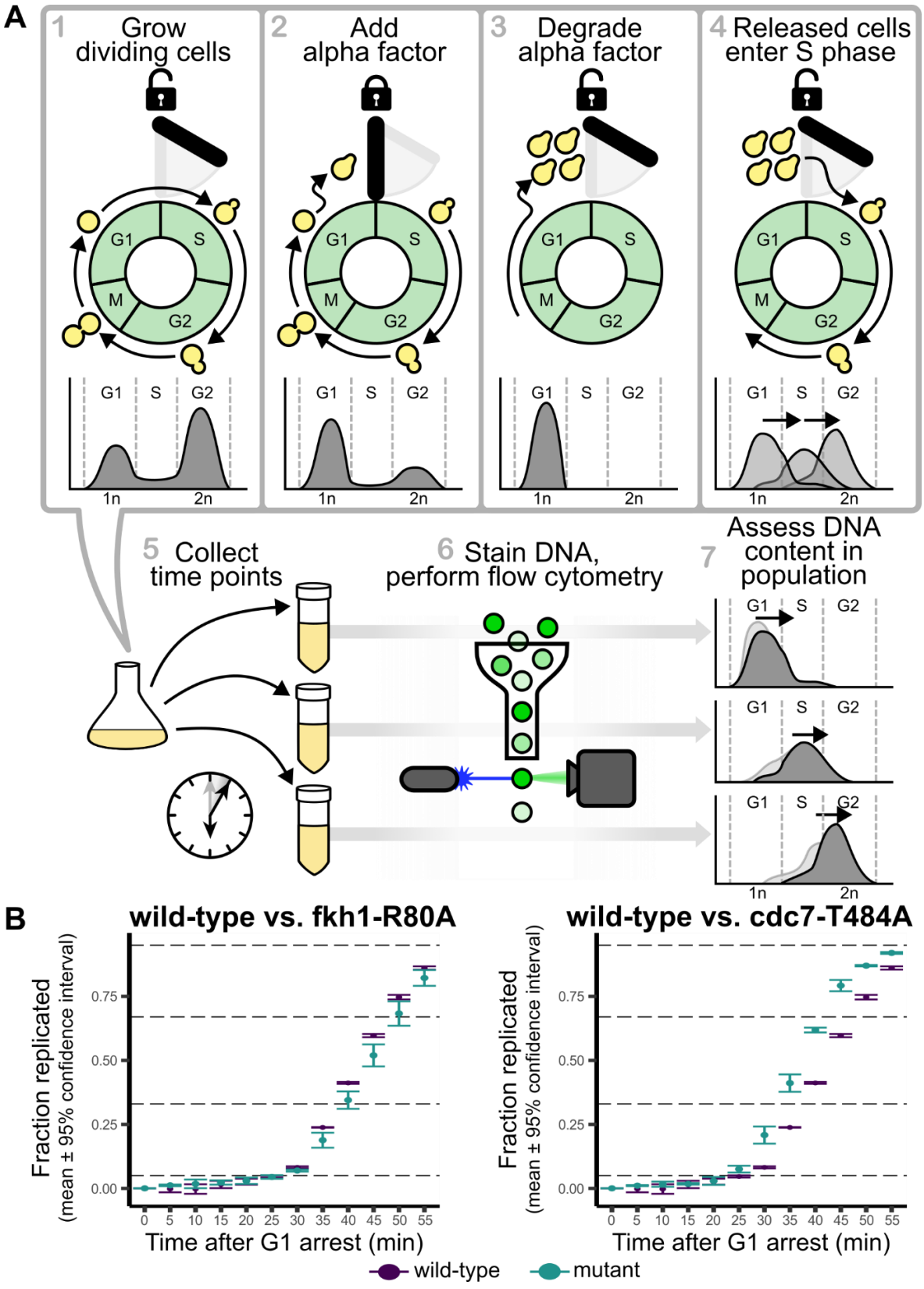


Figure 2.9. Flow cytometry on yeast cells progressing through synchronized S-phase.

A. Experimental schematic. Panels 1-4 depict "shmoo"-shaped cells arrested in G1-phase by alpha factor, which are then allowed to progress through S-phase upon degradation of alpha factor. Panels 5-6 depict the collection of S-phase samples and flow cytometry using a fluorescent DNA stain. Panel 7 shows mock flow cytometry histograms of populations progressing from 1n DNA content (before replication) to 2n DNA content (after replication).

B. S-phase profiles showing the fraction replicated at each time point, where 1n DNA content is 0% replicated and 2n DNA content is 100% replicated. Left: wild-type strain is compared with *fkh1-R80A* strain. Right: wild-type strain is compared with *cdc7-T484A* strain.

Figure 2.9.


A
1 Grow dividing cells
2 Add alpha factor
3 Degrade alpha factor
4 Released cells enter S phase
G1
S
G2
M
1n
2n
5 Collect time points
6 Stain DNA, perform flow cytometry
7 Assess DNA content in population
B
wild-type vs. fkh1-R80A
wild-type vs. cdc7-T484A
Fraction replicated
(mean ± 95% confidence interval)
0.00
0.25
0.50
0.75
0 5 10 15 20 25 30 35 40 45 50 55
Time after G1 arrest (min)
wild-type
mutant

To address this idea, cell cycle arrest and release experiments were performed with wild-type, *fkh1-R80A* and *cdc7-T484A* haploid cells such that S-phase progression could be monitored at high resolution while cells complete a synchronous S-phase (**Figure 2.9**). These experiments are necessary prerequisites to Sort-seq experiments in progress. Specifically, haploid cells are arrested in G1-phase and then released from arrest so that they can enter and proceed through S-phase synchronously (**Figure 2.9.a**). Progression through S-phase is measured by measuring DNA content by flow cytometry. Two independent biological replicates for each arrest and release experiment were performed, and the median value and variation are indicated in box-and-whiskers format. Data for the mutant strains *fkh1-R80A* and *cdc7-T484A* are shown in teal and are plotted alongside the data for wild-type cells (in purple) so that a mutant-to-wild-type comparison can be made (**Figure 2.9.b**). If the *cdc7-T484A* mutant explained the requirement for Fkh1's FHA domain in early origin function, the expectation was that the *cdc7-T484A* and *fkh1-R80A* mutants would produce similar DNA replication phenotypes. While higher-resolution Sort-seq data are required to address this issue definitively, these flow cytometry experiments can provide information on this point. In particular, the *fkh1-R80A* and *cdc7-T484A* mutant alleles have opposing effects on S-phase progression. While the *fkh1-R80A* mutant cells exhibit delayed S-phase progression in this assay, a phenotype that would be predicted based on the effect this mutant has on DNA replication origins (**Figure 2.1.b**), the *cdc7-T484A* mutant cells behave in the opposite fashion, showing an advanced in S-phase progression as measured by DNA content. This phenotype is more consistent with a defect in Rad53-mediated S-phase surveillance mechanisms than with Fkh1-mediated stimulation of early origin function.

2.4. Discussion

Once we learned that the canonical pT-specific protein binding function of Fkh1's FHA domain was required for promoting the function of early origins, the next critical question was obvious: Which pT-containing protein partner(s) explain the requirement for the Fkh1-FHA domain for early origin function? To answer this question, relevant direct Fkh1 FHA-dependent partner proteins must be identified. In addition, these partner proteins' pT-containing epitopes used to bind Fkh1's FHA domain must also be defined. This second step is critical: this knowledge allows us to generate a precise allele in the partner protein that abolishes the Fkh1-FHA interaction but leaves the partner protein's function otherwise intact. Such an allele is required to conclude that a defined protein-protein interaction mechanism is causally relevant to a specific molecular phenotype or process. It was this approach that allowed us to define the Fkh1-Mph1 interaction and explain the role of Mph1 in mating-type switching [20]. Importantly, while the relevant *MPH1* allele, *mph1-2TA*, acted equivalently to an *mph1Δ* (null) allele in mating-type switching, *mph1-2TA* acted equivalently to wild-type *MPH1* for other *mph1Δ* phenotypes, such as sensitivity to DNA damaging agents or an accelerated mutation rate [20].

2.4.1. FHA-pT peptide-specific binding mechanisms showed substantial dependence on pT-adjacent residues that remain poorly understood.

In vitro peptide arrays have been used to define the rules by which distinct FHA domains recognize their protein partners, but beyond the pT residue and an acidic residue at the +1 to +3 position, there is not enough information to make confident predictions regarding protein partners [108]. Part of the challenge is that both the FHA domain and its interacting partners are probably quite flexible, such that examining interactions with relatively short peptides like those

used on arrays limits the amount of binding surface that can be explored or that might be relevant *in vivo*.

Our study approached this issue from a much less comprehensive but more biologically driven direction that led us to conclude that there are substantial differences in FHA-pT binding mechanisms that involve pT-adjacent residues even between functionally and structurally similar FHA domains. Specifically, based on this biological reasoning, we addressed whether the known FHA-dependent Fkh1-Mph1 interaction or a postulated FHA-dependent Fkh1-Cdc7 interaction could account for the requirement for Fkh1's FHA domain in early origin function. We concluded that neither interaction could explain the Fkh1-FHA domain's ability to enhance early origin function based on *in vivo* experiments. However, during the course of our study, we learned that a specific C-terminal peptide on Cdc7 (containing T484) showed pT-specific binding to both the Fkh1-FHA and the Rad53-FHA1 domains, but the Cdc7-T484 peptide only showed a dependence on the conserved pT-anchoring arginine when binding to the Rad53-FHA1 domain. Conversely, the Mph1-T776 peptide, defined as a binding target for Fkh1-FHA domain, also showed pT-specific binding to both the Fkh1-FHA and the Rad53-FHA1 domains, but only showed a dependence on the conserved pT-anchoring arginine when binding to the Fkh1-FHA domain. Thus, despite the acidic nature and similarities of each pT peptide we examined, as well as the similarities of each FHA domain in this study, we observed differences in the peptide-binding mechanism between these two FHA domains for peptides derived from native substrates.

Given their common biological roles in genome duplication and stability, it was reasonable to think that the Fkh1-FHA domain and the Rad53-FHA1 domain would each recognize the Cdc7-pT484 peptide via a canonical FHA-dependent mechanism. Specifically, Cdc7 is the kinase subunit of the Dbf4-dependent kinase (DDK). Regulation of the DDK is a key

control point for both origin function (a role of Fkh1), and checkpoint control (a role of the Rad53 checkpoint kinase). Thus, because experimental evidence indicates that both of these FHA-domain containing proteins bind the DDK, it was reasonable to postulate that the Cdc7-T484 peptide was relevant to pT-dependent contacts with both FHA domain proteins [66,74]. However, our *in vitro* data suggest that the Cdc7-T484 peptide is more relevant to DDK interactions with Rad53, in terms of the canonical FHA-dependent binding mechanism. This canonical mechanism is likely to be the relevant *in vivo* mechanism because a pT-binding mechanism can be regulated by phosphorylation, meaning it can respond to the highly regulated events of S-phase checkpoint blocks.

While the Fkh1-DDK interaction remains poorly characterized, the Rad53-DDK interaction has been extensively investigated [74,107]. This interaction occurs via multiple protein-protein contacts, with the biochemically characterized interaction between Rad53's FHA1 domain and the Cdc7-pT484 epitope only accounting for a fraction of the stabilizing forces in the Rad53-DDK complex. In fact, to the best of our knowledge, phenotypic characterization of *cdc7-T484A* mutant yeast has not been reported, likely because the expectation was that *cdc7-T484A* would only produce a weak checkpoint defect, if it had any phenotype at all. In addition, because the interface between FHA domain and phosphothreonine-containing peptide is usually only one factor contributing to protein complex stability, it was not necessarily expected that the differences between Rad53-FHA1- and Fkh1-FHA-peptide binding observed in my experiments would even be biologically necessary. For example, in response to S-phase entry signals, phosphorylation of Cdc7's T484 residue, along with many other molecular events, could create a situation where Fkh1 recruitment of the DDK to early origins is favored. However, in response to DNA replication stressors, such as depletion of dNTPs, phosphorylation

of Cdc7's T484 could promote Rad53 binding to the DDK. Thus, given what we currently know, it was not hard to envision that the context of the larger molecular process in S-phase would have enough of an impact on controlling DDK recruitment to the relevant location that the differences in FHA-peptide binding that we observed would be less consequential.

2.4.2. Important but challenging: *in vivo* experiments to causally link specific protein-protein interactions to specific cellular processes.

A major goal of my work was to assign specific FHA-pT epitope interactions to discrete molecular processes that occur in yeast cells. We used the lab's previous study that assigned the Fkh1-FHA-Mph1 interaction to the role of controlling mating-type switching as a guide. In this study, I sought to define the Fkh1-FHA-dependent protein partner interactions that could explain why Fkh1's FHA domain, as shown by assessment of the defective *fkh1-R80A* mutant allele, was required to promote early origin function. The role of Fkh1's FHA domain in early origin function was revealed by examining the effect a *fkh1-R80A* allele had on origin activity using Sort-seq. Here we used Sort-seq to examine the effect of the *mph1-2TA* allele on origin activity. In the simplest case, if the Fkh1-Mph1 interaction explained the role of the Fkh1-FHA domain, the *mph1-2TA* allele should show altered early origin activity in the same way that the *fkh1-R80A* allele does. However, we did not observe this outcome. In fact, the *mph1-2TA* allele had no obvious effect on origin function. The simple conclusion is that the Fkh1-Mph1 interaction functions to promote HR and is not relevant to early origin function. However, it is still possible that the Fkh1-Mph1 interaction contributes to early origin function, but that it is only one of multiple Fkh1-protein interactions to do so. If multiple of Fkh1's protein partners contribute additively to early origin function, obtaining *in vivo* experimental evidence might require engineering a yeast strain with multiple mutant alleles in protein partners that each contribute

slightly to early origin function; reliably validating the slight contribution of each of these partners would be a challenging task.

Several lines of evidence discussed in this report made the Cdc7-pT484 peptide an attractive putative binding interface for the Fkh1-FHA domain. As discussed above, however, the *in vitro* peptide binding assays supported the idea that this peptide was probably more relevant to Rad53-DDK interactions than to Fkh1-DDK interactions. Nevertheless, it was important to examine the phenotype of the *cdc7-T484A* mutant allele. Using our flow cytometry assay, a prerequisite step for future Sort-seq experiments, we observed that the *cdc7-T484A* mutant allele caused a detectable acceleration in S-phase progression. In contrast, the *fkh1-R80A* mutant allele caused a slight delay in S-phase progression, the expected phenotype for a mutant that reduces early origin function. This *in vivo* experiment supports the conclusion we made from our *in vitro* peptide binding experiments: that the Cdc7-T484 peptide is not a key target of the Fkh1-FHA domain; rather, it is a relevant target of the Rad53-FHA1 domain. Specifically, an accelerated DNA replication phenotype, even in an S-phase unperturbed by experimentally induced stress, is consistent with a defect in checkpoint surveillance of DNA replication. Thus, the Cdc7-T484 peptide is likely monitored to some extent by Rad53 throughout S-phase in response to challenges that continuously arise during this complicated process. As we move toward Sort-seq analysis of these two alleles, the prediction is that the *cdc7-T484A* allele will produce a genome replication profile very different from that generated by *fkh1-R80A*.

2.5. Materials and methods

2.5.1. *In silico* analysis of phosphothreonine residues

Core replication proteins in yeast were analyzed using STRIP [78], which predicts pT probability using two tools (NetPhos 2.0 and DisPhos 1.3) and summarizes the output of five

kinase prediction tools (NetPhosK, PredPhospho, PPSP, ScanSite, and KinasePhos) to predict the relevant kinase that likely phosphorylates each residue. Peptides were considered likely to be phosphorylated if either pT probability value was greater than 0.500.

Cell cycle-linked kinases often belong to the CMGC kinase family, and many known replication proteins are targets of CMGC group kinases, so a CMGC kinase prediction increased a match's score in this analysis.

Databases (SGD, BioGRID) were searched to find experimental evidence of phosphorylation at each position. For most proteins in this list, the experimental evidence consisted of mass spectrometric evidence for a phosphorylated threonine residue at that position. However, for peptides from Mph1 and Cdc7, additional strong experimental evidence based on genetic and biochemical approaches exists [20,74].

Each candidate was aligned to each of the two Mph1 phosphoepitopes in Jalview 2.11.4.0 and the average per-residue conservation score was computed [109]. The higher of the two scores was divided by the maximum alignment score to yield a percentage, listed in the "alignment to Mph1 peptide" column. The cutoff for a "good match" was set at 75% or higher.

The surface exposure of candidate peptides was evaluated based on crystallographic structures of the relevant protein or complex when available, or AlphaFold models if no experimental structure was available [110].

Based on the five criteria listed here, candidates could receive a score of up to 5 stars. Candidates with a 5-star rating are highlighted in grey.

2.5.2. *In silico* structural analysis

ChimeraX's Matchmaker tool was used to align Rad53's FHA1 domain (chain A, residues 231-362; PDB:5T2S) to the AlphaFold model of Fkh1's FHA domain (residues 72-170)

[74,110,111]. The pairwise sequence alignment was used to identify aligned residues for a subsequent rigid-body alignment using the TM-align pairwise structure alignment tool [112]. A portion of Rad53's FHA1 domain centered on the key arginine residue (amino acids 265-274; PDB:5T2S) was aligned to the AlphaFold model of Fkh1's FHA domain, again centered on the key arginine residue (amino acids 72-81). PyMOL's "Find polar contacts" tool was used to identify potential hydrogen bonds/salt bridges between the Cdc7 peptide and either Rad53 or the Fkh1 FHA domain aligned to Rad53 by TM-align.

2.5.3. Prepare sgRNA and donor to generate mutant strains using CRISPR-Cas9

An all-in-one plasmid was constructed by inserting Cas9 from pML104 into pDB18 (Conrad A. Nieduszynski lab) and abolishing two BsaI sites, creating pCF4518 [113,114]. sgRNA oligo pairs were annealed to one another in a reaction containing 10 mM Tris pH 8.0, 25 mM NaCl, and 100 uM of each primer, which was incubated in a thermocycler at 80°C for 5 minutes, then the temperature was decreased by 0.1°C every 10 seconds until reaching a final temperature of 20°C.

Annealed oligos were inserted into the remaining BsaI site on pCF4518. A phosphorylation reaction was set up with 1.5 µl of annealed oligos diluted 1:200, 100 ng of BsaI-cut gel purified pCF4518, 1X T4 DNA ligase buffer, and 0.5 µl T4 PNK (NEB #M0201) in a final volume of 9.5 µl. The phosphorylation reaction was incubated at 37°C for 30 minutes, then the temperature was decreased to 16°C and 0.5 µl T4 DNA ligase (NEB #M0202) was added. The ligation reaction was incubated at 16°C for 16 hours, then enzymes were heat inactivated at 65°C.

Donor fragments were designed to create the desired mutation and a silent mutation to abolish the corresponding sgRNA's PAM site was also added. 120 bp homology arms on either side of these mutations were added, and the entire sequence was flanked by restriction sites and cloned into a pUC vector backbone. 4 minipreps' (NEB #1110) worth of plasmid DNA eluted in 200 µl of warmed 10 mM Tris pH 8.5 were digested with restriction enzymes (NEB high-fidelity versions) overnight at 37°C. Digests were ethanol precipitated, dissolved in 40 µl of 10 mM Tris pH 8.5, and run on an agarose gel. Bands corresponding to the donor fragment size were excised and gel purified (NEB #T1020) eluted in 8 µl of warmed 10 mM Tris pH 8.5.

2.5.4. *E. coli* transformation

DNA (100 ng plasmid or 5 µl ligation) was added to sterile 1.5 mL Eppendorf tubes. DH5α competent cells (Thermo) were thawed on ice, then 50-100 µl of cell suspension was added to each DNA tube and tubes were flicked to mix. Tubes were incubated on ice for 30 minutes, heat shocked at 42°C for 30 seconds, and incubated on ice for 5 minutes. SOC broth was added to 1 mL and tubes were incubated with shaking at 37°C for 1 hour. LB plates with the appropriate antibiotic were spread with 40 µl 100 mM IPTG and 40 µl 20 mg/mL (w/v) X-gal if blue-white screening was necessary. Recovered cells were plated on LB plates and incubated at 37°C for 14-24 hours.

2.5.5. Yeast transformation

S. cerevisiae were grown to OD600/mL = 0.4-0.6. For 4 transformations, 2 OD equivalents were pelleted at 3000xg for 30 seconds. Pellet was washed with 800 µl of TEL buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM LiOAc). Cells were pelleted at 3000xg for 1 minute and the supernatant was pipetted off. Pellet was resuspended in 160 µl of TEL buffer and 40 µl was aliquoted into each transformation tube. Transformation tubes contained DNA (for a

CRISPR transformation, 200 ng plasmid DNA and/or ~1 ug linear donor fragment were used). 150 µl PEG-TEL (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM LiOAc, 40% PEG 3500) was added to each transformation tube and tubes were inverted vigorously to mix. Tubes were incubated at 30°C for 30 minutes and 42°C for 20 minutes. Tubes were centrifuged at 1000xg for 30 seconds and supernatant was pipetted off. Pellets were resuspended in 200 µl YMD media (20g/L glucose, 134g/L yeast nitrogen base with ammonium sulfate (USBio #Y2025)) and the full volume was plated on selective media.

2.5.6. Purification of Fkh1 and Rad53-FHA1 proteins

Rosetta (DE3) cells (Novagen #70954) containing expression plasmids were grown for 8 hours in 3 mL LB media with 50 mg/L kanamycin and 34 mg/L chloramphenicol (LB+Kan+Cam media). 50 µl daylong culture was used to inoculate 50 mL LB+Kan+Cam media and grown for 14 hours. 50 mL of this culture was used to inoculate 800-1000 mL LB+Kan+Cam media in a 4 L flask. Cultures were grown until OD600/mL = 0.4-0.6, then induced by adding 1 mM IPTG. Cultures were incubated at 25°C for 3 hours. Cells were harvested and pellets were flash frozen in an ethanol/dry ice bath and stored at -80°C. Pellets were thawed on ice and resuspended in wash buffer (50 mM Tris, 5 mM $MgCl_2$, 5 mM ATP, 10% glycerol, 1 M NaCl, 5 mM 2-mercaptoethanol, 20 mM imidazole, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, pH 7.0). NEBExpress *E. coli* lysis reagent was added to 1:1 and DNase I (Thermo #EN0521) was added to 0.625 units per mL cell suspension. Cell suspension was rotated at room temperature for 1 hour. Lysate was clarified by spinning at 16000xg for 5 minutes and collecting supernatant.

50 µl Ni-NTA agarose slurry (Qiagen #30210) per gram wet cell weight was equilibrated in wash buffer, then added to clarified lysate. Lysate was rotated at 4°C for 1 hour, then applied

to a gravity column and flow-through was collected. Column was washed with 10 mL of wash buffer and proteins were eluted with 1.5 mL of elution buffer (50 mM Tris, 5 mM MgCl2, 5 mM ATP, 10% glycerol, 1 M NaCl, 5 mM 2-mercaptoethanol, 500 mM imidazole, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, pH 7.0). Dialysis was performed to change buffer to dialysis buffer (50 mM HEPES, 200 mM KCl, 10% v/v glycerol, 5 mM 2-mercaptoethanol, 1 mM EDTA, 1 mM EGTA, 5 mM MgOAc, 0.02% NP-40, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, pH 7.0). Aliquots were flash frozen in an ethanol/dry ice bath and stored at -80°C.

2.5.7. Protein immunoblotting

Samples were run on 6% or 8% polyacrylamide gels submerged in running buffer (0.1% w/v SDS, 25 mM Tris, 192 mM glycine) at 160 V for ~40 minutes. Transfer sandwiches were soaked in transfer buffer (25 mM Tris, 192 mM glycine, 20% v/v methanol, 0.005% w/v SDS), assembled with pre-soaked nitrocellulose membrane and placed in a BioRad Mini Trans-Blot apparatus, which was cooled with ice and placed at 4°C. Proteins were wet transferred at 100 V for 1 hour. Membranes were rinsed of methanol and cut to size, with an optional Ponceau staining (0.5% w/v Ponceau S, 1% v/v acetic acid) and imaging step. While rotating on an orbital shaker, membranes were blocked with 10 mL blocking buffer (5% skim milk, 20 mM Tris, 150 mM NaCl, 0.01% sodium azide). Blocking buffer was removed and primary antibody (1:1000 monoclonal mouse anti-Cdc7, 1:2000 polyclonal rabbit anti-Fkh1, or 1:250 mouse anti-FLAG (Cell Signaling #D6W5B)) diluted in 10 mL blocking buffer was added. Blots were incubated with primary antibody overnight. Blots were washed for 3x 15-minute washes with 20 mL TBST (20 mM Tris, 150 mM NaCl, 0.1% v/v Tween-20). Secondary antibody (usually 1:10000 anti-mouse-HRP or 1:20000 anti-rabbit-HRP) diluted in 10mL TBST was added and

blots were incubated with secondary antibody for 1 hour. Alternatively, primary and secondary antibodies were omitted and blots were instead incubated with 1:5000 anti-FLAG-HRP (Sigma #A8592) or 1:8000 streptavidin-HRP (Thermo #N100) for 1 hour. Blots were washed for 3x 15-minute washes with 20mL TBST. Blots were incubated with Thermo Fisher SuperSignal West Pico PLUS Chemiluminescent Substrate for 1 minute. Blots were imaged on an Azure C600 imager with manual exposure times of 1 second, 2 seconds, 5 seconds, 10 seconds, 30 seconds, 2 minutes, and 5 minutes. Where necessary, blots were stripped with mild strip buffer (1.5% w/v glycine, 1% w/v SDS, 1% Tween 20, pH 2.2), washed for 3x 15-minute washes with TBST, then proceeded starting at the blocking step.

2.5.8. *In vitro* peptide pulldown

Saturate beads with peptide: Lyophilized biotin-conjugated peptide was dissolved in 200 µl PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4). Peptide concentration was estimated via Nanodrop (Thermo). 20 µl well-mixed streptavidin bead slurry (Cell Signaling #5947) was added to a separate tube and the storage buffer was removed. Peptide and beads were combined and incubated overnight at 4°C with rotation. After brief low-speed centrifugation, beads were collected in a magnetic rack and supernatant was collected. Unbound peptide concentration was again estimated via Nanodrop. Beads were washed 5 times with 200ul PBS and stored in 200ul PBS.

Incubate protein with sssDNA: dialyzed protein was combined with 1.85 ug/ul sssDNA (Thermo #AM9680), 50 mM HEPES-KOH pH 7.0, 200 mM KCl, 10% v/v glycerol, 5 mM 2-mercaptoethanol, 1 mM EDTA, 1 mM EGTA, 5 mM MgOAc, 0.02% NP-40, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A. Sample was mixed well and stored on ice.

Pulldown: 50 µl peptide-saturated bead suspension was moved to a new tube. Protein+sssDNA mixture was diluted 1:10 in binding buffer (50 mM HEPES-KOH pH 7.0, 200 mM KCl, 10% v/v glycerol, 5 mM 2-mercaptoethanol, 1 mM EDTA, 1 mM EGTA, 5 mM MgOAc, 0.02% NP-40, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A) and a small sample was collected as "starting" material. The protein+sssDNA dilution was mixed with peptide-saturated beads and incubated for 2 hours at 4°C with rotation. Beads were collected on a magnetic rack and the "unbound" sample was collected. Beads were washed 5 times with 200 µl binding buffer. Beads were resuspended in 33.33 µl of 2X Laemmli buffer (20% v/v glycerol, 4% w/v SDS, 125 mM Tris-HCl pH 6.8, 10% v/v 2-mercaptoethanol, 0.1% w/v bromophenol blue). Input and unbound samples were mixed with 4X Laemmli buffer (40% v/v glycerol, 8% w/v SDS, 250 mM Tris-HCl pH 6.8, 20% v/v 2-mercaptoethanol, 0.2% w/v bromophenol blue) for a final 1X Laemmli buffer concentration. Samples were incubated at 95°C for 5 minutes.

2.5.9. Whole protein extract

Harvest 1.5 OD equivalents of *S. cerevisiae* and wash once with water. Resuspend in 1 mL 2 M LiOAc and incubate on ice for 5 minutes. Pellet and remove supernatant. Resuspend in 1mL 0.4 M NaOH and incubate on ice for 5 minutes. Pellet and remove supernatant. Resuspend in 100ul 1X SDS-PAGE sample buffer (60 mM Tris-HCl pH 6.8, 10% v/v glycerol, 2% w/v SDS, 5% v/v 2-mercaptoethanol, 0.0025% w/v bromophenol blue) and incubate at 95°C for 5 minutes. Store at -20°C. Before loading on gel, incubate at 95°C for 2 minutes and centrifuge at 10000xg for 1 minute.

2.5.10. Chromatin fractionation

When OD600/mL = 0.5-0.8, add sodium azide to 0.1% and collect 1.5 OD equivalents of *S. cerevisiae*. Pellet cells at 1500xg for 5 minutes and pipette off supernatant. Wash with water,

spin again and pipette off supernatant. Resuspend cells in 1.5mL prespheroplasting buffer (100 mM PIPES pH 9.4, 10 mM DTT) and incubate for 10 minutes at room temperature. Spin down at 1800xg for 1 minute and pipette off supernatant. Resuspend cells in 1 mL of ice-cold spheroplasting buffer (50 mM KH_2PO_4/K_2HPO_4 pH 7.5, 0.6 M sorbitol, 10 mM DTT). Add 1 µl Zymolyase-100T (Zymo Research) per OD equivalent and mix well. Incubate at 37°C for 15 minutes, inverting at 7.5 minutes. Pellet spheroplasts at 3000xg for 2 minutes and pipette off supernatant. Resuspend spheroplasts in 500 µl ice-cold wash buffer (20 mM NaH_2PO_4/Na_2HPO_4 pH 6.5, 0.6 M sorbitol, 1 mM $MgCl_2$, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A). Pellet spheroplasts at 4000xg for 1 minute at 4°C and pipette off supernatant. Resuspend spheroplasts in 60 µl EB buffer (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 2.5 mM $NaCl_2$, 0.1 mM $ZnSO_4$, 2 mM NaF, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A) and add Triton X-100 to a final concentration of 0.25% v/v. Incubate on ice for 5 minutes with gentle mixing. Overlay lysate on 30 µl of 30% w/v sucrose. Centrifuge at 12200xg for 10 minutes at 4°C. Supernatant represents cytoplasmic & nucleoplasmic fraction and pellet represents chromatin fraction. Use a P2 pipette tip stacked on a P200 pipette tip to carefully aspirate the supernatant and move it to a new tube. To each sample, add EBX buffer (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 2.5 mM $NaCl_2$, 0.1 mM $ZnSO_4$, 2 mM NaF, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, 0.25% v/v Triton X-100) for a final volume of 90 µl. Add 30 µl 4X Laemmli buffer (40% v/v glycerol, 8% w/v SDS, 250 mM Tris-HCl pH 6.8, 20% v/v 2-mercaptoethanol, 0.2% w/v bromophenol blue). Incubate at 95°C for 5 minutes. Before loading on gel, centrifuge at 10000xg for 1 minute.

2.5.11. Streptavidin shift assay

Adapted from [86]. Centrifuge sample at 10000xg for 1 minute to clear cell debris. In 0.2 mL tubes, add 0.3 µl of 10 ug/ul streptavidin (Sigma #189730) and 6ul sample. Incubate with rotation for 10 minutes at room temperature. Immediately load all of each sample on SDS-PAGE gel).

2.5.12. Nuclear extract

Spheroplast: Harvest *S. cerevisiae* (80-100 OD equivalents) by centrifugation at 2000xg for 5 minutes. Resuspend cells in 4 mL of prespheroplasting buffer (100 mM PIPES pH 9.4, 10 mM DTT) and incubate for 10 minutes at room temperature. Pellet cells at 1800xg for 1 minute and pipette off supernatant. Resuspend cells in 4 mL of ice-cold spheroplasting buffer (50 mM KH_2PO_4/K_2HPO_4 pH 7.5, 0.6 M sorbitol, 10 mM DTT). Add 0.5 µl Zymolyase-100T (Zymo Research) per OD equivalent and 0.675 µl glucanase (Sigma #67138) per OD equivalent and mix well. Incubate at 37°C for 30 minutes, inverting every 10 minutes. Pellet spheroplasts at 1800xg for 1 minute at 4°C and pipette off supernatant. Resuspend spheroplasts in 1 mL ice-cold wash buffer (20 mM NaH_2PO_4/Na_2HPO_4 pH 6.5, 0.6 M sorbitol, 1 mM $MgCl_2$, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A). Pellet spheroplasts at 1800xg for 1 minute at 4°C and pipette off supernatant. Resuspend spheroplasts again in 1 mL ice-cold wash buffer. Overlay the suspension on 2 mL cushion buffer (20 mM NaH_2PO_4/Na_2HPO_4 pH 6.5, 0.6 M sorbitol, 1 mM $MgCl_2$, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, 7.5% w/v Ficoll 400). Spin spheroplasts through the cushion buffer at 4000xg for 5 minutes at 4°C. Pipette off most of the supernatant, leaving about 200 µl. Add 2 mL of ice-cold wash buffer slowly, washing the sides of the tube. Then immediately aspirate off all the supernatant. Resuspend spheroplasts in 1 mL of ice-cold Ficoll buffer (18% w/v Ficoll 400, 20 mM

NaH_2PO_4/Na_2HPO_4 pH 6.5, 1 mM $MgCl_2$, 0.01% NP-40, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A).

Isolate nuclei: Drop spheroplast suspension into Dounce homogenizer containing Ficoll buffer. Homogenize the suspension for 20 strokes. Transfer lysate to fresh Eppendorf tubes and spin at 5000xg for 5 minutes at 4°C. Transfer this supernatant to fresh Eppendorf tubes and spin at 5000xg for 5 minutes at 4°C. Transfer this clarified supernatant to fresh Eppendorf tubes and spin at 16200xg for 20 minutes at 4°C to pellet nuclei. Pipette off supernatant (represents cytoplasmic fraction). Wash the sides of the tube with 400 µl ice-cold wash buffer and immediately aspirate off wash buffer.

Lyse nuclei: Resuspend nuclei in 20 µl EB buffer (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 2.5 mM $NaCl_2$, 0.1 mM $ZnSO_4$, 2 mM NaF, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A) and add Triton X-100 to a final concentration of 0.25% v/v. Incubate on ice for 10 minutes with gentle mixing. Add EBX buffer (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 2.5 mM $NaCl_2$, 0.1 mM $ZnSO_4$, 2 mM NaF, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, 0.25% v/v Triton X-100) to nuclei for a final volume of 100 µl. Sonicate for 2 rounds (3 pulses, 2x output, 30% duty cycle) with a 1 minute rest in between, keeping tube on ice at all times.

Size exclusion column: Twist off bottom closure of Zeba spin column (Thermo #89883) and loosen cap. Place column in collection tube. Centrifuge at 1500xg for 2 minutes at 4°C to remove storage solution. Mark the side of the column where the resin slants upward and place this mark on the outside every time. Add 300 µl buffer SB (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1% v/v Tween 20, 0.1% w/v SDS, 1 mM EDTA, 1 mM DTT) to column, then centrifuge at 15000xg for 1 minute at 4°C. Repeat buffer SB wash step 2 times. Apply sheared nuclei

sample to center of resin and centrifuge at 1500xg for 2 minutes at 4°C, collecting sample in a fresh Eppendorf tube.

2.5.13. Streptavidin bead pulldown

For nuclear extract sample, add 0.22 µl of well-mixed streptavidin magnetic beads (Cell Signaling #5947S) per OD equivalent. Incubate with rotating for 2 hours at 4°C. Place on magnetic rack and wait 1 minute to collect beads. Collect supernatant as "unbound" sample, add 4X Laemmli buffer (40% v/v glycerol, 8% w/v SDS, 250 mM Tris-HCl pH 6.8, 20% v/v 2-mercaptoethanol, 0.2% w/v bromophenol blue) for 1X final, and incubate at 95°C for 5 minutes before storing at -20°C. Wash magnetic beads 5 times: add 250 µl bead wash buffer (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 2.5 mM $NaCl_2$, 0.1 mM $ZnSO_4$, 2 mM NaF, 1 mM DTT, 0.2 ug/mL AEBSF, 3.6 ug/mL E-64, 10 ug/mL pepstatin A, 0.25% v/v Triton X-100, 100 mM NaCl, 5 mM EDTA), vortex on low speed to resuspend, place on magnetic rack for 1 minute, aspirate supernatant. Then resuspend bead pellet in 15 µl 1X SDS-PAGE sample buffer (60 mM Tris-HCl pH 6.8, 10% v/v glycerol, 2% w/v SDS, 5% v/v 2-mercaptoethanol, 0.0025% w/v bromophenol blue). Vortex on low speed to resuspend, centrifuge at low speed to collect any droplets, and incubate at 95°C for 5 minutes before storing at -20°C. When ready to use, heat at 95°C for 2 minutes and collect beads using magnetic rack.

2.5.14. Quantitation of protein immunoblot bands

Images were captured on an Azure C600 imager. Only images where no portion of the bands of interest were saturated were used for quantitation. In cases where bands were hard to resolve from neighboring bands, the blot image was imported into Inkscape, where a rectangle of consistent size was used to clip individual bands from the main blot image. The image was exported as a .tiff file and imported into ImageJ. The image was inverted, then each band was

selected using the Rectangle tool and its intensity was measured using the Measure function. The RawIntDen value was recorded and used for further calculations. Percent shift was calculated by dividing the RawIntDen value of a shifted band by the sum of the values of the shifted and unshifted bands.

2.5.15. Flow cytometry

To prepare genomic DNA samples for copy-number analyses, *MAT*a cells were grown in 25 mL YPD at 30°C to an OD_{600} of ~0.3 at which time α factor was added to a concentration of 5 μg/ml. Cultures continued to incubate at 30°C for 100 min to allow the majority of cells to reach G1-arrest. Cultures were then collected on a Whatman Nylon membrane (GE Healthcare #7404-004) using a Nalgene filtering apparatus (Thermo Scientific #300-4050) and washed with 100 ml YPD pre-warmed to 30°C. The cells were then immediately transferred to fresh pre-warmed 25 ml YPD supplemented with 200 μg/ml of pronase. 2.5 ml of culture was harvested at time 0 (G1-arrested sample) and at 35 min (mid S-phase) after release into YPD-pronase. Genomic DNA was purified from Zymolyase-treated yeast pellets using phenol–chloroform–isoamyl alcohol, followed by ethanol-based extraction. DNA was quantified on a Quantus fluorimeter. To determine cell-cycle distribution of each sample, cells were fixed with ethanol, stained with SYTOX Green and analyzed by flow cytometry (BD Accuri C6 Flow Cytometer).

2.5.16. Sort-seq

Yeast were grown at 30°C and processed as described in [115,116] except that a Sony MA900 was used to separate and collect S-phase and G2-phase cells. Genomic DNA was isolated as described [115]. Libraries for one wild-type sample were prepared at Northwestern University and single-end sequenced. The remaining libraries were paired-end sequenced in-house using a P2-100 flow cell and a NextSeq1000. The libraries for in-house sequencing were

made using the 1/10 dilution of the adaptor and amplified for seven cycles using the NEBNext Ultra II kit. Sequencing reads were mapped to the yeast genome, build sacCer3, using Bowtie2 [117] with the following options: bowtie2 -p 4 -X 750 -q –phred33. Mapped reads were converted to single nucleotide read coverages using Bedtools genomecov [118]. Single nucleotide coverages were normalized for sequencing depth and breadth [119] then summarized into 1 kbp bins. S/G2 ratios were determined for each independent cell population (three independent populations of *FKH1* cells and two independent populations of *fkh1-R80A* cells) using the 1 kbp bin values. Outlier S/G2 ratios were defined as any ratio that was over 1.5 IQRs below the first quartile (Q1) or above the third quartile (Q3) and were removed. The resulting set of ratios were next scaled to be between 1 and 2 using the following equation: scaled ratio = (ratio - min(ratio))/(max(ratio) - min(ratio)) + 1. The resulting scaled ratios were smoothed by fitting data to a cubic smoothing spline (R function: smooth.spline, default parameters). Smoothed data from a given genotype (three samples for *FKH1* and two samples for *fkh1-R80A*) were statistically summarized by finding the means and 95% confidence intervals.

Table 2.2. Yeast strains used in this study.

Name	Description	Plasmid (if applicable)	Source
CFY3533	*MATa, ADE2+, RAD5+*		[20]
CFY3552	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, fkh2Δ::HisG*		[20]
CFY3956	*MATa, ADE2+, RAD5+, fkh1-R80A*		[20]
CFY4068	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, MPH1-3xFLAG::HIS3*		[20]
CFY4069	*MATa, ADE2+, RAD5+, FKH1+, Mph1-3xFLAG::HIS3*		[20]
CFY4975	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, MPH1-3xFLAG::HIS3*	pCF4587	This study
CFY4976	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, MPH1-3xFLAG::HIS3*	pCF4588	This study
CFY4977	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, FDO1-3xFLAG::HIS3*	pCF4587	This study
CFY4979	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, mph1-2TA-3xFLAG::HIS3*	pCF4587	This study
CFY4980	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, mph1-2TA-3xFLAG::HIS3*	pCF4588	This study
CFY4981	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, CDC7-3xFLAG*	pCF4587	This study
CFY4982	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, CDC7-3xFLAG*	pCF4588	This study
CFY4983	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, MPH1-3xFLAG::HIS3*	pCF318	This study
CFY4984	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, cdc7-T484A*	pCF4587	This study
CFY4985	*MATa, ADE2+, RAD5+, FKH1+, CDC7-3xFLAG*		This study
CFY4986	*MATa, ADE2+, RAD5+, FKH1+, cdc7-T484A-3xFLAG*		This study
CFY4969	*MATa, ADE2+, RAD5+, FKH1+, cdc7-T484A*		This study

Table 2.3. Plasmids used in this study.

Name	Description	Source
pRS316	Yeast shuttle vector containing *URA3* marker	[120]
pCF942	*FKH1* in pRS316	[121]
pCF2608	*fkh1-R80A* in pRS316	[20]
pCF2968	Full-length *FKH1* with C-terminal 6xHis in pET28b vector in Rosetta(DE3)pLysS cells	[20]
pCF4587	Fkh1-TurboID flanked by *FKH1* pro/term in pRS316	This study
pCF4588	Fkh1^{R80A}-TurboID flanked by *FKH1* pro/term in pRS316	This study
pCF4621	Full-length *fkh1-R80A* with C-terminal 6xHis in pET28b vector in Rosetta(DE3)pLysS cells	This study
pCF4627	Rad53's FHA1 domain (aa22-162) with N-terminal 6xHis, 1xFLAG tags in pET28a vector in Rosetta(DE3)pLysS cells	This study
pCF4628	Rad53^{R70A}'s FHA1 domain (aa22-162) with N-terminal 6xHis, 1xFLAG tags in pET28a vector in Rosetta(DE3)pLysS cells	This study

Chapter 3

Examining the role of the *FDO1* gene with an emphasis on understanding Fdo1's interaction with Fkh1's forkhead-associated (FHA) domain in *Saccharomyces cerevisiae*

This chapter is in preparation for a publication in 2025, on which I will be first author. I (with the help of my advisor, Catherine Fox) conceived and designed all experiments. Timothy Hoggard conducted S-phase flow cytometry and Sort-seq experiments and developed R analysis pipelines to analyze the data. Timothy Hoggard also performed Y1000+ searches and developed R analysis pipelines to extract best hits and plot by species. Michael Sheets helped design Fkh1-TurboID constructs.

3.1. Abstract

Saccharomyces cerevisiae is a useful eukaryotic model organism in part because of its ability to perform homologous recombination (HR) efficiently. The yeast Fkh1 DNA binding protein, and particularly its conserved protein-protein interaction module, the forkhead-associated (FHA) domain, promotes homologous recombination during mating-type switching and during Cas9-mediated genome editing in yeast. The FHA domain specifically binds phosphothreonine (pT) residues (with additional substrate specificity determined by the residues immediately flanking the pT residue) in unstructured/flexible regions. In this report we examined an interaction between the Fkh1-FHA domain and the yeast protein Fdo1. Fdo1 encodes a non-conserved protein enriched in unstructured regions. Fdo1 orthologs can only be found in yeast species closely related to *S. cerevisiae*. We showed that a Fkh1-FHA-dependent interaction with Fdo1 required 16 distinct threonine residues on Fdo1, consistent with a classic FHA-pT recognition mechanism controlling the Fkh1-Fdo1 interaction. However, neither primary sequence context nor further mutant analysis has allowed us to narrow down a discrete pT-containing peptide region on Fdo1 that binds to the Fkh1-FHA domain. Further examination of Fdo1 using both AlphaFold structural prediction and primary sequence analyses revealed that Fdo1 is substantially unstructured as well as highly enriched in both asparagine and glutamine, amino acids whose presence in unstructured protein regions promotes less conventional mechanisms of protein-protein interaction through the formation of liquid-liquid phase separation (LLPS) or prion-like structural conformations. We analyzed Fdo1 and Fkh1 amino acid conservation across *Saccharomycotina* and observed a correlation between species that contained both orthologs and the ability to perform efficient HR. We established an assay to measure the efficiency of HR at the yeast *ADE2* locus and showed that the native *FKH1* gene contributes to efficient HR repair of a Cas9-generated double-strand break in a *ade2-1* mutant

allele. We also showed that *fdo1* mutants were defective in maintaining the normal rDNA copy number in the rDNA array, a phenotype consistent with a defect in either HR or DNA replication, and that a *fdo1* mutant incapable of interacting with Fkh1 through a pT-dependent mechanism accelerated the rate at which cells progressed through S-phase. This phenotype suggests a defect in S-phase checkpoint mechanisms. Together, these data support a model where the FHA-dependent Fkh1-Fdo1 interaction promotes normal DNA metabolism during S-phase in haploid yeast.

3.2. Introduction

Saccharomyces cerevisiae (budding yeast, henceforth referred to as "yeast") is a single-celled microbe that serves as a model for the eukaryotic cell. Before the genomic revolution, yeast was studied extensively for decades because of its simple life cycle and fast growth rate. Yeast offered many genetic advantages, including a small, compact genome of only 12 Mb with about ~6500 genes, most of which lack introns making them easy to functionally dissect. Yeast can easily grow as haploids or diploids, which made genetics easy. Yeast are also notable for their efficient and accurate homologous recombination (HR), which meant genes could be mapped and studied even without a lot of genome sequence information. Yeast was at the forefront of the genomic revolution, being the first model organism to have its genome completely sequenced, annotated thoroughly, and made fully accessible on the Saccharomyces Genome Database (SGD) [122]. The SGD was a powerful tool for even more gene-function dissection. However, despite all of these advantages, hundreds of genes in the yeast genome encode proteins with unknown molecular functions and biological roles [123]. Most researchers using yeast as a "general model for the eukaryotic cell" are focused on evolutionarily conserved genes that are reasonably believed to perform core eukaryotic cell functions. Thus, part of the

reason many of these SGD genes remain poorly studied is that many perform non-essential functions under laboratory conditions and are found only in fungal species closely related to *Saccharomyces cerevisiae*. *FDO1* (Fkh1 interacting protein involved in donor preference), previously named *YMR144W*, is one of these genes, and the focus of Chapter 3.

Fdo1 was isolated as a Fkh1-interacting protein in a 2-hybrid screen where Fkh1 was the bait protein [20]. Fkh1 is a sequence-specific DNA binding protein with orthologs across eukaryotes. Fkh1 has direct roles in several yeast chromosomal processes, including cell-cycle regulated transcription, DNA replication and homologous recombination (HR). The same 2-hybrid screen that identified *FDO1* also identified *MPH1.* In complete contrast to *FDO1*, *MPH1* is highly conserved, encoding a multifunctional DNA helicase that is the yeast ortholog of FANCM, the Fanconi anemia helicase in humans. This helicase has roles in stabilizing DNA replication intermediates and is important for genome stability. Mph1-Fkh1 interactions were discussed in Chapters 1 and 2. Fdo1 and Mph1 were selected for further focus for my thesis research for two major reasons. First, each of these proteins had also been identified as Fkh1 physical interactors in an independent proteome-wide biochemical screen [65]. This fact made me more confident that further experimental investment was likely to lead to biologically relevant insights. In addition, Mph1 and Fdo1 had not been implicated in direct mechanisms related to Fkh1's role in transcription, even though the transcription field is highly active. Therefore, we posited that Mph1 and Fdo1 were more likely to be directly relevant to the transcription-independent chromosomal processes controlled by Fkh1 that were the major goal of the lab and my research.

The 2-hybrid interaction between Mph1-Fkh1 was robust enough that it was useful for mapping regions of Mph1 sufficient for this interaction, and, as described in Chapters 1 and 2,

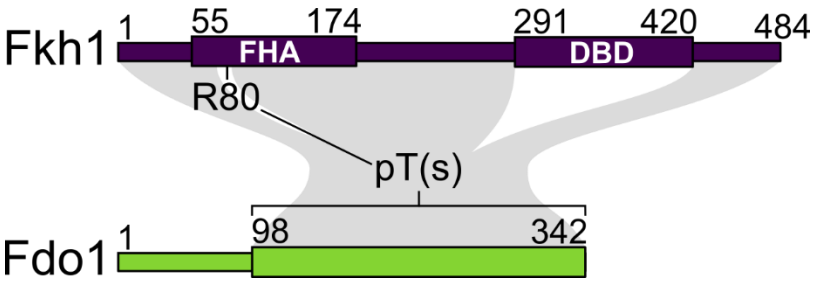
allowed us to define the direct Fkh1-Mph1 interaction peptide on Mph1 fairly easily [20]. However, the 2-hybrid interaction between Fdo1-Fkh1, while reproducible, was quite weak, making mapping of the protein surface(s) on Fdo1 important for the Fkh1-Fdo1 interaction challenging by this assay. In addition, while Fkh1's FHA domain was sufficient for the Mph1-Fkh1 interaction, Fkh1's FHA domain was *not* sufficient for the Fdo1-Fkh1 interaction, though it did contribute, suggesting that mapping the Fdo1-Fkh1 interaction and understanding its biological role might also be more complicated (**Figure 3.1**). However, a proteomic-scale localization screen indicated that Fdo1 is a nuclear protein, and a direct analysis of mating-type switching efficiency (i.e. *MAT*a locus recombining with *HML*α) revealed that Mph1 and Fdo1 contributed additively to this process and to similar extents [20,124].These observations led to a model where Fdo1 contributed to the HR event involved in mating-type switching by helping Fkh1 protein bound to the recombinational enhancer (RE) near the HMLα locus locate the double-strand break (DSB) at the *MAT*a locus (see **Figure 1.4**).

In this chapter, the Fkh1-FHA-dependent TurboID assay introduced in Chapter 2 was used to examine the interaction between Fkh1 and Fdo1. *In silico* and wet lab experiments were used to address a possible role for Fdo1 in general HR beyond its currently assigned role in the locus-specific HR event required for mating-type switching. The TurboID assay showed that the canonical pT-binding function of the Fkh1-FHA domain contributed to the Fkh1-Fdo1 interaction but was not sufficient, supporting conclusions made from the published 2-hybrid interaction experiments [20]. The FHA-dependent TurboID signal was robust enough that it was used to address whether any threonine residue(s) on Fdo1 were important for the Fkh1-Fdo1 interaction, as would be predicted if this interaction were direct.

Figure 3.1. Fdo1's interaction with Fkh1 is partially FHA-dependent.

Schematic showing relevant regions of Fdo1 and Fkh1. A full-length bait plasmid expressing Fkh1 (except where the Fkh1 DBD is replaced with a Gal4 DBD) isolated a prey plasmid expressing amino acids 98-342 of Fdo1. Fkh1's FHA domain uses arginine 80 to achieve a 2-hybrid interaction with one or more phosphorylated threonine residues on Fdo1. We do not know whether these interactions are direct as they have not been examined *in vitro* with purified proteins.

In addition, this Fkh1-Fdo1 2-hybrid interaction is only partially dependent on Fkh1's R80 residue, suggesting that an additional interaction interface involving another region of Fkh1 (represented by gray shading) is also important. Hence the entire Fkh1 protein, sans the Fkh1 DBD, is indicated as potentially interacting with Fdo1 98-342.



AlphaFold was used to predict the structural model of Fdo1, and this prediction was combined with analyses of Fdo1 threonine residues and their primary sequence context. No single threonine-containing peptide within Fdo1 shared strong/obvious similarity to the Mph1 pT-containing peptides that have been verified as direct binding targets for Fkh1-FHA [20]. Therefore, multiple T→A substitutions were generated across *FDO1*, and the mutant strains were assessed for interaction with Fkh1 by the TurboID assay. To date, a *fdo1-16TA* mutant that cannot support the FHA-dependent Fkh1-Fdo1 interaction has been identified, suggesting there is a requirement for one or more T residues for Fkh1 to bind Fdo1. However, thus far a discrete pT-containing peptide on Fdo1 required for the interaction with Fkh1's FHA domain has not been identified. Analysis of an AlphaFold model of Fdo1 and analyses of amino acid content suggest that Fdo1 contains highly unstructured (intrinsically disordered) regions rich in asparagine (N) and glutamine (Q) residues. These residues are often enriched in proteins that can form liquid-liquid phase separation (LLPS) compartments and proteins that can fold into heritable alternative structural conformations (such as prions).

Phenotypes caused by *fdo1* mutants were also explored using both *in silico* and wet-bench approaches. In an *in silico* experiment, we examined the conservation of the *FDO1* and *FKH1* genes within the subphylum *Saccharomycotina*, whose genome sequences and phylogenetic relationships were made accessible in [125]. *FDO1* could be located in only a few closely related yeasts within this subphylum. In contrast, at least a portion of the *FKH1* gene could be located across all reported species within *Saccharomycotina*. Using experimental data available in the literature, a correlation was observed between the ability to perform efficient HR and the presence of both a *FDO1* and a complete *FKH1* ortholog. These data were consistent with the hypothesis that the Fkh1-Fdo1 interaction was relevant to the proficiency of HR that helped

make *Saccharomyces cerevisiae* a famous eukaryotic model. To further explore this hypothesis, a simple colorimetric assay that reports on homologous recombination-mediated conversion of an *ade2-1* mutant locus to *ADE2* in yeast was established, and preliminary data suggest that a *fkh1* mutant reduces HR in this system by ~10%. The rDNA locus, a tandem array of ~200 rDNA gene cassettes in yeast, is sensitive to defects in HR and DNA replication [126–128]. Specifically, mutants causing defects in these DNA metabolic processes lead to contraction of the rDNA array (i.e. loss of rDNA gene copies, referred to as rDNA copy number (rDNA-CN) loss or reduction). Multiple independent isolates of *fdo1* mutants showed evidence of reduced rDNA-CN. Finally, cell cycle analyses by flow cytometry indicated that a *fdo1-19TA* mutant, which was unable to support the FHA-dependent Fkh1-Fdo1 interaction, led to cells accumulating in late S-phase. Together, these phenotypes suggest that that Fdo1 promotes normal chromosome metabolism in yeast.

3.3. Results and Discussion

3.3.1. The canonical function of the Fkh1-FHA domain contributed to a Fkh1-Fdo1 interaction.

I used the Fkh1-TurboID system introduced in Chapter 1 to assess a Fkh1-Fdo1 interaction (**Figure 3.2.a**). Specifically, I compared the level of Fkh1-FHA dependent biotinylation of Mph1 and Fdo1 using the streptavidin shift method described in **Figure 2.2.b**. The addition of streptavidin to denatured protein extract prior to the separation of proteins by SDS-PAGE and protein immunoblotting resulted in a detectable Mph1 molecular weight shift (**Figure 3.2.a**, compare lanes 1 and 2). The Mph1 shift also required the canonical function of the Fkh1-FHA domain, as indicated by the loss of shift when a Fkh1^{R80A}-TurboID experiment was performed (**Figure 3.2.a**, compare lanes 2 and 4). These results were expected based on the

data presented in Chapter 2. Next, Fdo1 was analyzed. A Fdo1 molecular weight shift was observed that was also fully streptavidin-dependent (**Figure 3.2.a**, compare lanes 5 and 6). This shift was partially dependent on the canonical function of the Fkh1-FHA domain (**Figure 3.2.a**, compare lanes 6 and 8), indicating that Fkh1 and Fdo1 are in close proximity *in vivo*, consistent with the two proteins' interaction in their more native context. The partial dependence on the Fkh1-FHA domain was also consistent with previously published 2-hybrid experiments that showed that the Fkh1-Fdo1 interaction was distinct among the proteins isolated in the 2-hybrid screen: the Fkh1-FHA domain was necessary but not sufficient to support this interaction [20]. However, in contrast to the 2-hybrid data that was challenging to interpret because of the low signal, these TurboID experiments indicated that a substantial Fkh1-Fdo1 interaction could occur in the absence of the canonical pT peptide-binding function of the Fkh1-FHA domain (**Figure 3.2**, compare lanes 6 and 8).

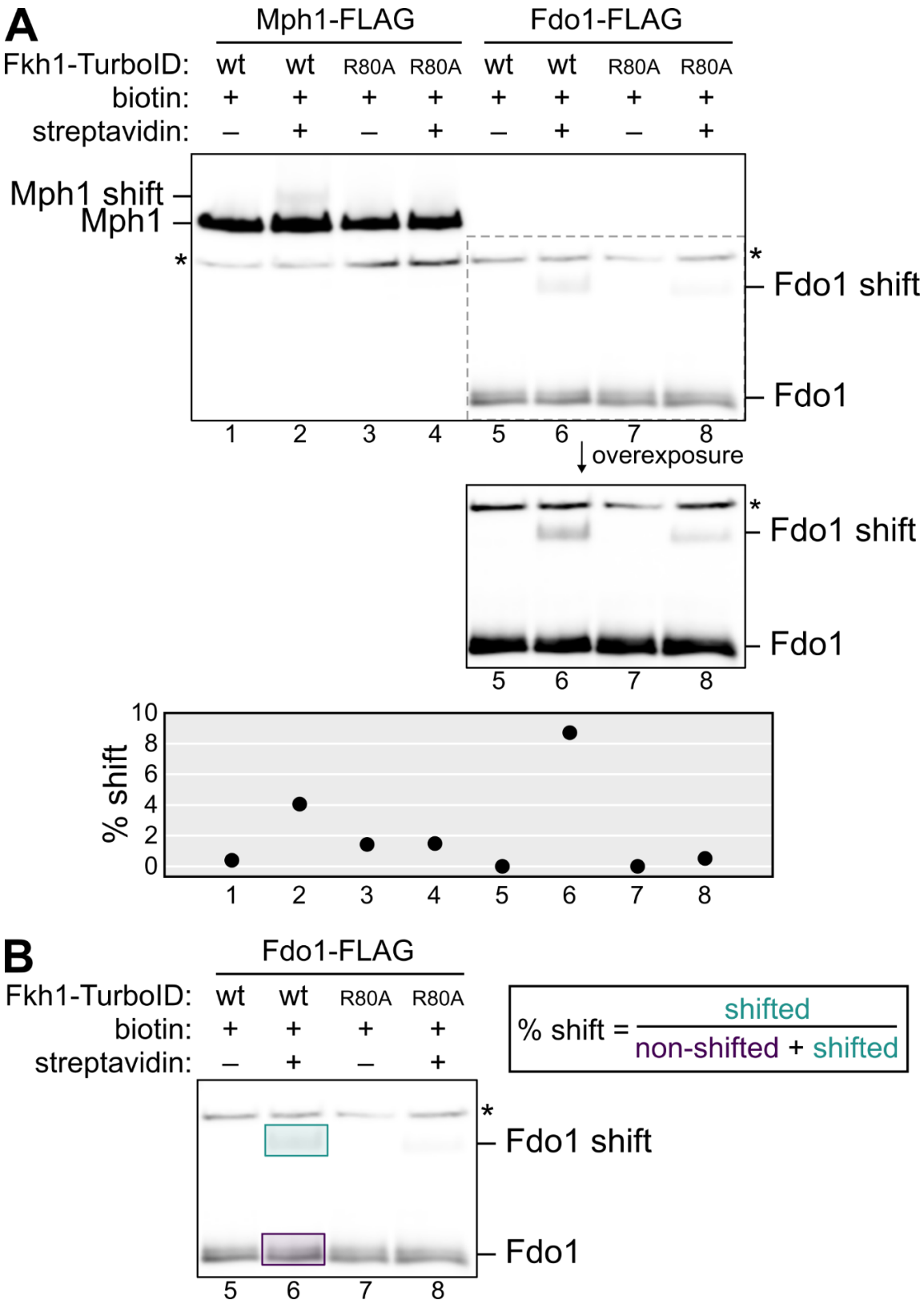
The goal was to use this assay to ask whether any of the threonine residues in Fdo1 could explain the Fkh1-FHA dependence of the Fkh1-Fdo1 interaction observed in **Figure 3.2.a,** as would be expected if the proximity labeling signal was reflective of a direct Fkh1-Fdo1 interaction. If the FHA-dependent Fkh1-Fdo1 interaction were direct, then alanine substitution(s) of the relevant threonine residue(s) within Fdo1 would yield a proximity labeling signal equivalent to that generated by Fkh1^{R80A}-TurboID (**Figure 3.2.a**, lane 8). The approach for quantifying this effect so that it could be compared across independent experiments is illustrated in **Figure 3.2.b**.

Figure 3.2. Fkh1-TurboID produces streptavidin shifts of Mph1 and Fdo1.

A. Protein immunoblot of streptavidin shift experiment to examine Mph1-FLAG and Fdo1-FLAG fusion proteins in the presence of the indicated components. Inset: longer exposure time of marked area. Below the gel, the estimated percentage of each protein that is present in the shifted band (referred to as % shift) has been quantified and calculated.

B. The calculations used to determine the % shift value in A. Highlighted regions mark example regions whose intensity is measured in ImageJ. These values have no biological meaning at this point, but they do allow us to compare % shift values between independent experiments more easily.

Figure 3.2.


A
Mph1-FLAG
Fdo1-FLAG
Fkh1-TurboID: wt wt R80A R80A wt wt R80A R80A
biotin: + + + + + + + +
streptavidin: – + – + – + – +
Mph1 shift
Mph1
*
Fdo1 shift
Fdo1
1 2 3 4 5 6 7 8
overexposure
% shift
B
Fdo1-FLAG
Fkh1-TurboID: wt wt R80A R80A
biotin: + + + +
streptavidin: – + – +
% shift = shifted / non-shifted + shifted

3.3.2. An AlphaFold model of Fdo1 revealed a protein lacking tertiary structure, limiting its usefulness in predicting candidate Fkh1-FHA-interacting pT peptides within Fdo1.

AlphaFold was used to predict a structural model for Fdo1 (**Figure 3.3.a**) [110]. This model predicts a distinctive structure in which the middle third (residues 107-224) of the 342-residue Fdo1 protein contains a long alpha helix. The N- and C-terminal ends of Fdo1 that flank this extended alpha helix are predicted to be largely unstructured. I will refer to these regions as Intrinsically Disordered Regions (IDRs) following the current nomenclature in the field that has focused on highly unstructured protein regions and their biochemical and biological roles [129,130]. Based on this AlphaFold structure, however, the entire protein is relatively solvent-exposed and thus accessible for protein-protein interactions, meaning that the structure itself was of little use in narrowing down the threonine residues worth considering as potential Fkh1-FHA interaction surfaces.

The Fdo1 protein sequence was used as a query for a Foldseek sequence similarity search [131]. The majority of protein matches were hypothetical, predicted, or uncharacterized proteins. There was one good structural match with an assigned function: *Homo sapiens* coiled-coil domain-containing protein 69. This protein is thought to act as a scaffold in the assembly and recruitment of spindle microtubules during mitosis [132]. Perhaps the distinct structure of Fdo1 lets it serve a similar scaffolding role in promoting HR, perhaps playing a role in the recruitment and assembly of other protein factors required for homologous recombination.

3.3.3. A mutant Fdo1 (*fdo1-16TA*) that contained 16 T→A substitutions phenocopied the effect of the *fkh1-R80A* mutant allele on the Fkh1-Fdo1 interaction.

In these experiments, I attempted to define the relevant threonine residue(s) on Fdo1 that could account for the observed FHA-dependent effect on Fdo1 biotinylation as measured by the protein proximity assay in **Figure 3.2**. In particular, assuming that the Fkh1-FHA dependent Fdo1-Fkh1 interaction was direct and occurred via the canonical mechanism, it should be possible to define one or some finite combination of T→A substitutions in Fdo1 that would produce a biotinylation readout in a Fkh1-TurboID experiment that was equivalent to (i.e. phenocopied) that produced by wild-type Fdo1 in a Fkh1^{R80A}-TurboID experiment (see **Figure 3.2.a**).

Based on current models and data for Fkh1-FHA (and Rad53-FHA1) as protein recruitment domains, I wanted to focus my efforts on T residues within Fdo1 that occur within unstructured, solvent-exposed regions. However, as discussed above, this approach was not useful because Fdo1 is entirely solvent exposed and largely unstructured (**Figure 3.3.a**). While the structure was certainly intriguing, it did not help narrow down high-confidence threonine residues worth targeting for substitution within Fdo1.

Next, I turned to the information we had about the primary sequence content of pT-containing peptides that were known to be Fkh1-FHA binding targets. Specifically, I used the experimental 2-hybrid data that assessed the validated key pT residues in the Fkh1-FHA-interacting Mph1 peptide (**Figure 3.3.b**) [20]. The Fkh1-FHA-interacting Mph1 peptide as originally examined contained two relevant threonine residues, T776 and T785. Phosphorylation of each of these residues contributed to the full stability of the Fkh1-FHA-Mph1 interaction *in vitro*.

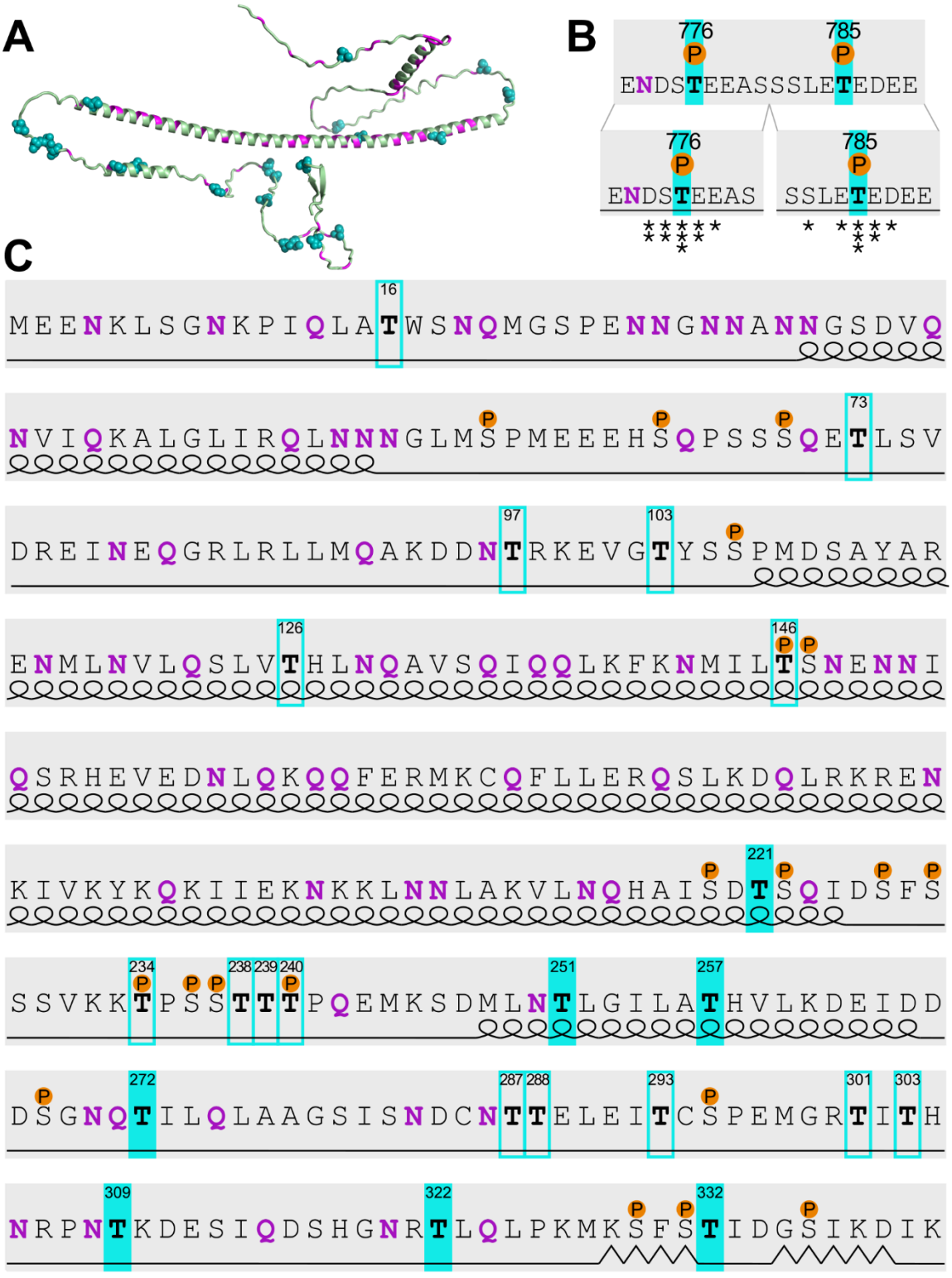
Figure 3.3. AlphaFold model of Fdo1.

A. AlphaFold model of Fdo1 [110]. Threonine residues are depicted as teal spheres. Asparagine and glutamine residues are shown in magenta.

B. Primary sequence of Mph1 peptides. Threonine residues are highlighted in teal. Phosphorylation marks are depicted with orange circles. Asparagine and glutamine residues are shown in magenta. Below: each 9-mer is pictured, with asterisks indicating the importance of each residue to the Fkh1-Mph1 interaction.

C. Primary sequence of Fdo1. Secondary structure is marked below each residue (loops = alpha helix, zig-zag = beta sheet). As in 3.3.B, threonine residues are highlighted in teal; however, threonine residues thus far found to be unnecessary for the Fkh1-Fdo1 interaction as measured by TurboID assay have empty teal outlines. Phosphorylation marks are depicted with orange circles. Asparagine and glutamine residues are shown in magenta.

Figure 3.3.


A
B
776
785
ENDSTEEASSSLETEDEE
776
ENDSTEEAS
785
SSLETEDEE
C
MEENKLSGNKPIQLATWSNQMGSPENNGNNANNGSDVQ
NVIQKALGLIRQLNNNGLMSPMEEEHSQPSSSQETLSV
DREINEQGRLRLLMQAKDDNTRKEVGTYSSPMDSAYAR
ENMLNVLQSLVTHLNQAVSQIQQLKFKNMILTSNENNI
QSRHEVEDNLQKQQFERMKCQFLLERQSLKDQLRKREN
KIVKYKQKIIEKNKKLNNLAKVLNQHAISDTSQIDSFS
SSVKKTPSSTTTPQEMKSDMLNTLGILATHVLKDEIDD
DSGNQTILQLAAGSISNDCNTTELEITCSPEMGRTITH
NRPNTKDESIQDSHGNRTLQLPKMKSFSTIDGSIKDIK

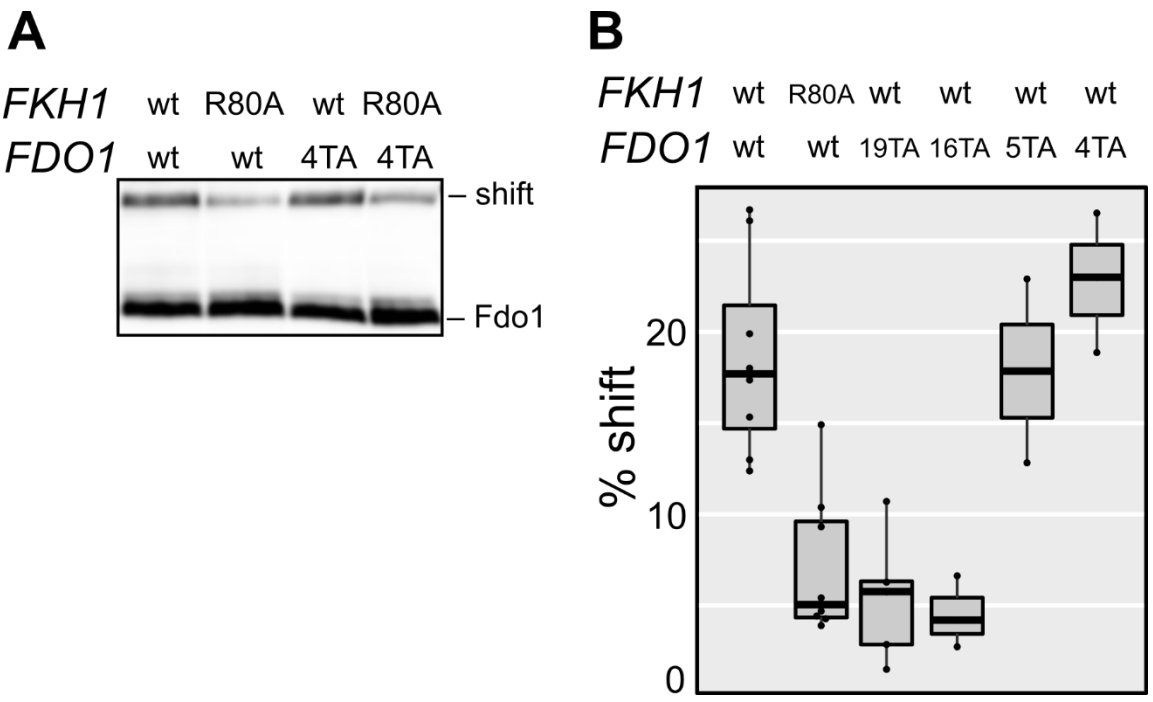
My experiments in Chapter 2 indicated that half of this peptide, one centered around pT776, was sufficient for canonical Fkh1-FHA-dependent binding. Therefore, I considered the Fkh1-FHA interaction surface on Mph1 as two distinct 9-mer peptides, each containing a distinct pT residue, either pT776 or pT785. (**Figure 3.3.b**). Each of these Mph1 peptides was generally acidic in nature, and each contained a glutamic acid (E) residue that was the second most critical residue (after the pT residue) for the Fkh1-Mph1 2-hybrid interaction [20] (**Figure 2.2.a**). Each peptide also contained two serine (S) residues and one aliphatic residue, either alanine (A) or leucine (L). Thus, I examined the threonine residues present within Fdo1, evaluating whether they existed within an obviously similar context (**Figure 3.3.c**). There was only one T residue, T288, that contained a +1 E, but the rest of the adjacent residues did not create a context with obvious similarity to either of the two Mph1 peptides in **Figure 3.3.b**. Next, I used published information from proteome-scale phosphorylation mapping to widen my consideration to T residues that might have a phosphorylated serine (that potentially mimics an aspartic acid residue) or phosphorylated threonine (that potentially mimics a glutamic acid residue) at the +1 position; of course, phosphorylation mapping could miss some rarer pSs and pTs. While incorporating this information into my analyses revealed a few more threonine residues that might be construed as having an acidic residue in the +1 position, like T146 with a pS at the +1 position and T239 with a pT at the +1 position, the rest of these peptides were not similar to the Mph1 peptides. In fact, what was most striking about the Fdo1 sequence with respect to this issue was that no single T within Fdo1 existed within a context that seemed even modestly similar to the Fkh1-FHA interacting peptides in Mph1, shown in **Figure 3.3.b**. Therefore, we took a "brute force" mutational approach.

Specifically, yeast were engineered to express a mutant Fdo1 protein in which all threonine residues in the relevant region sufficient for Fkh1 interaction via 2-hybrid (n = 19) were substituted with alanine (i.e. T103A-T333A). Mutations were made in the native *FDO1* gene in haploid yeast, changing it to an allele called *fdo1-19TA*. In a Fkh1-TurboID experiment, yeast expressing Fdo1^{19TA} showed a biotinylated signal in the presence of Fkh1-TurboID that phenocopied the signal on wild-type Fdo1 in the presence of Fkh1^{R80A}-TurboID (**Figure 3.4.a-b**). Thus, the *fdo1-19TA* mutant phenocopied this *fkh1-R80A* mutant phenotype. I then examined a *fdo1-16TA* mutant (i.e. T221A-T333A) and observed that its signal also phenocopied the *fkh1-R80A* mutant. This outcome was consistent with the conclusion that one or more of the T residues between T221 and T333 was a pT serving as an anchor for Fkh1-FHA binding. However, thus far, we have not narrowed down target threonine residues any further (**Figure 3.4.b**). Specifically, yeast cells were engineered to express one of two different *fdo1* mutants, *fdo1-5TA* (i.e. T272A-T303A) and *fdo1-4TA* (i.e. T234A-T240A), but neither of these more limited T→A mutants phenocopied the biotinylation effects of the mutant Fkh1^{R80A}-TurboID protein on wild-type Fdo1. Thus, if a Fkh1-FHA dependent Fdo1-Fkh1 interaction occurred via the direct canonical mechanism, the smallest subset of sufficient pT-containing interface(s) remains undefined. Nevertheless, the observation that the *fdo1-16TA* mutant, which is not expected to perturb the structure of Fdo1, was able to phenocopy the effect of the *fkh1-R80A* mutant on the Fkh1-Fdo1 interaction is still consistent with the idea that some form of phosphothreonine-containing interface is relevant for FHA binding.

Figure 3.4. Fkh1-TurboID streptavidin shift of Fdo1 T→A mutants.

A. Example protein immunoblot showing variations in the streptavidin shift of Fdo1-FLAG depending on whether Fkh1-TurboID or Fkh1^{R80A}-TurboID is present. Here, the shift is visually similar between wild-type and *fdo1-4TA* strains, and that shift remains FHA-dependent. Thus, we conclude that these four threonine residues do not have to be phosphorylated to support the FHA-dependent Fkh1-Fdo1 interaction.

B. Multiple trials were performed with strains of varying genotypes. The % shift for each trial was calculated by quantifying bands in ImageJ. Points representing one trial each were plotted on a box and whisker plot.



The lack of Fdo1 peptides with similarity to the Mph1 peptides that interact with Fkh1's FHA domain, our inability thus far to define a single or even a few clustered T residues that, when substituted with alanine, could phenocopy the *fkh1-R80A* mutant, and the substantial unstructured regions of Fdo1 predicted by the AlphaFold structure, have led us to consider whether the Fkh1-Fdo1 interaction might be occurring through a less standard protein-protein interaction mechanism. Two such possibilities were that Fdo1 and Fkh1 were interacting via a liquid-liquid phase separation (LLPS) mechanism or a prion-like mechanism, wherein heritable protein structural conformations can be adopted that can in turn create fibrils or substructures that create new interaction surfaces. In particular, large intrinsically disordered regions (IDRs), such as those present on the N- and C-terminal thirds of Fdo1, can promote both of these non-classic mechanisms [133]. In addition, IDRs enriched in asparagine (N) and/or glutamine (Q) have been linked to both mechanisms. For example, IDRs drive LLPS mechanisms linked to RNA processing granules, which includes the nucleolus, the site of rRNA synthesis [134]. The Fdo1 sequence contained many Q/N residues (**Figure 3.3.c**); a more quantitative assessment performed by Dr. Hoggard in the lab revealed that Fdo1 ranked in the top 2% of Q content and the top 3% of N content among the ~6700 yeast proteins. Thus, it is plausible that Fdo1 and Fkh1 interactions are occurring through a less-conventional mechanism discussed here.

Nevertheless, a *fdo1-16TA* mutant phenocopied a *fkh1-R80A* mutant according to my TurboID assay, suggesting that one or more threonine residues were important for this Fkh1-Fdo1 interaction detected *in vivo* (**Figure 3.4**). It will be worthwhile to systematically test more combinations of T→A substitutions to identify the minimal set of mutations that phenocopies *fkh1-R80A*. However, if a less conventional mechanism is relevant, we might expect that a distinct FHA-pT-binding peptide will not be identifiable. For a mechanism like LLPS, one

could imagine that phosphorylation of a certain quorum of threonine residues distributed across certain region(s) of Fdo1, rather than the phosphorylation of particular threonine residues, might be more relevant. In this scenario, the phosphorylation state across multiple Fdo1 threonine residues might be able to regulate formation of an LLPS. Phosphorylation has been shown to regulate an LLPS mechanism required for heterochromatin spreading [135].

3.3.4. The presence of both a Fdo1 ortholog and an intact Fkh1 ortholog within the subphylum *Saccharomycotina* correlated with efficient homologous recombination.

As discussed above, mating-type switching is a specialized form of homologous recombination involving the mating-type cassettes on chromosome III, but the primary distinguishing factor is the formation of a double-strand break (DSB) at the *MAT* locus by HO endonuclease [136]. Therefore, many of the factors that bind to and recognize the *MAT* DSB are probably also used in HR more generally, regardless of locus. Supporting this idea, a recent study showed that tethering the Fkh1 protein to a donor DNA template improved the efficiency of homologous recombination repair of a targeted CRISPR-Cas9-mediated double-strand break by over fivefold in yeast [15]. The requirement for Fdo1 in Cas9-initiated homologous recombination was not investigated.

FDO1 is a non-conserved gene found only in a few closely related yeast species, so in some ways one might not expect it to be involved in a process that is as basic and conserved as HR. However, *S. cerevisiae* is famous among model organisms in large part for its ability to perform highly efficient HR. We were therefore curious about whether there might be evolutionary sequence evidence for a link between Fdo1 and efficient HR. We used a Y1000+ database that contains draft genomes of over 1000 yeast species in the subphylum *Saccharomycotina*

(Opulente et al. 2023). We focused analyses on the 86 *Saccharomycotina* species discussed in [125] (**Figure 3.5.a**). Specifically, we looked for alignments in Y1000+ to *FDO1*, *FKH1* and *MPH1* using TBLAST-N algorithm. While alignments to regions of *MPH1* and *FKH1* could be found in these species, the ability to find any significant alignments to *FDO1* was confined to the species most closely evolutionarily linked to *Saccharomyces cerevisiae*.

The literature was then searched for experimental evidence concerning HR in these species, and experimental data for the 14 species marked in **Figure 3.5.a** is included in the Table presented in **Figure 3.5.b**. Five of the 14 species in **Figure 3.5.b** had strong matches to *FDO1* and to both the *FKH1* FHA and DBD domains, and four out of five were capable of efficient HR using donors with short homology arms of only 100 bp. Within this limited data set, the presence of both *FDO1* and an intact *FKH1* were most closely associated with the ability to perform efficient HR. Thus, *FDO1* may contribute to efficient HR but it is not the only feature in yeast responsible for this property. Based on these data, it is intriguing to speculate that a Fdo1-Fkh1 interaction may be a substantial contributing factor in yeast for enabling efficient HR.

However, it is important to acknowledge that this *in silico* work can only a reveal a correlation, and a correlation can occur without any cause-and-effect relationship. One confounding factor is that the close evolutionary relationship among the species examined simply makes it easier to identify sequence homologs of *FKH1* and *FDO1* in the first place. Species closely related to *S. cerevisiae* will carry fewer changes across all genes than more distantly related species will. This issue is particularly pertinent to Fdo1's amino acid sequence because if it indeed functions as a largely unstructured protein, it may tolerate many different primary sequences and still be able to perform its function. These facts may point to searching for structural homologs as a more informative strategy.

Figure 3.5. Fdo1 and Fkh1 may be linked to higher efficiency homologous recombination in *Saccharomycotina*.

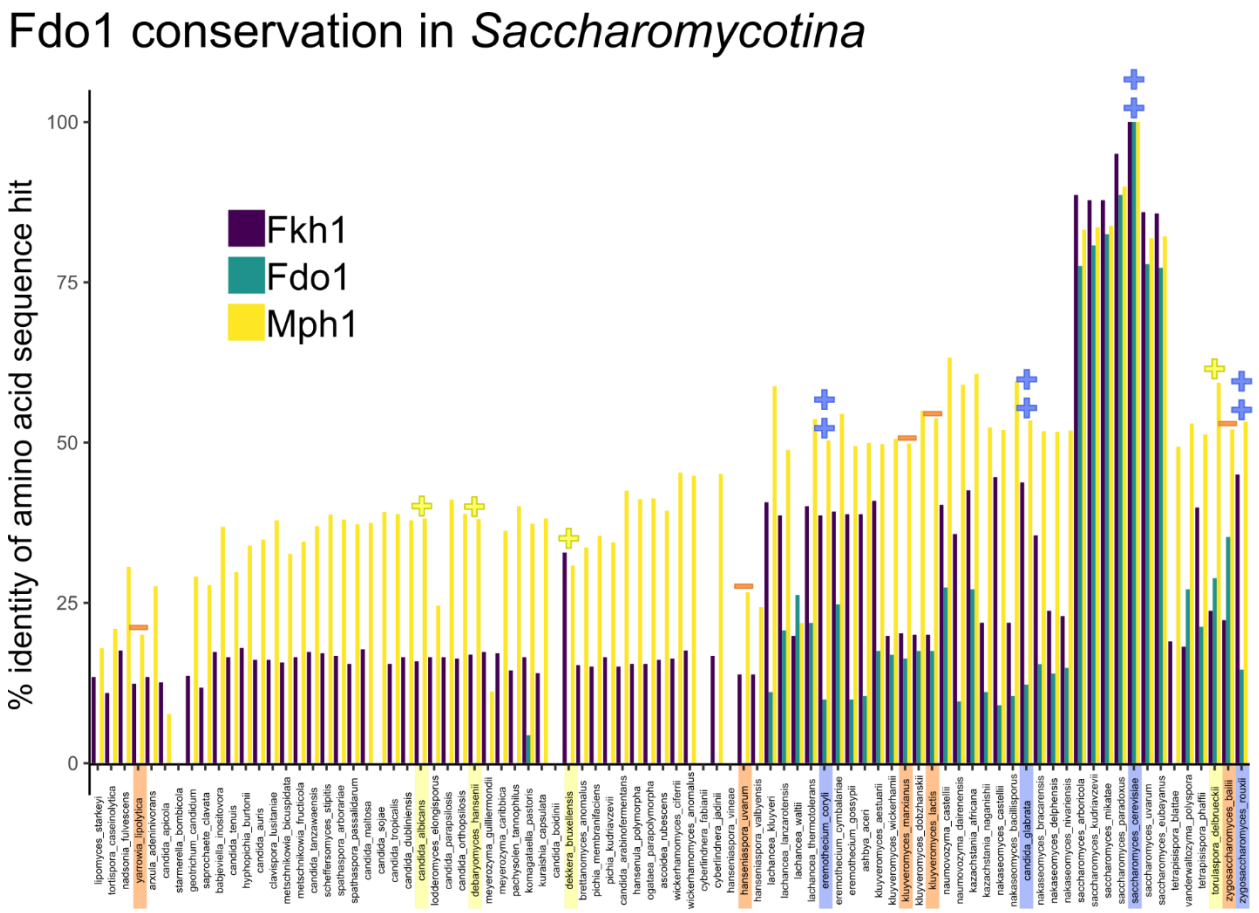
A. The full-length amino acid sequence of *S. cerevisiae* Fkh1 was used to search *Saccharomycotina* yeasts present in the Y1000+ database for matches that share high similarity to *S. cerevisiae* Fkh1. This histogram plots the percent identity between these matches and *S. cerevisiae* Fkh1. The search was repeated for Fdo1 and Mph1. Species are ordered according to their position on a phylogenetic tree [137].

B. Species listed in A were also searched for matches to each of Fkh1's domains individually: the FHA domain and the DNA binding domain (DBD). The presence of a match to each of these domains is noted, as well as whether a Fdo1 match is present.

A literature search was performed to examine whether species in this list can perform homologous recombination efficiently. To our knowledge, a systematic comparative study of homologous recombination efficiency across many yeast species has never been undertaken, so the data are limited and somewhat anecdotal.

For the purposes of this study, species able to integrate linear fragments with <100 bp homology arms were considered efficient at homologous recombination (++). Species requiring longer homology arms, between 100 and 400bp, were in the intermediate category (+). Species requiring homology arms of greater than 400bp or for whom homologous recombination is noted as inefficient (often with a marked preference for non-homologous end joining over HR) were noted for their poor HR efficiency (–).

Figure 3.5



Organism	FHA	DBD	Fdo1 match?	Required homology arm length:
Saccharomyces cerevisiae	yes	yes	yes	30bp [138]
Candida glabrata	yes	yes	yes	100bp [139]
Eremothecium coryli	yes	yes	yes	“short” [140]
Naumovozyma castellii	yes	yes	yes	200-230bp [141]
Zygosaccharomyces rouxii	yes	yes	yes	70-90bp [142]
Debaryomyces hansenii	yes	no	no	400bp [143]
Dekkera bruxellensis	yes	yes	no	165-220bp [144]
Torulaspora delbrueckii	yes	no	yes	320-345bp [145]
Candida albicans	yes	no	no	285-308bp [146]
Hanseniaspora uvarum	no	yes	no	~1000bp [147]
Kluyveromyces lactis	yes	no	yes	600bp [148]
Kluyveromyces marxianus	yes	no	yes	NHEJ predominant [149]
Yarrowia lipolytica	yes	no	no	500-1000bp [150]
Zygosaccharomyces bailii	yes	no	yes	400-700bp [151]

It is plausible that many proteins beyond Fkh1 and Fdo1 not examined in these analyses are contributing to the high efficiency of HR in *S. cerevisiae*. Thus, it is important to perform a direct experimental test of HR efficiency in various mutant backgrounds.

3.3.5. A colony color assay to measure HR efficiency in haploid yeast.

To measure whether mutations in Fkh1 or Fdo1 or defects in their interaction, as measured by my Fkh1-TurboID labeling assay, affected general HR efficiency in yeast, I established a simple colony colorimetric assay to measure HR efficiency taking advantage of the *ADE2* gene. The *ADE2* gene encodes a protein required for adenine biosynthesis. The *ade2-1* allele leads to a block in the step controlled by Ade2, which causes the accumulation of a biosynthetic intermediate that turns cells red, leading to pink or red colonies. I used CRISPR-Cas9 editing to target the *ade2-1* mutant gene, and provided a donor fragment that would repair the DSB break, convert the gene to *ADE2*, and disable Cas9 from targeting this repaired gene again. The expected outcomes of this assay are depicted in **Figure 3.6.a**. The bottom line is that the appearance of white colonies means that the *ade2-1* allele has been successfully repaired to *ADE2* by homologous recombination.

I have tested *FKH1*, *fkh1-R80A*, and *fkh1Δ* strains using this system (**Figure 3.6.b**). Homologous recombination efficiency was highest in the *FKH1* yeast, where about 50% the colonies were white. The *fkh1-R80A* and *fkh1Δ* strains showed reduced HR efficiency, where about 40% of colonies were white. For these experiments I used a 2432bp ADE2 fragment with a 5' homology arm of 300bp and a 3' homology arm of 2132bp on either side of the DSB. These arms are far longer than required for high efficiency homologous recombination in *S. cerevisiae*, so this assay can certainly be sensitized further (i.e. to see a bigger difference between *FKH1* versus *fkh1* mutants) by reducing the length of these homology arms. But even at the level the

assay is currently reporting, it will be worth examining *fdo1* mutants and *fdo1 fkh1* double mutants. A quantitative and more comprehensive analysis will be important for challenging the contributions of *FKH1* and *FDO1*, alone and in combination, to high-efficiency HR in yeast. While HR has been studied extensively in yeast, the major focus has been on larger defects than what I hypothesize might be relevant here. It may be useful to adapt the experiment to use HR to reconstitute a GFP reporter gene. With such an assay, many more cells could be examined in a short amount of time by flow cytometry, allowing for greater statistical power when dealing with subtle differences between phenotypes.

3.3.6. *FDO1* helped maintain the yeast rDNA array.

Another phenotype that can be assessed to address whether a mutation affects DNA replication or HR efficiency is the number of rDNA copies present in yeast rDNA array. The yeast rDNA array is a single contiguous locus on chromosome XII that encodes multiple tandem copies of a 9.1 kb gene cassette; this cassette includes all the information necessary to synthesize the essential ribosomal RNAs as well as additional elements that contribute to replication of the rDNA array and copy number maintenance (**Figure 3.7.a**). ~200-400 copies of this cassette are normally present in yeast, which constitutes between ~2-4 Mb. The yeast genome is only 12 Mb; therefore, the rDNA array can constitute ~15-25% of the total nuclear DNA. Thus, by simply assessing DNA content by flow cytometry, we can observe a preliminary hint if the rDNA array is expanded or contracted compared to a typical wild-type rDNA copy number. We therefore examined total DNA content by flow cytometry in three different genotypes: wild-type, a *fdo1Δ* mutant, and a 35-copy rDNA mutant (**Figure 3.7.b**). The 35-copy rDNA mutant is an engineered yeast with only 35 copies of the rDNA cassette that cannot be further expanded beyond this number during a normal mitotic cell cycle.

Figure 3.6. Colony color assay measures homologous recombination efficiency at *ADE2* locus.

A. Schematic showing multiple possible outcomes after yeast cells are transformed by a Cas9 enzyme with a guide RNA targeting *ade2-1* and a linear donor fragment derived from *ADE2*. A white colony will only arise by the generation of a double-strand break in *ade2-1* followed by repair by homologous recombination (HR) with the provided *ADE2* fragment. Other outcomes (repair by non-homologous end joining (NHEJ) or absence of a Cas9-generated double-strand break) can only result in red colonies. Unrepaired double-strand breaks pose some level of toxicity to cells, so cell death is also a possible outcome (row 4).

B. Example of resulting transformation plate. Only white and red colonies were used in calculating the fraction of white colonies; sectored, peach, and otherwise indeterminate colonies were not counted for this calculation. W, white colony. R, red colony. S, sectored colony. P, peach colony.

C. Equation used when calculating the fraction of white colonies.

D. Box plot showing the fraction of white colonies in each *FKH1* genotype over 4 trials.

Figure 3.6.

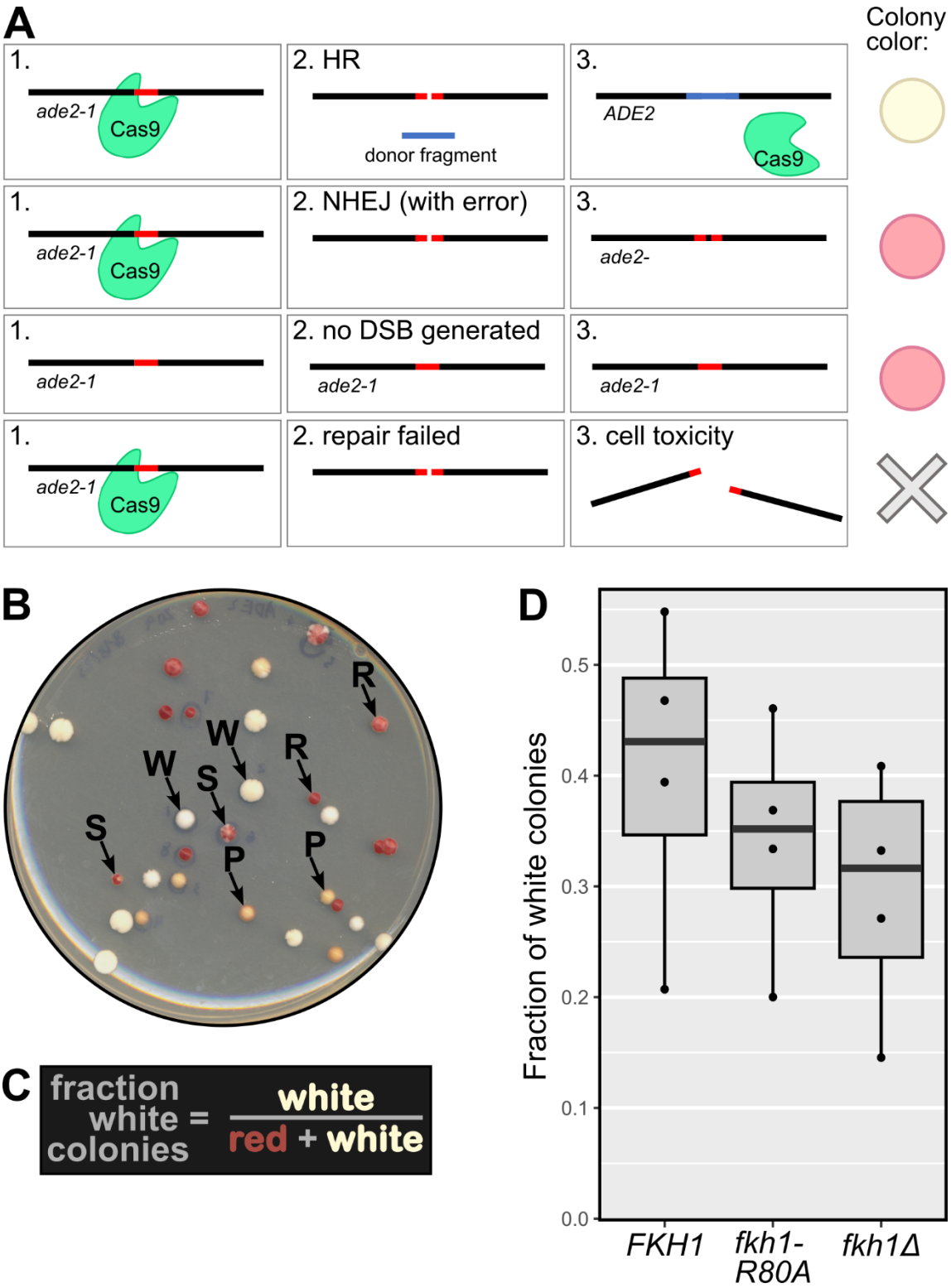
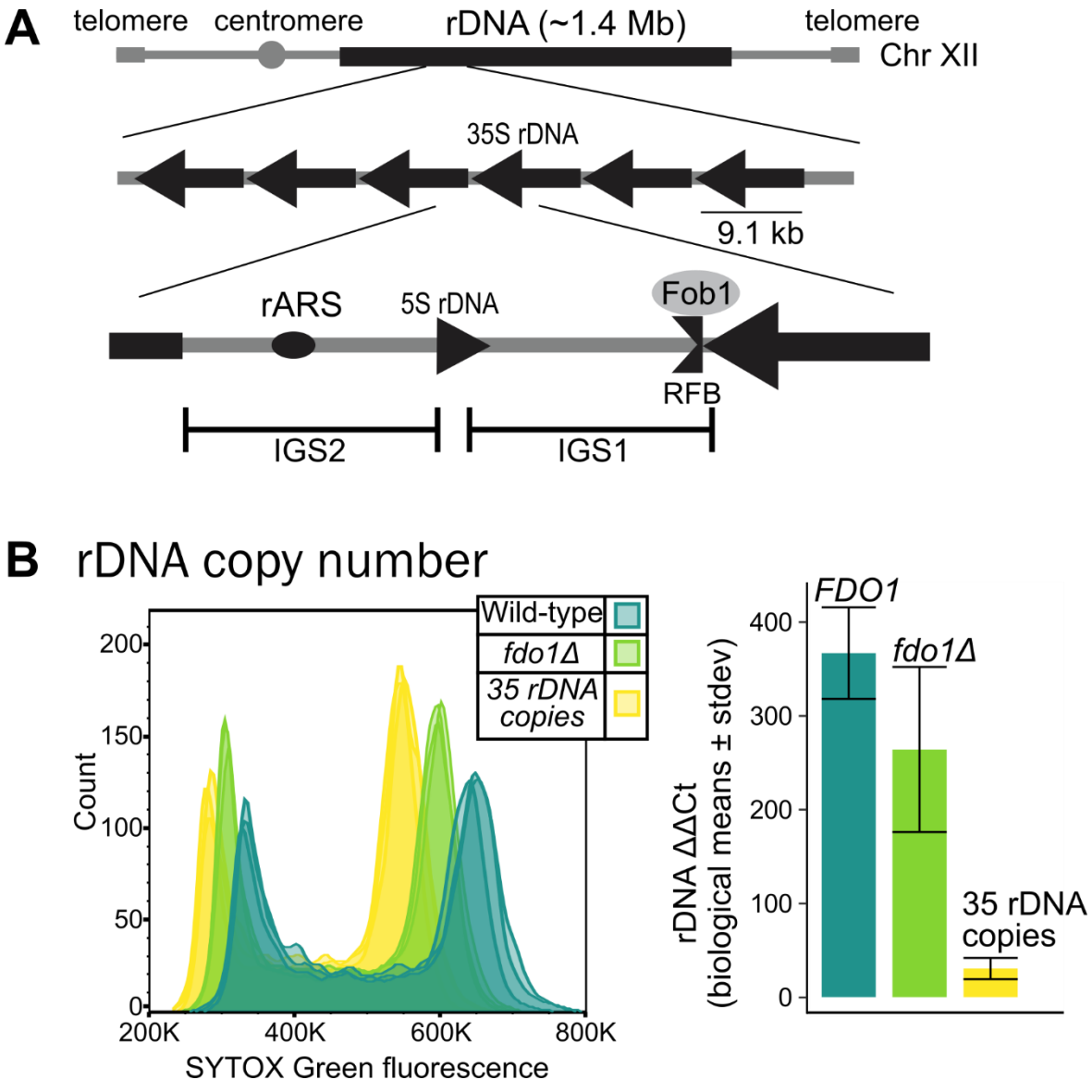

A
1.
ade2-1
Cas9
2. HR
donor fragment
3.
ADE2
Cas9
Colony color:
1.
ade2-1
Cas9
2. NHEJ (with error)
3.
ade2-
1.
ade2-1
2. no DSB generated
ade2-1
3.
ade2-1
1.
ade2-1
Cas9
2. repair failed
3. cell toxicity
B
R
W
W
S
R
S
P
P
C
fraction white colonies = white / red + white
D
Fraction of white colonies
0.0
0.1
0.2
0.3
0.4
0.5
FKH1
fkh1-R80A
fkh1Δ

Figure 3.7. rDNA copy number decreases in *fdo1Δ* strain.

A. Schematic showing the ribosomal DNA (rDNA) repeats and the mechanism of rDNA copy number expansion by replication fork blocking.

B. Left: Flow cytometry profiles showing the DNA content (SYTOX Green fluorescence intensity) of three strains. Right: rDNA copy number of these strains as measured by ddPCR.



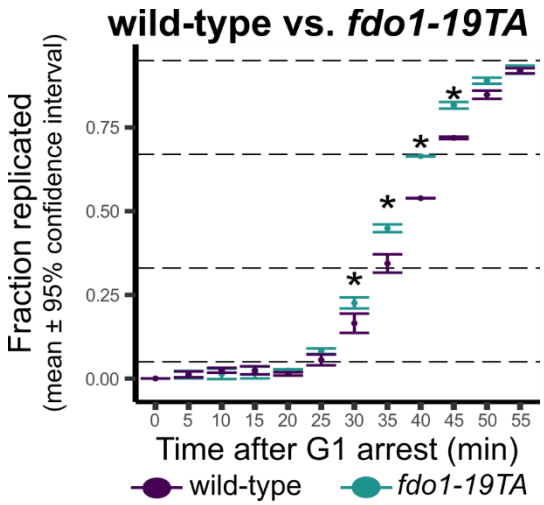
Based on analyses of multiple independent replicates of each of these strains, the *fdo1Δ* mutant contained fewer copies of the rDNA cassette than wild-type cells, but greater than ~35 copies. To measure rDNA copy number directly in these cells, droplet digital PCR was also performed (**Figure 3.7.c**). These data were consistent with the flow cytometry data. The observation that loss of the *FDO1* gene leads to contraction of the rDNA array (a reduction in the steady-state copy number) suggests that *FDO1* is important for DNA replication, HR, or both processes [152].

3.3.7. A *fdo-19TA* mutant progressed more rapidly through S-phase.

To address whether *FDO1* promoted normal DNA replication progression, we performed a cell cycle arrest and release experiment (see **Figure 2.9.a**). Briefly, wild-type (*FDO1*) or *fdo-19TA* mutant cells were arrested in G1-phase and then released into S-phase. The cells were assessed at discrete timepoints during this synchronous S-phase progression, and subsequently their DNA content was measured by flow cytometry (**Figure 3.8**). At each time point starting in early-to-mid S-phase, the *fdo1-19TA* mutant cells showed a greater DNA content relative to *FDO1* cells, suggesting that the DNA content of the mutant was increasing faster than that of the *FDO1* control. This phenotype is similar to that caused by the *cdc7-T484A* allele discussed in Chapter 2. As discussed in that chapter, this phenotype would be predicted or expected in cells that had a defect in the S-phase checkpoint (i.e. mechanisms that survey fidelity of DNA replication intermediates to prevent genome replication from forging ahead in the presence of genome-threatening challenges). The molecular nature of this phenotype calls for Sort-seq experiments to map replication intermediates across chromosomes as in Chapter 2. If Fdo1 only functioned to resolve DNA replication stress issues such as by promoting HR when DSBs arise, I would not predict a faster S-phase progression phenotype.

Figure 3.8. *fdo1-19TA* shows faster progression through S-phase.

S-phase profile showing the fraction replicated at each time point, where 1n DNA content is 0% replicated and 2n DNA content is 100% replicated. Wild-type strain is compared with *fdo1-19TA* strain. Two independent biological replicates were examined for each genotype.



Rather, I would expect a slowdown of DNA replication as cells tried to manage any DSBs that arise via less efficient means. Thus, this phenotype, together with the data above, suggests that *FDO1* may promote both efficient HR-mediated repair of DSBs as well as signaling to the S-phase checkpoint pathway. To address this issue, it will be important to assess key markers of the S-phase checkpoint, such as Rad53 phosphorylation, and to ask whether combining *fdo1* mutants with checkpoint-defective mutants causes synthetically lethal phenotypes.

3.4. Materials and methods

3.4.1. *In silico* analysis of phosphothreonine residues

Phosphothreonine residues in Fdo1 were analyzed using STRIP [78], which predicts pT probability using two tools (NetPhos 2.0 and DisPhos 1.3) and summarizes the output of five kinase prediction tools (NetPhosK, PredPhospho, PPSP, ScanSite, and KinasePhos) to predict the relevant kinase that likely phosphorylates each residue. Peptides were considered likely to be phosphorylated if either pT probability value was greater than 0.500.

Cell cycle-linked kinases often belong to the CMGC kinase family, and many known replication-related proteins are targets of CMGC group kinases, so a CMGC kinase prediction increased a match's score in this analysis.

Databases (SGD, BioGRID) were searched to find experimental evidence of phosphorylation at each position. For Fdo1 peptides, the experimental evidence consisted of mass spectrometric evidence for a phosphorylated threonine residue at that position. However, for peptides from Mph1, additional strong experimental evidence based on genetic and biochemical approaches exists [20].

Each candidate peptide was aligned to each of the two Mph1 phosphoepitopes in Jalview 2.11.4.0 and the average per-residue conservation score was computed [109]. The higher of the

two scores was divided by the maximum alignment score to yield a percentage, listed in the "alignment to Mph1 peptide" column. The cutoff for a "good match" was set at 75% or higher.

Based on the criteria listed here, candidates could receive a score of up to 4 stars.

3.4.2. Prepare sgRNA and donor to generate mutant strains using CRISPR-Cas9

An all-in-one plasmid was constructed by inserting Cas9 from pML104 into pDB18 (Conrad A. Nieduszynski lab) and abolishing two BsaI sites, creating pCF4518 [113,114]. sgRNA oligo pairs were annealed to one another in a reaction containing 10 mM Tris pH 8.0, 25 mM NaCl, and 100 uM of each primer, which was incubated in a thermocycler at 80°C for 5 minutes, then the temperature was decreased by 0.1°C every 10 seconds until reaching a final temperature of 20°C.

Annealed oligos were inserted into the remaining BsaI site on pCF4518. A phosphorylation reaction was set up with 1.5 µl of annealed oligos diluted 1:200, 100 ng of BsaI-cut gel purified pCF4518, 1X T4 DNA ligase buffer, and 0.5 µl T4 PNK (NEB #M0201) in a final volume of 9.5 µl. The phosphorylation reaction was incubated at 37°C for 30 minutes, then the temperature was decreased to 16°C and 0.5 µl T4 DNA ligase (NEB #M0202) was added. The ligation reaction was incubated at 16°C for 16 hours, then enzymes were heat inactivated at 65°C.

Donor fragments were designed to create the desired mutation and a silent mutation to abolish the corresponding sgRNA's PAM site was also added. 120 bp homology arms on either side of these mutations were added, and the entire sequence was flanked by restriction sites and cloned into a pUC vector backbone. 4 minipreps' (NEB #1110) worth of plasmid DNA eluted in 200 µl of warmed 10 mM Tris pH 8.5 were digested with restriction enzymes (NEB high-fidelity

versions) overnight at 37°C. Digests were ethanol precipitated, dissolved in 40 µl of 10 mM Tris pH 8.5, and run on an agarose gel. Bands corresponding to the donor fragment size were excised and gel purified (NEB #T1020) eluted in 8 µl of warmed 10 mM Tris pH 8.5.

3.4.3. *E.* coli transformation

DNA (100 ng plasmid or 5 µl ligation) was added to sterile 1.5 mL Eppendorf tubes. DH5α competent cells (Thermo) were thawed on ice, then 50-100 µl of cell suspension was added to each DNA tube and tubes were flicked to mix. Tubes were incubated on ice for 30 minutes, heat shocked at 42°C for 30 seconds, and incubated on ice for 5 minutes. SOC broth was added to 1 mL and tubes were incubated with shaking at 37°C for 1 hour. LB plates with the appropriate antibiotic were spread with 40 µl 100 mM IPTG and 40 µl 20 mg/mL (w/v) X-gal if blue-white screening was necessary. Recovered cells were plated on LB plates and incubated at 37°C for 14-24 hours.

3.4.4. Yeast transformation

S. cerevisiae were grown to OD600/mL = 0.4-0.6. For 4 transformations, 2 OD equivalents were pelleted at 3000xg for 30 seconds. Pellet was washed with 800 µl of TEL buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM LiOAc). Cells were pelleted at 3000xg for 1 minute and the supernatant was pipetted off. Pellet was resuspended in 160 µl of TEL buffer and 40 µl was aliquoted into each transformation tube. Transformation tubes contained DNA (for a CRISPR transformation, 200 ng plasmid DNA and/or ~1 ug linear donor fragment were used). 150 µl PEG-TEL (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM LiOAc, 40% PEG 3500) was added to each transformation tube and tubes were inverted vigorously to mix. Tubes were incubated at 30°C for 30 minutes and 42°C for 20 minutes. Tubes were centrifuged at 1000xg for 30 seconds and supernatant was pipetted off. Pellets were resuspended in 200 µl YMD media

(20g/L glucose, 134g/L yeast nitrogen base with ammonium sulfate (USBio #Y2025)) and the full volume was plated on selective media. Colonies were allowed to grow up to 1 week to allow for red color development.

3.4.5. Protein immunoblotting

Samples were run on 6% or 8% polyacrylamide gels submerged in running buffer (0.1% w/v SDS, 25 mM Tris, 192 mM glycine) at 160 V for ~40 minutes. Transfer sandwiches were soaked in transfer buffer (25 mM Tris, 192 mM glycine, 20% v/v methanol, 0.005% w/v SDS), assembled with pre-soaked nitrocellulose membrane and placed in a BioRad Mini Trans-Blot apparatus, which was cooled with ice and placed at 4°C. Proteins were wet transferred at 100 V for 1 hour. Membranes were rinsed of methanol and cut to size, with an optional Ponceau staining (0.5% w/v Ponceau S, 1% v/v acetic acid) and imaging step. While rotating on an orbital shaker, membranes were blocked with 10 mL blocking buffer (5% skim milk, 20 mM Tris, 150 mM NaCl, 0.01% sodium azide). Blocking buffer was removed and primary antibody (1:1000 monoclonal mouse anti-Cdc7, 1:2000 polyclonal rabbit anti-Fkh1, or 1:250 mouse anti-FLAG (Cell Signaling #D6W5B)) diluted in 10 mL blocking buffer was added. Blots were incubated with primary antibody overnight. Blots were washed for 3x 15-minute washes with 20 mL TBST (20 mM Tris, 150 mM NaCl, 0.1% v/v Tween-20). Secondary antibody (usually 1:10000 anti-mouse-HRP or 1:20000 anti-rabbit-HRP) diluted in 10mL TBST was added and blots were incubated with secondary antibody for 1 hour. Alternatively, primary and secondary antibodies were omitted and blots were instead incubated with 1:5000 anti-FLAG-HRP (Sigma #A8592) or 1:8000 streptavidin-HRP (Thermo #N100) for 1 hour. Blots were washed for 3x 15-minute washes with 20mL TBST. Blots were incubated with Thermo Fisher SuperSignal West Pico PLUS Chemiluminescent Substrate for 1 minute. Blots were imaged on an Azure C600

imager with manual exposure times of 1 second, 2 seconds, 5 seconds, 10 seconds, 30 seconds, 2 minutes, and 5 minutes. Where necessary, blots were stripped with mild strip buffer (1.5% w/v glycine, 1% w/v SDS, 1% Tween 20, pH 2.2), washed for 3x 15-minute washes with TBST, then proceeded starting at the blocking step.

3.4.6. Whole protein extract

Harvest 1.5 OD equivalents of *S. cerevisiae* and wash once with water. Resuspend in 1 mL 2 M LiOAc and incubate on ice for 5 minutes. Pellet and remove supernatant. Resuspend in 1mL 0.4 M NaOH and incubate on ice for 5 minutes. Pellet and remove supernatant. Resuspend in 100ul 1X SDS-PAGE sample buffer (60 mM Tris-HCl pH 6.8, 10% v/v glycerol, 2% w/v SDS, 5% v/v 2-mercaptoethanol, 0.0025% w/v bromophenol blue) and incubate at 95°C for 5 minutes. Store at -20°C. Before loading on gel, incubate at 95°C for 2 minutes and centrifuge at 10000xg for 1 minute.

3.4.7. Streptavidin shift assay

Adapted from [86]. Centrifuge sample at 10000xg for 1 minute to clear cell debris. In 0.2 mL tubes, add 0.3 µl of 10 ug/ul streptavidin (Sigma #189730) and 6ul sample. Incubate with rotation for 10 minutes at room temperature. Immediately load all of each sample on SDS-PAGE gel).

3.4.8. Quantitation of protein immunoblot bands

Images were captured on an Azure C600 imager. Only images where no portion of the bands of interest were saturated were used for quantitation. In cases where bands were hard to resolve from neighboring bands, the blot image was imported into Inkscape, where a rectangle of consistent size was used to clip individual bands from the main blot image. The image was exported as a .tiff file and imported into ImageJ. The image was inverted, then each band was

selected using the Rectangle tool and its intensity was measured using the Measure function. The RawIntDen value was recorded and used for further calculations. Percent shift was calculated by dividing the RawIntDen value of a shifted band by the sum of the values of the shifted and unshifted bands.

3.4.9. Y1000+ database search

The Y1000+ database was searched using TBLASTN with default parameters. Species present in [137] were considered. Peptides queried were Fkh1's FHA domain (aa50-202), Fkh1's DBD domain (aa291-420), and full-length Fdo1. We focused on the alignment with the most matching residues defined as the product of the alignment length and percent identity relative to *S. cerevisiae*. The percentage identities for each Fkh1 interactor in the various species were plotted based on each hit's 'full-length' percent identity relative to the *S. cerevisiae* sequence.

3.4.10. Flow cytometry

To prepare genomic DNA samples for copy-number analyses, *MAT*a cells were grown in 25 mL YPD at 30°C to an OD_{600} of ~0.5. To determine cell cycle distribution of each sample, cells were fixed with ethanol, stained with SYTOX Green and analyzed by flow cytometry (BD Accuri C6 Flow Cytometer).

3.4.11. ddPCR for rDNA copy number

ddPCR was performed following a published protocol [115] for examining yeast DNA replication using EvaGreen-based chemistry (Droplet Oil and Supermix; BioRad #1864005 and #1864033, respectively). ddPCR was performed with primers (oCF8765, oCF8766) listed in Appendix.

3.4.12. Sort-seq

Yeast were grown at 30°C and processed as described in [115,116] except that a Sony MA900 was used to separate and collect S-phase and G2-phase cells. Genomic DNA was isolated as described [115]. Libraries for one wild-type sample were prepared at Northwestern University and single-end sequenced. The remaining libraries were paired-end sequenced in-house using a P2-100 flow cell and a NextSeq1000. The libraries for in-house sequencing were made using the 1/10 dilution of the adaptor and amplified for seven cycles using the NEBNext Ultra II kit. Sequencing reads were mapped to the yeast genome, build sacCer3, using Bowtie2 [117] with the following options: bowtie2 -p 4 -X 750 -q –phred33. Mapped reads were converted to single nucleotide read coverages using Bedtools genomecov [118]. Single nucleotide coverages were normalized for sequencing depth and breadth [119] then summarized into 1 kbp bins. S/G2 ratios were determined for each independent cell population (three independent populations of *FKH1* cells and two independent populations of *fkh1-R80A* cells) using the 1 kbp bin values. Outlier S/G2 ratios were defined as any ratio that was over 1.5 IQRs below the first quartile (Q1) or above the third quartile (Q3) and were removed. The resulting set of ratios were next scaled to be between 1 and 2 using the following equation: scaled ratio = (ratio - min(ratio))/(max(ratio) - min(ratio)) + 1. The resulting scaled ratios were smoothed by fitting data to a cubic smoothing spline (R function: smooth.spline, default parameters). Smoothed data from a given genotype (three samples for *FKH1* and two samples for *fkh1-R80A*) were statistically summarized by finding the means and 95% confidence intervals.

Table 3.1. Yeast strains used in this study.

Name	Description	Plasmid (if applicable)	Source
CFY4606	*MATa, ADE2+, lys2Δ, fob1Δ::HIS3,* low rDNA copy number		This study
CFY4662	*MATa, ADE2+, lys2Δ, FOB1+, FDO1+*		This study
CFY4666	*MATa, ADE2+, lys2Δ, FOB1+, fdo1Δ*		This study
CFY4975	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, MPH1-3xFLAG::HIS3*	pCF4587	This study
CFY4976	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, MPH1-3xFLAG::HIS3*	pCF4588	This study
CFY4977	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, FDO1-3xFLAG::HIS3*	pCF4587	This study
CFY4978	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, FDO1-3xFLAG::HIS3*	pCF4588	This study
CFY4989	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, fdo1-4TA-3xFLAG::HIS3*	pCF4587	This study
CFY4991	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, fdo1-19TA-3xFLAG::HIS3*	pCF4587	This study
CFY4992	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, fdo1-16TA-3xFLAG::HIS3*	pCF4587	This study
CFY4993	*MATa, ADE2+, RAD5+, fkh1Δ::HisG, fdo1-5TA-3xFLAG::HIS3*	pCF4587	This study
CFY4994	*MATa, ADE2+, RAD5+, FKH1+*		This study
CFY4995	*MATa, ADE2+, RAD5+, fkh1-R80A*		This study
CFY4996	*MATa, ADE2+, RAD5+, fkh1Δ*		This study

Table 3.2. Plasmids used in this study.

Name	Description	Source
pASZ11	Vector containing *ADE2* gene	[153]
pDB18	CRISPR plasmid containing sgRNA	[114]
pML104	CRISPR plasmid containing Cas9	[113]
pCF4518	Modified pDB18 with Cas9 from pML104	This study
pCF4587	Fkh1-TurboID flanked by *FKH1* pro/term in pRS316	This study
pCF4588	Fkh1^{R80A}-TurboID flanked by *FKH1* pro/term in pRS316	This study
pCF4631	Modified pCF4518 CRISPR plasmid targeting *ade2-1*. Target: TACCCTTCTCCAGAAACAAT(CGG)	This study

Chapter 4: Future directions

4.1. Using the FHA-dependent Fkh1-TurboID system for discovering FHA-dependent Fkh1 protein partners

One of the major reasons I set up the Fkh1-TurboID system was to identify Fkh1-FHA-dependent partner proteins in a non-biased manner using mass spectrometry. Soon after establishing the Fkh1-TurboID system, I tried a mass spectrometry experiment in which I enriched for biotinylated proteins from samples enriched for chromatin and chromatin-associated proteins, which were generated from yeast expressing either Fkh1-TurboID or Fkh1^{R80A}-TurboID, the system I described in Chapter 2. The hope was that peptides present in the Fkh1-TurboID experiment but absent (or lower in abundance) in the Fkh1^{R80A}-TurboID experiment would represent strong candidates for FHA-dependent partners of Fkh1. However, I was not able to convincingly identify any peptides meeting this criterion. More importantly, I did not identify any peptides from known interaction partners of Fkh1, including Mph1 and Fdo1, so my important benchmarks were not identified. Instead, the majority of proteins identified were generally abundant yeast proteins. I think the reason for this lack of success was two-fold: 1. enriching for streptavidin-bound proteins from chromatin-enriched samples did not provide sufficient purification to avoid contamination by highly abundant proteins that were not relevant to Fkh1 biology; thus, the mass spectrometry data were swamped out by "background" peptides from abundant proteins in yeast, and 2. yeast contain a high number of biotinylated proteins naturally, including some abundant proteins, that complicate using a TurboID system for unbiased protein purification [154]. Yeast mutants and approaches have been generated recently to ameliorate the contamination caused by this biology, and it is possible that my Fkh1-TurboID system could be adapted for these systems. However, based on current progress in the lab and

reagents and strains we have in hand, I favor two different but not necessarily mutually exclusive approaches:

4.1.1. Using nuclear extracts as the starting material

I have observed that most of the background biotinylated protein signal is caused by proteins localizing outside of the nucleus. Because Fkh1 functions in the nucleus and we are most interested in nuclear processes, an obvious strategy is to purify yeast nuclei and use nuclear extracts for the streptavidin pulldown enrichment step prior to mass spectrometry. However, generating nuclear extracts from yeast is a challenging, labor-intensive procedure that is prone to loss of much of the starting material so it generates low numbers of purified nuclei. An alternative may be to modify the current chromatin enrichment protocol used in this first-pass experiment to include additional clarification steps and higher-stringency washes to remove most of the cytoplasmic contaminants present in the crude chromatin pellet. The chromatin could then be digested and the resulting extract centrifuged at high speed to remove any larger contaminants, such as membrane fragments or ribosomes, and leave chromatin-associated proteins in the soluble fraction. This soluble fraction would then be used for the streptavidin pulldown step. A technical alternative may be to use the mechanized higher-volume Dounce homogenizer available in Dr. Aaron Hoskins' laboratory to more easily and reproducibly release nuclei from cells, perhaps allowing me to generate larger amounts of nuclear extracts.

4.1.2. Using cell-cycle arrested cultures

One of the processes that I am most interested in is origin licensing and the earliest steps of origin remodeling prior to the initiation of DNA synthesis because these are the steps where the Fkh1-FHA domain is likely acting [60]. The potential partners of the Fkh1 FHA domain are likely contacting Fkh1 during this relevant point in the cell cycle, which I estimate to be mid-to-

late G1-phase and/or very early S-phase. I did attempt to arrest yeast in G1-phase with alpha factor but, perhaps interestingly, I could not observe the interaction between Fkh1 and Mph1 with my TurboID system in extracts made from these arrested cultures. I also could not observe this benchmark interaction in M-phase arrested yeast cultures. However, soon after cells were released from either the G1- or the M-phase arrests and began to move through the cell cycle again, I could observe the Fkh1-Mph1 interaction using the Fkh1-TurboID system. I do not yet know the explanation for these observations, but they discouraged me from investing a lot of time in making extracts from arrested cultures because the Fkh1-Mph1 reaction I had been relying upon would not be available as a benchmark to ensure my methods were working as expected.

A current method we have prepared to try in the lab involves synchronizing cells in M-phase using the temperature-sensitive *cdc15-2* allele. At the non-permissive temperature, *cdc15-2* cells arrest in late M-phase without completing cell division (cytokinesis). However, when released at permissive temperature, the cells synchronously pass through G1-phase and into S-phase. The goal would be to prepare extracts from cells as they traverse the late G1 → early S-phase boundary and identify biotinylated proteins from these extracts. This approach might also help me obtain better Fkh1-TurboID data for targeted experiments to examine specific components of the initiation machinery, such as Cdc7 or subunits of ORC. The challenge with these experiments is optimizing conditions for efficient arrest and synchronous release. While the methods are theoretically established, small changes, such as selective media growth required for the Fkh1-TurboID system, can alter the efficiency of both arrest and release.

4.2. Differential enrichment of pT-containing peptides by purified FHA domains followed by peptide identification by mass spectrometry

A key finding in Chapter 2 concerned the differences in FHA-pT peptide specificity I observed for the two similar FHA domains, the Fkh1-FHA domain and the Rad53-FHA1 domain. Specifically, the Mph1 peptide containing pT776 is a canonical partner of Fkh1's FHA domain, and the Cdc7 peptide containing pT484 is the canonical partner of Rad53's FHA1 domain. Despite the substantial similarity between these 9-mer peptides, and their abilities to bind to both of the above FHA domains in a pT-dependent manner, the requirement for the FHA-defining arginine, R80 in Fkh1 and R70 in Rad53, was only observed with the canonical peptide partner. This result was somewhat unexpected as the strong similarity between these peptides and the non-structured nature of the FHA-peptide interactions did not obviously suggest to us that these two FHA domains would show that great an ability to distinguish between protein partners with access to only short peptide epitopes. However, this result suggests that it may in fact be effective to use purified Fkh1-FHA and Rad53-FHA1, alongside their corresponding R→A mutants, to co-purify binding partners from a crude extract of native yeast nuclear proteins (or peptides derived from the same). In this approach, we could take advantage of the information encoded in native peptides that contain more sequence and structural information than short peptides used on synthetic arrays. We would take advantage of the ability of these two different FHA domains to discriminate their canonical protein partners from other potential partners by binding to the former set in an FHA-dependent manner.

4.3. Benefits and challenges of an *in silico*-targeted approach to identify Fkh1-FHA partners important for regulating DNA replication origins

As discussed in Chapter 2, reasonable matches to the Mph1 peptide can be found in many origin-regulatory proteins. One benefit to this *in silico* approach is that making the relevant mutations in yeast is usually straightforward and, in contrast to ordering synthetic peptides, is relatively inexpensive. However, until recently, a major challenge was phenotypic analysis. Specifically, the *fkh1-R80A* allele, though it causes a broad defect in early origin function, causes no easily observable cell growth phenotype. Thus, we would not expect a Fkh1-FHA interaction-defective allele in a gene encoding an origin regulatory protein to cause an easy-to-assess phenotype, such as temperature-sensitive growth. Unfortunately, while the Sort-seq assay we have used to observe a *fkh1-R80A* phenotype is information-rich, it is not a trivial or high-throughput assay for assessing mutant phenotypes. That is, Sort-seq is not an assay to use for a first-level screen of mutant alleles that might phenocopy the behavior of the *fkh1-R80A* allele in terms of origin activity.

We are considering two possible first-level assays for screening putative Fkh1-FHA partners by assessing allele behavior. First, there is a plasmid loss assay, in which a selected origin is tested for its ability to allow a plasmid to replicate. Specifically, several distinct FHA-dependent origins were defined that showed plasmid replication defects in *fkh1-R80A* cells but not *FKH1* cells [21]. Thus, a mutant allele that phenocopied a *fkh1-R80A* mutation should lose a certain plasmid at a similar rate as the *fkh1-R80A* allele does. If the defect is substantial enough, it should be measurable under selective growth conditions in which cells that do not retain the relevant plasmid cannot grow. Thus, simply measuring cell OD after an 8-hour growth phase may be sufficient to screen for a *fkh1-R80A*-like phenotype, as predicted for a mutation that

altered a Fkh1-FHA-partner interaction required for origin function. For phenotypes that are not quite so robust, there are other ways to measure more subtle changes in plasmid retention, which may be necessary if working with protein partner mutants that do not individually have quite so large an effect as *fkh1-R80A* itself does.

Second, the cell cycle arrest and release experiments described in Chapters 2 and 3, where a synchronous S-phase is monitored at high resolution by flow cytometry, may also provide an accessible first-level screen of mutant alleles that cause a defect in a Fkh1-FHA-dependent protein interaction required for origin function. Specifically, the *fkh1-R80A* allele causes a defect in early origin function as well as a delay, as predicted from this defect, in S-phase progression. We would predict that a mutant allele that blocks the Fkh1-FHA-protein interaction required for early origin function would phenocopy this effect and cause a delay in S-phase. In fact, that thinking is a major reason why we do not believe that the *cdc7-T484A* allele affects the Fkh1-FHA-interaction required for early origin function. Specifically, *cdc7-T484A* advances cell-cycle progression through S-phase. We would predict, therefore, that when we perform Sort-seq on *cdc7-T484A* mutants that early origins will be relatively unaffected and that late origins will be enhanced, thus accelerating genome duplication in S-phase. Regardless, the arrest and release experiment should be amenable to a higher throughput format, using a 48- or 96-well plate, for screening relevant mutant alleles in yeast for their ability to phenocopy the behavior of the *fkh1-R80A* allele.

A potential major limitation of the targeted approach is its assumption that a single Fkh1-FHA-partner interaction is sufficient to explain the role of Fkh1's FHA domain at early origins. Based on the limited studies of mating-type switching, this assumption may be incorrect. Instead at some, many or even all Fkh1-FHA-regulated early origins, the Fkh1-FHA domain may

interact with multiple partners, including epitopes within ATP-dependent chromatin remodelers recently implicated in origin licensing [48,49] and/or one or more core-origin regulatory proteins. If Fkh1 works in this way, using its FHA domain to transiently interact with multiple players required for origin activity, a single mutant allele in an interaction partner may not cause enough of a defect in early origin function to cause a detectable phenotype. If this is the case, and multiple mutant alleles must be combined to cause any phenotype, it will be considerably more challenging to link defined protein-protein interactions involving the Fkh1-FHA domain to specific molecular pathways or processes *in vivo*. However, there is also a precedent for the additive contribution of multiple Fkh1-FHA partners to a process (Mph1 and Fdo1 in mating-type switching) so perhaps other interactors could contribute additively to origin function, creating a subtle, yet measurable, phenotype.

Appendices

Figure A.1. Meier-Gorlin syndrome-associated Orc4^{Y232C} did not reduce ORC's stability on chromatin.

A Meier-Gorlin syndrome-associated allele was identified in *ORC4*. In this study [155], we created the equivalent mutation in *Drosophila melanogaster* and *Saccharomyces cerevisiae*. The below experiments were performed on *S. cerevisiae* protein extracts.

A-B. Protein immunoblots suggest that the MGS-associated *orc4-Y232C* mutation in yeast did not change the expression of ORC subunits Orc4 and Orc1.

C. By comparing chromatin-associated protein in the "pel" fractions and soluble proteins in the "sup" fractions in the presence of 0, 100, or 750 mM NaCl, it was observed that unlike *orc1-BAHΔ*, *orc4-Y232C* did not diminish ORC's ability to stably bind to chromatin in the presence of 100 mM NaCl.

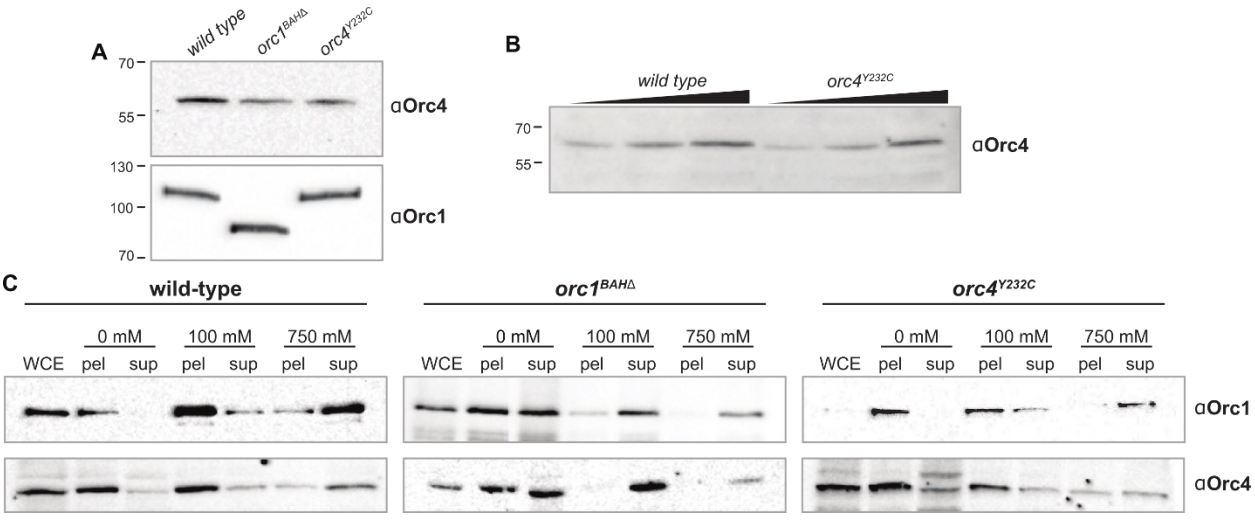
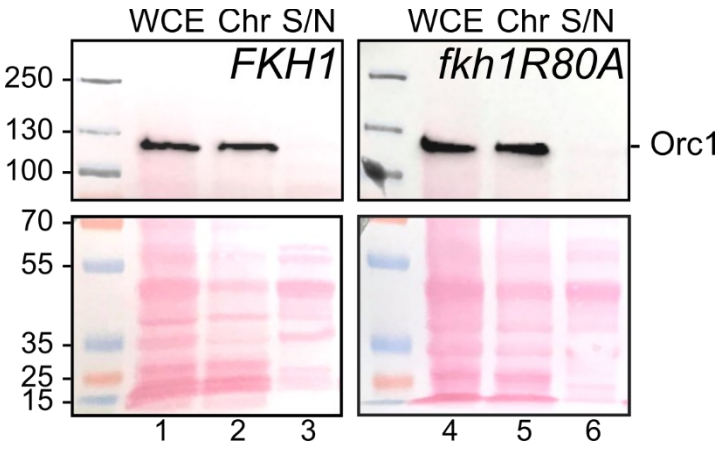


Figure A.2. The *fkh1-R80A* allele did not reduce ORC's stability on chromatin.

In this study [21], a protein immunoblot did not reveal a global change in ORC's association with chromatin in *fkh1-R80A* cells. In both *FKH1* and *fkh1-R80A* strains, ORC was able to stably associate with chromatin.



References

1. Weigel D, Jürgens G, Küttner F, Seifert E, Jäckle H. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. Cell. 1989;57: 645–658.
2. Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature. 1993;364: 412–420.
3. Wierstra I. FOXM1 (Forkhead box M1) in tumorigenesis: overexpression in human cancer, implication in tumorigenesis, oncogenic functions, tumor-suppressive properties, and target of anticancer therapy. Adv Cancer Res. 2013;119: 191–419.
4. Heslop JA, Duncan SA. FoxA factors: the chromatin key and doorstop essential for liver development and function. Genes & development. Cold Spring Harbor Laboratory; 2020. pp. 1003–1004.
5. Cuesta I, Zaret KS, Santisteban P. The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure. Mol Cell Biol. 2007;27: 7302–7314.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144: 646–674.
7. Zhou H, Shi B-J. New roles of DNA-binding and forkhead-associated domains of Fkh1 and Fkh2 in cellular functions. Cell Biochem Funct. 2022;40: 888–902.
8. Mahajan A, Yuan C, Lee H, Chen ES-W, Wu P-Y, Tsai M-D. Structure and function of the phosphothreonine-specific FHA domain. Sci Signal. 2008;1: re12.
9. Zhu G, Spellman PT, Volpe T, Brown PO, Botstein D, Davis TN, et al. Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. Nature. 2000;406: 90–94.
10. Hollenhorst PC, Bose ME, Mielke MR, Müller U, Fox CA. Forkhead genes in transcriptional silencing, cell morphology and the cell cycle. Overlapping and distinct functions for FKH1 and FKH2 in Saccharomyces cerevisiae. Genetics. 2000;154: 1533–1548.
11. Pic A, Lim FL, Ross SJ, Veal EA, Johnson AL, Sultan MR, et al. The forkhead protein Fkh2 is a component of the yeast cell cycle transcription factor SFF. EMBO J. 2000;19: 3750–3761.
12. Jorgensen P, Tyers M. The fork'ed path to mitosis. Genome Biol. 2000;1: REVIEWS1022.
13. Voth WP, Yu Y, Takahata S, Kretschmann KL, Lieb JD, Parker RL, et al. Forkhead proteins control the outcome of transcription factor binding by antiactivation. EMBO J. 2007;26: 4324–4334.
14. Aref R, Sanad MNME, Schüller H-J. Forkhead transcription factor Fkh1: insights into functional regulatory domains crucial for recruitment of Sin3 histone deacetylase complex. Curr Genet. 2021;67: 487–499.
15. Roy KR, Smith JD, Vonesch SC, Lin G, Tu CS, Lederer AR, et al. Multiplexed precision genome editing with trackable genomic barcodes in yeast. Nat Biotechnol. 2018;36: 512–520.
16. Sun K, Coïc E, Zhou Z, Durrens P, Haber JE. Saccharomyces forkhead protein Fkh1 regulates donor preference during mating-type switching through the recombination enhancer. Genes Dev. 2002;16: 2085–2096.

17. Li J, Coïc E, Lee K, Lee C-S, Kim J-A, Wu Q, et al. Regulation of budding yeast mating-type switching donor preference by the FHA domain of Fkh1. PLoS Genet. 2012;8: e1002630.
18. Knott SRV, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, et al. Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. Cell. 2012;148: 99–111.
19. Aparicio OM. Location, location, location: it's all in the timing for replication origins. Genes Dev. 2013;27: 117–128.
20. Dummer AM, Su Z, Cherney R, Choi K, Denu J, Zhao X, et al. Binding of the Fkh1 Forkhead associated domain to a phosphopeptide within the Mph1 DNA helicase regulates mating-type switching in budding yeast. PLoS Genet. 2016;12: e1006094.
21. Hoggard T, Hollatz AJ, Cherney RE, Seman MR, Fox CA. The Fkh1 Forkhead associated domain promotes ORC binding to a subset of DNA replication origins in budding yeast. Nucleic Acids Res. 2021;49: 10207–10220.
22. Wu X, Haber JE. A 700 bp cis-acting region controls mating-type dependent recombination along the entire left arm of yeast chromosome III. Cell. 1996;87: 277–285.
23. Rhind N. DNA replication timing: Biochemical mechanisms and biological significance. Bioessays. 2022;44: e2200097.
24. Poli J, Tsaponina O, Crabbé L, Keszthelyi A, Pantesco V, Chabes A, et al. dNTP pools determine fork progression and origin usage under replication stress. EMBO J. 2012;31: 883–894.
25. Choudhary R, Niska-Blakie J, Adhil M, Liberi G, Achar YJ, Giannattasio M, et al. Sen1 and Rrm3 ensure permissive topological conditions for replication termination. Cell Rep. 2023;42: 112747.
26. Bicknell LS, Bongers EMHF, Leitch A, Brown S, Schoots J, Harley ME, et al. Mutations in the pre-replication complex cause Meier-Gorlin syndrome. Nat Genet. 2011;43: 356–359.
27. Hiratani I, Gilbert DM. Replication timing as an epigenetic mark. Epigenetics. 2009;4: 93–97.
28. Stillman B. Cell cycle control of DNA replication. Science. 1996;274: 1659–1664.
29. Bell SP, Labib K. Chromosome duplication in Saccharomyces cerevisiae. Genetics. 2016;203: 1027–1067.
30. Speck C, Chen Z, Li H, Stillman B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nat Struct Mol Biol. 2005;12: 965–971.
31. Sun J, Fernandez-Cid A, Riera A, Tognetti S, Yuan Z, Stillman B, et al. Structural and mechanistic insights into Mcm2–7 double-hexamer assembly and function. Genes Dev. 2014;28: 2291–2303.
32. Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JFX. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell. 2009;139: 719–730.
33. Remus D, Diffley JFX. Eukaryotic DNA replication control: lock and load, then fire. Curr Opin Cell Biol. 2009;21: 771–777.
34. Green BM, Finn KJ, Li JJ. Loss of DNA replication control is a potent inducer of gene amplification. Science. 2010;329: 943–946.

35. Green BM, Morreale RJ, Ozaydin B, Derisi JL, Li JJ. Genome-wide mapping of DNA synthesis in Saccharomyces cerevisiae reveals that mechanisms preventing reinitiation of DNA replication are not redundant. Mol Biol Cell. 2006;17: 2401–2414.
36. Green BM, Li JJ. Loss of rereplication control in Saccharomyces cerevisiae results in extensive DNA damage. Mol Biol Cell. 2005;16: 421–432.
37. Nguyen VQ, Co C, Li JJ. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature. 2001;411: 1068–1073.
38. Moyer SE, Lewis PW, Botchan MR. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc Natl Acad Sci U S A. 2006;103: 10236–10241.
39. Rhind N, Gilbert DM. DNA replication timing. Cold Spring Harb Perspect Biol. 2013;5: a010132.
40. Rhind N. F = m*a: A framework for investigating the regulation of replication timing. Genes (Basel). 2022;13: 249.
41. Hawkins M, Retkute R, Müller CA, Saner N, Tanaka TU, de Moura APS, et al. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. Cell Rep. 2013;5: 1132–1141.
42. Weinreich M, Palacios DeBeer MA, Fox CA. The activities of eukaryotic replication origins in chromatin. Biochim Biophys Acta. 2004;1677: 142–157.
43. Palacios DeBeer MA, Fox CA. A role for a replicator dominance mechanism in silencing. EMBO J. 1999;18: 3808–3819.
44. Stevenson JB, Gottschling DE. Telomeric chromatin modulates replication timing near chromosome ends. Genes Dev. 1999;13: 146–151.
45. Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M. Histone acetylation regulates the time of replication origin firing. Mol Cell. 2002;10: 1223–1233.
46. Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. Conserved nucleosome positioning defines replication origins. Genes Dev. 2010;24: 748–753.
47. Müller P, Park S, Shor E, Huebert DJ, Warren CL, Ansari AZ, et al. The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. Genes Dev. 2010;24: 1418–1433.
48. Li S, Wasserman MR, Yurieva O, Bai L, O'Donnell ME, Liu S. Origin recognition complex harbors an intrinsic nucleosome remodeling activity. Proc Natl Acad Sci U S A. 2022;119: e2211568119.
49. Chacin E, Reusswig K-U, Furtmeier J, Bansal P, Karl LA, Pfander B, et al. Establishment and function of chromatin organization at replication origins. Nature. 2023;616: 836–842.
50. Mantiero D, Mackenzie A, Donaldson A, Zegerman P. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. EMBO J. 2011;30: 4805–4814.
51. Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. Curr Biol. 2011;21: 2055–2063.
52. Natsume T, Müller CA, Katou Y, Retkute R, Gierliński M, Araki H, et al. Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. Mol Cell. 2013;50: 661–674.
53. Dukaj L, Rhind N. The capacity of origins to load MCM establishes replication timing patterns. PLoS Genet. 2021;17: e1009467.

54. Bechhoefer J, Rhind N. Replication timing and its emergence from stochastic processes. Trends Genet. 2012;28: 374–381.
55. Hoggard T, Müller CA, Nieduszynski CA, Weinreich M, Fox CA. Sir2 mitigates an intrinsic imbalance in origin licensing efficiency between early- and late-replicating euchromatin. Proc Natl Acad Sci U S A. 2020;117: 14314–14321.
56. Pohl TJ, Brewer BJ, Raghuraman MK. Functional centromeres determine the activation time of pericentric origins of DNA replication in Saccharomyces cerevisiae. PLoS Genet. 2012;8: e1002677.
57. Ostrow AZ, Nellimoottil T, Knott SRV, Fox CA, Tavaré S, Aparicio OM. Fkh1 and Fkh2 bind multiple chromosomal elements in the S. cerevisiae genome with distinct specificities and cell cycle dynamics. PLoS One. 2014;9: e87647.
58. Lõoke M, Kristjuhan K, Värv S, Kristjuhan A. Chromatin-dependent and -independent regulation of DNA replication origin activation in budding yeast. EMBO Rep. 2013;14: 191–198.
59. Hoggard T, Shor E, Müller CA, Nieduszynski CA, Fox CA. A Link between ORC-origin binding mechanisms and origin activation time revealed in budding yeast. PLoS Genet. 2013;9: e1003798.
60. Hoggard T, Chacin E, Hollatz AJ, Kurat CF, Fox CA. The budding yeast Fkh1 Forkhead associated (FHA) domain promotes a G1-chromatin state and the activity of chromosomal DNA replication origins. PLoS Genet. 2024;20: e1011366.
61. Liang X, Van Doren SR. Mechanistic insights into phosphoprotein-binding FHA domains. Acc Chem Res. 2008;41: 991–999.
62. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989;340: 245–246.
63. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiol Rev. 1995;59: 94–123.
64. Park K, Yi SY, Lee C-S, Kim KE, Pai H-S, Seol D-W, et al. A split enhanced green fluorescent protein-based reporter in yeast two-hybrid system. Protein J. 2007;26: 107–116.
65. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams S-L, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 2002;415: 180–183.
66. Fang D, Lengronne A, Shi D, Forey R, Skrzypczak M, Ginalski K, et al. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. Genes Dev. 2017;31: 2405–2415.
67. May DG, Roux KJ. BioID: A method to generate a history of protein associations. Methods Mol Biol. 2019;2008: 83–95.
68. Cronan JE. Biotin protein ligase as you like it: Either extraordinarily specific or promiscuous protein biotinylation. Proteins. 2024;92: 435–448.
69. Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, et al. Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol. 2018;36: 880–887.
70. Marks AB, Fu H, Aladjem MI. Regulation of replication origins. Adv Exp Med Biol. 2017;1042: 43–59.

71. McConnell KH, Müller P, Fox CA. Tolerance of Sir1p/origin recognition complex-dependent silencing for enhanced origin firing at HMRa. Mol Cell Biol. 2006;26: 1955–1966.
72. Liang C, Weinreich M, Stillman B. ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. Cell. 1995;81: 667–676.
73. Ostrow AZ, Kalhor R, Gan Y, Villwock SK, Linke C, Barberis M, et al. Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in S. cerevisiae. Proc Natl Acad Sci U S A. 2017;114: E2411–E2419.
74. Almawi AW, Matthews LA, Larasati, Myrox P, Boulton S, Lai C, et al. "AND" logic gates at work: Crystal structure of Rad53 bound to Dbf4 and Cdc7. Sci Rep. 2016;6: 34237.
75. Reuter LM, Khadayate SP, Mossler A, Liebl K, Faull SV, Karimi MM, et al. MCM2-7 loading-dependent ORC release ensures genome-wide origin licensing. Nat Commun. 2024;15: 7306.
76. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, et al. A global protein kinase and phosphatase interaction network in yeast. Science. 2010;328: 1043–1046.
77. Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H. A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics. 2008;7: 1389–1396.
78. Aucher W, Becker E, Ma E, Miron S, Martel A, Ochsenbein F, et al. A strategy for interaction site prediction between phospho-binding modules and their partners identified from proteomic data. Mol Cell Proteomics. 2010;9: 2745–2759.
79. Helbig AO, Rosati S, Pijnappel PWWM, van Breukelen B, Timmers MHTH, Mohammed S, et al. Perturbation of the yeast N-acetyltransferase NatB induces elevation of protein phosphorylation levels. BMC Genomics. 2010;11: 685.
80. Soulard A, Cremonesi A, Moes S, Schütz F, Jenö P, Hall MN. The rapamycin-sensitive phosphoproteome reveals that TOR controls protein kinase A toward some but not all substrates. Mol Biol Cell. 2010;21: 3475–3486.
81. Holt LJ, Tuch BB, Villén J, Johnson AD, Gygi SP, Morgan DO. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science. 2009;325: 1682–1686.
82. Li X, Gerber SA, Rudner AD, Beausoleil SA, Haas W, Villén J, et al. Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. J Proteome Res. 2007;6: 1190–1197.
83. Randell JCW, Fan A, Chan C, Francis LI, Heller RC, Galani K, et al. Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. Mol Cell. 2010;40: 353–363.
84. Lanz MC, Yugandhar K, Gupta S, Sanford EJ, Faça VM, Vega S, et al. In-depth and 3-dimensional exploration of the budding yeast phosphoproteome. EMBO Rep. 2021;22: e51121.
85. Chen Y-C, Jiang P-H, Chen H-M, Chen C-H, Wang Y-T, Chen Y-J, et al. Glucose intake hampers PKA-regulated HSP90 chaperone activity. Elife. 2018;7. doi:10.7554/eLife.39925
86. Zhou X, Li W, Liu Y, Amon A. Cross-compartment signal propagation in the mitotic exit network. Elife. 2021;10. doi:10.7554/eLife.63645

87. Swaney DL, Beltrao P, Starita L, Guo A, Rush J, Fields S, et al. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat Methods. 2013;10: 676–682.
88. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol. 2002;20: 301–305.
89. Calzada A, Sánchez M, Sánchez E, Bueno A. The stability of the Cdc6 protein is regulated by cyclin-dependent kinase/cyclin B complexes in Saccharomyces cerevisiae. J Biol Chem. 2000;275: 9734–9741.
90. Honey S, Futcher B. Roles of the CDK phosphorylation sites of yeast Cdc6 in chromatin binding and rereplication. Mol Biol Cell. 2007;18: 1324–1336.
91. Perkins G, Drury LS, Diffley JF. Separate SCF(CDC4) recognition elements target Cdc6 for proteolysis in S phase and mitosis. EMBO J. 2001;20: 4836–4845.
92. Cross FR, Levine K. Molecular evolution allows bypass of the requirement for activation loop phosphorylation of the Cdc28 cyclin-dependent kinase. Mol Cell Biol. 1998;18: 2923–2931.
93. Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics. 2005;4: 310–327.
94. Lim HH, Loy CJ, Zaman S, Surana U. Dephosphorylation of threonine 169 of Cdc28 is not required for exit from mitosis but may be necessary for start in Saccharomyces cerevisiae. Mol Cell Biol. 1996;16: 4573–4583.
95. Ross KE, Kaldis P, Solomon MJ. Activating phosphorylation of the Saccharomyces cerevisiae cyclin-dependent kinase, cdc28p, precedes cyclin binding. Mol Biol Cell. 2000;11: 1597–1609.
96. Schaber M, Lindgren A, Schindler K, Bungard D, Kaldis P, Winter E. CAK1 promotes meiosis and spore formation in Saccharomyces cerevisiae in a CDC28-independent fashion. Mol Cell Biol. 2002;22: 57–68.
97. Thuret JY, Valay JG, Faye G, Mann C. Civ1 (CAK in vivo), a novel Cdk-activating kinase. Cell. 1996;86: 565–576.
98. Bodenmiller B, Wanka S, Kraft C, Urban J, Campbell D, Pedrioli PG, et al. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. Sci Signal. 2010;3: rs4.
99. Chen S, Bell SP. CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. Genes Dev. 2011;25: 363–372.
100. Gnad F, de Godoy LMF, Cox J, Neuhauser N, Ren S, Olsen JV, et al. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. Proteomics. 2009;9: 4642–4652.
101. Soufi B, Kelstrup CD, Stoehr G, Fröhlich F, Walther TC, Olsen JV. Global analysis of the yeast osmotic stress response by quantitative proteomics. Mol Biosyst. 2009;5: 1337–1346.
102. Masumoto H, Muramatsu S, Kamimura Y, Araki H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. Nature. 2002;415: 651–655.
103. Zegerman P, Diffley JFX. Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. Nature. 2010;467: 474–478.

104. Lopez-Mosqueda J, Maas NL, Jonsson ZO, Defazio-Eli LG, Wohlschlegel J, Toczyski DP. Damage-induced phosphorylation of Sld3 is important to block late origin firing. Nature. 2010;467: 479–483.
105. Chen Y-C, Kenworthy J, Gabrielse C, Hänni C, Zegerman P, Weinreich M. DNA replication checkpoint signaling depends on a Rad53-Dbf4 N-terminal interaction in Saccharomyces cerevisiae. Genetics. 2013;194: 389–401.
106. Matthews LA, Jones DR, Prasad AA, Duncker BP, Guarné A. Saccharomyces cerevisiae Dbf4 has unique fold necessary for interaction with Rad53 kinase. J Biol Chem. 2012;287: 2378–2387.
107. Matthews LA, Selvaratnam R, Jones DR, Akimoto M, McConkey BJ, Melacini G, et al. A novel non-canonical forkhead-associated (FHA) domain-binding interface mediates the interaction between Rad53 and Dbf4 proteins. J Biol Chem. 2014;289: 2589–2599.
108. Yongkiettrakul S, Byeon I-JL, Tsai M-D. The ligand specificity of yeast Rad53 FHA domains at the +3 position is determined by nonconserved residues. Biochemistry. 2004;43: 3862–3869.
109. Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci. 1993;9: 745–756.
110. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.
111. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30: 70–82.
112. Bittrich S, Segura J, Duarte JM, Burley SK, Rose Y. RCSB protein Data Bank: exploring protein 3D similarities via comprehensive structural alignments. Bioinformatics. 2024;40. doi:10.1093/bioinformatics/btae370
113. Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, et al. New vectors for simple and streamlined CRISPR-Cas9 genome editing in*Saccharomyces cerevisiae*. Yeast. 2015;32: 711–720.
114. CAN Lab protocols. [cited 28 Nov 2024]. Available: https://www.nieduszynski.org/methods/crispyCas9.php
115. Batrakou DG, Müller CA, Wilson RHC, Nieduszynski CA. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. Nat Protoc. 2020;15: 1255–1284.
116. Müller CA, Hawkins M, Retkute R, Malla S, Wilson R, Blythe MJ, et al. The dynamics of genome replication using deep sequencing. Nucleic Acids Res. 2014;42: e3.
117. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359.
118. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842.
119. Skene PJ, Henikoff S. A simple method for generating high-resolution maps of genome-wide protein binding. Elife. 2015;4: e09225.
120. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989;122: 19–27.
121. Casey L, Patterson EE, Müller U, Fox CA. Conversion of a replication origin to a silencer through a pathway shared by a Forkhead transcription factor and an S phase cyclin. Mol Biol Cell. 2008;19: 608–622.

122. Wong ED, Miyasato SR, Aleksander S, Karra K, Nash RS, Skrzypek MS, et al. Saccharomyces genome database update: server architecture, pan-genome nomenclature, and external resources. Genetics. 2023;224. doi:10.1093/genetics/iyac191
123. Peña-Castillo L, Hughes TR. Why are there still over 1000 uncharacterized yeast genes? Genetics. 2007;176: 7–14.
124. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, et al. Subcellular localization of the yeast proteome. Genes Dev. 2002;16: 707–719.
125. Shen X-X, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. G3 (Bethesda). 2016;6: 3927–3939.
126. Salim D, Bradford WD, Freeland A, Cady G, Wang J, Pruitt SC, et al. DNA replication stress restricts ribosomal DNA copy number. PLoS Genet. 2017;13: e1007006.
127. Kobayashi T, Sasaki M. Ribosomal DNA stability is supported by many 'buffer genes'- introduction to the Yeast rDNA Stability Database. FEMS Yeast Res. 2017;17. doi:10.1093/femsyr/fox001
128. Kwan EX, Foss EJ, Tsuchiyama S, Alvino GM, Kruglyak L, Kaeberlein M, et al. A natural polymorphism in rDNA replication origins links origin activation with calorie restriction and lifespan. PLoS Genet. 2013;9: e1003329.
129. Trivedi R, Nagarajaram HA. Intrinsically disordered proteins: An overview. Int J Mol Sci. 2022;23: 14050.
130. Youn J-Y, Dyakov BJA, Zhang J, Knight JDR, Vernon RM, Forman-Kay JD, et al. Properties of stress granule and P-body proteomes. Mol Cell. 2019;76: 286–294.
131. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2024;42: 243–246.
132. Pal D, Wu D, Haruta A, Matsumura F, Wei Q. Role of a novel coiled-coil domain-containing protein CCDC69 in regulating central spindle assembly. Cell Cycle. 2010;9: 4117–4129.
133. Michelitsch MD, Weissman JS. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. Proc Natl Acad Sci U S A. 2000;97: 11910–11915.
134. Ottoz DSM, Berchowitz LE. The role of disorder in RNA binding affinity and specificity. Open Biol. 2020;10: 200328.
135. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature. 2017;547: 236–240.
136. Haber JE. Mating-type genes and MAT switching in Saccharomyces cerevisiae. Genetics. 2012;191: 33–64.
137. Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, et al. Tempo and mode of genome evolution in the budding yeast subphylum. Cell. 2018;175: 1533-1545.e20.
138. Wach A. PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae. Yeast. 1996;12: 259–265.
139. Cormack BP, Falkow S. Efficient homologous and illegitimate recombination in the opportunistic yeast pathogen Candida glabrata. Genetics. 1999;151: 979–987.
140. Wendland J, Walther A. Chromosome number reduction in Eremothecium coryli by two telomere-to-telomere fusions. Genome Biol Evol. 2014;6: 1186–1198.

141. Astromskas E, Cohn M. Ends-in vs. ends-out targeted insertion mutagenesis in Saccharomyces castellii. Curr Genet. 2009;55: 339–347.
142. Solieri L. The revenge of Zygosaccharomyces yeasts in food biotechnology and applied microbiology. World J Microbiol Biotechnol. 2021;37: 96.
143. Strucko T, Andersen NL, Mahler MR, Martínez JL, Mortensen UH. A CRISPR/Cas9 method facilitates efficient oligo-mediated gene editing in Debaryomyces hansenii. Synth Biol. 2021;6: ysab031.
144. Di Canito A, Foschino R, Mazzieri M, Vigentini I. Molecular tools to exploit the biotechnological potential of Brettanomyces bruxellensis: A review. Appl Sci (Basel). 2021;11: 7302.
145. Pacheco A, Almeida MJ, Sousa MJ. Improved gene disruption method for Torulaspora delbrueckii. FEMS Yeast Res. 2009;9: 158–160.
146. Ng H, Dean N. Dramatic improvement of CRISPR/Cas9 editing in Candida albicans by increased single guide RNA expression. mSphere. 2017;2. doi:10.1128/mSphere.00385-16
147. Badura J, van Wyk N, Zimmer K, Pretorius IS, von Wallbrunn C, Wendland J. PCR-based gene targeting in Hanseniaspora uvarum. FEMS Yeast Res. 2023;23: foad034.
148. Kooistra R, Hooykaas PJJ, Steensma HY. Efficient gene targeting in Kluyveromyces lactis. Yeast. 2004;21: 781–792.
149. Abdel-Banat BMA, Nonklang S, Hoshida H, Akada R. Random and targeted gene integrations through the control of non-homologous end joining in the yeast Kluyveromyces marxianus. Yeast. 2010;27: 29–39.
150. Jang I-S, Yu BJ, Jang JY, Jegal J, Lee JY. Improving the efficiency of homologous recombination by chemical and biological approaches in Yarrowia lipolytica. PLoS One. 2018;13: e0194954.
151. Mollapour M, Piper P. Targeted gene deletion in Zygosaccharomyces bailii. Yeast. 2001;18: 173–186.
152. Kobayashi T. Ribosomal RNA gene repeats, their stability and cellular senescence. Proc Jpn Acad Ser B Phys Biol Sci. 2014;90: 119–129.
153. Stotz A, Linder P. The ADE2 gene from Saccharomyces cerevisiae: sequence and new vectors. Gene. 1990;95: 91–98.
154. Fenech EJ, Cohen N, Kupervaser M, Gazi Z, Schuldiner M. A toolbox for systematic discovery of stable and transient protein interactors in baker's yeast. Mol Syst Biol. 2023;19: e11084.
155. McDaniel SL, Hollatz AJ, Branstad AM, Gaskill MM, Fox CA, Harrison MM. Tissue-specific DNA replication defects in Drosophila melanogaster caused by a Meier-Gorlin syndrome mutation in Orc4. Genetics. 2020;214: 355–367.