U.PORTO

DOUTORAMENTO EM CIÊNCIAS BIOMÉDICAS

Molecular sensors of codon ambiguity in *Candida albicans* signaling pathways

Joana Sofia Fraga

D
2019

U.PORTO

Joana Sofia Fraga. Molecular sensors of codon ambiguity in ***Candida albicans*** signaling pathways

D.ICBAS **2019**

Molecular sensors of codon ambiguity in *Candida albicans* signaling pathways

Joana Sofia Fraga

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Joana Sofia Morais Fraga

Molecular sensors of codon ambiguity in *Candida albicans* signaling pathways

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Orientador – Sandra de Macedo Ribeiro

Categoria – Investigadora Principal

Afiliação – IBMC - Instituto de Biologia Molecular e Celular e I3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto

Co-orientador – Pedro José Barbosa Pereira

Categoria – Investigador Principal

Afiliação – IBMC - Instituto de Biologia Molecular e Celular e I3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto

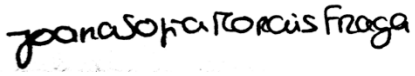
Co-orientador/Tutor – Vítor Costa

Categoria – Professor Associado

Afiliação – ICBAS - Instituto de Ciências Biomédicas Abel Salazar, IBMC - Instituto de Biologia Molecular e Celular e I3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto

Declaração de honra de acordo com o atual Código Ético de Conduta Académica da U.Porto, aprovado em dezembro de 2017, em particular considerando a obrigatoriedade decorrente do seu Artigo 14º:

Declaro que a presente tese é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico".



This work was funded by Fundação para a Ciência e a Tecnologia through PhD fellowship SFRH/BD/94403/2013

This work was funded by FEDER - Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020 - Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of project POCI-01-0145-FEDER-007274 (Institute for Research and Innovation in Health Sciences) and by FEDER through Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement in the framework of Project Norte-01-0145-FEDER-000008.

Agradecimentos

Ao longo deste percurso foram muitos aqueles que caminharam ao meu lado e me ajudaram a chegar até aqui.

Em primeiro lugar aos meus orientadores, Sandra Macedo-Ribeiro e Pedro Pereira pela oportunidade, por terem acreditado em mim, pelo apoio, pelo incentivo e atitude positiva, pelo que aprendi. Ao professor Vítor Costa pela ajuda e disponibilidade.

Às plataformas científicas do I3S, ao Frederico da B2Tech e ao Hugo Osório da Proteómica por todo o apoio. To Daren Hart and Philippe Mass for all the support and your welcoming in the lab during my short stay in Grenoble.

A todos os que fizeram do espaço que partilhamos diariamente um local agradável para se trabalhar: Jorge, Tatiana e José (pela ajuda no SAXS) pelas palavras de encorajamento; ao Pedro Martins, aos buddies Maria e Francisco, pela lufada de ar fresco e boa disposição; Zsuzsa, Xana e Inês, pelas longas e muitas conversas entre lágrimas e sorrisos, as palavras não chegam para agradecer toda a paciência, sem vocês não seria possível.

A todos aqueles que do "outro lado" tornam a minha vida um arco-íris, pela amizade, pelos sorrisos, por tudo aquilo que fica por escrever:
Ao Bruno, "o colega", pelos sábias palavras; à Palomita, não vejo o olhar cúmplice mas sinto-o nas frases que leio; à Joody pela partilha, imensa paciência e presença constante (pela nossa frase); à Ana e Irene pela leveza e pelas gargalhadas; ao João, pelo impulso e pelos empurrões que me fazem crescer. Ao Migas, Analuce, Joana, Silvina, Su e Pi, pelos encontros virtuais que aconchegam, pelos abraços que encaixam, pela acidez que ninguém mais entende, por resistirem ao crivo. Aos "Perdidos", Ana, Patrícia e Vânia por me terem acolhido de braços abertos e continuarem a colorir os meus dias (um dia alguém disse "no final, tudo se resolve"). Aos de sempre, Inês, Fifó, Tanocas, Sandra e aos mais pequenos pela história recheada que continuamos a escrever, pela resistência. Ao André, pela "partilha" e momentos de boa disposição, pela descoberta do Porto.

À minha família, o meu porto de abrigo, os melhores e mais genuínos sorrisos e abraços, a essência de tudo: ao meu padrinho, a quem dedico este trabalho, pela alegria que sentiria de o ver; aos meus tios e primos pelos momentos tão (e só) nossos; aos meus avós, pelo brilho no olhar, por me ensinarem que a curiosidade e vontade de aprender não tem idade; à Lara, pela verdade, o olhar que só tu entendes e tanto diz; aos meus pais pela compreensão, pelo apoio incondicional e por me ensinarem a descomplicar a vida.

A todos, o meu mais sincero e sentido obrigada.

Joana

Table of Contents

AGRADECIMENTOS VII

LIST OF FIGURES XIII

LIST OF TABLES XV

ABBREVIATIONS XVII

ABSTRACT XIX

RESUMO XXI

CHAPTER 1 1

1.1 THE HUMAN PATHOGEN *CANDIDA ALBICANS* 3

1.1.1 OVERVIEW OF CURRENT ANTIFUNGAL THERAPIES 4

1.2 *C. ALBICANS* BIOLOGY 7

1.2.1 *C. ALBICANS* GENETIC CODE AMBIGUITY 7

1.2.2 MORPHOGENESIS: YEAST-TO-HYPHAL TRANSITION 10

1.2.3 PHENOTYPIC SWITCHING: WHITE-OPAQUE TRANSITION 11

1.2.4 *C. ALBICANS* CELL WALL STRUCTURE 13

1.2.5 VIRULENCE FACTORS AND PATHOGENICITY IN *C. ALBICANS* 14

1.3 *C. ALBICANS* SIGNALING PATHWAYS 17

1.3.1 CEK1-MEDIATED PATHWAY 18

1.3.2 STRUCTURAL FEATURES AND REGULATION OF *S. CEREVISIAE* PROTEIN KINASE FUS3 21

1.4 OBJECTIVES 26

1.5 REFERENCES 29

CHAPTER 2 37

2.1 INTRODUCTION 39

2.2 MATERIALS AND METHODS 42

2.2.1 CLONING, EXPRESSION AND PURIFICATION OF CEK1 42

2.2.2 DYNAMIC LIGHT SCATTERING 45

2.2.3 CIRCULAR DICHROISM 45

2.2.4 DIFFERENTIAL SCANNING FLUORIMETRY 46

2.2.5 CEK1 PHOSPHORYLATION *IN VITRO* 46

2.2.6 CEK1 ACTIVITY ASSAY 47

2.2.7 MASS SPECTROMETRY 47

2.2.8 FLUORESCENCE MEASUREMENTS 48

2.2.9 CRYSTALLIZATION OF ΔN-CEK1 48

2.3 RESULTS 49

2.3.1 THE *C. ALBICANS* ΔN-CEK1_SER VARIANT IS MORE STABLE THAN ΔN-CEK1_LEU 49
2.3.2 AUTOPHOSPHORYLATION AT THE CONSERVED ^{231}TEY^{233} MOTIF OF ΔN-CEK1 REQUIRES A SERINE AT THE CUG-ENCODED POSITION 52
2.3.3 CODON DECODING AMBIGUITY AFFECTS CEK1 PROTEIN ACTIVITY 56
2.3.4 AUTOPHOSPHORYLATION OF CEK1 DOES NOT OCCUR IN TRANS 57
2.4 DISCUSSION 58
2.5 REFERENCES 62

CHAPTER 3 65

3.1 INTRODUCTION 67
3.2 MATERIALS AND METHODS 70
3.2.1 CLONING, EXPRESSION AND PURIFICATION 70
3.2.2 BIOPHYSICAL AND BIOCHEMICAL CHARACTERIZATION 71
3.2.3 SMALL ANGLE X-RAY SCATTERING MEASUREMENTS AND ANALYSIS 71
3.3 RESULTS 72
3.3.1 SEQUENCE AND STRUCTURAL ANALYSIS OF MAPK CEK1 72
3.3.2 THE PRESENCE OF THE N-TERMINAL TAIL INCREASES THE OVERALL STABILITY OF CEK1 VARIANTS 73
3.3.3 AUTOPHOSPHORYLATION OF FULL-LENGTH CEK1 75
3.3.4 NUCLEOTIDE AFFINITY IS NOT AFFECTED BY THE INSERTION OF A SERINE OR A LEUCINE AT THE CUG ENCODED POSITION 80
3.4 DISCUSSION 81
3.5 REFERENCES 84

CHAPTER 4 87

4.1 INTRODUCTION 89
4.2 MATERIAL AND METHODS 93
4.2.1 STRATEGIES FOR CLONING, EXPRESSING AND PURIFYING CEK1 INTERACTORS 93
4.2.2 ESPRIT CLONING AND PRODUCTION OF CST5 96
4.2.3 EXPRESSION AND PURIFICATION OF PROTEIN FROM THE CST5 LIBRARY 99
4.2.4 BIOPHYSICAL CHARACTERIZATION OF CST5 CONSTRUCTS 100
4.2.5 CRYSTALLIZATION OF CST5 101
4.2.6 BIO-LAYER INTERFEROMETRY 101
4.3 RESULTS 101
4.3.1 SEQUENCE AND STRUCTURE ANALYSIS OF THE SCAFFOLD PROTEIN CST5 101
4.3.2 CLONING AND EXPRESSION OF SCAFFOLD PROTEIN CST5 AND MAPKK HST7 102
4.3.3 IDENTIFICATION OF SOLUBLE CST5 CONSTRUCTS USING THE ESPRIT SYSTEM 103

4.3.4 BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF CST5 TRUNCATION CONSTRUCTS 105
4.3.5 INTERACTION OF CEK1 VARIANTS WITH CST5 109
4.4 DISCUSSION 109
4.5 REFERENCES 113

CHAPTER 5 115

5.1 GENERAL DISCUSSION 117
5.1.1 IMPACT OF CUG RESIDUE IDENTITY IN *C. ALBICANS* CEK1 117
5.1.2 CEK1 IN THE CONTEXT OF THE MAPK SIGNALING CASCADE 119
5.1.3 CEK1 N-TERMINAL TAIL 120
5.1.4 CONCLUDING REMARKS 121
5.2 REFERENCES 123
ANNEXES 125

List of figures

Figure 1.1 Mechanism of action of antifungal drugs. 5
Figure 1.2 Schematic illustration of the phylogeny of Saccharomycotina species, namely *Candida* species. 8
Figure 1.3 Distinct morphological states of *C. albicans*. 11
Figure 1.4 Structural organization of the fungal *C. albicans* cell wall. 14
Figure 1.5 General scheme of the main elements of MAPKs signal transduction pathways in *C. albicans*. 17
Figure 1.6 Schematic view of *C. albicans* MAPK Cek1-mediated pathway. 19
Figure 1.7 Amino acid sequence alignment of *C. albicans* Cek1 orthologues. 21
Figure 1.8 Structural analysis of *S. cerevisiae* Fus3. 23
Figure 1.9 Diagram of *S. cerevisiae* Ste5 interaction networks. 24
Figure 1.10 Overall structure of Fus3 in complex with Ste7 and Ste5 peptides. 25
Figure 2.1 Structural analysis of *C. albicans* Cek1. 41
Figure 2.2 Biophysical characterization of *C. albicans* Cek1 variants. 50
Figure 2.3 Microcrystals of *C. albicans* kinase-dead mutant ΔN-Cek1^{KR}_Ser. 51
Figure 2.4 ΔN-Cek1_Ser is autophosphorylated at Tyr233. 53
Figure 2.5 LC-MS/MS analysis of ΔN-Cek1 variants. 55
Figure 2.6 ΔN-Cek1_Ser displays higher catalytic activity than ΔN-Cek1_Leu. 57
Figure 2.7 The autophosphorylation of ΔN-Cek1 is an intramolecular event. 58
Figure 3.1 Schematic representation of truncated (ΔN-Cek1) and of full-length *C. albicans* Cek1. 67
Figure 3.2 Structural analysis of MAPK Cek1. 73
Figure 3.3 Thermal stability of full-length Cek1 kinase. 74
Figure 3.4 Structural analysis of full-length Cek1. 75
Figure 3.5 Immunoblot analysis of full-length Cek1_Ser and corresponding mutants fused to the Gb1 tag. 76
Figure 3.6 Full-length Cek1 is autophosphorylated at Tyr233. 77
Figure 3.7 LC-MS/MS analysis of the full-length-Cek1 variants. 79
Figure 3.8 Binding of TNP-ATP nucleotide to full-length Cek1 variants. 80
Figure 3.9 Cek1 N-terminal region repeat variability across different *C. albicans* strains 83
Figure 4.1 Diagram of the PPIs described in *C. albicans* Cek1 pathway. 90
Figure 4.2 Amino acid sequence analysis of *C. albicans* Cst5 scaffold protein and MAPKK Hst7. 92
Figure 4.3 Schematic representation of Cst5 construct. 97
Figure 4.4 CST5 inserts were cloned into the pESPRIT02 vector. 98
Figure 4.5 Cst5 structure prediction by homology modeling. 102

Figure 4.6 SDS-PAGE analysis of Hst7 expression. 103
Figure 4.7 Screening of Cst5 library. 105
Figure 4.8 Analysis of N-terminal truncated Cst5 constructs identified by screening of the truncation library. 106
Figure 4.9 Calibration of analytical SEC. 107
Figure 4.10 DLS of N-terminal truncation constructs of Cst5. 108
Figure 4.11 Representative thermal denaturation curves measured by DSF. 108
Figure 4.12 Interferometry analysis of the interaction between Cst5 and ΔN-Cek1_Ser protein 109

List of Tables

Table 2.1 Oligonucleotides used to amplify the Cek1 active domain and for site-directed mutagenesis. 43
Table 2.2 Extinction coefficient of Cek1 variants and Cek1_Ser mutants based on amino acid sequence. 45
Table 2.3 Thermal stability of ΔN-Cek1_Ser and ΔN-Cek1_Leu variants incubated with ATP. 51
Table 2.4 Thermal stability of ΔN-Cek1_Ser and ΔN-Cek1_Leu variants and ΔN-Cek1_Ser mutants. 53
Table 3.1 Summary of full-length Cek1 variants and respective mutants. Extinction coefficients for full-length Cek1 were calculated based on amino acid sequences with and without Gb1 tag. 71
Table 4.1 Oligonucleotides used for site-directed mutagenesis of Cst5. 94
Table 4.2 Extinction coefficient ($M^{-1} cm^{-1}$) of Cst5 constructs based on amino acid sequence 95
Table 4.3 pCoofy vectors with different N-terminal tags and respective primers used for cloning Hst7. 95
Table 4.4 Primers for vector amplification and complementary primer extensions for Hst7. 96
Table 4.5 Oligonucleotides used to design Cst5 library with N- and C- terminal truncations. 97
Table 4.6 Extinction coefficient of Cst5 constructs based on amino acid sequence. 100
Table 4.7 Yield of Cst5 construct expressed and purified. 106

Abbreviations

ADP Adenosine-diphosphate
AIDS Acquired-immunodeficiency-syndrome
ALS Agglutinin-Like-Sequence
APE Alanine-Proline-Glutamate
AppCp Adenosine-5'- [βγ-methylene] triphosphate
ATP Adenosine triphosphate
BAP Biotin acceptor peptide
BSA Bovine serum albumin
BLI Bio-Layer Interferometry
CD Circular dichroism
CWPs Cell wall proteins
C2H *C. albicans* two-hybrid
DFG Aspartate-Phenylalanine-Glycine
DLS Dynamic light scattering
DSF Differential scanning fluorimetry
DNA Deoxyribonucleic acid
Dnase Deoxyribonuclease I
DTT Dithiothreitol
EDTA Ethylenediamine tetraacetic acid
ESPRIT Expression of Soluble Proteins by Random Incremental Technology
FBD Fus3 binding domain
FELLS Fast Estimator of Latent Local Structure
Gb1 B1 domain of Protein G
GI Gastrointestinal
GPI Glycosyl-phophatidyl-inositol
GUT Gastrointestinally indUced Transition
His6 Hexahistidine
HOG High Osmolarity Glycerol
IC Invasive candidiasis
IDRs Intrinsically disordered regions
IL Inner layer
IMAC Immobilized metal affinity chromatography
IPTG Isopropyl β-D-1-thiogalactopyranoside
LeuRs Leucyl-tRNA sinthetase
MAPK Mitogen activated protein kinase
MAPKK Mitogen activated protein kinase kinase
MAPKKK Mitogen activated protein kinase kinase kinase
MBP Maltose binding protein
MS Mass spectrometry
MTL Mating type like
MW Molecular weight
OL Outer layer
Pd Polydispersity
PDB Protein Data Bank
PIR Internal repeats domains
PH Pleckstrin homology
PM Plasma membrane
PMSF Phenylmethylsulfonyl fluoride
polyQ Polyglutamine
polyN Polyasparagine
PPIs Protein-protein interactions
QS Quorum sensing
R_H Hydrodynamic radius
RNA Ribonucleic acid
RT Room temperature
SAPs Secreted aspartyl proteinases
SAXS Small-angle X-ray scattering
SDS-PAGE Sodium Dodecyl sulfate polyacrylamide gel electrophoresis
SEC Size exclusion chromatography
SerRs Seryl-tRNA synthetase
SLIC Sequence and Ligation Independent Cloning
TEV Tobacco etch virus
Tm Melting temperature
TNP-ATP 2',3'-O-Trinitrophenyl-adenosine-5'-triphosphate
Ve Elution volume
Vc Column volume
vWA von Willebrand type A
V0 Void volume
Y2H two hybrid assay

Abstract

Candida albicans is an opportunistic human fungal pathogen, normally residing as a commensal in more than 50% of the population. However, a high mortality rate in immunocompromised individuals is associated with severe *C. albicans* infections. The complex biology of *C. albicans* is reflected by its impressive morphological plasticity that, together with several additional virulence factors like phenotypic switching and modulation of the molecular cell wall properties, result in *C. albicans*-associated pathogenesis. *C. albicans* belongs to the *Candida* CTG clade, which ambiguously translates the standard leucine CUG codon predominantly as serine (97%) but also as leucine (3%). Although most CUG-residues are positioned in non-conserved protein sites, a few proteins contain these residues in strictly conserved regions. Some of these proteins are associated with Mitogen Activated Protein Kinase (MAPK) pathways, such as the Cek1-mediated pathway that is liked to *C. albicans* morphogenesis and virulence. Here we hypothesize that the genetic code alteration in *C. albicans* might have a functional impact on the MAPK Cek1 and in the regulation of the protein-protein interaction networks of this signaling cascade.

Our biochemical and biophysical data showed that CUG ambiguity affects the thermal stability and enzymatic activity of the active domain and full-length Cek1 protein. The incorporation of a serine residue at the CUG position of Cek1 induces autophosphorylation of the conserved tyrosine residue within the ^{231}TEY^{233} motif in the kinase activation loop and increases its intrinsic kinase activity. The serine variant of the active domain of Cek1 is able to interact with the scaffold protein Cst5. Moreover, the N-terminal tail of Cek1 enriched in amino acid repeats positively impacts the stability and activity of the full-length protein.

These results highlight the impact of codon ambiguity in a human pathogen *C. albicans*, and provide novel insights into the molecular mechanisms regulating the virulence-associated Cek1 MAPK signaling pathway.

Resumo

Candida albicans é um fungo oportunista que reside como comensal em mais de 50% da população. No entanto, infecções causadas por *C. albicans* estão associadas a uma elevada taxa de mortalidade em indivíduos imunodeprimidos. A complexa biologia de *C. albicans* reflete-se na sua plasticidade morfológica, que em conjunto com outros factores de virulência, tais como alterações fenotípicas e alterações das propriedades moleculares da parede celular, contribuem para a sua patogenicidade. Esta complexidade é incrementada pela capacidade de *C. albicans* traduzir de modo ambíguo o codão CUG, normalmente traduzido como leucina, predominantemente como serina (97%), mas também como leucina (3%). Embora a maior parte dos codões CUG estejam posicionados em locais não conservados das proteínas, algumas proteínas contêm estes resíduos em regiões muito conservadas. Algumas destas proteínas estão associadas a vias da proteína cinase ativada por mitogénio (MAPK), como por exemplo a via de sinalização mediada por Cek1, envolvida na regulação da morfogénese e virulência de *C. albicans*. Neste trabalho, foi estudado o impacto da ambiguidade genética de *C. albicans* na MAPK Cek1, assim como na regulação das interações proteína-proteína nesta cascata de sinalização.

Dados bioquímicos e biofísicos demostraram que a ambiguidade do codão CUG afecta a estabilidade térmica e a actividade enzimática da Cek1, efeito também observado numa versão truncada contendo apenas o domínio activo. A incorporação de serina na posição do codão CUG induz a autofosforilação da tirosina do motivo conservado TEY, presente no ”loop” de activação da cinase, aumentando a sua actividade. Adicionalmente, a variante serina do domínio activo da Cek1 interage com a proteína “scaffold” Cst5. A região N-terminal da Cek1, enriquecida em repetições de aminoácidos, tem um impacto positivo na estabilidade e actividade da proteína.

Estes resultados demonstram o impacto da ambiguidade do codão CUG em *C. albicans* e fornecem novos detalhes sobre o mecanismo molecular de regulação da via de sinalização da MAPK Cek1 associada à virulência deste patógeno humano.

Chapter 1

General Introduction

1.1 The human pathogen *Candida albicans*

Fungal pathogens that cause life-threatening human diseases are often harmless inhabitants on the surface of our mucosae. The commensal fungal populations can be the source of the opportunistic pathogens causing fungal infections, which are increasing in prevalence owing to more people living with suppressed immune systems, such as organ transplant recipients, human immunodeficiency virus and cancer patients (Brown *et al.* 2012a). Although it was estimated that only 0.1% of all fungi species are human pathogens (Brown *et al.* 2012b), life-threatening fungal systemic infections display mortality rates similar to tuberculosis and 3-fold higher than malaria (Bongomin *et al.* 2017). The main fungal pathogens causing most cases of serious fungal disease are *Candida*, *Aspergillus, Cryptococcus,* and *Pneumocystis* (Brown *et al.* 2012a), with *Candida* being responsible for more than 80% of all fungal infections (Vincent *et al.* 2009).

Candida species belong to the normal microbiota of healthy individuals and have exploded into prominence in recent years as opportunistic and nosocomial fungal pathogens, especially in immunocompromised patients. Invasive candidiasis (IC), including candidemia, is reported to be the fourth leading cause of nosocomial bloodstream infections in hospitals (Tortorano *et al.* 2006, Pfaller *et al.* 2010, Antinori *et al.* 2016). Among 200 different species of *Candida*, *C. albicans*, *C. glabrata*, *C. parapsilosis*, *C. tropicalis* and *C. krusei* are responsible for >95% of *Candida* infections (Butler *et al.* 2009). Despite the existence of epidemiological data, the lack of a uniform and consistent method to diagnose and describe the incidence of IC leads to insufficient and overly diverse epidemiological data. However, based on the increased incidence of the candidemia infections through the world it is critical to develop new therapeutic approaches and a better knowledge of this human fungal pathogen (Pfaller *et al.* 2010). *C. albicans* still remains the leading cause of candidemia worldwide, however recent epidemiological data suggested the emergence of non-albicans *Candida* strains due to rapidly growing resistance of *Candida* species and the use of antifungal drugs for prophylactic purposes (Pfaller *et al.* 2014, Li *et al.* 2015, Kaur *et al.* 2016, Lamoth *et al.* 2018). One example is *C. auris*, a new *Candida* species that has been reported as an emerging global health threat, which displays elevated resistance to all antifungal drugs, resulting in a high degree of mortality (Satoh *et al.* 2009, Clancy *et al.* 2017, Lamoth *et al.* 2018). Some controversial data of the *Candida* species distribution is dependent upon geographical location (Lamoth *et al.* 2018). In

Portugal, the incidence of candidemia was found to be 2.2 per 100,000 inhabitants, one of the lowest in European countries (Faria-Ramos *et al.* 2014, Bongomin *et al.* 2017, Sabino *et al.* 2017). Moreover, the distribution of *Candida* species in Portugal confirms the presence of *C. albicans* as the predominant species (Correia *et al.* 2004, Correia *et al.* 2016), followed by *C. parapsilosis* and *C. glabrata* (Costa-de-Oliveira *et al.* 2008, Faria-Ramos *et al.* 2014).

C. albicans is the most prevalent pathogen normally residing as a commensal inhabiting the mucosal oral cavity and gastrointestinal and genitourinary tracts in more than 50% of the human population (Pfaller *et al.* 2007). Clinical manifestations, from superficial to systemic infections, are particularly important in immunocompromised individuals, such as patients with acquired immunodeficiency syndrome and patients undergoing chemotherapy, and in individuals with multiple risk factors, subjected to invasive procedures and those who have implanted medical devices (e.g., catheters) (Pfaller *et al.* 2007, Noble *et al.* 2017). *C. albicans* infections are recognized as a serious public health challenge with high clinical and socio-economic importance. According to the Center for Disease Control, *C. albicans* is the sixth most common cause of nosocomial infections and has been associated with a high crude mortality rate of approximately 40% (Sardi *et al.* 2013).

C. albicans survives under the challenging and diverse conditions associated with various mucosal surfaces in the human body, demonstrating the great adaptability of this human pathogen to different host niches. Although the research over the last few decades shed some light on the underlying molecular mechanisms of *C. albicans* infectiousness, the difficulty to diagnose and effectively treat invasive or systemic candidiasis requires a concerted effort toward the discovery of novel antifungal therapies (Pfaller *et al.* 2007, Brown *et al.* 2012a).

1.1.1 Overview of Current Antifungal Therapies

The knowledge of *C. albicans* biology and virulence factors is essential not only to understand its pathogenic mechanisms but also to identify potential antifungal targets. An increment in resistance to traditional antifungals leads to the need to control *C. albicans* infections through early diagnosis and prevention of candidiasis. Only a few antifungal drugs are currently available to treat mucosal or systemic infections with *Candida*. The classes of antifungal agents that are mainly used include azoles, polyenes, and echinocandins (Figure 1.1). The mechanisms of action of azole and polyene antifungals are confined to targeting the fungal cell

membrane, whereas echinocandins act by disrupting the fungal cell wall (Pfaller 2012).

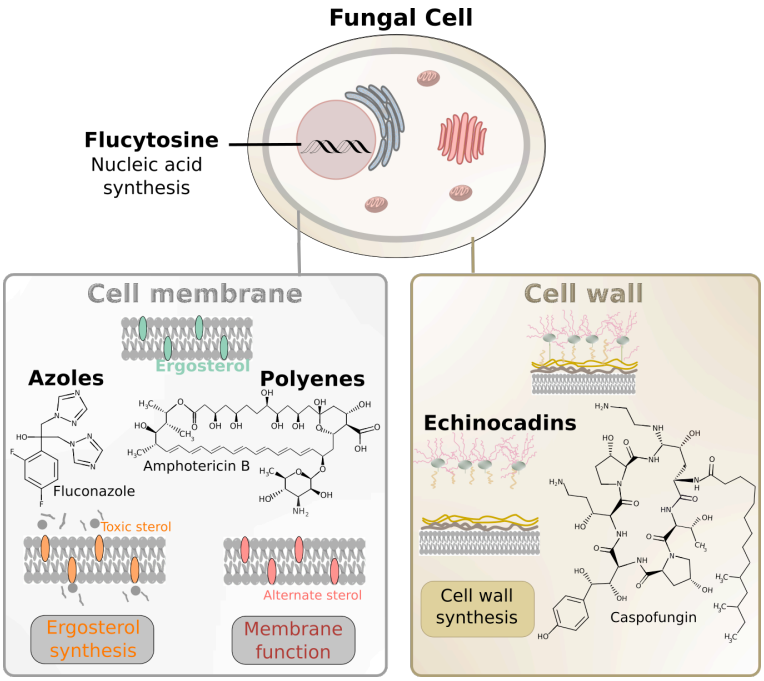


Figure 1.1 Mechanism of action of antifungal drugs. Azoles, such as fluconazole, target the ergosterol biosynthetic enzyme - lanosterol demethylase - blocking the production of ergosterol and causing the accumulation of a toxic sterol, which exerts severe membrane stress on the cell. Polyenes, like amphotericin B, exist primarily in the form of large, extramembranous aggregates that extract ergosterol from the cell membrane. Echinocadins, such as caspofungin, act as noncompetitive inhibitors of (1,3)-β-glucan synthase, causing loss of cell wall integrity and severe cell wall stress. Flucytosine interferes with DNA and RNA synthesis (Adapted from (Ostrosky-Zeichner *et al.* 2010, Shapiro *et al.* 2011).

The most commonly used class of antifungals agents, the azoles (fluconazole, itraconazole, voriconazole, posaconazole, and isavuconazole), target ergosterol biosynthesis, leading to the accumulation of toxic sterol intermediates and cell membrane stress (Ostrosky-Zeichner *et al.* 2010, Lv *et al.* 2016). Ergosterol, the main sterol in the fungal cell membrane, modulates structural and regulatory membrane features, such as fluidity and permeability and the function of many membrane-bound enzymes (Lv *et al.* 2016). The extensive prophylactic use of azoles and their fungistatic nature allow fungal survival, resulting in widespread *C. albicans* resistance to clinically relevant azoles.

Polyenes, such as amphotericin B, are amphipathic natural products that bind and extract ergosterol from cell membranes, creating drug-lipid complexes. These complexes disrupt the osmotic integrity of the fungal membrane, preventing ergosterol from exerting its many essential cellular functions, leading to cell lysis (Gray *et al.* 2012, Anderson *et al.* 2014). Structural similarities between ergosterol and cholesterol in the mammalian cell membrane represent the major use limitation associated with severe host toxicity (Shapiro *et al.* 2011). Echinocandins (caspofungin, micafungin, and anidulafungin) act by inhibiting β-1,3 glucan synthesis in the fungal cell wall, resulting in the loss of cell wall integrity and induction of cell wall stress. A significant roadblock in the elimination of *Candida* infections is the presence of drug-resistant biofilms. Fungal cells within the biofilms display resistance to azoles and polyenes, while echinocandins likely achieve better results against *Candida* biofilms (Taff *et al.* 2013, Jadhav *et al.* 2017). Furthermore, cell wall disturbing agents, such as Congo red, Calcofuluor white or zymolyase were demonstrated to induce cell wall stress by interfering with cell wall biogenesis (Sanz *et al.* 2017).

Targeting normal metabolic processes can be a viable strategy for drug development against fungal pathogens. The fluoropyrimidines, such as flucytosine (5-fluorocytosine), a systemic antifungal drug with limited activity spectrum, inhibit nucleic acid synthesis. Indeed, 5-fluorocytosine has no intrinsic antifungal activity; only after uptake by the fungal cells it is converted into compounds that block DNA synthesis or are incorporated into RNA, leading to the disruption of DNA replication and miscoding and disruption of protein synthesis (Vermes *et al.* 2000, Denning 2003, Hope *et al.* 2004, Cowen *et al.* 2008). These compounds resemble structurally key metabolites utilized by fungi but not by mammalian cells, which lack the enzymes to readily transport the drug into the cell and to convert it into toxic intermediates (Vermes *et al.* 2000). As a result of the existing resistant strains and the frequently acquired resistance during treatment, combination therapies with antifungal compounds are one of the pharmacological strategies that are beneficial in the treatment of – *Candida* - infections (Robbins *et al.* 2017). For example, due to a high incidence of primary or -acquired resistance to fluorocytosine, the use of this compound as monotherapy is restricted and it is commonly administered in combination with amphotericin B (Vermes *et al.* 2000).

Alternative therapies include the use of new active principles obtained from different sources, such as natural products, synthetic agents or polymeric materials, as well as vaccines and antibody-based immunotherapy (Castillo *et al.* 2018, de

Oliveira Santos *et al.* 2018, Nami *et al.* 2019). Recently, small-molecule inhibitors of filamentation were identified as promising candidates for the development of novel anti-virulence approaches to combat *C. albicans* infections (Romo *et al.* 2017). In summary, reducing *C. albicans* virulence by targeting its pathogenic processes represents a very attractive approach for the development of new antifungal drugs.

1.2 *C. albicans* biology

The most opportunistic human fungal pathogen, *C. albicans*, has some crucial and interesting features, which are involved in and affect its virulence. The phenotypic plasticity of *C. albicans* represents an important adaptation strategy to different environmental conditions and host niches. The capability to adopt different morphologies (e.g., yeast and hyphal forms, white and opaque cells, gray and Gastrointestinally indUced Transition – GUT - cells) represents a distinctive characteristic of *C. albicans* biology, linked to its pathogenicity during adhesion, invasion and bloodstream propagation. Moreover, the complexity of *C. albicans* biology expands with its ability to ambiguously translate the universal leucine CUG codon, predominantly as serine but also as leucine, in approximately 66% of *C. albicans* protein-coding genes (Rocha *et al.* 2011). The use of a non-universal genetic code leads to a diverse proteome, introduces different cell and colony morphologies, as well as distinct antifungal and immune responses. Understanding CUG mistranslation provides crucial insights into the physiological impact of genetic code ambiguity in *C. albicans*.

1.2.1 *C. albicans* genetic code ambiguity

The genetic code was once thought to be universal for all organisms. However, several genetic code alterations were found in mitochondrial and nuclear genomes, with a code change more pervasive in mitochondria than in nuclei (Knight *et al.* 2001, Swire *et al.* 2005). Two theories are evolving to explain the evolution of these alterations: the Codon Capture and the Ambiguous Intermediate theories. The codon capture model proposes that under directional mutational pressure (biased GC or AT pressure) certain codons lose the functionality of the corresponding tRNAs and are eliminated from the genome (Jukes *et al.* 1993). These unassigned codons can reappear and encode another amino acid, becoming reassigned into the genome by mutation of non-cognate misreading tRNAs. In contrast, in the strictly

neutral codon-disappearance hypothesis, selection can drive the process of codon reassignment with an intermediate stage in which translation is ambiguous (Schultz *et al.* 1996). According to this theory, the appearance of a mutant tRNA alters its decoding efficiency and specificity, leading to an expanded decoding capacity through an ambiguous decoding of a single codon by its cognate tRNA and mutant tRNA (Schultz *et al.* 1996).

Although several variations to the standard code supported by the ambiguous intermediate theory are described in yeast, here we focus on the nuclear genetic code change. This concept involves a sense-to-sense reassignment of the standard leucine CUG codon that occurs in several fungal species of the CTG clade (Figure 1.2), a group composed by *Debaryomyces hansenii* and several *Candida* species, which translate the CUG codon with distinct meanings (Santos *et al.* 1996, Miranda *et al.* 2007, Butler *et al.* 2009).

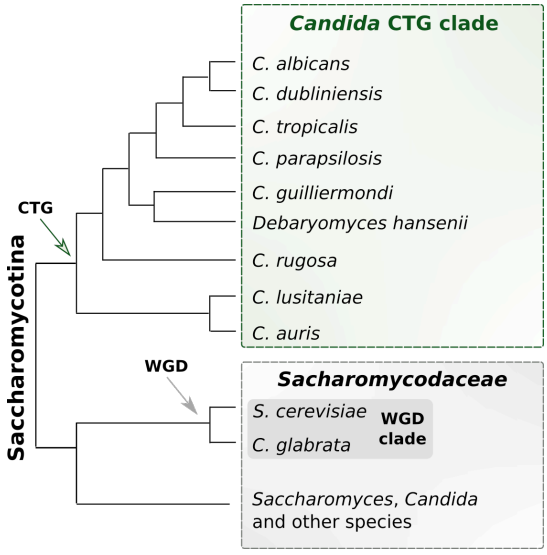


Figure 1.2 Schematic illustration of the phylogeny of Saccharomycotina species, namely *Candida* species. *C. albicans* and non-albicans *Candida* species (*C. dubliniensis, C. tropicalis, C. parapsilosis, C. guilliermondii, D. hansenii, C. rugosa, C. lusitaniae, C. auris*) are part of the *Candida* CTG clade and translate CTG codons as serine instead of leucine. In contrast, *C. glabrata* with *S. cerevisiae* belong to the *Saccharomycetaceae*, within the whole genome duplication (WGD) clade. The branch lengths are arbitrary.

This ambiguity is a consequence of the appearance of a mutant serine tRNA ($tRNA_{CAG}^{Ser}$) 270 million years ago, through the insertion of adenosine in the middle position of the CGA anticodon, creating the CAG anticodon sequence that matches the CUG codon sequence (Massey *et al.* 2003). The appearance of $tRNA_{CAG}^{Ser}$

produced a unique situation where the CUG codon became ambiguously decoded by two distinct tRNA species ($tRNA_{CAG}^{Ser}$ and the $tRNA_{CAG}^{Leu}$). During evolution, *Candida* species lost the original $tRNA_{CAG}^{Leu}$ and retained the mutant $tRNA_{CAG}^{Ser}$, changing the identity of the CUG codon from leucine to serine. On the other hand, *Saccharomyces* species, which lost the mutant $tRNA_{CAG}^{Ser}$, maintained CUG identity for leucine. In *C. albican*s the $tRNA_{CAG}^{Ser}$ has suffered mutations that further allowed the recognition by both seryl-tRNA (SerRS) and leucyl-tRNA (LeuRS) synthetases (Massey *et al.* 2003, Miranda *et al.* 2007).

Remarkably, *C. albicans,* the focus of this work*,* belongs to the *Candida* CTG clade together with *C. dubliniensis, C. tropicalis, C. guilliermondii* and many others which translate CUG codons ambiguously as both serine and leucine, while *C. cylindracea* decodes CUG codons only as serine (Santos *et al.* 1995, Suzuki *et al.* 1997, Gomes *et al.* 2007). Accordingly, non-standard codon usage in *C. albicans,* implies the translation of the leucine CUG codon predominantly as serine (97%), with only a small percentage (3%) of those codons still translated as leucine (Santos *et al.* 1993, Miranda *et al.* 2007). This CUG codon-dependent mechanism generates ambiguous proteins and expands the proteome of *C. albicans,* a mechanism that could lead to a serious decline in fitness due to the appearance of misfolded proteins and the synthesis of non-functional or toxic proteins (Santos *et al.* 1999, Santos *et al.* 2011). However, the alternative codon interpretation and the reassignment of the leucine CUG codon to serine resulted in *C. albicans* being highly tolerant to codon ambiguity and in certain conditions the ambiguous translation likely being advantageous (Santos *et al.* 1999, Silva *et al.* 2007, Butler *et al.* 2009, Moura *et al.* 2010, Rocha *et al.* 2011). Accordingly, *C. albicans* incorporated 3 to 5% leucine at CUG positions under variable stress conditions, namely temperature, low pH and oxidative stress (Gomes *et al.* 2007). In addition, the mis-incorporation rate could be artificially increased up to 28% without effects in *C. albicans* growth rate, but resulting in alterations in cell morphology (Gomes *et al.* 2007). Indeed, above 28% of leucine incorporation, *C. albicans* exhibited phenotypic diversity and genome-wide effects with expression of genes involved in cell adhesion, hyphal growth and an altered host immune response (Miranda *et al.* 2007, Bezerra *et al.* 2013, Miranda *et al.* 2013, Bezerra *et al.* 2015), suggesting that CUG ambiguity can be relevant in the process of *C. albicans* infection.

In order to understand the global impact of this ambiguous decoding, analysis of the CUG distribution in *C. albicans* ORFeome unveiled that *C. albicans*, as well as related CTG-clade species, are optimally adapted to ambiguity (Rocha *et*

al. 2011). In *C. albicans*, proteins containing CUG-residues are widely distributed by multiple functional categories, with a representative group involved in membrane and cell wall biogenesis and other virulence factors related to pathogenesis, like morphogenesis, biofilm formation and cell adhesion (Rocha *et al.* 2011). Contrarily to what happens in *Saccharomyces cerevisiae,* in which CUG-encoded residues are uniformly distributed in the protein sequence and predominantly buried in the protein core, in *C. albicans* ORFeome there has been a reassignment of the CUG codons, mainly distributed in less well-conserved regions of the proteins (90%) (Rocha *et al.* 2011). These CUG-encoded residues are critical for protein structure and folding, thereby reducing the probability of protein misfolding and explaining the tolerance of this microorganism to fluctuations in leucine incorporation levels (Rocha *et al.* 2011). Although most CUG-residues are positioned in non-conserved protein sites, in a small number of proteins the CUG-encoded residues are located in conserved active sites and are functionally relevant, suggesting an impact on structure and protein function (Rocha *et al.* 2011, Sárkány *et al.* 2014). Interestingly, many of these proteins are associated with signal transduction pathways, cell wall biosynthesis and virulence in *C. albicans* (Rocha *et al.* 2011). Functional differences were observed in eukaryotic translation initiation factor 4E upon incorporation of serine or leucine within a non-conserved site at the protein surface, demonstrating that CUG ambiguous decoding has an impact on modulating protein function (Feketova *et al.* 2010). Moreover, leucine incorporation at the conserved CUG positions of SerRS and LeuRS proteins does not affect their overall structure, but leads to a modulation of their enzymatic activity (Rocha *et al.* 2011, Zhou *et al.* 2013, Ji *et al.* 2016). Still, the in-depth analysis of the proteins affected by CUG ambiguous decoding and their impact on *C. albicans* commensalism and virulence dichotomy, remains undisclosed.

1.2.2 Morphogenesis: Yeast-to-hyphal transition

The complex biology of *C. albicans* is reflected by its impressive morphological plasticity - an important virulence attribute of this organism - in response to different environmental conditions, such as temperature, pH, CO_2 and nutrient availability (Sudbery *et al.* 2004, Shapiro *et al.* 2011). *C. albicans* is a highly polymorphic fungus that can undergo yeast-to-hypha transition, growing as unicellular budding yeast, in filamentous pseudohyphal and hyphal form and chlamydospores (Figure 1.3A). Yeast, also known as 'white cells' are characterized

by their elliptical form and asexual reproduction by budding, with a nuclear division at the junction between mother and daughter cell. Filamentous growth results in the production of elongated cells relative to the yeast form. The pseudohyphal cells are wider and form constrictions between elongated buds at the position of septa showing regular branching with daughter cells attached to the mother cells. In contrast, hyphal cells are narrower than pseudohyphal cells and have a parallel-sided wall, separated by septa with no constrictions (Sudbery *et al.* 2004, Sudbery 2011). Chlamydospores are round or oval refractive cells with a thicker cell wall and are larger than yeast cells (Sudbery 2011). Although both yeast and filamentous states are considered to be involved in the infection process, yeast cells are more relevant in the first state of dissemination in the bloodstream, whereas the hyphal form enables *C. albicans* to form biofilms and has been associated with greater adherence and tissue invasion (Saville *et al.* 2003).

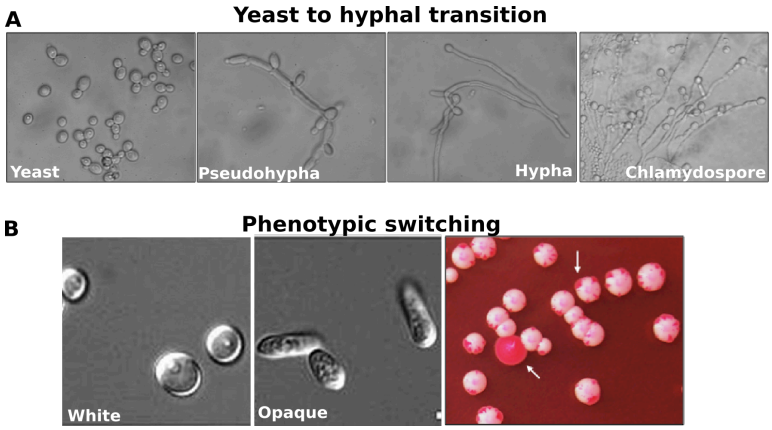


Figure 1.3 Distinct morphological states of *C. albicans*. (A) Microscopic appearance of the several morphologies display by *C. albicans* during yeast to hypha transitions: yeast, hyphal, pseudohyphal forms and chlamydospores. (B) Morphologies of two distinct phenotypes: white and opaque cells (left). Both morphotypes are easily detectable by their preferential staining with phloxine-B: opaque cells are identified by dark pink sectors (arrows in the right panel) (Adapted from (Bommanavar *et al.* 2017, Correia *et al.* 2017).

1.2.3 Phenotypic switching: white-opaque transition

Additional to yeast-to-hypha transition, *C. albicans* is able to form several more elongated yeast-like cell types (opaque, grey and GUT) that exhibit distinct properties and engage in different interactions with the host. Changes in cell and colony morphology – phenotypic switching – a spontaneous, reversible and low-

frequency transition process relevant to *C. albican*s pathogenicity, involves the interconversion between white domed colonies and opaque flat colonies (Slutsky *et al.* 1987, Soll 2014b) (Figure 1.3B). White colonies grow as typical round and smooth creamy colored colonies. In contrast, opaque cells are larger and more elongated, exhibit unique cell surface irregularities – pimples – and are the mating-competent cells. The two phenotypes vary in the ability to absorb the phloxine B dye, a stain that reflects alterations in cell wall permeability, readily used to discriminate between the white and opaque phases of *C. albicans.* The opaque cells form grey flattened colonies selectively stained dark pink with phloxine B (Anderson *et al.* 1987). White and opaque cells also differ in their gene expression profiles, mating competency, and virulence characteristics (Miller *et al.* 2002, Soll 2009). The white-opaque switching is highly sensitive to environmental conditions such as anaerobiosis (Ramirez-Zavala *et al.* 2008), the presence of CO_2 (Huang *et al.* 2009), *N*-acetylglucosamine (Huang *et al.* 2010) or acidic pH (Sun *et al.* 2015). However, glucose, low levels of CO_2 and alkaline pH promote the reverse switch to white cell (Slutsky *et al.* 1987). *C. albicans* white-opaque transition is regulated by genes present at the mating-type like (MTL) locus in response to pheromones (Miller *et al.* 2002). In nature, most strains are predominantly MTL-heterozygous (a/α), in which cells remain in the white state. When cells become MTL-homozygous (a/a or α/α), strains can undergo the transition to the opaque form (Miller *et al.* 2002, Soll 2009), the only mating-competent form cells of *C. albicans*. Moreover, mating-incompetent white cells can also display a pheromone response through biofilm formation (Soll 2014a). The white-opaque switch has important consequences on *C. albicans* pathogenicity, particularly on the mucosa, by affecting fungal adherence and biofilm formation. For instance, opaque cells are described to be more involved in cutaneous infections, while white cells are more virulent in systemic candidiasis (Soll 2009).

In addition to the white-opaque bistable transition, the 'white-gray-opaque' tristable phenotypic switching system was identified in *C. albicans* (Tao *et al.* 2014). Gray cells are an intermediate state, distinct in cellular and colony appearance from both the white and opaque phenotypes, that form smooth and dark colonies and also differ in global gene expression profile and virulence characteristics (Tao *et al.* 2014, Bommanavar *et al.* 2017). Furthermore, another phenotype has been also described in *C. albicans* during colonization of the gastrointestinary (GI) tract *in vivo* (White *et al.* 2007, Pande *et al.* 2013). During the passage through the mouse GI tract, *C. albicans* cells adopt an appearance of a novel yeast-like GUT

morphotype. While GUT cells are similar in appearance to opaque cells, these cells are characterized by the absence of pimple structures at the cell surface and show a lower mating efficiency, together with the expression of a transcriptome that is optimized for the adaptation to the digestive tract (Pande *et al.* 2013).

1.2.4 *C. albicans* cell wall structure

The *C. albicans* cell wall plays important roles in pathogenicity and virulence. It is an essential structure, which confers mechanical stability to the cell, maintaining its characteristic shape, morphology and integrity, essential for fungal adhesion and survival within the host (Klis *et al.* 2001). The cell wall also acts as a protective barrier and has attracted interest as a potential target for novel antifungal agents. In response to environmental and growth conditions, the cell wall can alter its structure or components, affecting the host immune response (Gow *et al.* 2012). The knowledge of its components and architecture contributes to understanding the cell wall assembly mechanisms, which are tightly linked to *C. albicans* pathogenicity, and offer a target to explore new strategies for its control.

C. albicans cell wall has a multi-layered structure with three major classes of macromolecules: mannans, β-glucans and chitin (Gow *et al.* 2012) (Figure 1.4). Two layers can be distinguished in the *C. albicans* cell wall: i) the outer layer (OL), composed by cell wall proteins (CWPs) highly O- and N-glycosylated (mannoproteins) that represent 35-40% of the total cell wall; and ii) the inner layer (IL) that contains the skeletal polysaccharides β-1,3 and β-1,6 glucans (50-60%) and chitin (1-2%). The CWPs located in the outer layer are attached to the inner layer via internal repeat domains (PIR) linkages or glycosyl-phosphatidyl-inositol (GPI), which is in turn linked to the skeleton through a more flexible β-1,6 glucan (Gow *et al.* 2012). While the polysaccharides localized in the inner layer – β-1,3 glucan and chitin – confer strength and shape to the cell wall, the less structured and permeable porous mannans of the outer layer do not influence cell shape. Instead, mannans affect the resistance of the cell wall to host molecules and the permeability to antifungal drugs.

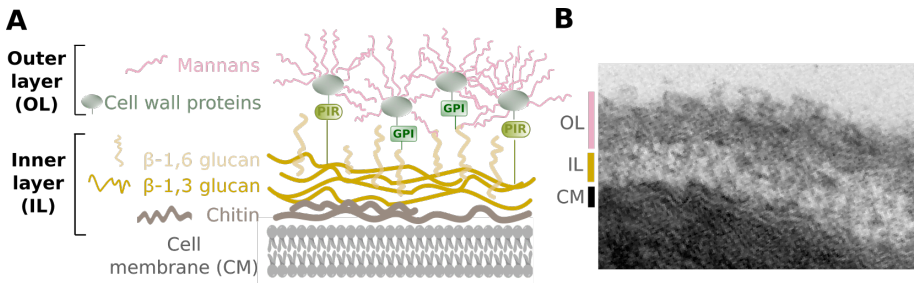


Figure 1.4 Structural organization of the fungal *C. albicans* cell wall. (A) Schematic representation of the two layers, IL and OL, distinguished in *C. albicans* cell wall with major structural components. The main structural components located in the inner layer are the skeletal polysaccharides chitin (brown) and β-glucans (β-1,3 glucan, dark yellow; β-1,6 glucan, light yellow). The outer layer is composed by mannans (pink) and cell wall proteins (gray) that are attached to the inner layer via GPI or PIR linkages (green). (B) Transmission electron micrograph sections of the *C. albicans* cell wall with main structural components identified (cell membrane, black; inner layer, yellow; outer layer, pink).

1.2.5 Virulence Factors and Pathogenicity in *C. albicans*

Several attributes and activities have been identified and considered virulence factors that contribute to the pathogenic potential of *C. albicans*. Virulence refers specifically to the ability of a pathogen to multiply and cause harm to its host (Casadevall *et al.* 2007, Mayer *et al.* 2013). *Candida* pathogenicity is due to the various virulence factors described above: morphogenesis (Sudbery *et al.* 2004), which is an important mechanism affecting the adhesion, invasion and dissemination of the fungus; phenotypic switching (Soll 2014b); and cell wall construction (Gow *et al.* 2012), which is crucial for the interaction with the host immune cells and for stress adaptation, especially relevant in an ever-changing host environment.

Furthermore, a number of other fungal attributes, such as expression of adhesion and invasion factors on the cell surface (Zhu *et al.* 2010), biofilm formation (d'Enfert 2006) and secretion of hydrolytic enzymes (proteases, phospholipases and lipases) (Hube 2006, Brown *et al.* 2007) are also implicated in *C. albicans* infection. *C. albicans* pathogenesis is defined by three major stages: adhesion, invasion and damage. The mechanism from attachment to invasion and damage is a dynamic process where it is possible to identify genes that are responsible for these distinct stages (Wachtler *et al.* 2011).

The phenomenon of adhesion is characterized by specialized surface proteins – adhesins – that mediate the attachment of fungal cells to host tissues and represents the initial stage of interaction in both commensalism and pathogenesis (Tronchin *et al.* 2008). Although both morphological forms of *C. albicans* are relevant during the various stages of the infectious process, filamentous forms are considered more adhesive due expression of numerous cell type-specific virulence factors including adhesins (Saville *et al.* 2003). One of the most studied groups of *C. albicans* adhesins is the Agglutinin-Like-Sequence (ALS), with members such as Als3, upregulated during infection of oral epithelial cells *in vitro* and during *in vivo* vaginal infection (Hoyer *et al.* 2008), and also the hypha-associated GPI-anchored proteins (Hwp1) (Sundstrom *et al.* 2002).

In addition to the stronger adherence capacity of hyphal cells at epithelial surface layers, they also exhibit greater invasiveness to tissues. Invasion of various tissues and resistance to attack by the host immune system is necessary for a pathogen to establish infection. Hyphae invasion is described to occur directly into cells by two different routes (Zakikhany *et al.* 2007): induced endocytosis and active penetration. Briefly, the induced endocytosis process is initiated by expression of fungal cell surface proteins – invasins. Als3, along with heat shock protein 1 (Ssa1), are the most important invasins that mediate the recognition and binding to host ligands such as cadherins (E-cadherin on epithelial cells and N-cadherin on endothelial cells) (Phan *et al.* 2007, Sun *et al.* 2010). Hyphal forms are also able to secret higher amounts of hydrolases, such as aspartyl proteases (SAPs), responsible for degradating a number of important host proteins like immunoglobulins, facilitating penetration and invasion into host tissues in addition to evading immune response (Naglik *et al.* 2003, Wachtler *et al.* 2012, Kumar *et al.* 2015). On the other hand, the factors that mediate the active penetration mechanism are still uncertain, although it is clearly a fungal-driven process, which requires viable *C. albicans* hyphae (Wachtler *et al.* 2012). Hyphal extension and adhesion, cell wall integrity, physical forces and SAPs have also been proposed to contribute to active penetration (Naglik *et al.* 2003).

Hyphae invasion into or through host cells induces host damage and characterizes the *C. albicans* pathogenic state affecting the response of the host immune defense. The role of the immune system and the host tissue environment contribute to the balance between host adaptation and *C. albicans* virulence. This interaction with host tissue in favor of *Candida* results in deep tissue penetration and the establishment of infection (Hube 2004).

Remarkably, an important virulence factor of *C. albicans* is the ability to form a multicellular biofilm on abiotic or biotic surfaces, which increases the potency of this fungus to convert from commensal into a pathogenic state (Jadhav *et al.* 2017). Biofilms are the major cause of nosocomial infections due to *C. albicans* capacity to grow as a biofilm in medical devices, such as catheters, dental implants, artificial joints and pacemakers. Accordingly, biofilm-associated infections are responsible for a high rate of mortality (Douglas 2003, Pfaller *et al.* 2007) and the increase resistance of biofilms to antifungal drugs causes an enhancement of treatment costs and represents a relevant threat from a clinical perspective.

The development of biofilms is initiated by different sequential steps: yeast cells adhere to a solid surface, tissues or devices (adherence), followed by proliferation of these yeast cells and the appearance and proliferation of pseudohyphal and hyphal cells (initiation). Consequently, biofilm maturation includes growth of hyphae and secretion of extracellular matrix components (maturation). Finally non-adherent yeast cells are released from the mature biofilm surface (dispersion) to seed new sites, which also contribute directly to virulence (Finkel *et al.* 2011, Lohse *et al.* 2018). The production of surface molecules like adhesins, as well as proteinases and phospholipases during biofilm formation, heightening adhesion, facilitates penetration and invasion in host cell membranes (Nobile *et al.* 2008). These factors are crucial mediators of contacts that promote biomass accumulation leading to increased virulence aid in establishing *C. albicans* infection.

Understanding the *C. albicans* pathogenicity mechanisms that allow *Candida* to survive and cause disease improves the potential for developing novel therapeutic strategies to reduce the high mortality rates of *C. albicans* infection. The cell wall provides a resilient framework, which changes continuously and allows the survival over a wide range of environmental conditions (Gow *et al.* 2012, Gow *et al.* 2017). Under conditions of cell wall damage, redundant pathways activate salvage mechanisms, increasing the synthesis of cell wall polymers. The assembly of the cell wall can be modified and adapted to handle stress, highlighting the importance of this structure for the growth, survival and morphology of *C. albicans*. Indeed, the activation of a common set of conserved and regulatory signal transduction pathways implicated in morphogenesis, cell wall remodeling and integrity allow the adaptation and response of *C. albicans* to environment changes. The identification of *C. albicans* signaling pathways that, strengthen the

pathogenesis of IC, is essential to understand the overall mechanism and to further contribute novel approaches to control *C. albicans* infection.

1.3 *C. albicans* signaling pathways

C. albicans signaling pathways are activated in response to different stimuli including osmotic, oxidative or nitrosative stress, pH or temperature alterations, nutrient starvation, and cell wall damage induced by certain drugs. Mitogen-activated protein kinase (MAPK) pathways are conserved signaling cascades in eukaryotes and represent one of the main mechanisms of adaptation to environmental changes in *C. albicans* (Alonso-Monge *et al.* 2006, Roman *et al.* 2007). Upon perception of the stimulatory signal, the signaling cascade composed of a conserved module of three kinases – the MAP kinase kinase kinase (MAPKKK), the MAP kinase kinase (MAPKK) and the MAPK – gets activated by sequential phosphorylation (Figure 1.5). The MAPK is activated by the phosphorylation of two adjacent residues, threonine and tyrosine, in a TxY motif in the activation loop, leading to an adaptive response via specific downstream transcription factors, which activate the corresponding target genes (Roman *et al.* 2007).

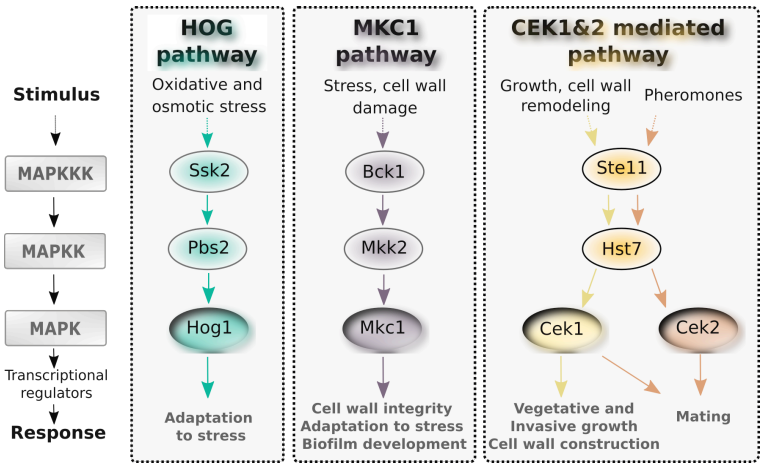


Figure 1.5 General scheme of the main elements of MAPKs signal transduction pathways in *C. albicans*. Different stimuli activate the core component of the cascade composed by a MAPKKK, a MAPKK, and a MAPK. Upon phosphorylation, the MAPK is translocated to the nucleus where it phosphorylates a transcription factor to promote the expression of the target genes leading to the adaptive response. The major MAPK pathways described in *C. albicans* are the High Osmolarity Glycerol (HOG), the Cell Wall Integrity (MKC) and the Filamentous growth/mating (Cek1/Cek2) (Adapted from (Alonso-Monge *et al.* 2006, Correia *et al.* 2017).

In *C. albicans* four MAPKs have been identified and characterized: Mkc1, Hog1 (High Osmolarity Glycerol), Cek1 and Cek2 (Alonso-Monge *et al.* 2006, Roman *et al.* 2007) (Figure 1.5). The Mkc1-mediated pathway (homolog to the Slt2/Mpk1 pathway in *S. cerevisiae*) responds to different stresses such as low temperature, osmotic and oxidative stress. Upon activation of the protein kinase C, occurs the sequential phosphorylation of the central MAPK core: the Bck1 (MAPKKK), the Mkk2 (MAPKK) and the Mkc1 (MAPK) (Navarro Garcia *et al.* 1995). This pathway mainly participates in cell wall integrity, stress adaptation and biofilm development (Kumamoto 2005, Heilmann *et al.* 2013). The HOG pathway (homolog to *Sc*Hog1) is associated predominantly with the adaptation and response to both osmotic and oxidative stress, and is triggered by Sln1, a protein kinase that acts as a sensor of stress and signals to Ypd1 and Ssk1. The signal converges to Ssk2 (MAPKKK) and Pbs2 (MAPKK) that activates the respective MAPK, Hog1, leading to the activation of downstream effectors involved in cell wall biogenesis, morphogenesis and virulence (San Jose *et al.* 1996, Alonso-Monge *et al.* 1999, Enjalbert *et al.* 2006). The others MAPK pathways described in *C. albicans* are centered on Cek1 and Cek2, homologues to *S. cerevisiae* Kss1 and Fus3, respectively, that share the upstream signaling components. While Cek2 is essentially related to mating (Chen *et al.* 2002), the Cek1-mediated pathway participates in morphogenesis, invasion and cell wall biogenesis (Csank *et al.* 1998, Roman *et al.* 2005), which is described below in more detail. Despite each signaling pathway responding to specific stimuli and generating a rather specialized response, it is clear that several cross-talk mechanisms exist between them.

1.3.1 Cek1-mediated pathway

The Cek1 MAPK cascade participates in cell wall biogenesis through the so-called Sterile Vegetative Growth pathway that promotes vegetative growth under basal conditions. The elements of this cascade include Ste11 (MAPKKK), Hst7 (MAPKK), Cek1 (MAPK) and also the scaffold protein Cst5 (Csank *et al.* 1998, Cote *et al.* 2011, Yi *et al.* 2011) (Figure 1.6). Cek1 can be regulated by growth signals requiring the upstream membrane sensors, such as Sho1, Msb2 and Opy2 (Roman *et al.* 2005, Roman *et al.* 2009b, Herrero de Dios *et al.* 2013) that lead to the cascade activation by sequential phosphorylation of the PAK kinase Cst20, and the MAPK module Ste11-Hst7-Cek1 (Csank *et al.* 1998), docked to the scaffold protein

Cst5. This module is upstream of the transcription factor Cph1 (Liu *et al.* 1994) involved in morphogenesis, hyphal formation, cell wall damage repair and glycosylation (Alonso-Monge *et al.* 2006, Ernst *et al.* 2011). Furthermore, the MAPK Cek1 pathway is also activated by the Ras1-GTPase that interacts with Cdc24/Cdc42 to promote the phosphorylation of Cst20 and consequently of Cek1, which promotes filamentous growth (Sudbery 2011). In addition, a balance between Cek1 activation and inactivation is an important step to manage the overall mechanism and *C. albicans* virulence. Similar to *S. cerevisiae* phosphatase Msg5, which dephosphorylates Fus3 (Doi *et al.* 1994), in *C. albicans* the phosphatase Cpp1 dephosphorylates Cek1 and plays a role in phenotypic switching (Csank *et al.* 1997, Guhad *et al.* 1998, Deng *et al.* 2018).

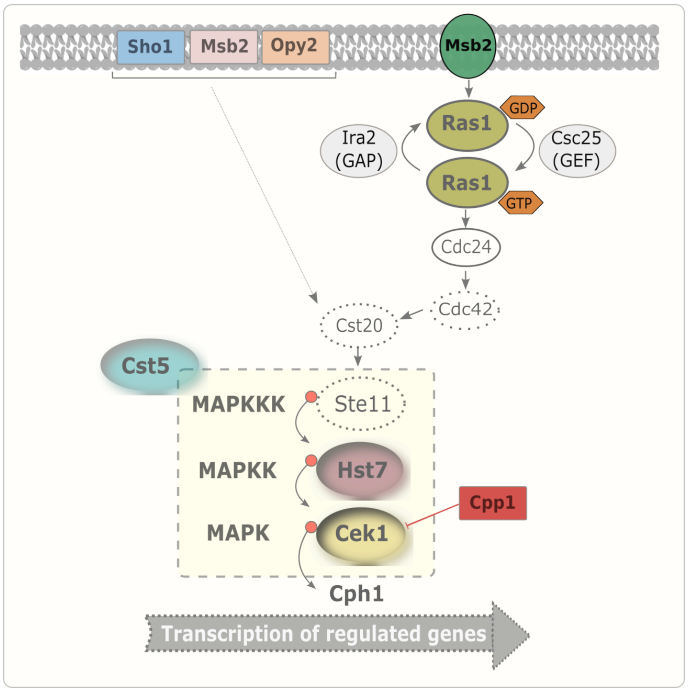


Figure 1.6 Schematic view of *C. albicans* MAPK Cek1-mediated pathway. Different stimuli are received through specific receptors at the cell membrane and transferred to the components of the MAPK cascade. The module of the MAPK central core Ste11-Hst7-Cek1 transmits the signal by sequential phosphorylation, activating Cek1 that is translocated to the nucleus and phosphorylates the transcription factor Cph1. Cph1 controls the expression of genes required for vegetative growth, morphogenesis and cell wall construction (Adapted from (Shapiro *et al.* 2011, Sárkány *et al.* 2014).

Moreover, Cek1 is a short-lived protein and can be regulated by quorum sensing (QS) molecules (Roman *et al.* 2009a), such as farnesol and tyrosol (Hornby *et al.* 2001, Chen *et al.* 2004). These molecules are secreted into the medium

during growth and are involved in morphogenesis. While farnesol inhibits hyphal and biofilm formation (Hornby *et al.* 2001, Ramage *et al.* 2002), tyrosol behaves in the opposite manner, stimulating *C. albicans* cell growth and hyphal development (Chen *et al.* 2004). Several studies indicate that the Cek1 pathway is involved in morphogenesis and filamentation. Strains with deletion of the *CEK1* gene, as well as of *CST20*, *HST7* or *CPH1*, present hyphal defects on solid medium in response to several hyphal-inducing conditions, and have a reduced virulence in animal models (Leberer *et al.* 1996, Lo *et al.* 1997, Whiteway 2000, Leberer *et al.* 2001). However, all mutants in the pathway are able to filament, which indicates that this route is not the only one leading to hyphae development but also reveals an important function of Cek1 beyond hyphae formation (Csank *et al.* 1998). In fact, the Cek1-mediated pathway is activated in response to damage to cell surface glycostructures that result in higher exposure of β-1,3 glucan and a consequent enhancement of dectin-1 mediated immune response, increasing phagocytosis and macrophage activation, revealing the importance of this pathway in *C. albicans* immune evasion mechanism (Galan-Diez *et al.* 2010). Moreover, Cek1 mutants and mutants of the upstream elements of the cascade show hypersensitivity to cell wall disturbing agents, such as calcofluor white and Congo red, demonstrating the role of Cek1 in cell wall biogenesis. The activation of the Cek1 pathway in response to drugs and consequently Cek1 phosphorylation support the relevance of Cek1 in cell wall remodeling, assembly and cell growth (Roman *et al.* 2005, Eisman *et al.* 2006, Roman *et al.* 2009a, Herrero de Dios *et al.* 2013, Román *et al.* 2016).

Besides the involvement of Cek1 on adhesion, invasion and damage, regulation of morphology and hyphal associated genes under certain conditions, it also participates in mating in response to pheromones (Chen *et al.* 2002, Magee *et al.* 2002). Thus, through the Cek1/Cek2 MAP kinase pathway these two proteins play complementary roles in *C. albicans* mating. Likewise, it is clear that several cross-talk mechanisms exist between the Cek1 and Mkc1 and HOG pathways that influence the *C. albicans* response and its virulence (Roman *et al.* 2007, Alonso-Monge *et al.* 2009).

1.3.2 Structural features and regulation of *S. cerevisiae* protein kinase Fus3

Protein kinases are components of signaling pathways that mediate cellular reactions to external and internal signals. They share a common structural topology that is a derivation of the typical protein kinase fold and is conserved in protein kinases, including in *C. albicans* Cek1 orthologue, which displays 63% amino acid sequence identity with *S. cerevisiae* Fus3 (Figure 1.7). Remarkably, Cek1 contains a CUG-encoded residue at position 199, structurally equivalent to Fus3 Ser141, within a conserved region closed to the nucleotide-binding pocket (Figure 1.7).

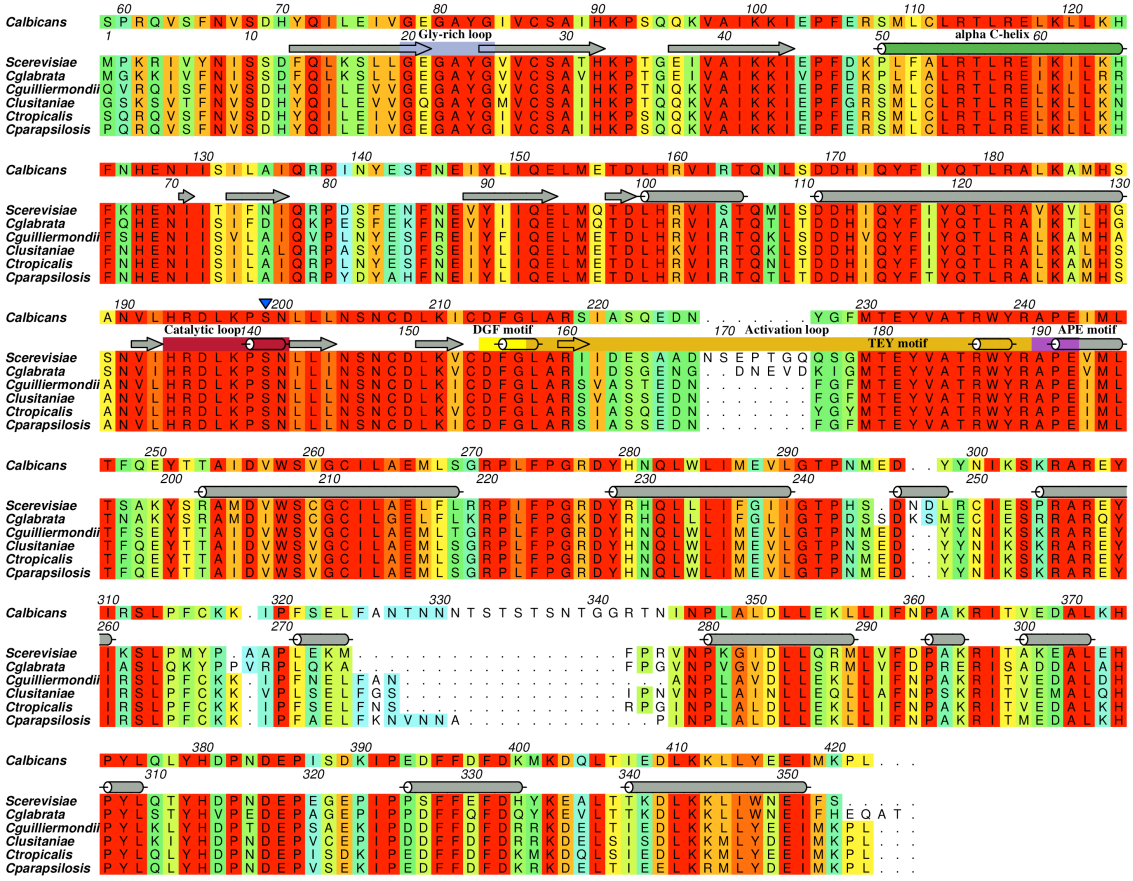


Figure 1.7 Amino acid sequence alignment of *C. albicans* Cek1 orthologues. Residues are colored according to a residue conservation scale (red: identical residues, orange to blue: decreasing conservation of amino acid properties; white: dissimilar residues). A blue triangle marks the CUG-encoded residue (Ser199) of *C. albicans* Cek1. *S. cerevisiae* Fus3 (PDB accession code 2B9F) secondary structure elements are represented above the corresponding amino acid sequence (cylinders, α-helices; arrows, β-sheets). Representative kinase structural elements are colored according to the Fus3 structure shown in Figure 1.8. Figure prepared with Aline (Bond *et al.* 2009).

The structure of *S. cerevisiae* Fus3 shown in Figure 1.8, highlights the key structural and functional elements of a protein kinase, which are essentially conserved in *C. albicans* Cek1 (Figure 1.7). The globular catalytic core is flanked by a small N-terminal lobe (mainly composed of β-strands) and a large C-terminal lobe (predominantly α-helix). The glycine-rich loop (G×G××G×V) and the αC-helix are structural elements of the N-lobe of kinases involved in nucleotide binding (Figure 1.8). The C-terminal lobe contains the substrate recognition elements and the catalytic residues (Asp137, Asn142) (Figure 1.8). The interface between these two lobes defines the active site cleft, where the adenine ring of ATP is deeply buried, with the γ-phosphate facing toward the active site opening, where protein substrates bind and the phosphotransfer reaction occurs (Zhang *et al.* 1994). Comparison of non-phosphorylated Fus3 (inactive form, Figure 1.8A) with Fus3 phosphorylated at tyrosine 182 (active form, Figure 1.8B) reveals no major differences, apart from a more flexible activation loop, which undergoes an order-to-disorder transition, allowing access of substrates to the binding pocket (Figure 1.8). This activation loop is part of a flexible extension (20–35 amino acid residues) identified between the two conserved tripeptide motifs Asp-Phe-Gly (DFG) and Ala-Pro-Glu (APE). In non-phosphorylated Fus3 only the N-terminal part of the activation loop is disordered (orange dots in Figure 1.8A), pointing to the flexible nature of this region, while the C-terminal part of the activation loop, containing the Thr180 and Tyr182 phosphorylatable residues in the conserved T×Y motif (Figure 1.7B orange sticks) is clearly structured and positioned in the substrate binding pocket. Tyr182 fits into a pocket outlined by residues at the N-terminal part of helix αC, where a bulkier phosphotyrosine side chain would not fit due to steric constraints (Remenyi *et al.* 2005, Bhattacharyya *et al.* 2006). In Fus3 phosphorylated at Tyr182, the activation loop is disordered, suggesting increased flexibility, allowing the access of substrates to the binding pocket and consequent phosphorylation (Figure 1.8B). Phosphorylation of Tyr182 induces the rotation of the N-terminal towards the C-terminal lobe, changes on the DFG motif and the rotation of helix αC, causing a conformational rearrangement of the catalytic cleft and the substrate binding site to achieve optimal kinase activity (Beenstock *et al.* 2016).

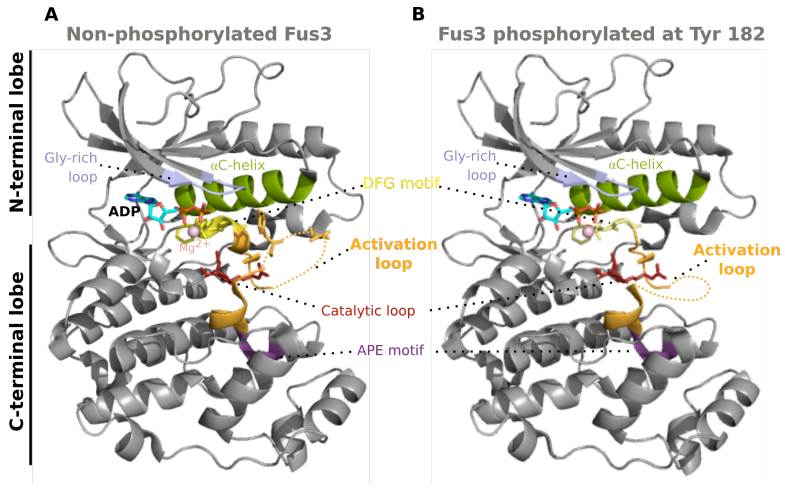


Figure 1.8 Structural analysis of *S. cerevisiae* Fus3. Overall structure of the non-phosphorylated form of Fus3 (PDB accession code 2B9F) (A) and of Fus3 phosphorylated at Tyr182 (PDB accession code 2F9G) (B). The N- and C-terminal lobes form a deep cleft that accommodates the nucleotide and Mg^{2+}. The main structural elements of Fus3 required for its catalytic activity are represented: glycine-rich loop (light blue), helix αC (green), activation loop (orange) with T×Y motif, between the DFG (yellow) and APE (violet) motifs and the catalytic loop (red). The region in the activation loop that is disordered and flexible is represented by orange dots. Figure prepared with PyMOL (Adapted from (Remenyi *et al.* 2005, Bhattacharyya *et al.* 2006)).

Activation of the *S. cerevisiae* Fus3 pathway results in the phosphorylation of $Fus3^{T180,Y182}$ by the upstream MAPKK Ste7 that requires interaction with the Ste5 scaffold protein (Remenyi *et al.* 2005, Bhattacharyya *et al.* 2006). Indeed, the Fus3 pathway has been linked directly to Ste5, which plays a direct role in the protein kinase regulation though interactions between signaling components, membrane receptors and associated kinases by its allosteric control and feedback mechanism of Ste5 phosphorylation by Fus3 (Figure 1.9).

In *S. cerevisiae* mating is mediated by a receptor-associated heterotrimeric G-protein (Gαβγ) in response to pheromones leading to the dissociation of the Gα subunit from the Gβγ complex (Ste4/Ste18), which activate the MAPK cascade through the binding of the Gβ subunit (Ste4) to the Ste5 RING domain, located near the N-terminal region (Inouye *et al.* 1997) (Figure 1.9). The Ste5 RING triggers a rapid recruitment of this scaffold protein to the membrane domain, and is also involved in the Ste5 self-association (Inouye *et al.* 1997). The recruitment and stabilization of the Ste5 interaction with the plasma membrane (PM) is also dependent of its PM domain and the pleckstrin homology (PH) domain, which is

critical for initiation of downstream signaling (Figure 1.9). The Ste5 PH domain allows interaction with the MAPKKK Ste11, which is activated by membrane-bound Ste20 (Garrenton *et al.* 2006) (Figure 1.9). Upon MAPK pathway activation, the signal is transmitted from Ste11 to Ste7 and Fus3. This mechanism is mediated by the assembly of Ste7 through the von Willebrand type A (vWA) domain of Ste5, which by allosteric interaction allows the activation of Fus3 by Ste7 (Bhattacharyya *et al.* 2006, Coyle *et al.* 2013).

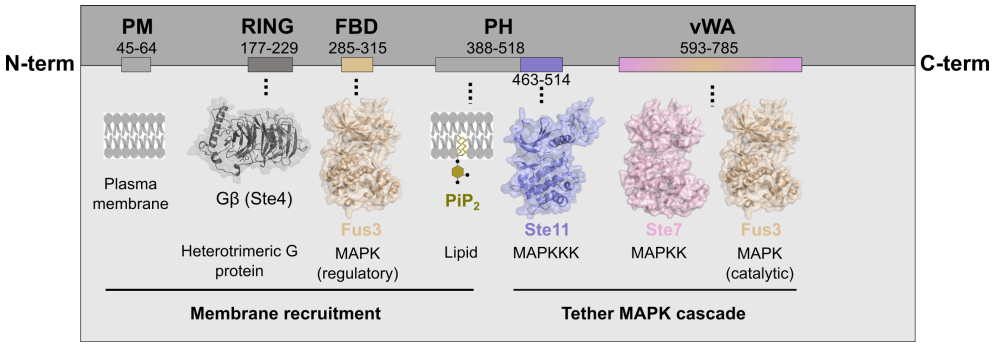


Figure 1.9 Diagram of *S. cerevisiae* Ste5 interaction networks. The modular Ste5 scaffold protein is composed of a PM interaction domain, a RING domain, a FBD domain, a PH domain and vWA domain. These domains mediate the membrane recruitment of Ste5 and the three-tiered mating MAPK signaling pathway. The 3D structure of Fus3 (PDB accession code 2B9F) and the homology models of Ste4, Ste11 and Ste7 (created using the Swiss-Model online server (Arnold *et al.* 2006)) were represented with PyMOL.

Fus3 interacts with Ste7 through a canonical MAPK docking sites at the N-terminal of Ste7 (10-residue peptides, with consensus sequence $(R/K)_{1,2}\times_{4-6}L\times L$) in the opposite side of Fus3 active site (Figure 1.10A) (Remenyi *et al.* 2005). Furthermore, activation of Fus3 by the scaffold protein Ste5 requires a short 30-amino acid peptide that binds to both N- and C-terminal kinase lobes via a flexible linker region, enhancing autophosphorylation of Tyr182 in the Fus3 activation loop (Figure 1.10B) (Remenyi *et al.* 2005, Coyle *et al.* 2013). Interestingly, this Fus3 binding domain (FBD) appears to be important for regulatory feedback phosphorylation of Ste5 by Fus3, rather than for core signal transmission through the MAPK cascade (Bhattacharyya *et al.* 2006, Good *et al.* 2009). Phosphorylation of at least four sites on the Ste5 scaffold is dependent on recruitment of Fus3 through the Ste5-FBD, which inhibits pathway activity (Malleshaiah *et al.* 2010). Moreover, an intramolecular interaction between the PH domain and vWA domain of Ste5 scaffold protein also inhibits MAPK activation (Zalatan *et al.* 2012). Recently, the N-terminal

cluster of phosphorylation sites in Ste5 was demonstrated to be important for Ste5 membrane localization and regulation of signal output (Winters *et al.* 2005, Repetto *et al.* 2018). In addition to Ste5, other regulatory phospho-inhibition mechanisms, such as the phosphatase Msg5 (Nagiec *et al.* 2015) and the Gγ subunit of Ste18 are also involved in the negative regulatory mechanism by preventing Ste5 association to PM (Choudhury *et al.* 2018).

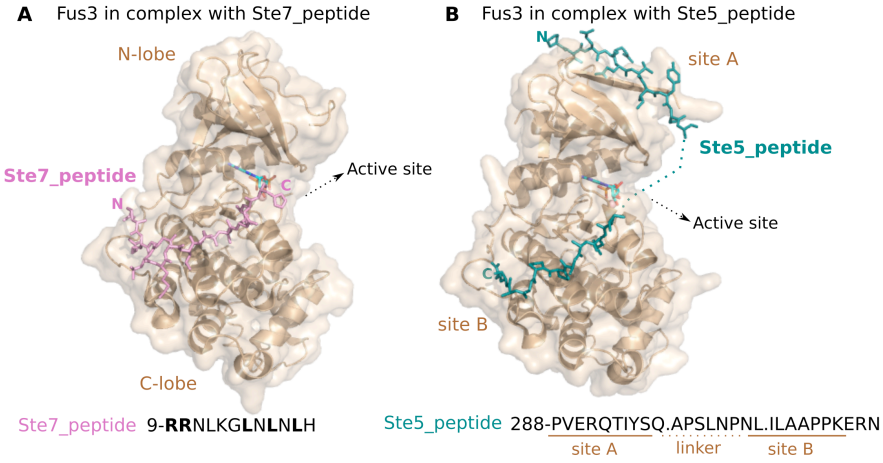


Figure 1.10 Overall structure of Fus3 in complex with Ste7 and Ste5 peptides. (A) Structure of Fus3 with docking motif from Ste7 (Ste7_peptide, purple; PDB accession code 2B9H). (B) Structure of Fus3 in complex with Ste5_peptide (cyan, PDB accession code 2F49), which binds to Fus3 in a bipartite manner (site A and B). Figure prepared with PyMOL. (Adapted from (Remenyi *et al.* 2005, Bhattacharyya *et al.* 2006)).

Besides the specific phosphorylation by upstream kinases, some protein kinases are activated by autophosphorylation of their activation segment, either through an intramolecular (*cis*) or intermolecular (*trans*) reaction (Oliver *et al.* 2007, Lochhead 2009). In the case of Fus3, autophosphorylation at Tyr182 occurs in *cis* and is relevant for producing different signaling outcomes (Bhattacharyya *et al.* 2006). Despite the high sequence similarity between *S. cerevisiae* Fus3, *C. albicans* Cek1 and many of the components of these MAPK pathways, the mechanism by which Cek1 phosphorylation and activation occurs remains unclear. In particular, the presence of proteins containing CUG-encoded residues in the Cek1-mediated pathway, as well as the disappearance of the vWA domain in Cst5, compared to *S. cerevisiae* Ste5, might challenge signal transduction and the complex interplay of protein interaction networks suggesting a reshaped molecular mechanism in Cek1 with potential implications for *C. albicans* virulence.

1.4 Objectives

C. albicans is an opportunistic human fungal pathogen, which can cause from superficial infections of the oral and genital tract to systemic disease in individuals with a compromised immune system. The high mortality rate associated with *C. albicans* infections is mainly due to the resistance to antifungal agents, making this polymorphic fungus the most prevalent causative agent of candidemia and a significant global health problem. An intriguing feature of *C. albicans* biology is the non-standard usage of the CUG codon, which is mainly decoded as serine (97%) but also translated with the standard meaning as leucine (3%), an ambiguous interpretation of the genetic code that might have impact for virulence. The ability to adapt to changing environmental conditions is ensured by signaling networks that include the MAPK signaling pathway centered on Cek1, enriched in proteins containing CUG residues. The Cek1-mediated pathway, which is involved in filamentous growth and in the biogenesis of the fungal cell wall, encompasses the upstream MAP kinases Ste11, Ste7 and also the scaffold protein Cst5. In order to understand and contribute new insights into the impact of CUG translational ambiguity for *C. albicans* virulence and pathogenesis, the main goal of the thesis was to explore the consequences of the CUG residue identity on the structure, function and macromolecular interactions of key proteins within the MAPK Cek1 pathway. To accomplish this central goal, the work includes three main sections:

In the first part, described in Chapter 2, we aimed to characterize the active domain of the Cek1 protein kinase, by analyzing the role of CUG residue identity on the stability, function and structure of this target protein. The Cek1 kinase contains a CUG residue in a strictly conserved position, close to the nucleotide binding pocket, where a serine or a leucine incorporation could modulate its kinase activity and substrate recognition.

In Chapter 3, following the observation that the Cek1 N-terminal tail is enriched in amino acid repeats, we wanted to understand the role of this non-globular region of the protein on the structure and catalytic function of this MAPK. To address this question, full-length Cek1 variants containing a serine or a leucine at the CUG residue were characterized to provide insights on the interplay between CUG ambiguity and the amino acid repeats in protein structure and function.

In Chapter 4 the objective was to explore the molecular details of the interactions between Cek1 and different components of the MAPK cascade: the

scaffold protein Cst5 and the upstream MAPKK Hst7. Moreover, a specific goal was to understand if those protein-protein interactions could be regulated by CUG residue identity. Considering this aim, we characterized the scaffold protein Cst5 and MAPKK Hst7 using biophysical and structural methods, a step towards understanding the macromolecular interaction networks and the consequences of CUG ambiguous decoding in rewiring protein interactions in the Cek1 signaling cascade linked to *C. albicans* virulence traits.

Throughout this thesis we combined biophysical and molecular biology approaches to achieve the goals established. The first challenge of the project was to produce enough high-quality recombinant proteins (Cek1, Hst7 and Cst5) in *Escherichia coli*, a heterologous protein expression system. In order to circumvent the difficulties producing Cst5, a library-based construct screening method, termed ESPRIT (Expression of Soluble Proteins by Random Incremental Truncation) was used. The recombinant proteins used were purified by standard chromatographic techniques and further analyzed in terms of structural integrity and homogeneity by several methods: dynamic light scattering (DLS) that provides information on the size characteristics and the homogeneity of proteins; circular dichroism spectroscopy (CD) that affords an estimation of the content of secondary structural features, by comparison with reference datasets from proteins with known three-dimensional structures; differential scanning fluorimetry (DSF), which relies on an environmentally sensitive fluorescent dye (Sypro Orange), was used to determine the melting temperatures (Tm) of the proteins and infer their thermal stability.

Moreover, the phosphorylation state of Cek1 was analyzed by immunoblot with specific antibodies and confirmed by mass spectrometry analysis. The kinase activity of Cek1 was monitored by the determination of ADP production in a bioluminescent kinase assay (ADP-Glo). In addition, a fluorescence method using an ATP analog (2',3´-O-(2,4,6-trinitrophenyl) adenosine 5-triphosphate (TNP-ATP), was utilized to explore its nucleotide binding affinity. TNP-ATP displays minimal fluorescence in solution by itself, but upon binding to a protein a significant fluorescence increase is observed. Besides, to uncover the macromolecular dynamics of Cek1, the interaction between Cek1 and its molecular partners was assessed using the biolayer interferometry system (BLI). Finally, to complement the work, the crystallization of the target proteins (Cek1, Hst7 and Cst5) was attempted to allow the determination of their three-dimensional structure by X-ray crystallography, either free or in complex with their molecular partners.

As a whole, the work developed in this thesis provided new insights into the impact of the CUG residue identity on the dynamics and regulation of the Cek1 signaling cascade, which may facilitate the design of new therapeutic approaches to the opportunistic fungal pathogen *C. albicans*.

1.5 References

Alonso-Monge, R., F. Navarro-Garcia, *et al.* (1999). Role of the mitogen-activated protein kinase Hog1p in morphogenesis and virulence of Candida albicans.*J Bacteriol* 181: 3058-68.

Alonso-Monge, R., E. Roman, *et al.* (2009). Fungi sensing environmental stress.*Clin Microbiol Infect* 15 Suppl 1: 17-9.

Alonso-Monge, R., E. Roman, *et al.* (2006). The MAP kinase signal transduction network in *Candida albicans*.*Microbiology* 152: 905-12.

Anderson, J. M. and D. R. Soll (1987). Unique phenotype of opaque cells in the white-opaque transition of *Candida albicans*.*J Bacteriol* 169: 5579-88.

Anderson, T. M., M. C. Clay, *et al.* (2014). Amphotericin forms an extramembranous and fungicidal sterol sponge.*Nat Chem Biol* 10: 400-6.

Antinori, S., L. Milazzo, *et al.* (2016). Candidemia and invasive candidiasis in adults: A narrative review.*Eur J Intern Med* 34: 21-28.

Arnold, K., L. Bordoli, *et al.* (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.*Bioinformatics* 22: 195-201.

Beenstock, J., N. Mooshayef, *et al.* (2016). How Do Protein Kinases Take a Selfie (Autophosphorylate)?*Trends Biochem Sci* 41: 938-953.

Bezerra, A. R., A. R. Guimaraes, *et al.* (2015). Non-Standard Genetic Codes Define New Concepts for Protein Engineering.*Life (Basel)* 5: 1610-28.

Bezerra, A. R., J. Simoes, *et al.* (2013). Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification.*Proc Natl Acad Sci U S A* 110: 11079-84.

Bhattacharyya, R. P., A. Remenyi, *et al.* (2006). The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway.*Science* 311: 822-6.

Bommanavar, S. B., S. Gugwad, *et al.* (2017). Phenotypic switch: The enigmatic white-gray-opaque transition system of *Candida albicans*.*J Oral Maxillofac Pathol* 21: 82-86.

Bond, C. S. and A. W. Schuttelkopf (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments.*Acta Crystallogr D Biol Crystallogr* 65: 510-2.

Bongomin, F., S. Gago, *et al.* (2017). Global and Multi-National Prevalence of Fungal Diseases-Estimate Precision.*J Fungi (Basel)* 3.

Brown, A. J. P., F. C. Odds, *et al.* (2007). Infection-related gene expression in *Candida albicans*.*Current Opinion in Microbiology* 10: 307-313.

Brown, G. D., D. W. Denning, *et al.* (2012a). Hidden killers: human fungal infections.*Sci Transl Med* 4: 165rv13.

Brown, G. D., D. W. Denning, *et al.* (2012b). Tackling human fungal infections.*Science* 336: 647.

Butler, G., M. D. Rasmussen, *et al.* (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes.*Nature* 459: 657-62.

Casadevall, A. and L. A. Pirofski (2007). Accidental virulence, cryptic pathogenesis, martians, lost hosts, and the pathogenicity of environmental microbes.*Eukaryot Cell* 6: 2169-74.

Castillo, H. A. P., L. N. M. Castellanos, *et al.* (2018). Nanoparticles as New Therapeutic Agents against *Candida albicans*. Candida Albicans. D. Sandai.

Chen, H., M. Fujita, *et al.* (2004). Tyrosol is a quorum-sensing molecule in *Candida albicans*.*PNAS* 101: 5048-5052.

Chen, J., J. Chen, *et al.* (2002). A conserved mitogen-activated protein kinase pathway is required for mating in *Candida albicans*.*Molecular Microbiology* 46: 1335-1344.

Choudhury, S., P. Baradaran-Mashinchi, *et al.* (2018). Negative Feedback Phosphorylation of Ggamma Subunit Ste18 and the Ste5 Scaffold Synergistically Regulates MAPK Activation in Yeast.*Cell Rep* 23: 1504-1515.

Clancy, C. J. and M. H. Nguyen (2017). Emergence of Candida auris: An International Call to Arms.*Clin Infect Dis* 64: 141-143.

Correia, A., P. Sampaio, *et al.* (2004). Study of molecular epidemiology of candidiasis in portugal by PCR fingerprinting of Candida clinical isolates.*J Clin Microbiol* 42: 5899-903.

Correia, I., D. Prieto, *et al.* (2017). The MAP Kinase Network As the Nervous System of Fungi. Reference Module in Life Sciences.
Correia, I., E. Roman, *et al.* (2016). Complementary roles of the Cek1 and Cek2 MAP kinases in Candida albicans cell-wall biogenesis.*Future Microbiol* 11: 51-67.
Costa-de-Oliveira, S., C. Pina-Vaz, *et al.* (2008). A first Portuguese epidemiological survey of fungaemia in a university hospital.*Eur J Clin Microbiol Infect Dis* 27: 365-74.
Cote, P., T. Sulea, *et al.* (2011). Evolutionary reshaping of fungal mating pathway scaffold proteins.*mBio* 2: e00230-10.
Cowen, L. E. and W. J. Steinbach (2008). Stress, drugs, and evolution: the role of cellular signaling in fungal drug resistance.*Eukaryot Cell* 7: 747-64.
Coyle, S. M., J. Flores, *et al.* (2013). Exploitation of latent allostery enables the evolution of new modes of MAP kinase regulation.*Cell* 154: 875-87.
Csank, C., C. Makris, *et al.* (1997). Derepressed hyphal growth and reduced virulence in a VH1 family-related protein phosphatase mutant of the human pathogen Candida albicans.*Mol Biol Cell* 8: 2539-51.
Csank, C., K. Schroppel, *et al.* (1998). Roles of the Candida albicans mitogen-activated protein kinase homolog, Cek1p, in hyphal development and systemic candidiasis.*Infect Immun* 66: 2713-21.
d'Enfert, C. (2006). Biofilms and their Role in the Resistance of Pathogenic Candida to Antifungal Agents.*Current Drug targets* 7: 465-470.
de Oliveira Santos, G. C., C. C. Vasconcelos, *et al.* (2018). Candida Infections and Therapeutic Strategies: Mechanisms of Action for Traditional and Alternative Agents.*Front Microbiol* 9: 1351.
Deng, F. S. and C. H. Lin (2018). Cpp1 phosphatase mediated signaling crosstalk between Hog1 and Cek1 mitogen-activated protein kinases is involved in the phenotypic transition in Candida albicans.*Med Mycol* 56: 242-252.
Denning, D. W. (2003). Echinocandin antifungal drugs.*Lancet* 362: 1142-51.
Doi, K., A. Gartner, *et al.* (1994). MSG5, a novel protein phosphatase promotes adaptation to pheromone response in S. cerevisiae.*EMBO J* 13: 61-70.
Douglas, L. J. (2003). Candida biofilms and their role in infection.*Trends Microbiol* 11: 30-6.
Eisman, B., R. Alonso-Monge, *et al.* (2006). The Cek1 and Hog1 Mitogen-Activated Protein Kinases Play Complementary Roles in Cell Wall Biogenesis and Chlamydospore Formation in the Fungal Pathogen Candida albicans.*Eukaryot Cell* 5: 347-358.
Enjalbert, B., D. A. Smith, *et al.* (2006). Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen Candida albicans.*Mol Biol Cell* 17: 1018-32.
Ernst, J. F. and J. Pla (2011). Signaling the glycoshield: maintenance of the Candida albicans cell wall.*Int J Med Microbiol* 301: 378-83.
Faria-Ramos, I., J. Neves-Maia, *et al.* (2014). Species distribution and in vitro antifungal susceptibility profiles of yeast isolates from invasive infections during a Portuguese multicenter survey.*Eur J Clin Microbiol Infect Dis* 33: 2241-7.
Feketova, Z., T. Masek, *et al.* (2010). Ambiguous decoding of the CUG codon alters the functionality of the Candida albicans translation initiation factor 4E.*FEMS Yeast Res* 10: 558-69.
Finkel, J. S. and A. P. Mitchell (2011). Genetic control of Candida albicans biofilm development.*Nat Rev Microbiol* 9: 109-18.
Galan-Diez, M., D. M. Arana, *et al.* (2010). Candida albicans beta-glucan exposure is controlled by the fungal CEK1-mediated mitogen-activated protein kinase pathway that modulates immune responses triggered through dectin-1.*Infect Immun* 78: 1426-36.
Garrenton, L. S., S. L. Young, *et al.* (2006). Function of the MAPK scaffold protein, Ste5, requires a cryptic PH domain.*Genes Dev* 20: 1946-58.
Gomes, A. C., I. Miranda, *et al.* (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans.*Genome Biol* 8: R206.
Good, M., G. Tang, *et al.* (2009). The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation.*Cell* 136: 1085-97.
Gow, N. A. and B. Hube (2012). Importance of the Candida albicans cell wall during commensalism and infection.*Curr Opin Microbiol* 15: 406-12.
Gow, N. A. R., J. P. Latge, *et al.* (2017). The Fungal Cell Wall: Structure, Biosynthesis, and Function.*Microbiol Spectr* 5.

Gray, K. C., D. S. Palacios, *et al.* (2012). Amphotericin primarily kills yeast by simply binding ergosterol.*Proc Natl Acad Sci U S A* 109: 2234-9.
Guhad, F. A., C. Csank, *et al.* (1998). Reduced pathogenicity of a Candida albicans MAP kinase phosphatase (CPP1) mutant in the murine mastitis model.*APMIS* 106: 1049-55.
Heilmann, C. J., A. G. Sorgo, *et al.* (2013). Surface stress induces a conserved cell wall stress response in the pathogenic fungus Candida albicans.*Eukaryot Cell* 12: 254-64.
Herrero de Dios, C., E. Roman, *et al.* (2013). The transmembrane protein Opy2 mediates activation of the Cek1 MAP kinase in Candida albicans.*Fungal Genet Biol* 50: 21-32.
Hope, W. W., L. Tabernero, *et al.* (2004). Molecular mechanisms of primary resistance to flucytosine in Candida albicans.*Antimicrob Agents Chemother* 48: 4377-86.
Hornby, J. M., E. C. Jensen, *et al.* (2001). Quorum Sensing in the Dimorphic Fungus Candida albicans Is Mediated by Farnesol.*Applied and Environmental Microbiology* 67: 2982-2992.
Hoyer, L. L., C. B. Green, *et al.* (2008). Discovering the secrets of the Candida albicans agglutinin-like sequence (ALS) gene family--a sticky pursuit.*Med Mycol* 46: 1-15.
Huang, G., T. Srikantha, *et al.* (2009). CO2 Regulates White-Opaque Switching in *Candida albicans*.*Curr Biol.* 19: 330-334.
Huang, G., S. Yi, *et al.* (2010). N-acetylglucosamine induces white to opaque switching, a mating prerequisite in Candida albicans.*PLoS Pathog* 6: e1000806.
Hube, B. (2004). From commensal to pathogen: stage- and tissue-specific gene expression of Candida albicans.*Curr Opin Microbiol* 7: 336-41.
Hube, B. (2006). Infection-associated genes of Candida albicans.*Future Microbiol* 1: 209-18.
Inouye, C., N. Dhillon, *et al.* (1997). Ste5 RING-H2 domain: role in Ste4-promoted oligomerization for yeast pheromone signaling.*Science* 278: 103-6.
Jadhav, A. and S. M. Karuppayil (2017). Candida Albicans Biofilm as a Clinical Challenge. Developments in Fungal Biology and Applied Mycology. T. S. e. al., Springer Nature Singapore Pte Ltd.
Ji, Q. Q., Z. P. Fang, *et al.* (2016). C-terminal Domain of Leucyl-tRNA Synthetase from Pathogenic Candida albicans Recognizes both tRNASer and tRNALeu.*J Biol Chem* 291: 3613-25.
Jukes, T. H. and S. Osawa (1993). Evolutionary changes in the genetic code.*Comp Biochem Physiol B* 106: 489-94.
Kaur, R., M. S. Dhakad, *et al.* (2016). Emergence of non-albicans Candida species and antifungal resistance in intensive care unit patients.*Asian Pac J Trop Biomed* 6: 455-460.
Klis, F. M., P. de Groot, *et al.* (2001). Molecular organization of the cell wall of Candida albicans.*Med Mycol* 39 Suppl 1: 1-8.
Knight, R. D. L. L. F. and M. Yarus (2001). How Mitochondria Redefine the Code.*J Mol Evol* 53: 299-313.
Kumamoto, C. A. (2005). A contact-activated kinase signals Candida albicans invasive growth and biofilm development.*Proc Natl Acad Sci U S A* 102: 5576-81.
Kumar, R., D. Saraswat, *et al.* (2015). Novel Aggregation Properties of *Candida albicans* Secreted Aspartyl Proteinase Sap6 Mediate Virulence in Oral Candidiasis.*Infection and Immunity* 83: 2614-2626.
Lamoth, F., S. R. Lockhart, *et al.* (2018). Changes in the epidemiological landscape of invasive candidiasis.*J Antimicrob Chemother* 73: i4-i13.
Leberer, E., D. Harcus, *et al.* (1996). Signal transduction through homologs of the Ste20p and Ste7p protein kinases can trigger hyphal formation in the pathogenic fungus *Candida albicans*.*Microbiology* 93: 13217-13222.
Leberer, E., D. Harcus, *et al.* (2001). Ras links cellular morphogenesis to virulence by regulation of the MAP kinase and cAMP signalling pathways in the pathogenic fungus Candida albicans.*Mol Microbiol* 42: 673-87.
Li, R., S. Puri, *et al.* (2015). Candida albicans Cek1 mitogen-activated protein kinase signaling enhances fungicidal activity of salivary histatin 5.*Antimicrob Agents Chemother* 59: 3460-8.
Liu, H., J. Kohler, *et al.* (1994). Suppression of hyphal formation in Candida albicans by mutation of a STE12 homolog.*Science* 266: 1723-6.
Lo, H. J., J. R. Kohler, *et al.* (1997). Nonfilamentous C. albicans mutants are avirulent.*Cell* 90: 939-49.
Lochhead, P. A. (2009). Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation.*Sci Signal* 2: pe4.

Lohse, M. B., M. Gulati, *et al.* (2018). Development and regulation of singleand multi-species Candida albicans biofilms.*Microbiology* 16.
Lv, Q. Z., L. Yan, *et al.* (2016). The synthesis, regulation, and functions of sterols in Candida albicans: Well-known but still lots to learn.*Virulence* 7: 649-59.
Magee, B., M. Legrand, *et al.* (2002). Many of the genes required for mating in *Saccharomyces cerevisiae* are also required for mating in *Candida albicans*.*Molecular Microbiology* 46: 1345-1351.
Malleshaiah, M. K., V. Shahrezaei, *et al.* (2010). The scaffold protein Ste5 directly controls a switch-like mating decision in yeast.*Nature* 465: 101-5.
Massey, S. E., G. Moura, *et al.* (2003). Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp.*Genome Res* 13: 544-57.
Mayer, F. L., D. Wilson, *et al.* (2013). Candida albicans pathogenicity mechanisms.*Virulence* 4: 119-28.
Miller, M. G. and A. D. Johnson (2002). White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating.*Cell* 110: 293-302.
Miranda, I., R. Rocha, *et al.* (2007). A genetic code alteration is a phenotype diversity generator in the human pathogen Candida albicans.*PLoS One* 2: e996.
Miranda, I., A. Silva-Dias, *et al.* (2013). Candida albicans CUG Mistranslation Is a Mechanism To Create Cell Surface Variation.*MBio* 4: e00285-13.
Moura, G. R., J. A. Paredes, *et al.* (2010). Development of the genetic code: insights from a fungal codon reassignment.*FEBS Lett* 584: 334-41.
Nagiec, M. J., P. C. McCarter, *et al.* (2015). Signal inhibition by a dynamically regulated pool of monophosphorylated MAPK.*Mol Biol Cell* 26: 3359-71.
Naglik, J. R., S. J. Challacombe, *et al.* (2003). Candida albicans secreted aspartyl proteinases in virulence and pathogenesis.*Microbiol Mol Biol Rev* 67: 400-28, table of contents.
Nami, S., A. Aghebati-Maleki, *et al.* (2019). Current antifungal drugs and immunotherapeutic approaches as promisingstrategies to treatment of fungal diseases.*Biomedicine & Pharmacotherapy* 110: 857-868.
Navarro Garcia, F., M. Sanchez, *et al.* (1995). Functional Characterization of the MKC1 Gene of Candida albicans,
Which Encodes a Mitogen-Activated Protein Kinase
Homolog Related to Cell Integrity.*Molecular and Cell Biology* 14: 2197-2206.
Nobile, C. J., H. A. Schneider, *et al.* (2008). Complementary adhesin function in C. albicans biofilm formation.*Curr Biol* 18: 1017-24.
Noble, S. M., B. A. Gianetti, *et al.* (2017). Candida albicans cell-type switching and functional plasticity in the mammalian host.*Nat Rev Microbiol* 15: 96-108.
Oliver, A. W., S. Knapp, *et al.* (2007). Activation segment exchange: a common mechanism of kinase autophosphorylation?*Trends Biochem Sci* 32: 351-6.
Ostrosky-Zeichner, L., A. Casadevall, *et al.* (2010). An insight into the antifungal pipeline: selected new molecules and beyond.*Nature reviews* 9: 719-727.
Pande, K., C. Chen, *et al.* (2013). Passage through the mammalian gut triggers a phenotypic switch that promotes Candida albicans commensalism.*Nat Genet* 45: 1088-1091.
Pfaller, M. A. (2012). Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment.*Am J Med* 125: S3-13.
Pfaller, M. A., D. R. Andes, *et al.* (2014). Epidemiology and Outcomes of Invasive Candidiasis Due to Non-albicans Species of Candida in 2,496 Patients: Data from the Prospective Antifungal Therapy (PATH) Registry 2004–2008.*PLoS One* 9: 1-12.
Pfaller, M. A. and D. J. Diekema (2007). Epidemiology of invasive candidiasis: a persistent public health problem.*Clin Microbiol Rev* 20: 133-63.
Pfaller, M. A. and D. J. Diekema (2010). Epidemiology of invasive mycoses in North America.*Crit Rev Microbiol* 36: 1-53.
Phan, Q. T., C. L. Myers, *et al.* (2007). Als3 Is a Candida albicans Invasin That Binds to Cadherins and Induces Endocytosis by Host Cells.*Plos Biology* 5: 0543-0557.
Ramage, G., S. P. Saville, *et al.* (2002). Inhibition of Candida albicans biofilm formation by farnesol, a quorum-sensing molecule.*Appl Environ Microbiol* 68: 5459-63.
Ramirez-Zavala, B., O. Reuss, *et al.* (2008). Environmental induction of white-opaque switching in Candida albicans.*PLoS Pathog* 4: e1000089.

Remenyi, A., M. C. Good, *et al.* (2005). The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network.*Mol Cell* 20: 951-62.
Repetto, M. V., M. J. Winters, *et al.* (2018). CDK and MAPK Synergistically Regulate Signaling Dynamics via a Shared Multi-site Phosphorylation Region on the Scaffold Protein Ste5.*Mol Cell* 69: 938-952 e6.
Robbins, N., T. Caplan, *et al.* (2017). Molecular Evolution of Antifungal Drug Resistance.*Annu Rev Microbiol* 71: 753-775.
Rocha, R., P. J. Pereira, *et al.* (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen.*Proc Natl Acad Sci U S A* 108: 14091-6.
Roman, E., R. Alonso-Monge, *et al.* (2009a). The Cek1 MAPK is a short-lived protein regulated by quorum sensing in the fungal pathogen Candida albicans.*FEMS Yeast Res* 9: 942-55.
Roman, E., D. M. Arana, *et al.* (2007). MAP kinase pathways as regulators of
fungal virulence.*TRENDS in Microbiology* 15: 181-90.
Román, E., I. Correia, *et al.* (2016). The Cek1-mediated MAP kinase pathway regulates exposure of α-1,2 and β-1,2-mannosides in the cell wall of Candida albicans modulating immune recognition.*Virulence* 7: 558-577.
Roman, E., F. Cottier, *et al.* (2009b). Msb2 signaling mucin controls activation of Cek1 mitogen-activated protein kinase in Candida albicans.*Eukaryot Cell* 8: 1235-49.
Roman, E., C. Nombela, *et al.* (2005). The Sho1 adaptor protein links oxidative stress to morphogenesis and cell wall biosynthesis in the fungal pathogen Candida albicans.*Mol Cell Biol* 25: 10611-27.
Romo, J. A., C. G. Pierce, *et al.* (2017). Development of Anti-Virulence Approaches for Candidiasis via a Novel Series of Small-Molecule Inhibitors of Candida albicans Filamentation.*MBio* 8.
Sabino, R., C. Verissimo, *et al.* (2017). Serious fungal infections in Portugal.*Eur J Clin Microbiol Infect Dis* 36: 1345-1352.
San Jose, C., R. A. Monge, *et al.* (1996). The mitogen-activated protein kinase homolog HOG1 gene controls glycerol accumulation in the pathogenic fungus Candida albicans.*J Bacteriol* 178: 5850-2.
Santos, M. A., C. Cheesman, *et al.* (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp.*Mol Microbiol* 31: 937-47.
Santos, M. A., A. C. Gomes, *et al.* (2011). The genetic code of the fungal CTG clade.*C R Biol* 334: 607-11.
Santos, M. A., G. Keith, *et al.* (1993). Non-standard translational events in Candida albicans mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon.*EMBO J* 12: 607-16.
Santos, M. A. S., V. M. Perreau, *et al.* (1996). Transfer RNA structural change is a key element in the reassignment of the CUG codon in Candida albicans.*The EMBO journal* 15: 5060-5068.
Santos, M. A. S. and M. F. Tuite (1995). The CUG codon is decoded in vivo as serine and not leucine in Candida albicans.*Nucleic Acids Res* 23: 1481-1486.
Sanz, A. B., R. García, *et al.* (2017). The CWI Pathway: Regulation of the Transcriptional Adaptive Response to Cell Wall Stress in Yeast.*Journal of Fungi* 4.
Sardi, J. C., L. Scorzoni, *et al.* (2013). Candida species: current epidemiology, pathogenicity, biofilm formation, natural antifungal products and new therapeutic options.*J Med Microbiol* 62: 10-24.
Sárkány, Z., A. Silva, *et al.* (2014). Ser or Leu: structural snapshots of mistranslation in Candida albicans.*Front Mol Biosci* 1: 1-14.
Satoh, K., K. Makimura, *et al.* (2009). Candida auris sp. nov., a novel ascomycetous yeastisolated from the external ear canal of an inpatientin a Japanese hospital.*Microbiol Immunol* 53: 41-44.
Saville, S. P., A. L. Lazzell, *et al.* (2003). Engineered control of cell morphology in vivo reveals distinct roles for yeast and filamentous forms of Candida albicans during infection.*Eukaryot Cell* 2: 1053-60.
Schultz, D. W. and M. Yarus (1996). On malleability in the genetic code.*J Mol Evol* 42: 597-601.

Shapiro, R. S., N. Robbins, *et al.* (2011). Regulatory circuitry governing fungal development, drug resistance, and disease.*Microbiol Mol Biol Rev* 75: 213-67.
Silva, R. M., J. A. Paredes, *et al.* (2007). Critical roles for a genetic code alteration in the evolution of the genus Candida.*EMBO J* 26: 4555-65.
Slutsky, B., M. Staebell, *et al.* (1987). White-Opaque Transition: a Second High-Frequency Switching System in *Candida albicans*.*Journal of Bacteriology* 169: 189-197.
Soll, D. R. (2009). Why does *Candida albicans* switch?*FEMS Yeast Res* 9: 973-989.
Soll, D. R. (2014a). The evolution of alternative biofilms in an opportunistic fungal pathogen: An explanation for how new signal transduction pathways may evolve.*Infection, Genetics and Evolution* 22: 235-243.
Soll, D. R. (2014b). The role of phenotypic switching in the basic biology and pathogenesis of Candida albicans.*J Oral Microbiol* 6.
Sudbery, P., N. Gow, *et al.* (2004). The distinct morphogenic states of Candida albicans.*Trends Microbiol* 12: 317-24.
Sudbery, P. E. (2011). Growth of Candida albicans hyphae.*Nat Rev Microbiol* 9: 737-48.
Sun, J. N., N. V. Solis, *et al.* (2010). Host cell invasion and virulence mediated by Candida albicans Ssa1.*PLoS Pathog* 6: e1001181.
Sun, Y., C. Cao, *et al.* (2015). pH Regulates White-Opaque Switching and Sexual Mating in Candida albicans.*Eukaryot Cell* 14: 1127-34.
Sundstrom, P., J. E. Cutler, *et al.* (2002). Reevaluation of the role of HWP1 in systemic candidiasis by use of *Candida albicans* strains with selectable marker URA3 targeted to the ENO1 locus.*Infect and Immun* 70: 3281-3283.
Suzuki, T., T. Ueda, *et al.* (1997). The 'polysemous' codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity.*The EMBO Journal* 16: 1122-1134.
Swire, J., O. P. Judson, *et al.* (2005). Mitochondrial genetic codes evolve to match amino acid requirements of proteins.*J Mol Evol* 60: 128-39.
Taff, H. T., K. F. Mitchell, *et al.* (2013). Mechanisms of Candida biofilm drug resistance.*Future Microbiol* 8: 1325-37.
Tao, L., H. Du, *et al.* (2014). Discovery of a "white-gray-opaque" tristable phenotypic switching system in candida albicans: roles of non-genetic diversity in host adaptation.*PLoS Biol* 12: e1001830.
Tortorano, A. M., C. Kibbler, *et al.* (2006). Candidaemia in Europe: epidemiology and resistance.*Int J Antimicrob Agents* 27: 359-66.
Tronchin, G., M. Pihet, *et al.* (2008). Adherence mechanisms in human pathogenic fungi.*Med Mycol* 46: 749-72.
Vermes, A., H. J. Guchelaar, *et al.* (2000). Flucytosine: a review of its pharmacology, clinical indications, pharmacokinetics, toxicity and drug interactions.*J Antimicrob Chemother* 46: 171-9.
Vincent, J. L., J. Rello, *et al.* (2009). International study of the prevalence and outcomes of infection in intensive care units.*JAMA* 302: 2323-9.
Wachtler, B., F. Citiulo, *et al.* (2012). Candida albicans-epithelial interactions: dissecting the roles of active penetration, induced endocytosis and host factors on the infection process.*PLoS One* 7: e36952.
Wachtler, B., D. Wilson, *et al.* (2011). From attachment to damage: defined genes of Candida albicans mediate adhesion, invasion and damage during interaction with oral epithelial cells.*PLoS One* 6: e17046.
White, S. J., A. Rosenbach, *et al.* (2007). Self-regulation of Candida albicans population size during GI colonization.*PLoS Pathog* 3: e184.
Whiteway, M. (2000). Transcriptional control of cell type and morphogenesis in Candida albicans.*Curr Opin Microbiol* 3: 582-8.
Winters, M. J., R. E. Lamson, *et al.* (2005). A membrane binding domain in the ste5 scaffold synergizes with gbetagamma binding to control localization and signaling in pheromone response.*Mol Cell* 20: 21-32.
Yi, S., N. Sahni, *et al.* (2011). Utilization of the mating scaffold protein in the evolution of a new signal transduction pathway for biofilm development.*MBio* 2: e00237-10.
Zakikhany, K., J. R. Naglik, *et al.* (2007). In vivo transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination.*Cellular Microbiology* 9: 2938-2954.

Zalatan, J. G., S. M. Coyle, *et al.* (2012). Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation.*Science* 337: 1218-22.
Zhang, F., A. Strand, *et al.* (1994). Atomic structure of the MAP kinase ERK2 at 2.3 A resolution.*Nature* 367: 704-11.
Zhou, X. L., Z. P. Fang, *et al.* (2013). Aminoacylation and translational quality control strategy employed by leucyl-tRNA synthetase from a human pathogen with genetic code ambiguity.*Nucleic Acids Res* 41: 9825-38.
Zhu, W. and S. G. Filler (2010). Interactions of Candida albicans with epithelial cells.*Cell Microbiol* 12: 273-82.

Chapter 2

Genetic code ambiguity modulates the activity of a *Candida albicans* MAP kinase linked to cell wall remodeling

Part of this chapter was published in:

Joana Sofia Fraga, Zsuzsa Sárkány, Alexandra Silva, Inês Correia, Pedro J.B. Pereira, Sandra Macedo-Ribeiro S. (2019), Genetic code ambiguity modulates the activity of a *C. albicans* MAP kinase linked to cell wall remodeling, BBA - Proteins and Proteomics 1867: 654-661. doi: 10.1016/j.bbapap.2019.02.004.

Work contributions:
Cloning, expression and purification protocols of the active domain of Cek1_Ser variant were established in collaboration with Zsuzsa Sárkány in the scope of Vanessa Carvalho MSc. Mass spectrometry measurements were performed by Hugo Osório at the i3S Proteomics Scientific Platform.

2.1 Introduction

Over the years it has been repeatedly demonstrated that the genetic code is not immutable and the presence of non-standard translation in evolutionarily distant organisms was frequently observed (Knight *et al.* 2001a). Variations in codon assignment, including several sense-to-sense codon alterations, have been identified in numerous mitochondrial genomes (Knight *et al.* 2001b), which are smaller than nuclear genomes and where genetic code changes are less likely to be harmful (Bezerra *et al.* 2015). Alterations in the canonical genetic code involving nonsense-to-sense codon modifications, such as the reassignment of standard termination codons to different amino acids occurred frequently in mitochondrial genomes (Bezerra *et al.* 2015), possibly through a mechanism preceded by codon-anticodon mismatches and dual function of stop codons (Seligmann 2018). Reassignment of termination codons to glutamine, cysteine or tryptophan also occurs in nuclear genomes (Bezerra *et al.* 2015), and recent studies showed that in some ciliates termination codons can be ambiguously recognized as either sense or stops in a context-dependent manner (Swart *et al.* 2016). Furthermore, ambiguous meaning of translation stop codons is at the basis of natural expansion of the genetic code, which occurs by insertion of non-canonical amino acids such as selenocysteine at UGA sites in prokaryotes and eukaryotes, and pyrrolysine at UAG codons in Methanosarcinaceae methyltransferase coding genes (Ambrogelly *et al.* 2007).

The reassignment of the “universal” leucine CUG codon to serine or alanine is the only known case of alteration of sense codon meaning in nuclear genomes and has been demonstrated in different yeast species with biotechnological or biomedical relevance. The first evidence for the change in the meaning of the CUG codon was found in yeast species belonging to the CTG-clade that comprise the most pathogenic *Candida* species, including *C. albicans* (Santos *et al.* 1993). The yeast *Pachysolen tannophilus* translates CUG codons as alanine instead of the canonical leucine, with implications in biotechnological applications (Muhlhausen *et al.* 2016). Some yeasts ambiguously decode the CUG codon as leucine and serine through different mechanisms, an intriguing feature considering that accuracy during translation of the genetic code is critical to ensure adequate cell function and mistranslation can negatively impact cell survival (Kapur *et al.* 2018, Santos *et al.* 2018). However, recent data show that cells are relatively tolerant to errors in protein translation, and that increased mistranslation (e.g., under stress conditions)

enhances microbial fitness and modulates host-microbe interactions (Evans *et al.* 2018).

C. albicans translates the CUG codon predominantly as serine (97%) with only a small percentage (3%) being translated as leucine (Santos *et al.* 1995, Miranda *et al.* 2007). In response to environmental and stress conditions, such as low pH or during oxidative stress, the incorporation of leucine increases to up to 5% (Gomes *et al.* 2007). The levels of leucine insertion at CUG codons can be artificially increased to between 28% and 98% in engineered *C. albicans* strains, resulting in colonies with high morphological diversity, enhanced expression of genes involved in cell adhesion, as well as increased resistance to antifungal agents, oxidative stress and decreased susceptibility to phagocytosis by macrophages (Miranda *et al.* 2007, Bezerra *et al.* 2013, Miranda *et al.* 2013). Those results suggested that proteome expansion resulting from ambiguous CUG translation might result in phenotypic diversity, increase fitness and modulate host-pathogen interactions. Interestingly, increased incorporation of leucine at CUG codons up to 28% did not result in significant changes in growth rates (Gomes *et al.* 2007, Miranda *et al.* 2007), demonstrating the high tolerance of *C. albicans* to CUG ambiguity. Indeed, in *C. albicans* and other CTG-clade species the distribution of residues encoded by CUG codons is prevalent in non-conserved and partially exposed regions at the protein surface, where both leucine and serine can be incorporated without major impact on protein structure or function (Rocha *et al.* 2011). However, in a few proteins, the residues encoded by CUG codons are located in functionally relevant positions, where a serine or other polar amino acids are generally conserved in homologous proteins. In contrast, in *Ascoideia asiatica* the CUG codon is translated either as leucine or serine at approximately equal proportions in a stochastic manner, and CUG-encoded residues in key structural and functional sites are strictly avoided (Muhlhausen *et al.* 2018).

Some studies have demonstrated that CUG ambiguity impacts the activity and regulation of *C. albicans* proteins with no major changes on the global protein structure. Examples are *C. albicans* eukaryotic translation initiation factor 4E that contains a CUG-encoded residue located at the protein surface (Feketova *et al.* 2010), and the seryl- and leucyl-tRNA synthetases that contain a single CUG-encoded residue each at highly conserved amino acid sequence positions (Rocha *et al.* 2011, Zhou *et al.* 2013, Ji *et al.* 2016). Previous analysis of the impact of CUG ambiguous translation in the *C. albicans* proteome revealed that CUG codons are over-represented on key signaling pathways, such as the Mitogen Activated Protein

Kinase (MAPK) pathway, that respond to various external cues and mediate the regulation of morphogenesis and virulence (Rocha *et al.* 2011). In particular, Cek1, a key MAPK, contains one CUG-encoded residue in the nucleotide binding pocket in a position where a serine is strictly conserved in homologous proteins, and ambiguity is predicted to have functional impact (Rocha *et al.* 2011, Sárkány *et al.* 2014) (Figure 2.1). Cek1 is activated by sequential phosphorylation by upstream MAPK cascade activators (Cst20-Ste11-Ste7) in response to extracellular stimuli, leading to the phosphorylation of the transcription factor Cph1 (Maiti *et al.* 2015). This kinase is involved in distinct essential mechanisms of *C. albicans*: invasive hyphal growth (Csank *et al.* 1998, Roman *et al.* 2009, Gow *et al.* 2012), biofilm formation (Yi *et al.* 2011), cell wall integrity, host recognition (Galan-Diez *et al.* 2010, Herrero-de-Dios *et al.* 2014), and response to oxidative and osmotic stress (Alonso-Monge *et al.* 2006), and therefore it is a key target to explore CUG ambiguity-mediated morphological changes.

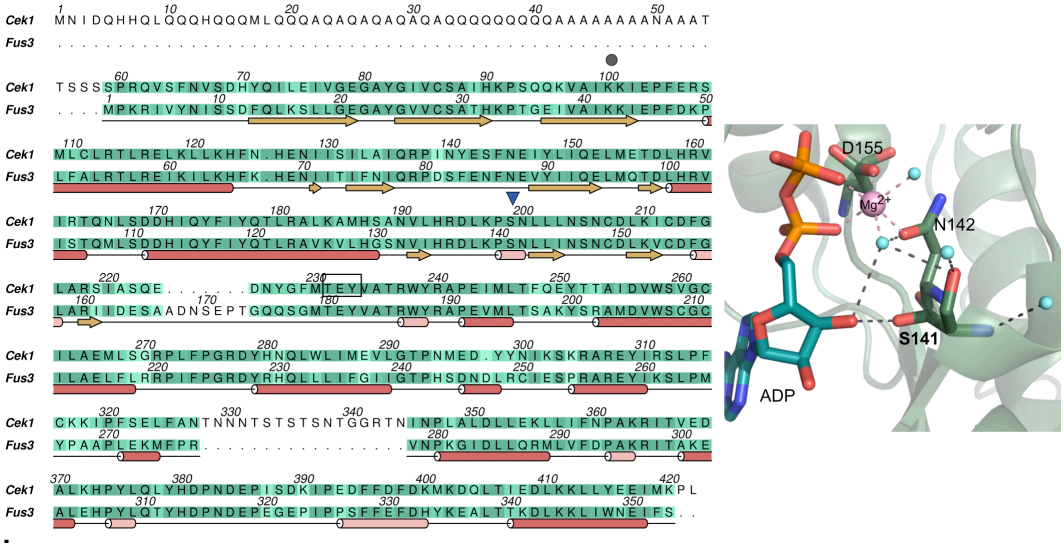


Figure 2.1 Structural analysis of *C. albicans* Cek1. Amino acid sequence alignment of *C. albicans* Cek1 and of *S. cerevisiae* orthologue Fus3. The secondary structure elements of *S. cerevisiae* Fus3 (PDB accession code 2B9F) are represented below the corresponding amino acid sequence (red cylinders, α-helices; pink cylinders, 3_{10} helices; yellow arrows, β-sheets). Important residues are marked: CUG-encoded residue (Ser199; blue triangle), catalytic residue (Lys100; black circle) and ^{231}TEY^{233} motif (black box). Detailed view of the ATP-binding pocket of *S. cerevisiae* Fus3: Ser141 (shown as sticks) of Fus3 is structurally equivalent to the *C. albicans* CUG-encoded residue 199. Hydrogen bonds are represented as dashed lines. The magnesium ion is represented as a pink sphere and blue spheres represent ordered water molecules.

To assess the impact of CUG ambiguity on Cek1 kinase, two recombinant variants of the active domain of Cek1, having a serine or a leucine residue at the CUG-encoded position (ΔN-Cek1_Ser or ΔN-Cek1_Leu) were expressed and purified for structural and functional studies. Biochemical and biophysical data show that although the incorporation of leucine at this position does not induce Cek1 unfolding, it reduces the protein's thermal stability and decreases its enzymatic activity. In addition, we show for the first time that only the Cek1_Ser variant from *C. albicans* is autophosphorylated *in vitro* at the tyrosine residue of the conserved threonine-glutamate-tyrosine (^{231}TEY^{233}) motif within the kinase activation loop.

2.2 Materials and Methods

2.2.1 Cloning, expression and purification of Cek1

A synthetic gene encoding full-length *C. albicans* Cek1 (serine variant for CUG codon at position 199) with codon usage optimized for expression in *Escherichia coli* was obtained from a commercial supplier (Eurofins Scientific). The active domain of Cek1 (amino acids 58-421) was amplified by PCR and cloned into expression vector pETMBP (EMBL dabatase) in frame with a N-terminal hexahistidine (His6) tag, the maltose binding protein (MBP) solubility tag and a tobacco etch virus (TEV) protease recognition site. This tagged form of the active domain of Cek1 is herein named ΔN-Cek1. The ΔN-Cek1_Ser (serine variant – UCC replacing the CUG codon) expression plasmid was used as template to produce a plasmid expressing ΔN-Cek1_Leu (leucine variant – UUA replacing the CUG codon) by site-directed mutagenesis (QuikChange, Agilent Technologies) (Table 2.1). To prepare the Cek1 phosphorylation-null mutants, Thr231 and Tyr233 of the conserved ^{231}TEY^{233} phosphorylation motif were mutated to valine and phenylalanine (ΔN-Cek1^{TV}, ΔN-Cek1^{YF}, ΔN-Cek1$^{TV/YF}$), respectively. Furthermore, using the same strategy a phosphomimetic mutant where Thr231 was mutated to aspartate (ΔN-Cek1^{TD}) and the kinase-dead mutant, with Lys100 replaced by arginine (ΔN-Cek1^{KR}) was produced (Table 2.1). All residue numbers refer to the full-length amino acid sequence of *C. albicans* Cek1.

Table 2.1 Oligonucleotides used to amplify the Cek1 active domain and for site-directed mutagenesis.

Oligonucleotide	Sequence (5´ to 3´)	Use
Cek1_For	CCGGCCATGGGCAGTCCACGTCAGGTCAGC	gene amplification
Cek1_Rev	CCGGGGTACCTTATTTCATGATTTCTTCATACAG	gene amplification
Cek1_Leu_For	TGGAGTTCAGCAACAAGTTTAAGGGCTTCAGATCGC GATG	mutagenesis
Cek1_Leu_Rev	CATCGCGATCTGAAGCCCTTAAACTTGTTGCTGAAC TCCA	mutagenesis
Cek1_TV_For	GAGGACAATTATGGCTTTATGGTCGAATACGTTGCA ACCCGTTG	mutagenesis
Cek1_TV_Rev	CAACGGGTTGCAACGTATTCGACCATAAAGCCATAA TTGTCCTC	mutagenesis
Cek1_YF_For	CAATTATGGCTTTATGACCGAATTCGTTGCAACCCG	mutagenesis
Cek1_YF_Rev	CGGGTTGCAACGAATTCGGTCATAAAGCCATAATTG	mutagenesis
Cek1_TV/YF_For	GGAGGACAATTATGGCTTTATGGTCGAATTCGTTGC AACCCGTTGGTA	mutagenesis
Cek1_TV/YF_Rev	TACCAACGGGTTGCAACGAATTCGACCATAAAGCCA TAATTGTCCTCC	mutagenesis
Cek1_KR_For	GCAACAGAAAGTTGCGATAAGGAAAATCGAACCGTT TGAAC	mutagenesis
Cek1_KR_Rev	GTTCAAACGGTTCGATTTTCCTTATCGCAACTTTCT GTTGC	mutagenesis
Cek1_TD_For	CAACGGGTTGCAACGTATTCGTCCATAAAGCCATAA TTGTCCTC	mutagenesis
Cek1_TD_Rev	GAGGACAATTATGGCTTTATGGACGAATACGTTGCA ACCCGTTG	mutagenesis

Cek1 variants (ΔN-Cek1_Ser and ΔN-Cek1_Leu) and ΔN-Cek1_Ser mutants were overexpressed in *E. coli* strain BL21 StarTM (DE3) (Stratagene). Following transformation, *E. coli* cells were grown in Luria-Bertani medium at 37˚C until an OD_{600} of 0.6-0.8. At this point, the temperature was decreased to 20˚C and protein expression was induced with addition of 0.9 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and allowed to proceed overnight. Cells were harvested by centrifugation at 4˚C and 4000 g for 30 min, resuspended in lysis buffer (40 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 5mM 2-β-mercaptoethanol) supplemented with 100 μg/ml lysozyme and stored at -20˚C.

Prior to lysis by sonication (SONOPULS HD2200, Bandelin Electronic), phenylmethylsulfonyl fluoride (PMSF) was added to the cell suspension. The crude

cell extract was clarified by centrifugation at 4˚C and 35,000 g for 45 min and loaded onto a HisTrap HP column (5 ml, GE Healthcare) pre-equilibrated in buffer A (40 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 5% (v/v) glycerol, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol). Bound proteins were eluted with buffer A supplemented with 75 mM imidazole. Cek1-containing fractions were pooled and further purified by size exclusion chromatography (SEC) on a HiPrep 26/60 Sephacryl S-100 column (GE Healthcare) equilibrated with protein buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 5 mM $MgCl_2$, 1 mM DTT). Fractions from the preparative size exclusion chromatography were analyzed by SDS-PAGE and those containing pure Cek1 were pooled and concentrated on centrifugal ultrafiltration devices (10 kDa cutoff; Merck Millipore). Final protein concentration was estimated by measuring the absorbance of the samples at 280nm using the theoretical extinction coefficient for each variant (Table 2.2) prior to storing the samples at -80˚C. Additionally, the purity and oligomeric state of the purified proteins were analyzed by SEC on a Superdex 200 10/300 GL column (GE Healthcare) using protein buffer as mobile phase. The average yields for the purified ΔN-Cek1_Ser and ΔN-Cek1_Leu variants were 5 and 1 mg per liter of *E. coli* culture, respectively.

In order to remove the fusion tags (His6 and MBP tags), purified ΔN-Cek1 was digested with TEV protease overnight at 4˚C in protein buffer. The buffer of the sample was then exchanged to 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% (v/v) glycerol using a HiPrep 26/10 desalting column (GE Healthcare), prior to a second immobilized metal affinity chromatography (IMAC) step using a HisTrap HP column equilibrated with buffer A. Untagged Cek1 present in the HisTrap column flow through was further purified on a Superdex 75 10/300 GL column equilibrated with protein buffer and eluted protein fractions containing pure Cek1, as assessed by SDS-PAGE, were pooled, concentrated and stored at -80˚C until used.

Table 2.2 Extinction coefficient of Cek1 variants and Cek1_Ser mutants based on amino acid sequence.

Protein	CUG residue	Mutation	Extinction coefficient (M^{-1} cm^{-1}) (tag/no tag)
ΔN-Cek1_Ser	serine	-	109,670/41,830
ΔN-Cek1_Leu	leucine	-	109,670/41,830
ΔN-Cek1^{TV}_Ser	serine	threonine to valine	109,670/41,830
ΔN-Cek1^{YF}_Ser	serine	tyrosine to phenylalanine	108,180/40,340
ΔN-Cek1$^{TV/YF}$_Ser	serine	threonine to valine and tyrosine to phenylalanine	108,180/40,340
ΔN-Cek1^{KR}_Ser	serine	lysine to arginine	109,670/41,830
ΔN-Cek1^{TD}_Ser	serine	Threonine to aspartate	109,670/41,830

2.2.2 Dynamic light scattering

Molecular size measurements were carried out in a Zetasizer Nano ZS DLS system (Malvern Instruments) using a DTS 2112 cuvette. ΔN-Cek1 was diluted in protein buffer to 1 mg/ml and three independent measurements were made for each sample at 20˚C. The intensity size distribution obtained by dynamic light scattering (DLS) is a plot of relative intensity of light scattered by particles in different size classes. However, when there is more than one peak in this plot, the intensity size distribution must be converted to a volume size distribution for a more realistic view of the data (Stetefeld *et al.* 2016), as was the case for Cek1 protein samples. All data were analyzed using DLS (nano) software (Malvern Instruments) to determine the polydispersity (Pd) (relative standard deviation) and the mean hydrodynamic radius (R_H) values from the volume distributions, used to estimate the molecular weight (assuming a globular protein) from a built-in empirical calibration graph.

2.2.3 Circular dichroism

The overall secondary structure of ΔN-Cek1 variants (ΔN-Cek1_Ser and ΔN-Cek1_Leu) and of selected mutants was assessed by far-UV circular dichroism (CD) using a J815 circular dichroism spectrometer (JASCO). Measurements of samples at 0.1 mg/ml were performed at 20˚C with 1 nm bandwidth, 1 s response, 100 nm/min scanning speed and 6 accumulations. The circular dichroism spectra (190-

260 nm) were analyzed using the JASCO software (JASCO) to determine the secondary structure content of the samples.

2.2.4 Differential scanning fluorimetry

The thermal stability of the ΔN-Cek1 (0.3 mg/ml) assessed by differential scanning fluorimetry (DSF) was determined by following SYPRO Orange (Invitrogen) fluorescence using an iCycler iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad). The melting curves (excitation / emission, 470 nm / 570 nm) were obtained by increasing the temperature from 25 to 85°C in 0.5°C steps with 30 s hold time. Data were analyzed with CFX Manager software (Bio-Rad), which calculates the melting temperature (Tm) from the inflection point of the melting curve. The thermal denaturation assay was performed in triplicate for each protein and data represent the average of three independent experiments with standard error. A statistical analysis was performed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ with Prism7 (GraphPad Software).

2.2.5 Cek1 phosphorylation *in vitro*

The immunoblot analysis of ΔN-Cek1 proteins was performed using polyclonal anti-Cek1 serum produced in-house by immunizing rabbits with the untagged active domain of ΔN-Cek1. Purified proteins were separated on 12% (w/v) polyacrylamide SDS-PAGE gels and electrotransferred onto nitrocellulose membranes. The membranes were blocked with 5% (w/v) BSA in TBS-Tween (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20) for 1 h at room temperature (RT), incubated overnight at 4°C with rabbit anti-Cek1 serum (1:5000), followed by an incubation with anti-rabbit IgG HRP-conjugated antibody (1:10000) for 1h at RT. ΔN-Cek1 samples were detected with ECL Plus reagent (GE Healthcare). The phosphorylation state of recombinant ΔN-Cek1 proteins was evaluated by immunoblot analysis using a phospho-specific antibody. After blocking, membranes were incubated with rabbit anti-phospho p44/42 MAPK antibody (1:2000, Cell Signaling Technology) overnight at 4°C, and then with anti-rabbit IgG HRP-conjugated antibody (1:10000) for 1h at RT. Phosphorylated samples were detected with ECL Plus reagent.

Cek1 autophosphorylation activity was probed by mixing ΔN-Cek1_Ser or the phosphomimetic mutant ΔN-Cek1^{TD}_Ser with the untagged catalytically inactive

mutant (untagged ΔN-Cek1^{KR}_Ser, used as substrate, 2:1 molar ratio). The reaction at RT was initiated by addition of 1mM ATP and the phosphorylation status of the catalytically inactive mutant at different time points was evaluated by immunoblot with rabbit anti-phospho p44/42 MAPK antibody and anti-Cek1 serum, as described above.

2.2.6 Cek1 activity assay

The activity of purified ΔN-Cek1 was monitored by measure the production of ADP using a chemiluminescent ADP detection assay (ADP-Glo kit, Promega). The reactions were performed in 384-well plates, kept in the dark, using 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 0.1 mg/ml BSA as assay buffer. The standard curves of ATP to ADP conversion were determined using different ATP/ADP concentrations according to the manufacturer's protocol. The reactions were started by the addition of 1 mM ATP to 10 μM ΔN-Cek1, and stopped by addition of ADP-Glo™ Reagent after 1h incubation. Measurements were performed using an endpoint luminescence protocol in a Synergy 2 plate reader (BioTek).

2.2.7 Mass spectrometry

Protein identification and quantitation was performed by nanoscale liquid chromatography coupled to tandem mass spectrometry (nano LC-MS/MS), using an equipment composed by an Ultimate 3000 liquid chromatography system coupled to a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Samples were loaded onto a trapping cartridge (Acclaim PepMap C18, Thermo Scientific) and after 3 min loading, trap column was switched in-line to a EASY-Spray column (PepMap RSLC, C18, Thermo Scientific). Separation was achieved using a formic acid (FA: 0.1% (v/v)) and acetonitrile (AC: 80% (v/v)) gradient (10 min (2.5% to 10% AC), 180 min (10% to 35% AC), 30 min (35% to 99% AC) and 30 min (hold 99% AC)). Subsequently, the column was equilibrated with 2.5% AC for 40 min. Data acquisition was controlled with Xcalibur 4.0 and Tune 2.8 software (Thermo Scientific). The mass spectrometer was operated in data-dependent (dd) positive acquisition mode alternating between a full scan (m/z 380-1580) and subsequent HCD MS/MS of the 10 most intense peaks from full scan (normalized collision energy of 27%). The ESI spray voltage was 1.9 kV.

The raw data were processed using Proteome Discoverer 2.2.0.388 software (Thermo Scientific) and searched against the SwissProt/UniProt database. The Sequest HT search engine was used to identify tryptic peptides. The ion mass tolerance was 10 ppm for precursor ions and 0.02 Da for fragment ions. The maximum allowed number of missing cleavage sites was set to 2. Cysteine carbamidomethylation was defined as constant modification. Methionine oxidation, phosphorylation (S, T and Y) and protein N-terminus acetylation were defined as variable modifications. Peptide confidence was set to high. The processing node Percolator was enabled with the following settings: maximum delta Cn 0.05; decoy database search target FDR 1%, validation was based on q-value.

2.2.8 Fluorescence measurements

The nucleotide binding assay was performed with the fluorescent analog of ATP, TNP-ATP (Jena Bioscience). The fluorescence emission scanning spectra were obtained with a Fluoromax-4 spectrofluorometer (HORIBA) in protein buffer at 20˚C using a quartz cuvette (10x10 mm). Samples were excited at 410 nm and emission spectra scanned from 500-600 nm, with excitation and emission slits set at 6 nm. For fluorescence titrations, ΔN-Cek1 variants (5 μM) were mixed with TNP-ATP (0-25 μM) and emission was recorded at 547 nm wavelength. The fluorescence of TNP-ATP was subtracted from the total fluorescence (enzyme plus TNP-ATP) to yield the specific fluorescence enhancement and data were normalized to the maximum fluorescence and analyzed using the nonlinear curve fitting to determine TNP-ATP binding affinity ($K_{d}^{\text{TNP-ATP}}$) with Prism7 (GraphPad Software).

2.2.9 Crystallization of ΔN-Cek1

Initial crystallization trials were performed in-house by sitting-drop vapour diffusion method at 20˚C in 24-well plates. Drops consisted of equal volumes (1μL) of ΔN-Cek1_Ser variant protein (9 mg/ml in protein buffer: 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 5 mM $MgCl_2$, 1 mM DTT) and precipitant solution, equilibrated against 300 μl of reservoir solution. The following crystallization commercial screens were used: Morpheus HT96, PACT premier (Molecular Dimensions), Grid Screen Ammonium Sulfate and Grid Screen Sodium Malonate (Hampton Research).

Additional crystallization conditions were screened at the High-throughput Crystallization Laboratory (HTX lab, EMBL, Grenoble, France) using the crystallization kits Index (Hampton Research), JCSG, PEGs-I, Classics-Suite (Qiagen) Wizard I&II (Rigaku Reagents) and Morpheus (Molecular Dimensions). In this case, the inactive kinase-dead mutant ΔN-Cek1^{KR}_Ser (18 mg/ml) together with the untagged ΔN-Cek1^{KR}_Ser (5 mg/ml) and the phosphorylation-null mutant ΔN-Cek1^{TV}_Ser (9.5 mg/ml) were used. Moreover, crystallization of ΔN-Cek1^{KR}_Ser and ΔN-Cek1^{TV}_Ser were also achieved in complex with 1mM of ADP (Adenosine diphosphate) and AppCp (Adenosine-5'-((β,γ)-methyleno) triphosphate, sodium salt) (Jena Bioscience).

2.3 Results

2.3.1 The *C. albicans* ΔN-Cek1_Ser variant is more stable than ΔN-Cek1_Leu

In order to evaluate the structural impact of serine or leucine insertion at the CUG-encoded position of Cek1, a thorough biophysical characterization was performed. Analytical SEC showed no differences on the elution profile of the two ΔN-Cek1 variants, with elution volumes compatible with monomeric proteins (85 kDa; Figure 2.2A). In accordance, DLS measurements indicated that both ΔN-Cek1 variants were monodisperse, with hydrodynamic radii consistent with a monomeric organization (Figure 2.2B). Additionally, the presence of leucine or serine at the CUG-encoded position had no major impact on the fold of the ΔN-Cek1 kinase, since no major structural differences were observed by CD spectroscopy (Figure 2.2C). Indeed, the CD spectra revealed no significant differences in secondary structure content between the two forms (Cek1_Ser: 32% α-helix, 43% β-sheet, 1.6% β-turn; Cek1_Leu: 34% α-helix, 41% β-sheet, 1.6% β-turn), indicating a preservation of the overall structure. Interestingly, the incorporation of serine at the CUG-encoded position leads to an increase (+3°C) in the Tm of ΔN-Cek1, as determined by DSF (Figure 2.2D).

The stability of the two ΔN-Cek1 variants was unaltered by the addition of nucleotides, since their Tm remained unchanged upon incubation with ATP (Table 2.3). Altogether, these results indicate that the presence of a leucine residue at the single CUG-encoded position did not induce protein misfolding but significantly

reduced the thermal stability of ΔN-Cek1, which was not due to differences in the phosphorylation states of the two variants, as shown below.

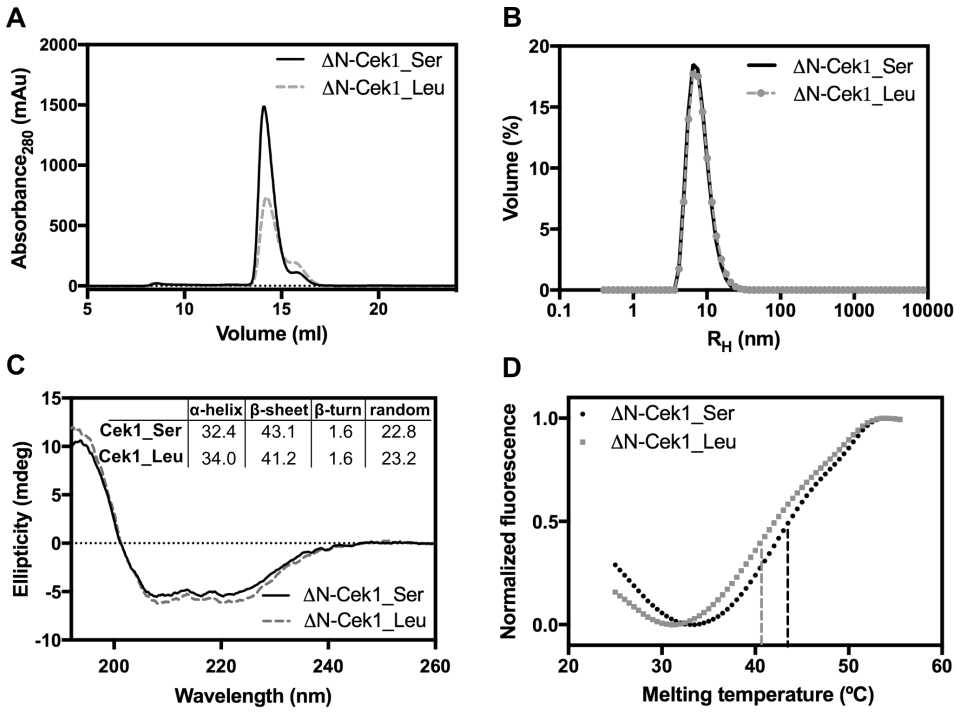


Figure 2.2 Biophysical characterization of *C. albicans* Cek1 variants. (A) SEC of ΔN-Cek1_Ser (solid line) and Cek1_Leu (dashed line). Both variants elute as monomers (85 kDa). Downward triangles indicate the elution volume of protein standards: 1 - Aldolase (158 kDa), 2 – Conalbumin (75 kDa) and 3 – Ovalbumin (43 kDa). (B) The two ΔN-Cek1 variants are monodisperse (Pd=26.8% ΔN-Cek1_Ser; Pd=29.3% ΔN-Cek1_Leu) and monomeric (R_H ΔN-Cek1_Ser=4.05nm; R_H ΔN-Cek1_Leu=4.18nm, in good agreement with the estimated molecular mass of Cek1), as assessed by DLS. (C) CD spectra of ΔN-Cek1_Ser and ΔN-Cek1_Leu revealing no significant differences in overall secondary structure. (D) Representative thermal denaturation measured by DSF for both variants of ΔN-Cek1. The normalized fluorescence intensity is plotted as a function of the temperature (˚C). The temperature at the inflection point of the unfolding transition defines the Tm value for each protein (Tm ΔN-Cek1_Ser=43.1 ± 0.7˚C; Tm ΔN-Cek1_Leu=40.5 ± 0.3˚C).

Table 2.3 Thermal stability of ΔN-Cek1_Ser and ΔN-Cek1_Leu variants incubated with ATP. Data shown are mean ± standard error of the mean (n=3). Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ for a 95% confidence interval comparing each variant, ΔN-Cek1_Ser or ΔN-Cek1_Leu with ATP ($p>0.05$ non-significant (ns), *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

Protein	Melting temperature (°C) (mean ± SD)	Significance (adjusted *p* value)
ΔN-Cek1_Ser	43.1 ± 0.7	ns (0.9999)
ΔN-Cek1_Ser with ATP	43.2 ± 0.8	
ΔN-Cek1_Leu	40.5 ± 0.3	ns (0.9999)
ΔN-Cek1_Leu with ATP	39.9 ± 0.9	

Crystallization experiments were performed with ΔN-Cek1_Ser variant, unfortunately without success. Therefore, due to higher thermal stability of the inactive kinase-dead mutant ΔN-Cek1^{KR}_Ser, this protein was selected to crystallization trials (Table 2.4). An initial crystallization condition with microcrystals was obtained with ΔN-Cek1^{KR}_Ser using the Morpheus screening kit (0.12 M Alcohols, 0.1 M Imidazole/MES pH 6.5 and 30% PEG 500 MME/PEG 20.000) (Figure 2.3). This first crystallization condition is currently being further optimized to obtain diffraction-quality crystals to further determine the three-dimensional structures by X-ray crystallography (currently in progress).

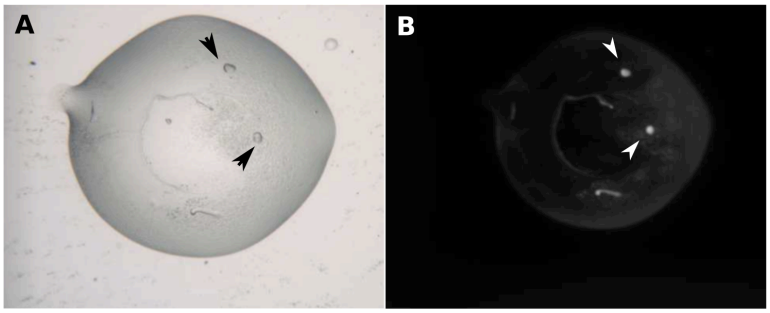


Figure 2.3 Microcrystals of *C. albicans* kinase-dead mutant ΔN-Cek1^{KR}_Ser. (A) Visible light. (B) UV light. Microcrystals are marked with arrow.

2.3.2 Autophosphorylation at the conserved ^{231}TEY^{233} motif of ΔN-Cek1 requires a serine at the CUG-encoded position

MAP kinases, including the *Saccharomyces cerevisiae* MAPK Fus3, are often found to be autophosphorylated when expressed in *E. coli* (Bhattacharyya *et al.* 2006). The phosphorylation status of recombinant ΔN-Cek1 variants was assessed using anti-phospho p44/42, a specific antibody that recognizes the phosphorylation of both threonine (Thr231) and tyrosine (Tyr233) residues of the ^{231}TEY^{233} motif (Nagiec *et al.* 2015). Immunoblot analysis of both ΔN-Cek1 variants with anti-phospho p44/42 revealed that only recombinant ΔN-Cek1_Ser was autophosphorylated *in vitro*, and no phosphorylation was detected in the ΔN-Cek1_Leu variant using this approach (Figure 2.4A). The autophosphorylation activity of ΔN-Cek1_Ser was further explored using two single mutants (ΔN-Cek1^{TV}_Ser; ΔN-Cek1^{YF}_Ser) and a double mutant (ΔN-Cek1$^{TV/YF}$_Ser) of the putative phosphorylation sites (Thr231 mutated to valine and Tyr233 mutated to phenylalanine), as well as a kinase-dead mutant (ΔN-Cek1^{KR}_Ser, Lys100 mutated to arginine) and the putative phosphomimetic variant (ΔN-Cek1^{TD}_Ser, Thr231 mutated to aspartate). As for the ΔN-Cek1_Ser and ΔN-Cek1_Leu variants, these specific mutations had no effect on the fold and overall secondary structure content of the enzyme, as assessed by CD (Figure 2.4B). Mutation of Thr231 to valine (ΔN-Cek1^{TV}_Ser) at the ^{231}TEY^{233} motif had no impact on the phosphorylation signal of ΔN-Cek1_Ser (Figure 2.4A). In contrast, replacement of Tyr233 by phenylalanine (ΔN-Cek1^{YF}_Ser) completely abolished the autophosphorylation of recombinant ΔN-Cek1_Ser, indicating that ΔN-Cek1_Ser autophosphorylation occurs on Tyr233 (Figure 2.4A). Therefore, when Thr231 and Tyr233 were mutated to valine and phenylalanine (ΔN-Cek1$^{TV/YF}$_Ser), respectively, recognition by anti-phospho p44/42 was also completely abolished. As previously described, the substitution of the catalytic lysine residue by an arginine decreased the kinase activity by preventing transfer of the phosphate group from the nucleotide phosphate to the acceptor protein (Carrera *et al.* 1993). Thus, mutation of the catalytic lysine to arginine (ΔN-Cek1^{KR}_Ser) abolished the autophosphorylation signal, while the putative phosphomimetic mutant (ΔN-Cek1^{TD}_Ser) displayed an enhanced autophosphorylation signal in the immunoblot assay, as expected (Figure 2.4A).

Analysis of the thermal stability of the ΔN-Cek1^{YF}_Ser and ΔN-Cek1$^{TV/YF}$_Ser mutants (Table 2.4), which are not autophosphorylated, showed no significant differences in the melting temperatures compared to wild-type ΔN-Cek1_Ser,

underscoring that differences in the melting temperatures of ΔN-Cek1_Leu and ΔN-Cek1_Ser do not result from ΔN-Cek1_Ser autophosphorylation.

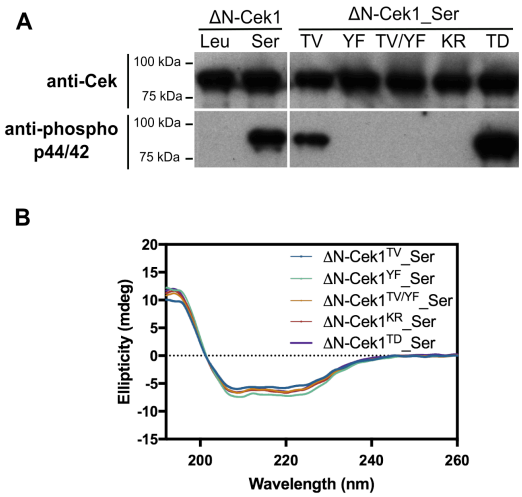


Figure 2.4 ΔN-Cek1_Ser is autophosphorylated at Tyr233. (A) Immunoblot analysis of ΔN-Cek1_Leu, ΔN-Cek1_Ser and ΔN-Cek1_Ser mutants (ΔN-Cek1^{TV}_Ser, ΔN-Cek1^{YF}_Ser, ΔN- ΔN-Cek1$^{TV/TF}$_Ser, ΔN-Cek1^{KR}_Ser, ΔN-Cek1^{TD}_Ser) probed with rabbit anti-Cek serum (upper panel) and rabbit anti-phospho p44/42 (bottom panel). Autophosphorylation was detected in ΔN-Cek1_Ser and ΔN-Cek1^{TV}_Ser, indicating that autophosphorylation occurs on Tyr233. (B) CD spectra of ΔN-Cek1_Ser mutants (ΔN-Cek1^{TV}_Ser, ΔN-Cek1^{YF}_Ser, ΔN-Cek1$^{TV/YF}$_Ser, Cek1^{KR}_Ser, Cek1^{TD}_Ser). The mutations did not induce any major differences on the secondary structure content of the enzyme.

Table 2.4 Thermal stability of ΔN-Cek1_Ser and ΔN-Cek1_Leu variants and ΔN-Cek1_Ser mutants. Data shown are mean ± standard error of the mean (n=3). Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ for a 95% confidence interval comparing all protein with ΔN-Cek1_Ser ($p>0.05$ non-significant (ns), *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

Protein	Melting temperature (˚C) (mean ± SD)	Significance (adjusted *p* value)
ΔN-Cek1_Ser	43.1 ± 0.7	-
ΔN-Cek1_Leu	40.5 ± 0.3	** (0.007)
ΔN-Cek1^{TV}_Ser	43.0 ± 0.7	ns (0.9999)
ΔN-Cek1^{YF}_Ser	43.3 ± 0.5	ns (0.9999)
ΔN-Cek1$^{TV/YF}$_Ser	44.4 ± 0.6	ns (0.9999)
ΔN-Cek1^{KR}_Ser	47.2 ± 0.6	**** (0.0001)
ΔN-Cek1^{TD}_Ser	46.4 ± 0.2	** (0.0014)

To further confirm that ΔN-Cek1_Ser was autophosphorylated on Tyr233 of the ^{231}TEY^{233} motif, as suggested by the immunoblot data, both variants of ΔN-Cek1 were analyzed by mass spectrometry (Figure 2.5). Protein samples were analyzed as-purified and after incubation with 1mM ATP. Interestingly, using this approach we were not able to detect phosphorylated peptides in the as-purified ΔN-Cek1 variants. However, upon incubation with ATP, a peptide (SIASQEDNYGFMTEpYVATR) was identified in the ΔN-Cek1_Ser variant sample, displaying a mass compatible with a single phosphorylation (Figure 2.4A). Although the experimental procedure used is not quantitative, the measured signal for the phosphorylated peptide was approximately 20-fold lower than for the non-phosphorylated peptide (Figure 2.5A). This result further supported the immunoblot data for the ΔN-Cek1_Ser mutants, suggesting that the autophosphorylation of ΔN-Cek1_Ser occurred at Tyr233 (Fig 2.4A). In agreement with the immunoblot results, no phosphorylated peptide could be detected by mass spectrometry on the ΔN-Cek1_Leu variant sample (Figure 2.5B), suggesting that a leucine residue at the CUG-encoded position had a negative impact on the autophosphorylation activity of ΔN-Cek1. In summary, recombinant ΔN-Cek1_Ser expressed in *E. coli* was readily autophosphorylated at Tyr233 of the ^{231}TEY^{233} motif, suggesting that the presence of a serine at the CUG-encoded position was required for protein autophosphorylation and may be crucial for regulating the kinase activity (Fig 2.5C).

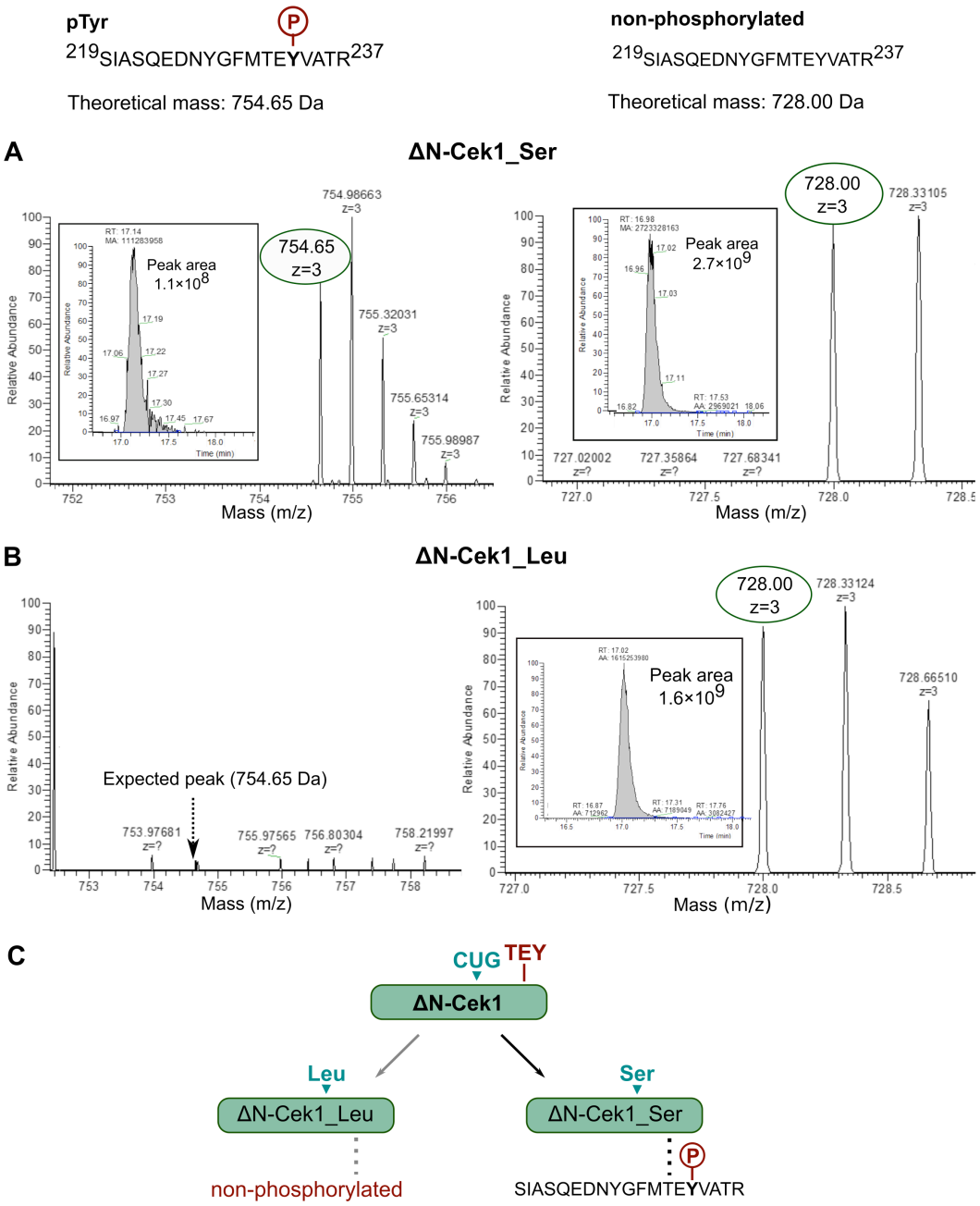


Figure 2.5 LC-MS/MS analysis of ΔN-Cek1 variants. Representative extracted ion chromatogram with the peak area and the respective m/z for the peptide of interest (SIASQEDNYGFMTEYVATR). The integrated peak areas (insets) of the phosphorylated peptide and of its equivalent non-phosphorylated version were used to calculate an apparent percentage of phosphorylation of ΔN-Cek1. An arrow marks the position of the peptide ion corresponding to the expected mass of a single phosphopeptide on the tyrosine residue, whose theoretical mass (754.65), as well as that of its non-phosphorylated counterpart (728.00), was calculated using the online server ProteinProspector (MS-isotope). (A) MS spectrum for the ΔN-Cek1_Ser variant: Left panel - phosphopeptide (SIASQEDNYGFMTEpYVATR, peak area 1.1×10^{8}); Right panel - the equivalent non-

phosphorylated peptide (peak area 2.7×10^9). Comparative analysis of the peak areas corresponding to phosphorylated and non-phosphorylated peptide fragments, suggested that only a small percentage of ΔN-Cek1_Ser was autophosphorylated on tyrosine 233. (B) MS analysis of the ΔN-Cek1_Leu variant revealed absence of the phosphorylated peptide (Left panel), while the non-phosphorylated peptide was detectable (Right panel). (C) Schematic representation of phosphorylation results of ΔN-Cek1 variants obtained by immunoblot analysis combined with mass spectrometry: phosphorylation was not detected in ΔN-Cek1_Leu, while in ΔN-Cek1_Ser a phosphorylated peptide was identified, indicating that autophosphorylation occurred on Tyr233.

2.3.3 Codon decoding ambiguity affects Cek1 protein activity

In contrast to ΔN-Cek1_Leu, ΔN-Cek1_Ser is autophosphorylated *in vitro*. To investigate the influence of phosphorylation on the enzymatic activity of Cek1 and unveil the role of CUG ambiguity on its function, the catalytic activity of ΔN-Cek1 variants was measured. Using a luminescent ADP-detection assay, ΔN-Cek1_Ser displayed a 10-fold higher activity than ΔN-Cek1_Leu (Figure 2.6A). As expected, the ΔN-Cek^{KR}_Ser mutant displayed no catalytic activity, while the activity of mutant ΔN-Cek1^{TD}_Ser was significantly increased (~2-fold; Figure 2.6A). Considering that the CUG-encoded residue is located in the ATP binding pocket, we reasoned that the differences in the autophosphorylation activity observed for the two ΔN-Cek1 variants could result from dissimilarities in ATP binding affinity. With this aim, we used TNP-ATP, a fluorescent ATP analog that binds to the ATP binding pocket of protein kinases (LaConte *et al.* 2017), to assess the affinity of both variants to the nucleotide. This assay showed that TNP-ATP bound similarly to both ΔN-Cek1 variants (5.7 μM for ΔN-Cek1_Ser and 8 μM for ΔN-Cek1_Leu) (Figure 2.6B) and that differences in enzyme activity did not seem to be correlated with alterations in ATP binding affinity. However, the fluorophore bound to the nucleotide might mask small differences in ATP binding.

Overall, these results demonstrated that ΔN-Cek1_Ser had intrinsic autophosphorylation activity and was more active than ΔN-Cek1_Leu *in vitro*. Indeed, the ΔN-Cek1^{TD}_Ser mutant, mimicking the phosphorylation of the ^{231}TEY^{233} motif threonine, in the activation loop, displayed significantly enhanced catalytic activity.

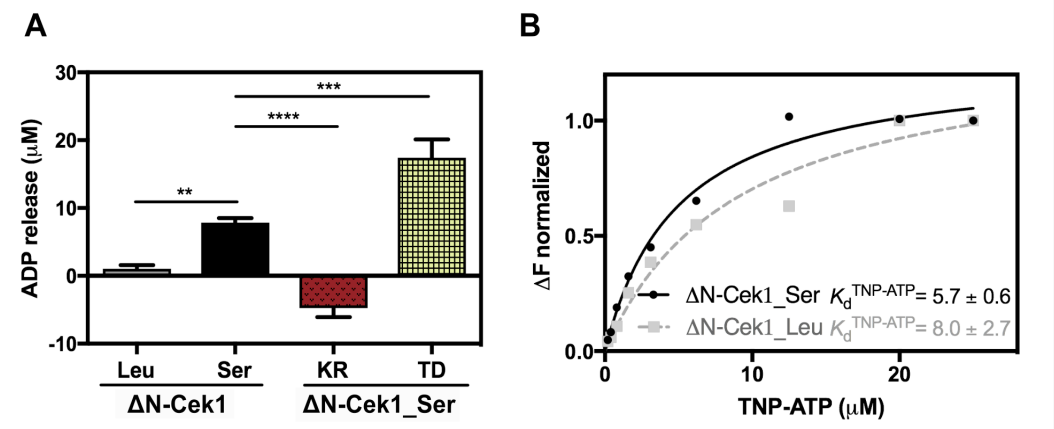


Figure 2.6 ΔN-Cek1_Ser displays higher catalytic activity than ΔN-Cek1_Leu. (A) Activity assay by quantification of ADP release of ΔN-Cek1 variants, inactive (ΔN-Cek1^{KR}_Ser, negative control) and phosphomimetic (ΔN-Cek1^{TD}_Ser, positive control) mutants. ΔN-Cek1_Ser released significantly more ADP compared to ΔN-Cek1_Leu, while ΔN-Cek1^{TD}_Ser was the most active. Data are the average of three determinations and error bars represent standard deviation. Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ for a 95% confidence interval, $p>0.05$ non-significant (ns), *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$). (B) Binding of TNP-ATP to ΔN-Cek1 variants. Both ΔN-Cek1 variants bind the nucleotide with comparable affinity. The data are representative of three independent experiments.

2.3.4 Autophosphorylation of Cek1 does not occur in trans

In order to explore if the autophosphorylation of ΔN-Cek1_Ser occurred through a cis- (intramolecular) or trans-autophosphorylation (intermolecular) mechanism (Beenstock *et al.* 2016), the inactive kinase-dead mutant (ΔN-Cek1^{KR}_Ser) was used as substrate of ΔN-Cek1_Ser or of its phosphomimetic mutant, ΔN-Cek1^{TD}_Ser. Although both ΔN-Cek1_Ser (Figure 2.7A) and ΔN-Cek1^{TD}_Ser (Figure 2.7B) were autophosphorylated, neither was able to phosphorylate the substrate, suggesting that ΔN-Cek1 kinase is unable to phosphorylate its inactive form, supporting the *cis*-autophosphorylation mechanism.

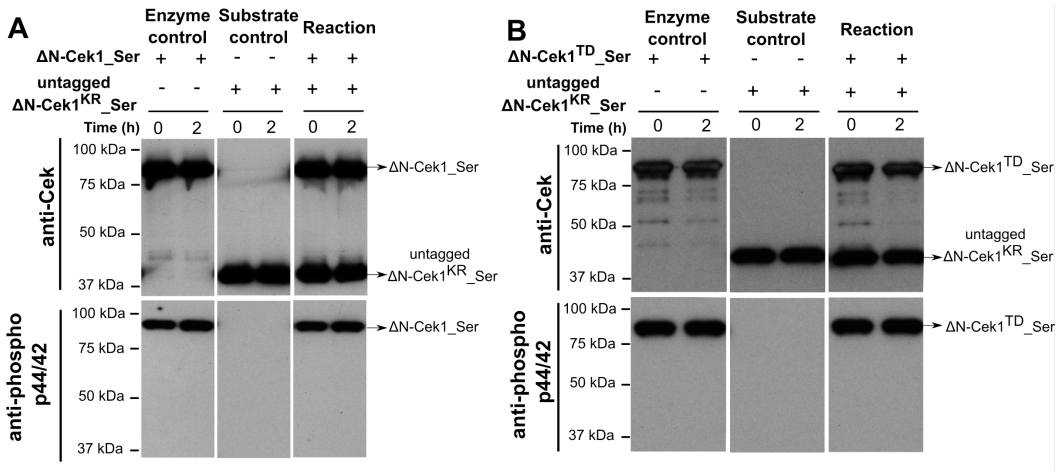


Figure 2.7 The autophosphorylation of ΔN-Cek1 is an intramolecular event. Transphosphorylation assay of the inactive mutant of ΔN-Cek1_Ser (untagged ΔN-Cek1^{KR}_Ser, substrate, 42 kDa) by (A) ΔN-Cek1_Ser variant (85 kDa) and (B) ΔN-Cek1^{TD}_Ser phosphomimetic mutant (85 kDa), probed with rabbit anti-Cek serum (upper panel) and rabbit anti-phospho p44/42 (bottom panel). Phosphorylation of untagged ΔN-Cek1^{KR}_Ser was not observed in the reaction, only the autophosphorylation of the two enzymes was detected.

2.4 Discussion

C. albicans is a human commensal that, under specific conditions, is able to cause disease that ranges from superficial mucosal to life-threatening systemic infections (Sudbery 2011). This opportunistic human pathogen has the remarkable ability to quickly adapt to varied host niches and threats elicited by host immune defenses. Hence, in response to multiple stress factors and/or damage induced by antifungal drugs, the activation of highly coordinated signaling pathways generates the diversity required for survival and virulence (Shapiro *et al.* 2011). Interestingly, many of the enzymes involved in signal transduction cascades in *C. albicans* are encoded by genes that contain CTG codons, stochastically translated into serine (97%) and leucine (3%) residues often located in highly conserved sites (Rocha *et al.* 2011, Sárkány *et al.* 2014). One of these enzymes is the MAPK Cek1 (with one CUG-encoded residue) that is central for *C. albicans* morphogenesis, cell wall biosynthesis and host interaction (Galan-Diez *et al.* 2010, Herrero-de-Dios *et al.* 2014). Accordingly, we hypothesized that the fluctuations in serine/leucine incorporation levels induced by different stress conditions might impact the function of Cek1 with downstream consequences on *C. albicans* virulence traits.

To assess the role of the identity of the CUG-encoded residue in Cek1 function, we characterized the structural and functional features of the ΔN-Cek1_Ser and ΔN-Cek1_Leu variants of the kinase active domain. Cek1 displays high amino acid sequence identity (63%) with the structurally characterized Fus3 (Remenyi *et al.* 2005, Sárkány *et al.* 2014), the MAPK that mediates the mating pheromone response in *S. cerevisiae*. The unique conserved CUG-encoded residue in the nucleotide binding pocket of Cek1 (Ser199) is structurally equivalent to a serine residue in Fus3 that stabilizes a water molecule from the coordination sphere of the magnesium ion that interacts with the phosphate moiety of the nucleotide (Rocha *et al.* 2011, Sárkány *et al.* 2014). Consequently, the presence of a leucine residue at this position is expected to interfere with the nucleotide binding and/or phosphate transfer activity of Cek1. Our results showed that although the ambiguity at the CUG-encoded position does not affect the overall secondary structure of ΔN-Cek1, incorporation of leucine at position 199 decreases considerably the thermal stability of the kinase. This difference likely results from the expected loss of one or two hydrogen bonds upon replacement of the polar serine residue by the hydrophobic leucine (Rees *et al.* 2001, Pace *et al.* 2014). Although the CUG-encoded residue faces the ATP binding pocket, binding to TNP-ATP did not seem to be altered by the identity of the amino acid at position 199.

Autophosphorylation is a widespread feature among protein kinases and is also observed in Fus3 (Bhattacharyya *et al.* 2006), a protein that is only fully activated via a kinase cascade that results in dual phosphorylation on the tyrosine and threonine residues of the activation loop TEY motif (Gartner *et al.* 1992). Interestingly, following expression in *E. coli*, only ΔN-Cek1_Ser is recognized by the anti-phospho p44/42 antibody. Our results showed that ΔN-Cek1_Ser autophosphorylation occurs on the conserved Tyr233 of the ^{231}TEY^{233} motif within the activation loop. In agreement, *in vitro* activity assays showed that the catalytic activity of ΔN-Cek1_Ser is significantly higher than that of ΔN-Cek1_Leu. Although our mass spectrometry data suggested the occurrence of autophosphorylation on Tyr233 upon incubation with ATP, we cannot exclude the possibility that the *in vitro* kinase activity assays, based on the quantification of ADP release, report on the combined effects of ΔN-Cek1 autophosphorylation and intrinsic ATPase activity, as previously described for some protein kinases (Rominger *et al.* 2007, Ahmad *et al.* 2013). Further, our autophosphorylation results using the kinase dead mutant ΔN-Cek^{KR}_Ser as substrate, suggested that Cek1 autophosphorylation did not occur via a trans-autophosphorylation mechanism, and likely involves an intramolecular

reaction as proposed for *S. cerevisiae* Fus3 (Bhattacharyya *et al.* 2006). In yeast MAPKs, enzyme activation requires an order-to-disorder transition, meaning that in the inactive state part of the activation loop, including the phosphorylatable tyrosine, is positioned at the substrate binding pocket and obstructs access of substrate peptides to the active site (Remenyi *et al.* 2005, Bhattacharyya *et al.* 2006, Lochhead 2009). Therefore, kinase activity is inhibited via a pseudo-substrate mechanism, prior to full activation by the upstream kinase, which phosphorylates both residues in the TEY motif resulting in its displacement from the catalytic site. In the resting state, positioning of the tyrosine side chain close to the kinase catalytic cleft, highly conserved in *C. albicans* Cek1, may explain the identification of a small subpopulation of monophosphorylated protein in recombinant ΔN-Cek1. Therefore, in our Cek1 samples the enzyme catalytic site might be predominantly inaccessible to substrate peptides or proteins, explaining the inability of ΔN-Cek1_Ser or ΔN-Cek1^{TD}_Ser to phosphorylate the generic substrate myelin basic protein (Bhattacharyya *et al.* 2006) and the synthetic Cph1 peptide (Maiti *et al.* 2015) (data not shown).

Although our current data do not allow to infer the precise autophosphorylation mechanism of Cek1 and why this activity is compromised in the ΔN-Cek1_Leu variant, we can speculate that the presence of a leucine at the CUG-encoded position impacts negatively the phosphoryl transfer activity of Cek1 either by directly diminishing the catalytic activity or by modifying the dynamics of the activation loop and therefore altering the residence time of the phosphoryl-accepting tyrosine at the kinase catalytic site. Cek1 forms part of a highly regulated multienzyme assembly, whose activation is triggered in response to different external signals (Cote *et al.* 2011). In *S. cerevisiae*, the scaffold protein Ste5 enhances Fus3 autophosphorylation on the tyrosine residue of the TEY motif (Bhattacharyya *et al.* 2006) and, together with the phosphatase Msg5, maintains the monophosphorylated form of Fus3, which has an inhibitory role on the signaling pathway (Nagiec *et al.* 2015). Interestingly, the allosteric Fus3 autophosphorylation mechanism mediated by bipartite binding of the scaffold protein Ste5 is likely not conserved in Cek1, since this region is structurally distinct in Cst5, the *C. albicans* MAPK scaffold protein (Cote *et al.* 2011). In *C. albicans*, CUG translational ambiguity might play a role in rewiring the collective behavior of these macromolecular assemblies with consequences for the flow of information across these signaling cascades.

In this study, we evaluate the effect of CUG translational ambiguity on the MAP kinase Cek1, a key protein of the signaling pathway related to morphogenesis and virulence in *C. albicans*. Biochemical and biophysical data revealed that ΔN-Cek1_Ser variant induces an increase on the protein thermal stability compared to leucine variant, without any changes on the overall secondary structure. Moreover, only the ΔN-Cek1_Ser variant was autophosphorylated *in vitro* at the tyrosine residue of the ^{231}TEY^{233} motif of the activation loop. Accordingly, incorporation of a serine residue at the CUG position seems to induce a more active protein with possible implications on the Cek1 function. Further studies will be required to understand the molecular details of CUG codon ambiguous translation in the context of the Cek1 full-length protein and for *C. albicans* host recognition and pathogenicity.

2.5 References

Ahmad, S., M. A. Hughes, *et al.* (2013). Development and validation of a high-throughput intrinsic ATPase activity assay for the discovery of MEKK2 inhibitors.*J Biomol Screen* 18: 388-99.

Alonso-Monge, R., E. Roman, *et al.* (2006). The MAP kinase signal transduction network in *Candida albicans*.*Microbiology* 152: 905-12.

Ambrogelly, A., S. Palioura, *et al.* (2007). Natural expansion of the genetic code.*Nat. Chem. Biol* 3: 29-35.

Beenstock, J., N. Mooshayef, *et al.* (2016). How Do Protein Kinases Take a Selfie (Autophosphorylate)?*Trends Biochem Sci* 41: 938-953.

Bezerra, A. R., A. R. Guimaraes, *et al.* (2015). Non-Standard Genetic Codes Define New Concepts for Protein Engineering.*Life (Basel)* 5: 1610-28.

Bezerra, A. R., J. Simoes, *et al.* (2013). Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification.*Proc Natl Acad Sci U S A* 110: 11079-84.

Bhattacharyya, R. P., A. Remenyi, *et al.* (2006). The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway.*Science* 311: 822-6.

Carrera, A. C., K. Alexandrov, *et al.* (1993). The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring ATP.*Proc Natl Acad Sci U S A* 90: 442-6.

Cote, P., T. Sulea, *et al.* (2011). Evolutionary reshaping of fungal mating pathway scaffold proteins.*mBio* 2: e00230-10.

Csank, C., K. Schroppel, *et al.* (1998). Roles of the Candida albicans mitogen-activated protein kinase homolog, Cek1p, in hyphal development and systemic candidiasis.*Infect Immun* 66: 2713-21.

Evans, C. R., Y. Fan, *et al.* (2018). Errors during Gene Expression: Single-Cell Heterogeneity, Stress Resistance, and Microbe-Host Interactions.*mBio* 9: e01018-18.

Feketova, Z., T. Masek, *et al.* (2010). Ambiguous decoding of the CUG codon alters the functionality of the Candida albicans translation initiation factor 4E.*FEMS Yeast Res* 10: 558-69.

Galan-Diez, M., D. M. Arana, *et al.* (2010). Candida albicans beta-glucan exposure is controlled by the fungal CEK1-mediated mitogen-activated protein kinase pathway that modulates immune responses triggered through dectin-1.*Infect Immun* 78: 1426-36.

Gartner, A., K. Nasmyth, *et al.* (1992). Signal transduction in Saccharomyces cerevisiae requires tyrosine and threonine phosphorylation of FUS3 and KSS1.*Genes Dev* 6: 1280-92.

Gomes, A. C., I. Miranda, *et al.* (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans.*Genome Biol* 8: R206.

Gow, N. A., F. L. van de Veerdonk, *et al.* (2012). Candida albicans morphogenesis and host defence: discriminating invasion from colonization.*Nat Rev Microbiol* 10: 112-22.

Herrero-de-Dios, C., R. Alonso-Monge, *et al.* (2014). The lack of upstream elements of the Cek1 and Hog1 mediated pathways leads to a synthetic lethal phenotype upon osmotic stress in Candida albicans.*Fungal Genet Biol* 69: 31-42.

Ji, Q. Q., Z. P. Fang, *et al.* (2016). C-terminal Domain of Leucyl-tRNA Synthetase from Pathogenic Candida albicans Recognizes both tRNASer and tRNALeu.*J Biol Chem* 291: 3613-25.

Kapur, M. and S. L. Ackerman (2018). mRNA Translation Gone Awry: Translation Fidelity and Neurological Disease.*Trends Genet* 34: 218-231.

Knight, R. D., S. J. Freeland, *et al.* (2001a). Rewiring the keyboard: evolvability of the genetic code.*Nat Rev Genet* 2: 49-58.

Knight, R. D. L. L. F. and M. Yarus (2001b). How Mitochondria Redefine the Code.*J Mol Evol* 53: 299-313.

LaConte, L. E. W., S. Srivastava, *et al.* (2017). Probing Protein Kinase-ATP Interactions Using a Fluorescent ATP Analog.*Methods Mol Biol* 1647: 171-183.

Lochhead, P. A. (2009). Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation.*Sci Signal* 2: pe4.

Maiti, P., P. Ghorai, *et al.* (2015). Mapping of functional domains and characterization of the transcription factor Cph1 that mediate morphogenesis in Candida albicans.*Fungal Genet Biol* 83: 45-57.
Miranda, I., R. Rocha, *et al.* (2007). A genetic code alteration is a phenotype diversity generator in the human pathogen Candida albicans.*PLoS One* 2: e996.
Miranda, I., A. Silva-Dias, *et al.* (2013). Candida albicans CUG Mistranslation Is a Mechanism To Create Cell Surface Variation.*MBio* 4: e00285-13.
Muhlhausen, S., P. Findeisen, *et al.* (2016). A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes.*Genome Res* 26: 945-55.
Muhlhausen, S., H. D. Schmitt, *et al.* (2018). Endogenous Stochastic Decoding of the CUG Codon by Competing Ser- and Leu-tRNAs in Ascoidea asiatica.*Curr Biol* 28: 2046-2057 e5.
Nagiec, M. J., P. C. McCarter, *et al.* (2015). Signal inhibition by a dynamically regulated pool of monophosphorylated MAPK.*Mol Biol Cell* 26: 3359-71.
Pace, C. N., H. Fu, *et al.* (2014). Contribution of hydrogen bonds to protein stability.*Protein Sci* 23: 652-61.
Rees, D. C. and A. D. Robertson (2001). Some thermodynamic implications for the thermostability of proteins.*Protein Sci* 10: 1187-1194.
Remenyi, A., M. C. Good, *et al.* (2005). The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network.*Mol Cell* 20: 951-62.
Rocha, R., P. J. Pereira, *et al.* (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen.*Proc Natl Acad Sci U S A* 108: 14091-6.
Roman, E., R. Alonso-Monge, *et al.* (2009). The Cek1 MAPK is a short-lived protein regulated by quorum sensing in the fungal pathogen Candida albicans.*FEMS Yeast Res* 9: 942-55.
Rominger, C. M., M. D. Schaber, *et al.* (2007). An intrinsic ATPase activity of phospho-MEK-1 uncoupled from downstream ERK phosphorylation.*Arch Biochem Biophys* 464: 130-7.
Santos, M., P. M. Pereira, *et al.* (2018). Codon misreading tRNAs promote tumor growth in mice.*RNA Biol* 15: 773-786.
Santos, M. A., G. Keith, *et al.* (1993). Non-standard translational events in Candida albicans mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon.*EMBO J* 12: 607-16.
Santos, M. A. S. and M. F. Tuite (1995). The CUG codon is decoded in vivo as serine and not leucine in Candida albicans.*Nucleic Acids Res* 23: 1481-1486.
Sárkány, Z., A. Silva, *et al.* (2014). Ser or Leu: structural snapshots of mistranslation in Candida albicans.*Front Mol Biosci* 1: 1-14.
Seligmann, H. (2018). Alignment-based and alignment-free methods converge with experimental data on amino acids coded by stop codons at split between nuclear and mitochondrial genetic codes.*Biosystems* 167: 33-46.
Shapiro, R. S., N. Robbins, *et al.* (2011). Regulatory circuitry governing fungal development, drug resistance, and disease.*Microbiol Mol Biol Rev* 75: 213-67.
Stetefeld, J., S. A. McKenna, *et al.* (2016). Dynamic light scattering: a practical guide and applications in biomedical sciences.*Biophys Rev* 8: 409-427.
Sudbery, P. E. (2011). Growth of Candida albicans hyphae.*Nat Rev Microbiol* 9: 737-48.
Swart, E. C., V. Serra, *et al.* (2016). Genetic Codes with No Dedicated Stop Codon: Context-Dependent Translation Termination.*Cell* 166: 691-702.
Yi, S., N. Sahni, *et al.* (2011). Alternative mating type configurations (a/alpha versus a/a or alpha/alpha) of Candida albicans result in alternative biofilms regulated by different pathways.*PLoS Biol* 9: e1001117.
Zhou, X. L., Z. P. Fang, *et al.* (2013). Aminoacylation and translational quality control strategy employed by leucyl-tRNA synthetase from a human pathogen with genetic code ambiguity.*Nucleic Acids Res* 41: 9825-38.

Chapter 3

The MAPK Cek1 N-terminal tail homorepeats and its role in the kinase activity

Work contributions:
Cloning, expression and purification protocols of the full-length Cek1_Ser variant were established in collaboration with Inês Correia in the scope of Marta Silva internship. Mass spectrometry measurements were performed by Hugo Osório at the i3S Proteomics Scientific Platform. Small-angle X-ray scattering experiment was done in collaboration with José António Manso at the European Synchrotron Radiation Facility (Grenoble, France).

3.1 Introduction

Candida albicans is a highly adapted commensal organism that senses environmental changes through its membrane sensors that trigger quick responses through various signaling pathways, including Mitogen-Activated Protein Kinase (MAPK) pathways. The MAPK Cek1 is a short-lived protein involved in fungal morphogenesis and virulence (Galan-Diez *et al.* 2010, Herrero-de-Dios *et al.* 2014). Cek1 encompasses a CUG-encoded residue that faces the kinase ATP binding pocket, and the incorporation of a serine or a leucine at this position was predicted to impact on its kinase activity and downstream signaling functions (Rocha *et al.* 2011, Sárkány *et al.* 2014). We have previously demonstrated that the presence of serine at the CUG position of the N-terminally truncated Cek1 (ΔN-Cek1, residues 60-422) (Figure 3.1), containing the globular kinase domain, increases its thermal stability and enzymatic activity, when compared to the leucine variant (Fraga *et al.* 2019). The observed differences between the serine and leucine variants of Cek1 do not result from changes in the overall protein structure. However, the role of CUG ambiguous translation in the context of the full-length Cek1 protein, containing 60 additional N-terminal amino acids (Figure 3.1), remained unexplored.

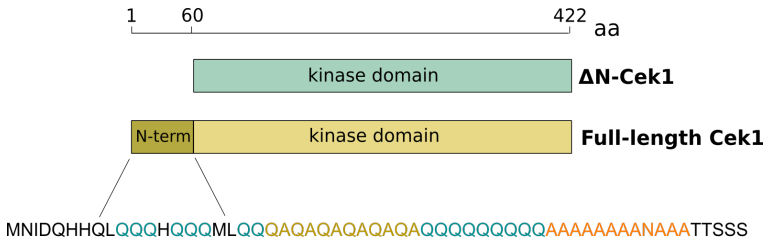


Figure 3.1 Schematic representation of truncated (ΔN-Cek1) and of full-length *C. albicans* Cek1. The N-terminal region (1-59 aa) and the globular kinase domain are represented (60-422). The sequence of the N-terminal tail is colored based on different homorepeats: glutamines (blue), glutamines and alanines (yellow), alanines (orange). The amino acid sequence length is represented at the top of the figure.

The variable and often conformationally flexible tails in protein kinases are known to adopt different conformations that enhance their structural and functional diversity. Those segments modulate the enzymatic activity of the globular catalytic domain and enable protein-protein interactions with regulatory roles in signaling kinase interaction networks (Kathiriya *et al.* 2014, Wright *et al.* 2015). In fact, the intrinsically disordered N-terminal tail of the Src kinase was demonstrated to be

important for its dimerization and in the modulation of its kinase activity (Spassov *et al.* 2018). On the other hand, in the mycobacterial tyrosine kinase A (PtkA) the unstructured N-terminal tail has an inhibitory effect on the enzymatic activity, which is regulated by phosphorylation (Niesteruk *et al.* 2018a). Moreover, disordered segments were also described in the C-terminal region of eukaryotic protein kinases and shown to have a crucial role for substrate recruitment (Sun *et al.* 2010), autophosphorylation and signal propagation (Keppel *et al.* 2017).

An interesting feature of Cek1 N-terminal tail is the over-representation of amino acid repeats, including glutamine, alanine and a combined glutamine/alanine containing segments (Figure 3.1). These low complexity regions containing little diversity in amino acid composition and few hydrophobic amino acids are generally unstructured a.k.a. intrinsically disordered regions – IDRs (van der Lee *et al.* 2014, Kumari *et al.* 2015). These IDRs do not fold into stable secondary structures and often contain small segments that act in synergy with structured domains, frequently adopting a specific structure upon binding to other proteins or into large multi-protein complexes (Faux *et al.* 2005, Luo *et al.* 2014). Therefore, these disordered structural elements, by acquiring an ensemble of different conformations, play critical roles in modulation transient protein-protein interactions and the formation of dynamic signaling complexes whose assembly can be modulated by post-translational modifications with implications in protein function (Babu *et al.* 2012, van der Lee *et al.* 2014).

Glutamine repeats (polyQ) are one the most common homorepeats in eukaryotic proteomes (Schaefer *et al.* 2012). The normal length of the polyQ region appears to be specific for each protein family and its expansion beyond a certain threshold was associated with protein unfolding, misfolding and aggregation leading to neurodegenerative diseases (Zoghbi *et al.* 2000, Chavali *et al.* 2017a, Mier *et al.* 2018). However, under physiological conditions polyQ repeats play important roles in protein-protein interactions, signaling pathways and transcriptional regulation (Schaefer *et al.* 2012, Gemayel *et al.* 2015, Totzeck *et al.* 2017). Homorepeats in budding yeast were preferentially found in regulatory proteins, important for cell fitness with impact on morphological phenotypes. The diverse molecular interactions described in these regions provide evidences that they can be beneficial to organisms and protect cells, contributing for the adaptability in response to environmental disturbances (Chavali *et al.* 2017b). Another interesting property of homorepeats is their high repeat-length polymorphism. In the fungus *C. glabrata*, the polyQ polymorphism of the general

transcription co-repressor Tup1A and the polyN polymorphism of the Ssk2, a MAPKKK of HOG signaling pathway, were correlated with decreased antifungal susceptibility (Challa *et al.* 2018). The high genetic plasticity and the functional versatility of these regions contribute for fungal adaptation to diverse types of environmental stresses, by modulating diverse molecular interaction networks with possible fitness advantages. In *C. albicans,* a significant percentage of the genome contains polymorphisms and a high variability in the number of repeats among strains (Whiteway *et al.* 1992, Butler *et al.* 2009). In fact, a repeat-length variation (11 to 49 repeats) within the polyQ segment of the transcription factor Rlm1 in different *C. albicans* clinical isolates was demonstrated to increase resistance to stress conditions (Sampaio *et al.* 2009).

In this work, we aim to unveil the role of the Cek1 N-terminal tail in modulating the biochemical and structural properties of this MAPK, as well to understand its combined effect with the ambiguous translation of the CUG-encoded residue as serine or leucine. Previous studies have shown that the N-terminal flexible tail of Cek1 is not relevant for its phosphorylation upon resumption of growth but it influences *C. albicans* susceptibility to cell wall disturbing agents (Correia *et al.* 2016). These results suggest that the N-terminal tail of Cek1 might play a role in the regulation of downstream phosphorylation events. Considering the differences in enzymatic activity and thermal stability observed for truncated Cek1 variants we wanted to explore the effect of the N-terminal flexible tail on the full-length protein. Based on that, two variants of full-length Cek1 were produced with a serine or a leucine at the CUG-encoded position. Purified recombinant proteins were thoroughly characterized by complementary biochemical and biophysical methodologies. The results suggest that the presence of the N-terminal tail increases the thermal stability of both full-length Cek1 variants. Nevertheless, and in accordance with the data on truncated Cek1 variants, the incorporation of a serine at the CUG position conferred higher thermal stability and enzymatic activity to the full-length Cek1 kinase, when compared to the leucine variant. Interestingly, autophosphorylation of Tyr233, belonging to the ^{231}TEY^{233} motif of Cek1, was observed for both variants of the full-length protein.

Together these findings show that Cek1 N-terminal region plays an important role on the kinase stability and activity. Moreover, it positively influences the previously observed beneficial effect of the incorporation of a serine at the CUG position. Hence, we can hypothesize that CUG identity in *C. albicans* controls Cek1

intrinsic kinase activity and ultimately could interfere with MAPK signaling pathway, influencing *C. albicans* morphogenesis and virulence.

3.2 Materials and Methods

3.2.1 Cloning, expression and purification

All the methods used for cloning, expression and purification of full-length Cek1 protein were performed as previously described for the truncated variants of Cek1 protein (ΔN-Cek1) (Chapter 2, section 2.2.1). Briefly, a synthetic gene encoding full-length Cek1 (Eurofins Scientific) was cloned into expression vector pETGb1 (EMBL dabatase) in frame with a N-terminal hexahistidine tag (His6), the B1 domain of Protein G (Gb1) and a tobacco etch virus (TEV) protease recognition site. The abbreviations used for full-length Cek1 follow what was early defined for the truncated Cek1 active domain, namely full-length Cek1_Ser (serine variant – UCC replacing the CUG codon) and full-length Cek1_Leu (leucine variant – UUA replacing the CUG codon) and all the mutants of both variants for the phosphorylation residues of the TEY motif (full-length Cek1^{TV}, full-length Cek1^{YF} and full-length Cek1$^{TV/YF}$) and the kinase dead mutant (full-length Cek1^{KR}). The full-length Cek1 variants and mutants were overexpressed in *E. coli* strain Rosetta (Stratagene) and induced with addition of 0.9 mM IPTG at 18˚C overnight.

Purification was performed according to the protocol previously described (Fraga *et al.* 2019) using an immobilized metal affinity chromatography (IMAC) by a HisTrap HP column (GE Healthcare) and a preparative size exclusion chromatography (SEC) Hiprep 26/60 Sephacryl S-100 (GE Healthcare) column. Purified protein (in 20 mM Tris pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 5 mM $MgCl_2$, 1 mM DTT) was analyzed by SDS-PAGE and concentrated on centrifugal ultrafiltration devices (10 kDa cutoff; Merck Millipore). Final protein concentration was estimated by measuring the absorbance of the samples at 280nm using the theoretical extinction coefficient for each variant (Table 3.1) prior to storing the samples at -80˚C.

Table 3.1 Summary of full-length Cek1 variants and respective mutants. Extinction coefficients for full-length Cek1 were calculated based on amino acid sequences with and without Gb1 tag.

Protein	CUG residue	Mutation	Extinction coefficient (M^{-1} cm^{-1}) (tag/no tag)
Full-length Cek1_Ser	serine	-	53,290/41,830
Full-length Cek1^{TV}_Ser	serine	threonine to valine	53,290/41,830
Full-length Cek1^{YF}_Ser	serine	tyrosine to phenylalanine	51,800/ 40,340
Full-length Cek1$^{TV/YF}$_Ser	serine	threonine to valine and tyrosine to phenylalanine	51,800/ 40,340
Full-length Cek1^{KR}_Ser	serine	lysine to arginine	53,290/41,830
Full-length Cek1_Leu	leucine	-	53,290/41,830
Full-length Cek1^{TV}_Leu	leucine	threonine to valine	53,290/41,830
Full-length Cek1^{YF}_Leu	leucine	tyrosine to phenylalanine	51,800/ 40,340
Full-length Cek1$^{TV/YF}$_Leu	leucine	threonine to valine and tyrosine to phenylalanine	51,800/ 40,340
Full-length Cek1^{KR}_Leu	leucine	lysine to arginine	53,290/41,830

3.2.2 Biophysical and biochemical characterization

The biophysical and biochemical analyses of full-length Cek1 by differential scanning fluorimetry (DSF), circular dichroism (CD), mass spectrometry (MS), *in vitro* phosphorylation (immunoblot), activity assay (ADP-Glo) and fluorescence measurements (TNP-ATP) were performed as detailed in Chapter 2 (section 2.2). The full-length Cek1 kinase variants fused to a Gb1 tag were used in all biophysical and biochemical assays except in the immunoblot assay where the Gb1 tag was removed. When the tagged protein was used, the presence of Gb1 is specifically indicated in each figure.

3.2.3 Small angle X-ray scattering measurements and analysis

Small angle X-ray scattering (SAXS) experiments were performed at the BioSAXS beamline BM29 (Pernot *et al.* 2013) of the European Synchrotron Radiation Facility (Grenoble, France), in collaboration with José António Manso (Macromolecular Structure group, i3S). SAXS data were collected using continuous-flow size-exclusion chromatography (SEC-SAXS). Samples in 20 mM Tris pH 7.5, 150

mM NaCl, 5% (v/v) glycerol, 5 mM $MgCl_2$, 1 mM DTT were injected onto a pre-equilibrated Yarra 3μm SEC-3000 column (Phenomenex). The images (frames) were collected over a scattering vector from 0.0035 to 0.5 $Å^{-1}$ (q= (4πsinθ)/λ, where 2θ is the scattering angle) using a Pilatus 1M detector (Dectris) at a radiation wavelength (λ) of 0.9919 Å. Data were processed and analyzed with the ATSAS 2.8 package (Franke *et al.* 2017)(Franke 2017) and the Guinier analysis was performed with the program PRIMUS/qt (Petoukhov *et al.* 2012).

3.3 Results

3.3.1 Sequence and structural analysis of MAPK Cek1

A detailed analysis of protein sequence and structural features, including the identification of specific structural and sequence motifs, is essential for a comprehensive understanding of its properties and to obtain clues about the protein function. The analysis of *C. albicans* MAPK Cek1 protein sequence using Fast Estimator of Latent Local Structure (FELLS) (Piovesan *et al.* 2017) predicts that Cek1 protein is essentially α-helical (60%), including several small regions with predicted random coil structural features (29%) (Figure 3.2A). The N-terminal tail of Cek1 (residues 1-59) includes a small hydrophobic cluster (residues 1-17), and is predicted to be highly disordered although with a propensity to acquire an α-helical structure (Figure 3.2A).

The structural homology model of Cek1 (residues 60-419) (Figure 3.2B), obtained using the online server Swiss-Model (Arnold *et al.* 2006), used as a template the 3D structure of the *S. cerevisiae* ortholog Fus3 (PDB accession code: 2B9F, 63% of amino acid sequence identity), which lacks the N-terminal tail of *C. albicans* Cek1 protein. Based on the structure analysis, the conserved CUG-encoded residue (serine 199) faces the nucleotide binding pocket (Chapter 2, Figure 2.1) (Rocha *et al.* 2011, Sárkány *et al.* 2014), and we previously demonstrated that a serine at the CUG-encoded position altered the thermal stability and auto-phosphorylation activity of the truncated Cek1 protein (ΔN–Cek1) (Fraga *et al.* 2019).

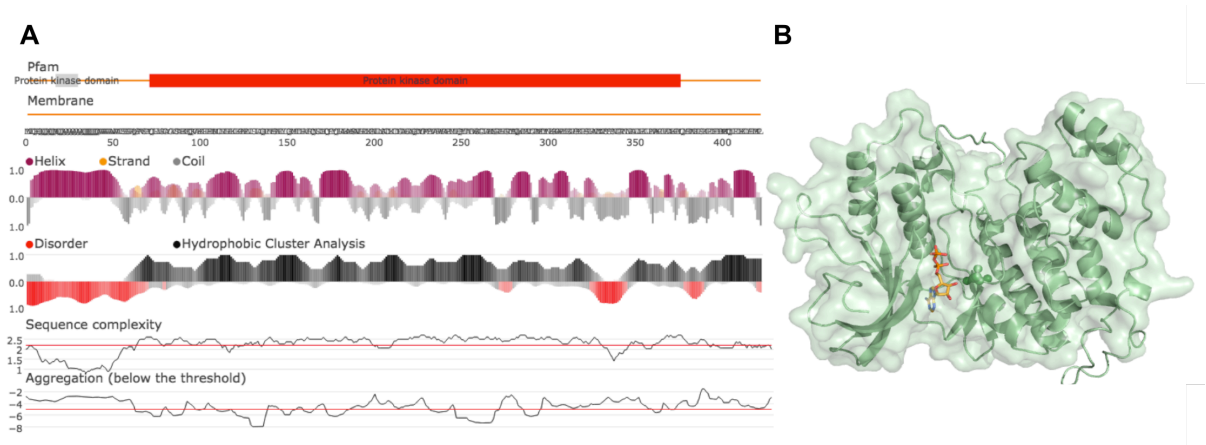


Figure 3.2 Structural analysis of MAPK Cek1. (A) Sequence analysis of full-length Cek1 protein by FELLS: Pfam domains, secondary structure prediction and disorder prediction of the Cek1 amino acid sequence. The sequence complexity and aggregation propensities are represented as continuous lines with the lower threshold in red. (B) Homology model of *C. albicans* Cek1 obtained using the online server Swiss-Model (Arnold *et al.* 2006) using *S. cerevisiae* Fus3 3D structure as a template.

3.3.2 The presence of the N-terminal tail increases the overall stability of Cek1 variants

In order to understand the effect of the N-terminal tail on the overall structure and thermostability of Cek1 kinase, the full-length Cek1 variants were used for DSF analysis. Data showed that full-length Cek1 containing a serine at the CUG-encoded position (full-length Cek1_Ser) was more stable (+3˚C) when compared with the leucine variant (full-length Cek1_Leu) (Figure 3.3A). These results are in agreement with those obtained for the truncated forms of Cek1 (ΔN-Cek1_Ser and ΔN-Cek1_Leu), revealing that a serine at the CUG position increased the stability of both Cek1 constructs. In summary, the thermostability results also demonstrate that the N-terminal low complexity region increases the overall stability of the protein independently of the identity of residue incorporated at the CUG-position (Figure 3.3B).

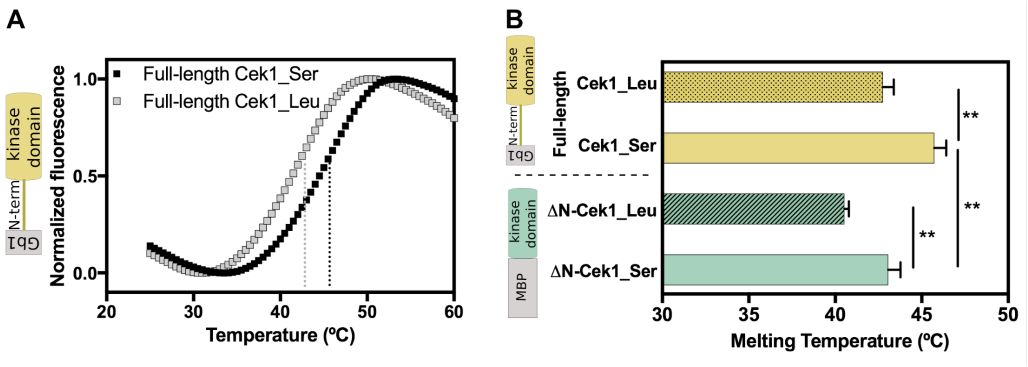


Figure 3.3 Thermal stability of full-length Cek1 kinase. (A) Representative thermal denaturation curve measured by DSF for both variants of full-length-Cek1. The temperature at the inflection point of the unfolding transition defines the Tm value for each protein (full-length Cek1_Ser=45.7 ± 0.7˚C; full-length Cek1_Leu=42.8 ± 0.6˚C). (B) Comparative values of Tm of the full-length Cek1 and truncated Cek1 (ΔN_Cek1_Ser=43.1 ± 0.7 ˚C; ΔN_Cek1_Leu=40.5 ± 0.3˚C) variants showed an increased in the Tm of the full-length Cek1 compared to ΔN-Cek1. Data shown are mean ± standard error of the mean (n=3). Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) p<0.05 for a 95% confidence interval (p>0.05 non-significant (ns), *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

The differences in the thermal stability of full-length Cek1 variants are not a consequence of significant structural conformational changes as confirmed by analysis of the variants by CD spectroscopy (Figure 3.4A). The data analysis of the CD revealed no differences on the global secondary structure composition of both variants (Figure 3.4A). Moreover, information on the full-length Cek1 variants structure in solution at low resolution, provided by SAXS, supported this observation (Figure 3.4B-C). The calculated scattering profiles of the full-length-Cek1 variants matched the experimental SAXS curve. The linearity of the Guinier plot of the scattering data profile indicates that both variants of full-length Cek1 were monodisperse and monomeric (Figure 3.4C). The Guinier radii of gyration (R_g) determined from the modified Guinier plot was similar for both protein variants (full-length Cek1_Ser 40±1 Å; full-length Cek1_Leu 38±1 Å), suggesting that insertion of a leucine or serine at the CUG encoded position did not affect the overall shape of the protein. Nonetheless, these results do not rule out the presence of local structural changes, not detected by SAXS, within the full-length Cek1 variants.

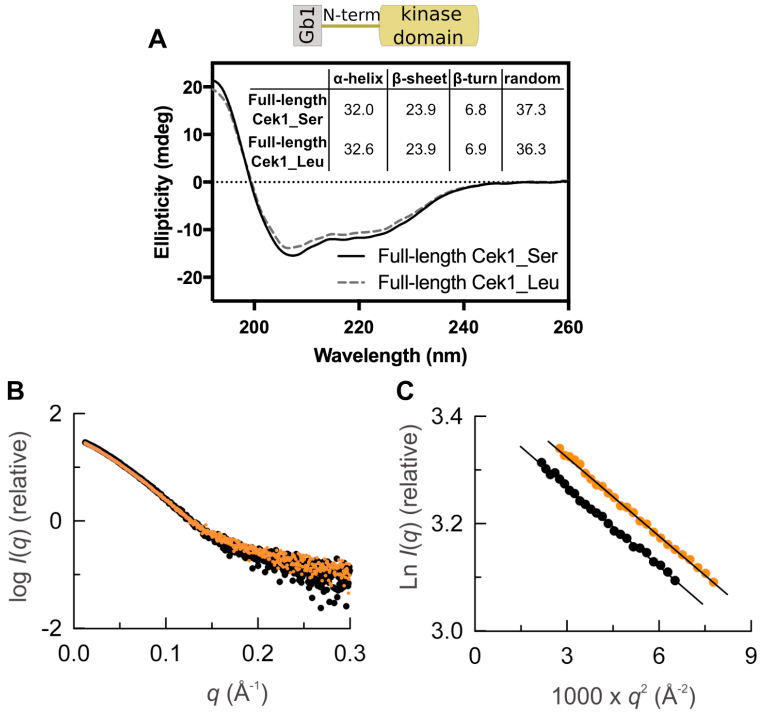


Figure 3.4 Structural analysis of full-length Cek1. (A) Far-UV CD spectra of full-length-Cek1_Ser and full-length-Cek1_Leu revealing no significant differences in the overall secondary structure of the variants. (B) SAXS analysis of the full-length-Cek1_Ser and full-length-Cek1_Leu. SAXS profiles extrapolated to infinite dilution of the full-length-Cek1_Ser (black) and full-length-Cek1_Leu (orange). Curves are offset on the log scale. (C) Guinier plots of the data in B.

3.3.3 Autophosphorylation of full-length Cek1

In order to assess the phosphorylation state of the full-length Cek1 variants, an immunoblot using the anti-phospho p44/42 was performed. It is important to remark that the recombinant full-length Cek1 variants (with a Gb1 and a His6 tag at the N-terminal region of the proteins) were previously digested with TEV protease to remove the tags. Gb1 is the highly stable and soluble B1 domain of protein G, an immunoglobulin-binding protein, which is widely used as an epitope tag for over-expression of proteins in *E. coli* (Huth *et al.* 1997, Cheng *et al.* 2004). Hence, Gb1 contains, at its surface, binding sites for the C-terminal fragment of the heavy chain of immunoglobulin G allowing its recognition by the antibodies used in the immunoblots, producing a 'false positive' signal (Figure 3.5). Therefore, all the immunoblot experiments to detect the phosphorylation of full-length Cek1 variants

were performed after removal of the Gb1 tag to avoid non-specific signal of the anti-phospho p44/42 antibody.

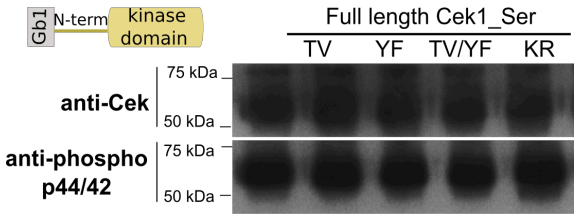


Figure 3.5 Immunoblot analysis of full-length Cek1_Ser and corresponding mutants fused to the Gb1 tag. Immunoblot probed with rabbit anti-Cek serum (upper panel) and rabbit anti-phospho p44/42 (bottom panel). A string signal was detected for all proteins, indicating a non-specific reaction of the anti-phospho p44/42 antibofy with the solubility tag Gb1.

As observed for the ΔN-Cek1_Ser, full-length Cek1_Ser variant was also auto-phosphorylated during expression in *E. coli*, as determined by immunoblot analysis using the phospho-specific p44/42 antibody (Figure 3.6A). However, the full-length Cek1_leucine variant showed a lower phosphorylation signal, even though the total amount of Cek1 protein loaded was equivalent for both variants as assessed using the anti-Cek serum (Figure 3.6A). Accordingly, the quantification of the enzyme activity of full-length Cek1 variants by the measurement of ADP release showed that full-length Cek1_Ser was significantly more active than full-length Cek1_Leu (Figure 3.6B). The serine variant showed a 6-fold higher activity when compared to leucine, suggesting that a leucine at the CUG-encoded position affects the kinase activity of the full-length Cek1, in agreement with what was previously shown for the truncated Cek1 active domain (Figure 3.6C).

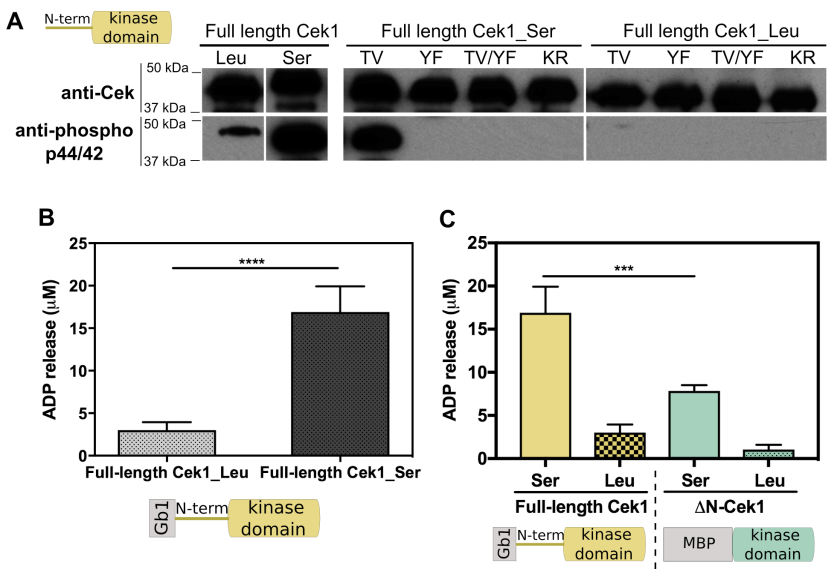


Figure 3.6 Full-length Cek1 is autophosphorylated at Tyr233. The serine variant displays higher catalytic activity than the leucine variant. (A) Immunoblot analysis of full-length Cek1_Leu and full-length Cek1_Ser and of the corresponding mutants for both variants, probed with rabbit anti-Cek serum (upper panel) and rabbit anti-phospho p44/42 (bottom panel). Autophosphorylation was detected in both full-length Cek1 variants, indicating that autophosphorylation occurs on Tyr233. (B) Activity assay by quantification of ADP release of full-length Cek1 variants: full-length Cek1_Ser released significantly more ADP when compared to full-length Cek1_Leu. (C) Comparative data on the ADP release of full-length Cek1 and truncated Cek1 variants. (B-C) Data shown are mean ± standard error of the mean (n=3). Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ for a 95% confidence interval ($p>0.05$ non-significant (ns), *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

Immunoblot analysis of the full-length Cek1 mutants indicated that autophosphorylation of the ^{231}TEY^{233} motif also occurs at Tyr233 (Figure 3.6A). The autophosphorylation signal detected with anti-phospho p44/42 was only observed in the full-length Cek1_Ser mutant where the Thr231 was substituted for a valine (full-length Cek1^{TV}_Ser) (Figure 3.6A). Regarding the additional mutants of the ^{231}TEY^{233} motif, where the Tyr233 was replaced by phenylalanine (full-length Cek1^{YF}) and the double mutant where the Thr231 and Tyr233 were mutated to valine and phenylalanine (full-length Cek1$^{TV/YF}$), respectively, no autophosphorylation was detected, supporting the evidence that autophosphorylation occurs at Tyr233 (Figure 3.6A). As expected, the substitution of the catalytic lysine by an arginine in the inactive kinase-dead mutant (full-length Cek1^{KR}) abolished the autophosphorylation signal (Figure 3.6A). To further confirm these results, MS

analysis identified a phosphorylated peptide with a mass compatible with a single phosphorylation (SIASQEDNYGFMTEpYVATR) on the full-length Cek1_Ser validating the findings that autophosphorylation occurs at Tyr233 (Figure 3.7A). Nonetheless, in the full-length Cek1_Leu an equivalent phosphorylated peptide was found indicating the same result but the identity of phosphorylated residue remains unclear due to the low MS signal detected that does not allow the accurate identification of the residue modified by phosphorylation (Figure 3.7B). However, MS analysis together with immunoblot data revealed that an insertion of a serine at the CUG position positively impacted on the autophosphorylation of recombinant expressed full-length Cek1 protein (Figure 3.7C) and increased its catalytic activity.

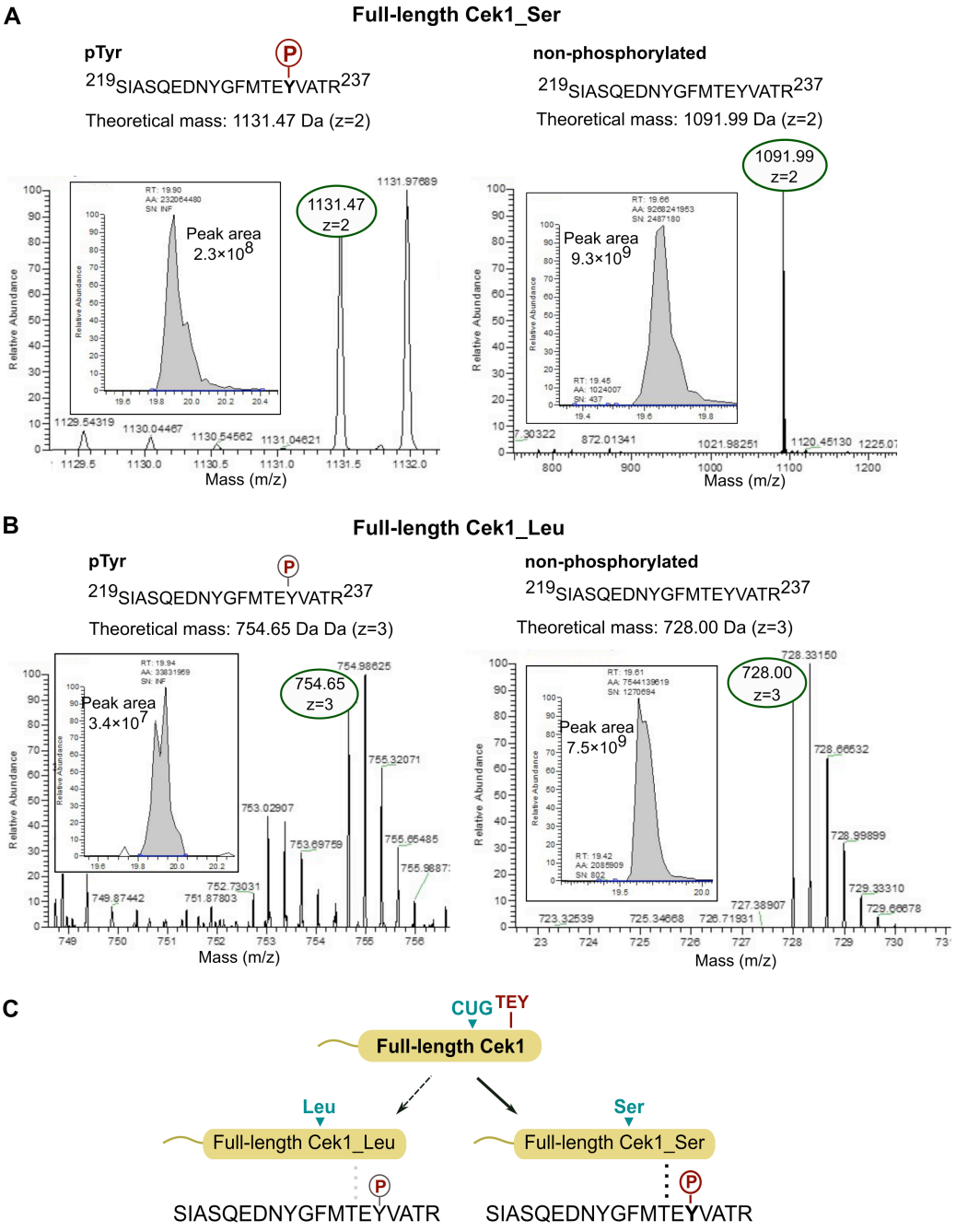


Figure 3.7 LC-MS/MS analysis of the full-length-Cek1 variants. Representative extracted ion chromatogram with the peak area and the respective m/z for the peptide of interest (SIASQEDNYGFMTEYVATR). The integrated peak areas (insets) of the phosphorylated peptide and of its equivalent non-phosphorylated version were used to calculate an apparent percentage of phosphorylation of full-length Cek1. The theoretical mass of the peptides found was calculated using the online server ProteinProspector (MS-isotope). (A) MS spectrum for the full-length Cek1_Ser variant: Left panel - phosphopeptide (SIASQEDNYGFMTEpYVATR, peak area 2.3×10^{8}); Right panel: the equivalent non-phosphorylated peptide (peak area 9.3×10^{9}). Comparative analysis of the peak areas corresponding to phosphorylated and non-phosphorylated peptide fragments, suggested

that only a small percentage of full-length Cek1_Ser was autophosphorylated on tyrosine 233. (B) MS spectrum for the full-length Cek1_Leu variant: Left panel - phosphopeptide (SIASQEDNYGFMTEpYVATR, peak area 3.4×10^{7}); Right panel - the equivalent non-phosphorylated peptide (peak area 7.5×10^{9}). (C) Schematic representation of phosphorylation results on the full-length Cek1 variants. Data obtained by immunoblot analysis combined with mass spectrometry indicate that Tyr233 was modified by autophosphorylation in both variants.

3.3.4 Nucleotide affinity is not affected by the insertion of a serine or a leucine at the CUG encoded position

Given that a serine residue at the CUG position, located in the ATP binding pocket of Cek1, increased the stability and the kinase activity of full-length Cek1, we further explored the mechanism behind these differences. One of the hypotheses was the ATP binding affinities of the full-length Cek1 variants could be different. To test this hypothesis a TNP-ATP nucleotide, a fluorescent ATP analog, was used to evaluate its binding affinity of each full-length Cek1 variant. Results show that both variants bind the nucleotide with similar affinities (3.4 µM for full-length Cek1_Ser and 3.5 µM for full-length Cek1_Leu) (Figure 3.8). Thus, the incorporation of a serine or a leucine at the CUG residue did not affect its nucleotide binding affinity, as we previously observed for the truncated active domain of Cek1 (Chapter 2, section 2.3.3, Figure 2.6B).

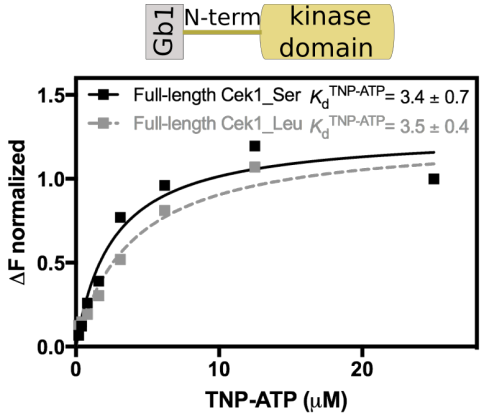


Figure 3.8 Binding of TNP-ATP nucleotide to full-length Cek1 variants. Both variants show equivalent affinities for TNP-ATP (full-length Cek1_Ser $K_d^{TNP\text{-}ATP}$=3.4 ± 0.7; full-length Cek1_Leu $K_d^{TNP\text{-}ATP}$=3.5 ± 0.4). The data are representative of three independent experiments.

3.4 Discussion

In the previous chapter, we have demonstrated that the identity of the CUG residue in the N-terminal truncated *C. albicans* MAPK Cek1, containing only the structurally conserved kinase domain, modulates its thermal stability and autophosphorylation activity. In order to evaluate if the differences observed between serine and leucine insertion at the CUG-encoded residue of the truncated Cek1 also occur in the full-length protein, that includes the N-terminal tail with homorepeats, we characterized biochemically and biophysically the full-length Cek1 variants.

Our results confirm that the replacement of a polar serine by an aliphatic leucine residue has a negative impact on the enzymatic activity and stability of full-length Cek1, without compromising its overall structure. As observed for the truncated Cek1 variants, those differences do not result from variations in the nucleotide binding affinities. Interestingly, both full-length Cek1 variants were autophosphorylated as detected by the anti-phospho p44/42 antibody. However, the autophosphorylation signal detected with full-length Cek1_Leu was less intense, in accordance with activity assays, in which the catalytic activity of the leucine variant was significantly lower than the serine variant. This result was in contrast to what was uncovered by the truncated Cek1 protein, where a leucine at the CUG encoded position completely abolished the autophosphorylation of the Cek1 kinase. Moreover, as we previously observed in the truncated Cek1_Ser, the autophosphorylation of full-length Cek1 occurs on the conserved Tyr233 of the ^{231}TEY^{233} motif within the kinase activation loop.

Hence, our results reinforce previous evidences that the identity of the CUG codon modulates Cek1 stability and enzymatic activity. In this way, we could confirm our initial hypothesis that a leucine at the CUG position interferes with Cek1 activity and function by reducing, in the full-length Cek1, or abolishing, in the truncated Cek1, the autophosphorylation signal. Interestingly, the presence of the N-terminal tail enhances the overall thermal stability and enzymatic activity of the full-length Cek1 variants. In the full-length leucine-containing variant, the inclusion of the N-terminal tail increases the Tm (42.8 ˚C) to values similar to those for the more stable truncated serine-containing Cek1 (ΔN-Cek1_Ser=43.1 ˚C). This suggests that the presence of this low complexity region partly compensates the loss of stability imparted by the incorporation of leucine at the CUG site. For the serine variant, the Cek1 full-length protein displays a Tm that is 2.6 ˚C higher than for the

truncated construct (ΔN-Cek1_Ser). Nevertheless, the negative impact of leucine insertion at the CUG position for enzymatic activity is not fully compensated by the inclusion of the N-terminal tail. Therefore, the influence of the CUG residue identity for the phosphorylation of specific peptide substrates by the full-length protein still needs to be explored. However, preliminary assays suggest that full-length Cek1 variants were unable to phosphorylate the synthetic Cph1 peptide, as we previous observed for the truncated Cek1 protein. Altogether, flexible N-terminal tail of Cek1 seems to have a positive effect on Cek1 thermal stability and enzymatic activity, and potentiates Cek1 autophosphorylation *in vitro*.

Our results showed the N-terminal low complexity region positively impacts on the stability, enzymatic activity and phosphorylation status of the full-length Cek1, but the exact molecular mechanisms underlying the observed effects remain to be determined. Flexible regions in protein kinases are known to undergo conformational exchange and their highly dynamic properties may mediate transient interactions between structured domains facilitating the interaction with binding partners (Mollica *et al.* 2016) or directly regulate enzyme activity (Niesteruk *et al.* 2018b). For example, the N-terminus of PtkA modulates, via a "fly-casting" like mechanism of transient interactions, the accessibility of the active site within its conserved kinase domain (Niesteruk *et al.* 2018b). Although it is tempting to speculate that in Cek1 the N-terminal tail might directly affect the dynamic properties of the structured kinase domain, differences on the expression tags used in both constructs (MBP for the truncated Cek1 and Gb1 for the full-length protein) do not allow a direct and readable explanation for the observed dissimilarities. Previous *in vivo* results with *C. albicans* strains demonstrated that the N-terminal region was not essential for Cek1 activation by upstream protein kinases, since deletion of this region still enabled its phosphorylation (Correia *et al.* 2016).

An interesting feature of low complexity regions, in particular those containing homorepeats, is the potential for expansion or contraction of the repeated amino acid segment. *C. albicans* showed significant intra-strain variability in the number of amino acid repeats (Whiteway *et al.* 1992, Butler *et al.* 2009) which could modulate protein function, that we speculate might offer some advantages for *C. albicans* rapid adaptation to the different host microenvironments (Sampaio *et al.* 2009). Preliminary comparison of Cek1 polymorphisms in various *C. albicans* strains showed a significant variability in the size of the polyQ repeats in its N-terminal tail (personal communication by Dr.

Chavali, Indian Institute of Science Education and Research, Tirupati, India) (Figure 3.9).

Figure 3.9 Cek1 N-terminal region repeat variability across different *C. albicans* strains (provided by Dr. Chavali)**.**

We postulate that variable flexible segments enriched in polyQs create new regulatory opportunities by modulating protein function and/or complex assembly. Those regulatory networks might impact on the flow information across *C. albicans* signaling pathways, with consequences for adaptation to the host and pathogenicity. For instance, the polyQ tract at the N-terminal region of the *C. albicans* Tpk2 is involved in its invasive properties (Bockmuhl *et al.* 2001). Additionally, the polyQ tract in the transcription factor Efg1 (at the N- and C-terminal regions) is required for the specific morphogenetic switching (Noffz *et al.* 2008) and in the Cph1 (two polyQ at the C-terminal) is essential for the transcriptional activation (Maiti *et al.* 2015).

Altogether, our biochemical and biophysical results with full-length Cek1 give the first insight on the role of the CUG ambiguity in the context of full-length *C. albicans* Cek1. Further studies will be required in order to understand the complex mechanism that mediates the correlation between CUG codon ambiguity and the presence of amino acid homorepeats in *C. albicans* proteins with implications in virulence and pathogenicity.

3.5 References

Arnold, K., L. Bordoli, *et al.* (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.*Bioinformatics* 22: 195-201.

Babu, M., J. Vlasblom, *et al.* (2012). Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae.*Nature* 489: 585-9.

Bockmuhl, D. P., S. Krishnamurthy, *et al.* (2001). Distinct and redundant roles of the two protein kinase A isoforms Tpk1p and Tpk2p in morphogenesis and growth of Candida albicans.*Mol Microbiol* 42: 1243-57.

Butler, G., M. D. Rasmussen, *et al.* (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes.*Nature* 459: 657-62.

Challa, K., T. Edlind, *et al.* (2018). Polymorphism of Polymeric Amino Acid Regions in Fungal Proteins and Correlation with Altered Echinocandin and Azole Susceptibility.*Antimicrob Agents Chemother* 62.

Chavali, S., P. L. Chavali, *et al.* (2017a). Constraints and consequences of the emergence of amino acid repeats in eukaryotic proteins.*Nat Struct Mol Biol* 24: 765-777.

Chavali, S., A. Gunnarsson, *et al.* (2017b). Intrinsically Disordered Proteins Adaptively Reorganize Cellular Matter During Stress.*Trends Biochem Sci* 42: 410-412.

Cheng, Y. and D. J. Patel (2004). An efficient system for small protein expression and refolding.*Biochem Biophys Res Commun* 317: 401-5.

Correia, I., E. Roman, *et al.* (2016). Complementary roles of the Cek1 and Cek2 MAP kinases in Candida albicans cell-wall biogenesis.*Future Microbiol* 11: 51-67.

Faux, N. G., S. P. Bottomley, *et al.* (2005). Functional insights from the distribution and role of homopeptide repeat-containing proteins.*Genome Res* 15: 537-51.

Fraga, J. S., Z. Sarkany, *et al.* (2019). Genetic code ambiguity modulates the activity of a C. albicans MAP kinase linked to cell wall remodeling.*Biochim Biophys Acta Proteins Proteom* 1867: 654-661.

Franke, D., M. V. Petoukhov, *et al.* (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.*J Appl Crystallogr* 50: 1212-1225.

Galan-Diez, M., D. M. Arana, *et al.* (2010). Candida albicans beta-glucan exposure is controlled by the fungal CEK1-mediated mitogen-activated protein kinase pathway that modulates immune responses triggered through dectin-1.*Infect Immun* 78: 1426-36.

Gemayel, R., S. Chavali, *et al.* (2015). Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity.*Molecular Cell* 59: 615–627.

Herrero-de-Dios, C., R. Alonso-Monge, *et al.* (2014). The lack of upstream elements of the Cek1 and Hog1 mediated pathways leads to a synthetic lethal phenotype upon osmotic stress in Candida albicans.*Fungal Genet Biol* 69: 31-42.

Huth, J. R., C. A. Bewley, *et al.* (1997). Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR.*Protein Sci* 6: 2359-64.

Kathiriya, J. J., R. R. Pathak, *et al.* (2014). Presence and utility of intrinsically disordered regions in kinases.*Mol Biosyst* 10: 2876-88.

Keppel, T. R., K. Sarpong, *et al.* (2017). Biophysical Evidence for Intrinsic Disorder in the C-terminal Tails of the Epidermal Growth Factor Receptor (EGFR) and HER3 Receptor Tyrosine Kinases.*J Biol Chem* 292: 597-610.

Kumari, B., R. Kumar, *et al.* (2015). Low complexity and disordered regions of proteins have different structural and amino acid preferences *Molecular BioSyatems* 11: 585-594.

Luo, H. and H. Nijveen (2014). Understanding and identifying amino acid repeats.*Brief Bioinform* 15: 582-91.

Maiti, P., P. Ghorai, *et al.* (2015). Mapping of functional domains and characterization of the transcription factor Cph1 that mediate morphogenesis in Candida albicans.*Fungal Genet Biol* 83: 45-57.

Mier, P. and M. A. Andrade-Navarro (2018). Glutamine Codon Usage and polyQ Evolution in Primates Depend on the Q Stretch Length.*Genome Biol Evol* 10: 816-825.

Mollica, L., L. M. Bessa, *et al.* (2016). Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment.*Front Mol Biosci* 3: 52.
Niesteruk, A., M. Hutchison, *et al.* (2018a). Structural characterization of the intrinsically disordered domain of Mycobacterium tuberculosis protein tyrosine kinase A.*FEBS Lett* 592: 1233-1245.
Niesteruk, A., H. R. A. Jonker, *et al.* (2018b). The domain architecture of PtkA, the first tyrosine kinase from Mycobacterium tuberculosis, differs from the conventional kinase architecture.*J Biol Chem* 293: 11823-11836.
Noffz, C. S., V. Liedschulte, *et al.* (2008). Functional mapping of the Candida albicans Efg1 regulator.*Eukaryot Cell* 7: 881-93.
Pernot, P., A. Round, *et al.* (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution.*J Synchrotron Radiat* 20: 660-4.
Petoukhov, M. V., D. Franke, *et al.* (2012). New developments in the ATSAS program package for small-angle scattering data analysis.*J Appl Crystallogr* 45: 342-350.
Piovesan, D., I. Walsh, *et al.* (2017). FELLS: fast estimator of latent local structure.*Bioinformatics* 33: 1889-1891.
Rocha, R., P. J. Pereira, *et al.* (2011). Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen.*Proc Natl Acad Sci U S A* 108: 14091-6.
Sampaio, P., E. Nogueira, *et al.* (2009). Increased number of glutamine repeats in the C-terminal of Candida albicans Rlm1p enhances the resistance to stress agents.*Antonie Van Leeuwenhoek* 96: 395-404.
Sárkány, Z., A. Silva, *et al.* (2014). Ser or Leu: structural snapshots of mistranslation in Candida albicans.*Front Mol Biosci* 1: 1-14.
Schaefer, M. H., E. E. Wanker, *et al.* (2012). Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks.*Nucleic Acids Res* 40: 4273-87.
Spassov, D. S., A. Ruiz-Saenz, *et al.* (2018). A Dimerization Function in the Intrinsically Disordered N-Terminal Region of Src.*Cell Rep* 25: 449-463 e4.
Sun, T., X. Yang, *et al.* (2010). Cellular abundance of Mps1 and the role of its carboxyl terminal tail in substrate recruitment.*J Biol Chem* 285: 38730-9.
Totzeck, F., M. A. Andrade-Navarro, *et al.* (2017). The Protein Structure Context of PolyQ Regions.*PLoS One* 12: e0170801.
van der Lee, R., M. Buljan, *et al.* (2014). Classification of intrinsically disordered regions and proteins.*Chem Rev* 114: 6589-631.
Whiteway, M., D. Dignard, *et al.* (1992). Dominant negative selection of heterologous genes: isolation of Candida albicans genes that interfere with Saccharomyces cerevisiae mating factor-induced cell cycle arrest.*Proc Natl Acad Sci U S A* 89: 9410-4.
Wright, P. E. and H. J. Dyson (2015). Intrinsically disordered proteins in cellular signalling and regulation.*Nat Rev Mol Cell Biol* 16: 18-29.
Zoghbi, H. Y. and H. T. Orr (2000). Glutamine repeats and neurodegeneration.*Annu Rev Neurosci* 23: 217-47.

Chapter 4

Assembly and regulation of the Cek1 interaction network in *Candida albicans*

Work contributions:

Expression of Soluble Proteins by Random Incremental Technology (ESPRIT) was performed in collaboration with Philippe Mas and Darren Hart at the Institute of Structural Biology (Grenoble, France) through a project financed by Instruct-ERIC (Structural Biology European Research Infrastructure). The Bio-Layer Interferometry and data analysis was done by Carla Almeida at X-PROT, Biocant (Cantanhede).

4.1 Introduction

The cellular functions are critically dependent on the correct assembly of proteins to become functional multiprotein complexes. Protein-protein interactions (PPIs) are essential to most biological processes and the major challenge is to understand its complex network properties and dynamic interchanges in response to various signals (Westermarck *et al.* 2013). PPIs often define the specificity in signal transduction and might differ in the physiological and pathological conditions. The correct functioning of signaling pathways requires multiple sequential and transient interactions between upstream and downstream components of the signal cascade.

Several approaches have been used to uncover the protein interaction networks in *Candida albicans*, such as the detection of binary interactions by the commonly used methodology yeast two-hybrid assay (Y2H) (Ratushny *et al.* 2008, Cote *et al.* 2011). Nevertheless, as previously mentioned, the *C. albicans* alternative genetic code usage poses one the biggest challenges regarding large-scale protein interaction studies in this human pathogen, mostly because of the requirement to develop unique tools adapted to *C. albicans* (Schoeters *et al.* 2019). Protein-protein interactions in *C. albicans* have been identified by the vesicle capture interaction assay (Boysen *et al.* 2009) or tandem-affinity purification (Kaneko *et al.* 2004, Blackwell *et al.* 2009). Another approach for PPIs detection is the expression of the bait protein with an unnatural photo-cross-linker amino acid, that promotes the formation of covalent links with interacting proteins in its proximity (Palzer *et al.* 2013). Recently, a bimolecular fluorescence complementation assay (Subotic *et al.* 2017) and the adapted yeast two-hybrid system for use in *C. albicans* (C2H) (Stynen *et al.* 2010) were developed providing new insights into molecular mechanisms of PPIs and signaling in *C. albicans*. Furthermore, this C2H improved for a high-throughput screen disclosed thousands of PPIs in *C. albicans* (Schoeters *et al.* 2018).

Many biological processes relevant for *C. albicans* survival in the human host as a commensal or as a pathogen, such as biofilm formation and filamentous growth, depend on the dynamic association and dissociation of regulatory macromolecular interaction networks (Das *et al.* 2019). The MAPK signaling cascades, in particular, respond to external cues by assembling functional protein complexes whose function is regulated by highly specific PPIs. The Cek1 MAPK pathway is highly enriched in proteins containing CUG-encoded residues, in

particular Cek1 (1 CUG), Hst7 (4 CUGs) and Cst5 (3 CUGs). Therefore, the CUG ambiguity might modulate and reshape the PPIs in this *C. albicans* signal transduction cascade linked to morphogenesis and virulence. In agreement, discrepancies were observed in PPIs using *S. cerevisiae* Y2H assay (Cote *et al.* 2011), in which all CUG codons are translated as leucine, and the *C. albicans* C2H (Stynen *et al.* 2010). This C2H methodology revealed known and novel protein interactions, namely the interaction between Hst7 and Cek1, that was not identified by the Y2H system (Stynen *et al.* 2010) (Figure 4.1, continuous line), suggesting that the CUG residue identity is relevant for the Hst7-Cek1 interaction. In addition, Cek1 interaction with downstream transcription factor Cph1 was also identified (Stynen *et al.* 2010) (Figure 4.1, continuous line). Despite the disadvantage of the Y2H assay that could miss some of the PPI in *C. albicans*, some interactions among these regulatory proteins were discovered by this method and confirmed by co-immunoprecipitation experiments (Cote *et al.* 2011, Yi *et al.* 2011). Hence, it was demonstrated that MAPK Cek1 is able to interact with upstream kinases (Ste11 and Hst7) and the scaffold protein Cst5 (Figure 4.1, dashed lines). Moreover, *C. albicans* Cst5 binds to the MAP kinases Ste11 and Hst7, which also interact with Ste11 (Cote *et al.* 2011, Yi *et al.* 2011) (Figure 4.1, dashed lines).

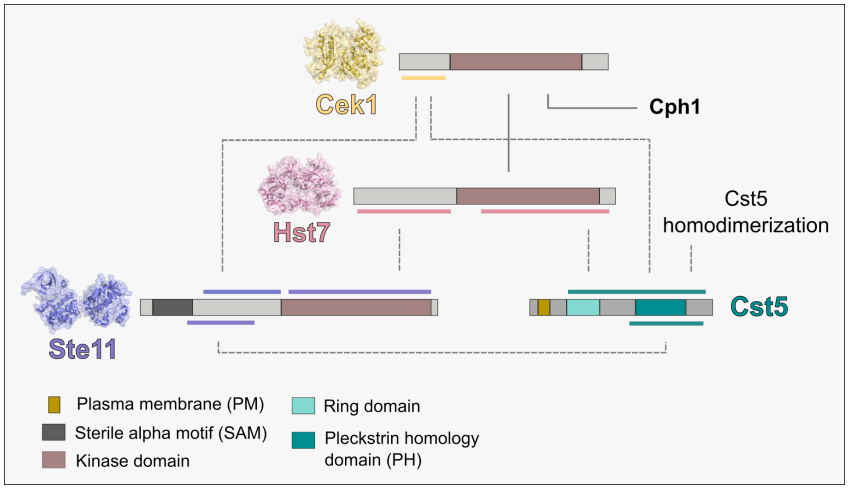


Figure 4.1 Diagram of the PPIs described in *C. albicans* Cek1 pathway. The scheme maps the binding regions between MAP kinases Ste11-Hst7-Cek1, the scaffold protein Cst5, and the downstream transcription factor Cph1. The locations of the interacting regions mapped for each interaction are underlined. Dashed lines represent the interactions identified by Y2H system and continuous line the interaction by C2H methodology.

Docking interactions between MAPKs and their interaction partners are important for directing the signal transduction flow. The presence of CUG-encoded residues together with differences in domain organization of the scaffold protein Cst5, which is crucial for the regulation of signal transduction, suggest alterations in the docking protein-protein interactions of this Cek1-dependent MAPK network (Cote *et al.* 2011). The scaffold protein Cst5 shares 24% of amino acid sequence identity with *S. cerevisiae* Ste5 and has less than half its size (Figure 4.2A). Moreover, while both scaffold proteins contain the RING and pleckstrin homology (PH) domain, the disappearance of the von Willebrand type A (vWA) domain from the Cst5 protein (Figure 4.2A-B), suggest a modified assembly mechanism regulating the Cek1-mediated signal transduction cascade (Cote *et al.* 2011, Yi *et al.* 2011).

The Ste5 RING domain interacts with Gβ subunit of G protein to localize it to the membrane when the pathway is activated, and contains a cysteine-rich region involved in zinc coordination, that is well conserved in *C. albicans* Cst5. The RING-H2 motif of *S. cerevisiae* Ste5 is a variant of the class RING domains, with a second histidine in place of the cysteine usually found at position 5. In *C. albicans* Cst5, the RING domain displays a histidine residue at position 5, but contains at position 4 an aspartate (conserved in *C. albicans* orthologs) replacing the histidine residue found in *S. cerevisiae* Ste5 (Figure 4.2A). As in Ste5, the Cst5 RING domain mediates its homodimerization (Cote *et al.* 2011). The PH region of Ste5, well conserved in Cst5, binds to phosphoinositol, facilitating its membrane docking, and also to Ste11 allowing the association with its activator Ste20 and MAPK cascade initiation (Inouye *et al.* 1997, Garrenton *et al.* 2006). In contrast, the Fus3-binding domain (FBD) is not totally conserved in Cst5 (Figure 4.2A). This region is not required for Fus3 activation but instead serves a feedback regulatory role (Bhattacharyya *et al.* 2006, Good *et al.* 2009), a molecular mechanism that might not be fully replicated in the Cst5-Cek1 interaction.

In *S. cerevisiae*, the Ste5 vWA domain, missing in Cst5, mediated the interaction with Ste7, acting as a co-activator that allosterically unlocks the Fus3 activation loop facilitating its phosphorylation by Ste7 (Bhattacharyya *et al.* 2006, Good *et al.* 2009). Ste7 displays 39% of amino acid sequence identity with *C. albicans* Hst7 and contains two docking motifs, conserved in Hst7, that bind to Fus3 (Figure 4.2C) (Remenyi *et al.* 2005).

Nevertheless, it seems that the smaller size of Cst5 and the lack of the vWA domain, compared to *S. cerevisiae* Ste5, do not hamper Cst5 homodimerization and the tethering of the MAPKKK Ste11, MAPKK Hst7 and MAPK Cek1. The scaffold

protein Cst5 is an essential protein in the activation and regulation of *C. albicans* signaling cascade but the fine details of the molecular interactions mediating its tethering function appear to have rewritten (Stynen *et al.* 2010, Cote *et al.* 2011, Yi *et al.* 2011).

Figure 4.2 Amino acid sequence analysis of *C. albicans* Cst5 scaffold protein and MAPKK Hst7. (A) Amino acid sequence alignment of *C. albicans* Cst5 and *S. cerevisiae* orthologue Ste5. The three CUG-encoded residues (Ser226, 237 and 254) of *C. albicans* Cst5 are marked with blue triangle. The main structural elements of Cst5 (above amino acid sequence) and Ste5 (below amino acid sequence) are represented: RING domain (light blue), PH domain (teal), vWA domain (orange), FBD (yellow), and the cysteine and histidine

conserved zinc atom coordination motifs of the RING-H2 finger domain. (B) Schematic representation of the domain organization of Cst5 and Ste5, according to the amino acid sequence alignment shown in A. (C) Amino acid sequence alignment of *C. albicans* Hst7 and *S. cerevisiae* orthologue Ste7. The four CUG-encoded residues (Ser71, 227, 264 and 284) of *C. albicans* Ste7 are marked with blue triangle. Ste7 peptides docking motifs that bind to Fus3 (pink). (A, C) Residues are colored according to a residue conservation scale (red: identical residues, orange to blue: decreasing conservation of amino acid properties; white: dissimilar residues). Figure prepared with Aline (Bond *et al.* 2009).

In order to map the protein-protein interactions regulating the *C. albicans* Cek1 signal transduction pathway, we aimed to characterize the upstream kinase of Cek1, Hst7 and the scaffold protein Cst5. To this end, different Cst5 truncation constructs were successfully produced using the ESPRIT (Expression of Soluble Proteins by Random Incremental Technology) for structural and functional studies. Remarkably, our results disclosed a direct interaction between the active domain of Cek1 and a Cst5 construct lacking the N-terminal RING domain. Nonetheless, additional studies are essential to unravel the molecular details of this interaction and the effect of CUG residue identity on the assembly of the kinase modules on the Cst5 scaffold.

4.2 Material and Methods

4.2.1 Strategies for cloning, expressing and purifying Cek1 interactors

A synthetic gene encoding *C. albicans* Cst5 (CUG codons mutated to a serine residue) was purchased (Genscript) with the codons optimized for *E. coli* expression and cloned into pPR-IBA2 vector (IBA Biotecnology) with restriction sites NdeI and EcoRI. Cst5 (Strep_Cst5_His) was in frame with a N-terminal Strep-tag and a tobacco etch virus (TEV) protease recognition site and a C-terminal with TEV recognition site and hexahistidine (His6) tag. In addition, a new construct of Cst5 (Strep_Cst5) was produced by site-directed mutagenesis (QuickChange, Agilent Tecnologies) to introduce a STOP codon before the TEV recognition site and His6 tag at the C-terminal (Table 4.1).

Table 4.1 Oligonucleotides used for site-directed mutagenesis of Cst5.

Oligonucleotide	Sequence (5´ to 3´)
Cst5_stop1277 for	TGCTGACCTCGGACCCGTAAGAGAATCTTTATTTTCAG
Cst5_stop1277 rev	CTGAAAATAAAGATTCTCTTACGGGTCCGAGGTCAGCA

Expression of Cst5 was assessed at small scale combining five different *E. coli* strains (BL21(DE3), BL21Star(DE3), BL21(DE3)pLysS, Origami2(DE3) and Rosetta(DE3) and temperatures (20 ˚C and 25 ˚C) obtaining ten different expression conditions. Protein expression was induced by the addition of Isopropyl-β-D-thigalactoside (IPTG) to a final concentration of 0.5 mM and incubating for 4 hours or overnight depending on temperature, 25 or 20 ˚C respectively. After protein expression, the cultures were harvested at 4000 g, 4 ˚C for 15 minutes and pellets were ressuspended in lysis buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 5% (v/v) glycerol, 5 mM 2-mercaptoethanol). Cells were lysed by sonication (Sonoplus HD2200, Bandelin Electronic) and centrifuged at 13000 g (Eppendorf 5414) for 15 min at 4 ˚C. The supernatants (soluble fractions) and pellets (insoluble fractions) were resuspended in lysis buffer and analyzed by SDS-PAGE.

Soluble expression of Strep_Cst5_His was found in BL21(DE3)pLysS at 25 ˚C, after 4 hours on inducing protein expression with IPTG, while for Strep_Cst5 in Rosetta(DE3) soluble protein expression was observed following an overnight induction at 20 ˚C. Purification of both constructs was performed in a StrepTrap HP column (5 ml, GE Healthcare) pre-equilibrated in buffer A (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1mM EDTA, 5% glycerol, 5mM β-mercaptoethanol) and proteins eluted with buffer A supplemented with 2.5 mM *d*-desthiobiotin. Further, the eluted fraction was purified by size exclusion chromatography (SEC) on a HiPrep 16/60 Sephacryl S100 column (GE Healthcare) equilibrated with protein buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1mM EDTA, 5% glycerol, 1 mM DTT). Fractions containing protein were concentrated on centrifugal ultrafiltration devices (10 kDa cutoff; Merck Millipore) and final concentration was estimated measuring the absorbance at 280 nm using the theoretical extinction coefficient (Table 4.2).

Table 4.2 Extinction coefficient (M^{-1} cm^{-1}) of Cst5 constructs based on amino acid sequence.

Protein	Extinction coefficient (M^{-1} cm^{-1}
Strep_Cst5_His	47,900
Strep_Cst5	46,410

The synthetic gene coding for *C. albicans* Hst7 (GenScript) (CUG codons mutated to a serine residue) was cloned using the Sequence and Ligation Independent Cloning (SLIC) technique. SLIC cloning is sequence independent and based on homologous recombination of vector and insert with 15–25 bp homologies on both DNA ends. The EMBL pETM-vectors were modified to obtain the vectors – pCoofy – with different tags followed by protease cleavage sites (Table 4.3) (Scholz *et al.* 2013). The HRV 3C protease cleavage site, located downstream of the N-terminal purification or solubility enhancing tag, can be used for tag removal, and the 3' end of the toxic ccdB gene, serves as primer binding sites common to all of the pCoofy vectors.

Table 4.3 pCoofy vectors with different N-terminal tags and respective primers used for cloning Hst7.

pCoofy	N-tag	fw primer	Rev primer
pCoofy1	His6	3C	ccdB
pCoofy18	His10	3C	ccdB
pCoofy2	Trx-His6	3C	ccdB
pCoofy4	His6-MBP	3C	ccdB
pCoofy16	His10-NusA	3C	ccdB
pCoofy5	His6-Sumo1	Sumo1	ccdB
pCoofy6	His6-Sumo3	Sumo3	ccdB

Specific primers based on the SLIC strategy were designed for PCR amplification of pCoofy vectors and Hst7. Primers were composed of Hst7 sequence plus an extension complementary to vectors primers LP1 forward (corresponds to the HRV 3C protease site) and LP2 reverse (located at the C-terminus of ccdB) (Table 4.4). The addition of adapter sequences at the end of each PCR product alters the desired insert sequence, important for cloning of multiple PCR fragments into a single vector. The SLIC reaction was carried out by recombinase enzyme RecA to

anneal vector and insert to obtain high levels of cloning efficiency followed the *E. coli* transformation and sequence analysis.

Table 4.4 Primers for vector amplification and complementary primer extensions for Hst7. Underlined sequences correspond to the sequence of primers LP1 and LP2.

Oligonucleotide	Sequence (5´ to 3´)
3C_Hst7_fw	CTGGAAGTTCTGTTCCAGGGGCCCATGGGCACCCGCACCAC CCGTA
Sumo1_Hst7_fw	GTCTACCAGGAACAAACCGGTGGAATGGGCACCCGCACCAC CCGTA
Sumo3_Hst7_fw	GTGTTCCAGCAGCAGACCGGTGGAATGGGCACCCGCACCAC CCGTA
ccdB_Hst7_rv	CCCCAGAACATCAGGTTAATGGCGTAAAACCTTGCACTTGCG GATTTTCTTCGC

The positive clones of pCoofy_Hst7 obtained from SLIC strategy, with N-terminal tags: His6, His10, Trx-His6, His10-NusA, His6-SUMO1 and His6-SUMO3 were tested in a small-scale expression. Soluble expression of Hst7 in fusion with these N-terminal tags was assessed in *E. coli* strains BL21(DE3), BL21(DE3)pLysS and Rosetta(DE3) induced with addition of 0.9 mM of IPTG and tested with different conditions such as temperature and incubation time: 25-30 ˚C during 3-6 hours or 20 ˚C overnight. After protein expression, cultures were centrifuged at 4000 g, 4 ˚C for 15 minutes and pellets ressuspended in lysis buffer (20 mM sodium phosphate pH 7.5, 500 mM NaCl, supplemented with 100 µg/ml lysozyme). Cells were lysed by sonication (Sonoplus HD2200 200 W sonicator, Bandelin Electronic) and then centrifuged at 15000 g (Eppendorf 5414R) for 15 minutes at 4 ˚C. Pellets (insoluble fractions) and the supernatants (soluble fractions) were analyzed by SDS-PAGE.

4.2.2 ESPRIT cloning and production of Cst5

The Esprit technology was used at the Institute for Structural Biology (Genoble, France) in collaboration with Philippe Mas and Darren Hart through a project financed by Instruct-ERIC (Structural Biology European Research Infrastructure). Before starting the experiments, an analysis with Phyre^2 online software (www.sbg.bio.ic.ac.uk/phyre2) was used to predict and identify domains of Cst5 protein and design a truncation library: for the 5' deletion libraries - N-terminal truncated – N-terminal full length (N-term fl) plus the position 10 aa downstream of

the modeled PH domain (N-term LPAP, Figure 4.3 purple box) were fixed; for 3' deletion library the C-terminal aa (C-term fl) plus position 10 aa upstream of the modeled ring domain (C-term FQTI, orange box) were fixed (Figure 4.3).

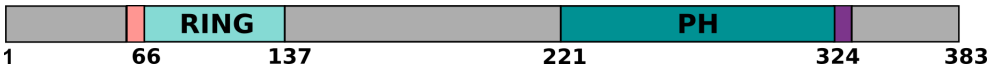


Figure 4.3 Schematic representation of Cst5 construct. The cyan box represents the ring domain 66-137 aa and teal box the PH domain 221-324 aa, predicted using the Phyre[2] software. To design the truncation library, 10 aa upstream the ring domain (pink box) as well as 10 aa downstream the PH domain (purple box) were marked.

ESPRIT technique was performed according to the procedure previously described in detail by Mas *et al.* (Mas *et al.* 2017). The synthetic gene of Cst5 (Genscript) was amplified by PCR using the designed primers (Table 4.5) and cloned into pESPRIT02 vector, in frame with N-terminal TEV-cleavable hexahistidine tag plus a C-terminal sequence encoding biotin acceptor peptide (BAP), using the restriction sites AscI/NsiI (N-terminal) or AatII/NotI (C-terminal) according to the exonuclease III/mung bean nuclease protocol (Figure 4.4). The obtained vectors containing Cst5 fragments with the size range of interest were sequenced prior to starting the ESPRIT screening strategy.

Table 4.5 Oligonucleotides used to design Cst5 library with N- and C- terminal truncations.

Oligonucleotide	Sequence (5´ to 3´)
Cst5 N-ter for	GGATCTTGGGCGCGCCTAAATGCTGCAATCAACGCCGAAATCCTG
Cst5 fl rev	GGATCTTGATGCATTCGGGTCCGAGGTCAGCAGAGCG
Cst5 LPAP rev	GGATCTTGATGCATTCGGTGCCGGCAGTTGCAGATGTG
Cst5 fl for	GGATCTTGGACGTCGAATGCTGCAATCAACGCCGAAATCCTG
Cst5 FQTI for	GGATCTTGGACGTCGATTTCAGACCATTAACCCGGGCCTGAATC
Cst5 rev	GGATCTTGGCGGCCGCTTACGGGTCCGAGGTCAGCAGAGCG

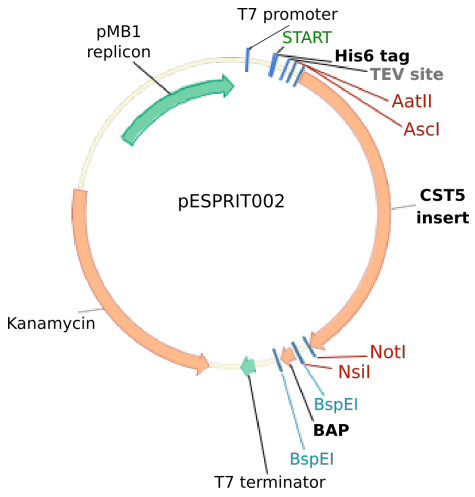


Figure 4.4 CST5 inserts were cloned into the pESPRIT02 vector. The truncation library was constructed using an exonuclease III/mung bean nuclease protocol. For 5' deletion libraries, inserts were subcloned in-frame with the sequence encoding the BAP using AscI and NsiI. For 3' deletion libraries, inserts were in-frame with the His6 and TEV encoding sequence using AatII and NotI.

The ESPRIT plasmids with Cst5 inserts were cut at the gene terminus to be truncated, adjacent to the His6 or BAP sequence, using the restriction enzymes AscI or NotI, that leave a 5' overhang (exonuclease III substrate) and AatII or NsiI resulting in 3' overhangs (exonuclease III resistant). The exonuclease truncation reaction was performed to generate the linearized pESPRIT02 plasmids containing truncated Cst5 inserts that were size selected by agarose gel electrophoresis and transformed in *E. coli*. After random Cst5 construct libraries clones were analyzed and validated.

The Cst5 constructs were transformed in arabinose-inducible *E. coli* BL21 strain (BL21 AI RIL) and titrated onto agar trays by an automated process using Kbiosystems K3 colony picker in 384 well plates. Afterwards, a colony array of protein expression was performed on nitrocellulose membranes probed with fluorescent streptavidin, to detect the biotinylation of the C-terminal BAP (green), an indicator of solubility, and a monoclonal antibody used against the N-terminal His6 tag (red). Clones exhibiting both N- and C-terminal tags fluorescence signal (double tag, yellow) were selected by colony blot analysis. The positive clones from the array-based screen were analyzed in a small scale through the expression and purification screening. After the detection of the soluble clones by electrophoresis and western blot, Cst5 constructs were selected to scale up (Mas *et al.* 2017).

4.2.3 Expression and Purification of protein from the Cst5 library

Cst5 constructs were overexpressed in *E. coli* strain BL21 AI RIL and grown at 37 ˚C until OD_{600} reached 0.6-0.8. The expression was induced at 25 ˚C overnight with 0.2 % (w/v) arabinose and 50 µM of biotin. Cells were harvested by centrifugation at 4 ˚C, 4000g for 25 minutes (Beckman Avanti J-26 XP) and ressuspended in lysis buffer (20 mM sodium phosphate pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, supplemented with 100 µg/ml lysozyme) and store at -20˚C.

Bacterial pellets were lysed on ice with agitation for 45-60 minutes in the presence of protease inhibitor cocktail and 20 µg/ml DNAse (supplemented with 1 mM $MgCl_2$). The cell lysate was clarified by centrifugation (45 minutes at 35,000 g; Beckman Avanti J-26 XP) and loaded onto a HisTrap HP column (GE Healthcare) equilibrated with binding buffer (20 mM sodium phosphate pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 5mM imidazole). Fractions were eluted by a step gradient with elution buffer (20 mM sodium phosphate pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 500 mM imidazole). Elution fractions were analyzed by SDS-PAGE and selected for further purification by SEC using a Hiprep 26/60 Sephacryl S-100 column (GE Healthcare) equilibrated with protein buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 1 mM EDTA, 1 mM DTT). Fractions from the preparative SEC were analyzed by SDS-PAGE and further concentrated on centrifugal ultrafiltration devices (10kDa cutoff) (Merck Millipore). Protein concentration of each purified Cst5 construct was estimated by measuring the absorbance at 280nm using the corresponding extinction coefficient (Table 4.6).

Table 4.6 Extinction coefficient of Cst5 constructs based on amino acid sequence.

Cst5 construct	Extinction coefficient (M^{-1} cm^{-1})
Cst5_ΔN71 Cst5_ΔN74	42,400
Cst5_ΔN163 Cst5_ΔN171 Cst5_ΔN176 Cst5_ΔN183 Cst5_ΔN184 Cst5_ΔN189 Cst5_ΔN205 Cst5_ΔN207	35,410
Cst5_ΔN211	33,920
Cst5_ΔC109	13,980
Cst5_ΔC127 Cst5_ΔC131 Cst5_ΔC152 Cst5_ΔC209	15,470

4.2.4 Biophysical characterization of Cst5 constructs

Biophysical characterization of Cst5 by size dynamic light scattering (DLS) and differential scanning fluorimetry (DSF) was performed according to protocol previously described for Cek1 protein (detailed in Chapter 2 section 2.2).

The purity and oligomeric state of the Cst5 constructs were analyzed by analytical SEC on a Superdex 75 10/300 GL column (GE Healthcare) in protein buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 1 mM EDTA, 1 mM DTT). Previously, the column was calibrated using a gel filtration calibration kit containing highly purified and well-characterized protein standards (GE Healthcare): blue dextran 2000 (used for determination of the void volume - Vo) and a mixture of the following proteins, conalbumin (75 kDa), ovalbumin (44 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). The calibration curve was prepared by measuring the elution volumes (Ve) of the standard proteins, calculating their corresponding Kav values and plotting them versus the logarithm of their respective molecular weight (MW). The Kav parameter was determined according to the equation: Kav=(Ve-V0)/(Vc-V0), where Vc represents the total bed volume.

4.2.5 Crystallization of Cst5

Crystallization conditions of Cst5 constructs (Cst5_ΔN176 and Cst5_ΔN205) were screened at 20 ˚C using sitting-drop vapour-diffusion method with commercial crystallization screens: Index, PEG/Ion 2 (Hampton research), Proplex and Morpheus (Molecular dimensions). Experiments were set up in 96-well using an Oryx4 protein crystallization robot (Douglas Instruments). The drops consisted of equal volumes (100 ηl) of Cst5 (13 mg/ml of Cst5_ΔN176 and 5 mg/ml of Cst5_ΔN205) and crystallization solution, equilibrated against 200 µl of reservoir solution.

4.2.6 Bio-Layer Interferometry

The binding assays were performed by Dr. Carla Almeida (X-PROT, Biocant-Cantanhede) on an Octet Red96 biolayer interferometry (BLI) system (FortéBio Inc) using Streptavidin A coated biosensors. The assays were performed at 30 ˚C, in a total reaction volume of 200 µL and using 96-well plates (Greiner Bio). For the Cts5 binding both variants of ΔNCek1 were tested. The biotinylated Cts5 peptide (ΔN184) was loaded at 10 µg/mL, in 1x kinetic buffer (PBS pH 7.4, 0.1% BSA, 0.02% Tween 20) onto seven biosensors and then incubated with different concentrations of ΔNCek1_Ser or ΔNCek1_Leu. The analyte was diluted also in the same buffer and tested for different concentration in a range from 0 ηM to 5.60 µM, by half serial dilutions. The association step was carried out over 400 seconds while the dissociation step was followed for 600 seconds. The Data Analysis Software (version 9.0 –Forte Bio) was used to process and analyze the data using a buffer blank as reference and a 1:1 binding interaction model with global fitting.

4.3 Results

4.3.1 Sequence and structure analysis of the scaffold protein Cst5

The Cst5 sequence was analyzed with online software Phyre2, one of the most widely used methods for protein structure prediction based on the comparison of a protein sequence of interest with a large database of structurally characterized sequences (Kelley *et al.* 2015). The Cst5 secondary structure was predicted with a high percentage of disorder (51%) and a prediction of 21% and 16% for α-helix and β–strand, respectively. The Phyre2 displays the predicted secondary structure of Cst5, with 201 amino acids (52%) of Cst5 modeled at >90% confidence

using multiple templates. More than one solution of the secondary structure models, which align in two regions of the Cst5 sequence, was obtained. Based on the Phyre2 criteria, the first two templates were selected in which Cst5 protein regions have been modeled, a RING domain (aa 66-137, 18% identity to template sequence, 97.5% confidence) and a PH domain (aa 221-324, 12% identity to template sequence, 97.6% confidence) (Figure 4.5). To complement the analysis, the structural homology models of Cst5 domains were obtained using the online server Swiss-Model (Arnold *et al.* 2006). The 3D structure of the RING (aa 77-117, 20.9% identity) and PH domains (aa 220-331, 13.5% identity) were determined using as a template the isolated RING domain structure (PDB 2MT5) and the PH region of the Rho guanine nucleotide exchange factor 9 (PDB 4MT7) (Figure 4.5).

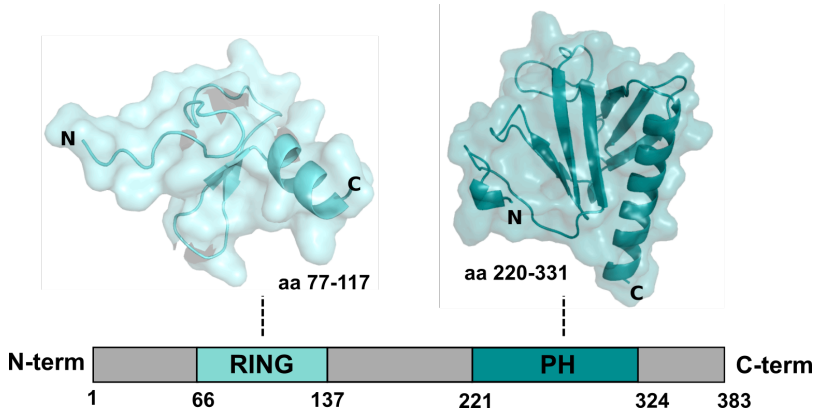


Figure 4.5 Cst5 structure prediction by homology modeling. Schematic representation of Cst5 sequence with the RING domain (cyan blue) and the PH domain (teal) and three-dimensional homology models of these regions.

4.3.2 Cloning and expression of scaffold protein Cst5 and MAPKK Hst7

The first approach to express Cst5 was achieved using the expression vector pPR-IBA2 (Strep_Cst5_His). The expression patterns were analyzed by SDS-PAGE, but Cst5 expression was only observed in some conditions and found exclusively in inclusion bodies without visible expression in the soluble fraction. In order to overcome this issue and improve the Cst5 solubility, some strategies were tested, such as the addition of L-arginine in the Luria-Bertani (LB) medium to *E. coli* cells during expression (Tsumoto *et al.* 2004), although they were not efficient to enhance the protein solubility. Since the presence of a His6 tag at the C-terminal was shown to have an effect on protein stability (Cereija *et al.* 2019), a new

construct (Strep_Cst5) in which the His6 tag was removed was used as an alternative to test Cst5 expression with same conditions described above. However, no significant recovery of Cst5 in the soluble fraction was observed. Nevertheless, Cst5 was expressed using two conditions, namely BL21(DE3)pLysS *E. coli* cells at 25 ˚C, after 4 hours induction of protein expression (Strep_Cst5_His) and Rosetta(DE3) after an overnight induction at 20 ˚C. However, the amount of Cst5 protein recovered after purification was very low and not enough to be used in further biochemical and biophysical experiments.

The pCoofy expression vectors (His6, His10, Trx-His6, His10-NusA, His6-SUMO1 and His6-SUMO3) were successfully cloned with Hst7 by SLIC technology. These Hst7 constructs were used to test protein expression in different conditions. The expression test patterns showed that the soluble expression of Hst7 was obtained with His10 tag (MW 68 kDa) in *E. coli* Rosetta(DE3) at 25 ˚C after induction of protein expression for 4 hours (Figure 4.6). This expression condition was selected for future large scale Hst7 production.

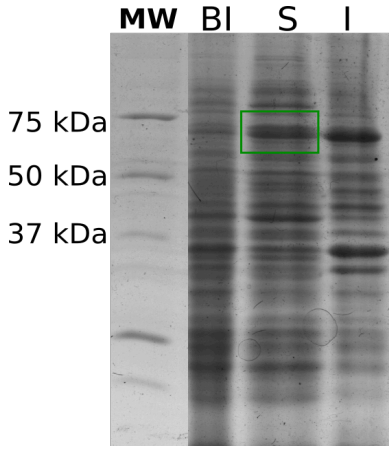


Figure 4.6 SDS-PAGE analysis of Hst7 expression. His10_Hst7 (68kDa) expression pattern in *E. coli* Rosetta(DE3) cells at 25 ˚C during 4 hour (green box). MW – molecular weight, BI – sample before induction, S- soluble fraction, I- insoluble fraction.

4.3.3 Identification of soluble Cst5 constructs using the ESPRIT system

Initial trials described above, using two constructs of the scaffold protein Cst5 in different expression conditions, yielded no soluble forms of the Cst5 protein. Therefore, the ESPRIT technique, a random incremental truncation library approach was used to produce a large library of random genetic constructs of the

Cst5 and screen them to identify soluble expression. According to sequence and structural analysis, N- and C-terminally truncated variants were defined and selected for the exonuclease III and mung bean nuclease truncation protocol to generate a random library of Cst5 (Mas *et al.* 2017). The N-terminal truncated library was performed fixing the C-terminal amino acid plus 10 amino acids downstream the modeled PH domain (Figure 4.3). The same strategy was followed to the C-terminal truncated variants, where the N-terminal amino acid plus 10 amino acids upstream the RING domain were fixed (Figure 4.3). The protocol generates a random library of Cst5 constructs cloned in-frame with N-terminal cleavable His6 tag and C-terminal BAP, which is used as a marker for protein solubility and stability during colony screening.

For each library, Cst5 fragments were selected by expected size on agarose gel and colony PCR analysis showed the insert lengths divided in two pools - 5' and 3' deletions. In total, 18.432 randomly truncated clones (6000 for each sub library) were picked into 384-well plates and screened by construction of colony arrays. The solubility screening was employed in three steps: the robotic spotting of clones on nitrocellulose membrane, growing colonies and inducing protein expression by shifting membranes to agar plates containing IPTG. Putative soluble expression clones were analyzed on the nitrocellulose membrane hybridized with fluorescent probes. The detection of biotinylation of a C-terminally fused BAP together with the presence of the N-terminal His6 tag, suggests the construct being expressed by that truncation mutant is both soluble and intact. The 95 Cst5 clones exhibiting the strongest signal and possessing both BAP and His6 tags were selected to proceed with small-scale protein expression tests, Ni-NTA purifications, SDS-PAGE and western blot.

Sixteen constructs (11 for N-terminal and 5 for C-terminal truncations library) with an in frame Cs5 sequence showed a diversity of sizes and the encoded protein fragments were mapped onto the Cst5 primary sequence (Figure 4.7A). Sizes of the Cst5 clones were identified after DNA sequencing of the inserts (including tags). The clones were further selected based on the ability to express soluble protein in small-scale approaches and sequenced to identify construct boundaries that covered structurally uncharacterized regions of the Cst5 protein (Figure 4.7A). Cst5 constructs from the N-terminal truncation library range from 24 to 40 kDa, however only constructs that do not contain the RING domain were expressed in the soluble fraction (ΔN163, ΔN171, ΔN176, ΔN183, ΔN183, ΔN189, ΔN205, ΔN207, ΔN211) (Figure 4.7B). The remaining N-terminal (ΔN71, ΔN74) and C- terminal (ΔC109,

ΔC127, ΔC131, ΔC209) Cst5 constructs containing the RING domain were not selected for further studies due to lack of expression (Figure 4.7B). All of the nine Cst5 constructs expressed in the soluble fraction were selected from the N-terminal truncation library and subsequently expressed and purified in large scale cultures (Figure 4.7B).

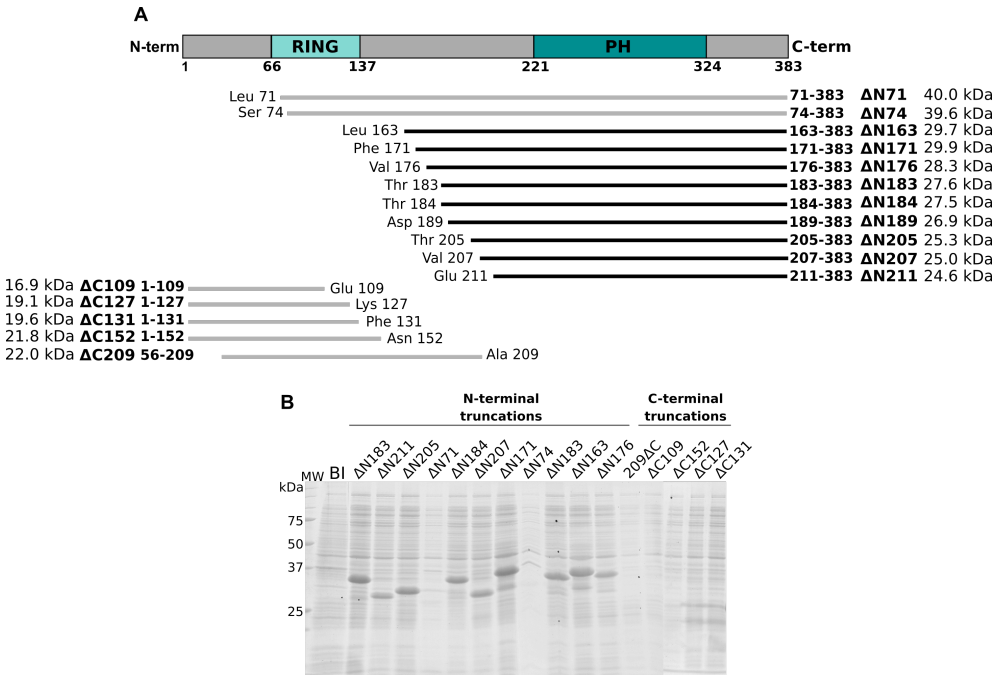


Figure 4.7 Screening of Cst5 library. (A). Schematic representation of the Cst5 protein sequence with RING and PH domains (above). N- and C-terminal truncation libraries with corresponding theoretical molecular weight of Cst5 fragments are shown below. Cst5 fragments selected for *E. coli* expression (black lines) represent Cst5 constructs that are expressed in the soluble fraction. (B) 12.5% SDS-PAGE analysis of small-scale expression of Cst5 constructs. Lanes BI – before induction and expression of Cst5 after induction at 25 ˚C overnight with 0.2 % (w/v) arabinose and 50 μM of biotin.

4.3.4 Biochemical and biophysical characterization of Cst5 truncation constructs

N-terminally truncated Cst5 constructs exhibiting detectable soluble protein expression in small-scale experiments were further purified with yields between 2 mg and 20 mg of purified protein per liter of *E. coli* culture (Table 4.7).

Table 4.7 Yield of Cst5 construct expressed and purified.

Cst5_ ΔN	163	171	176	183	184	189	205	207	211
Yield (mg/L)	24	6	13	16	15	13	5	2	4

All the Cst5 constructs were analyzed by SDS-PAGE (Figure 4.8A) and eluted earlier than expected for monomeric proteins, considering their theoretical molecular weights. Three Cst5 constructs (ΔN205, ΔN207, ΔN211) were purified as a single band above 25kDa on SDS-PAGE gel (Figure 4.8A). The six larger constructs with theoretical molecular weight around 27-30kDa (ΔN163, ΔN171, ΔN176, ΔN183, ΔN184, ΔN189) showed a single band below 37kDa on the SDS-PAGE gel (Figure 4.8A). The characterization of the Cst5 constructs by analytical size exclusion (SEC) chromatography showed that all Cst5 proteins eluted in a volume corresponding to proteins with a molecular weight higher than expected for monomeric proteins of spherical shape (Figure 4.8B-D). These results were confirmed with the estimated molecular weight calculated from the calibration curve revealing an apparent molecular weight compatible with dimeric assemblies (Figure 4.9).

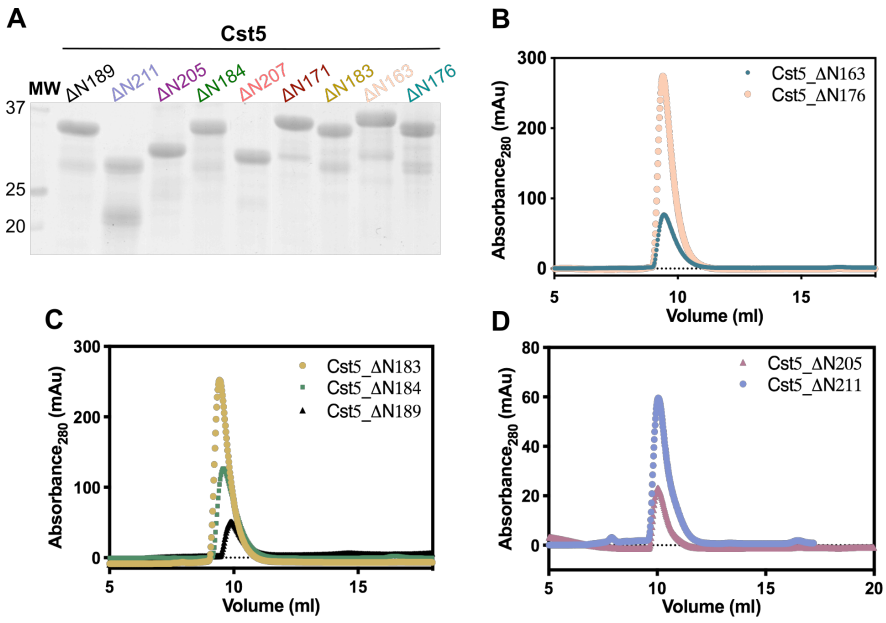
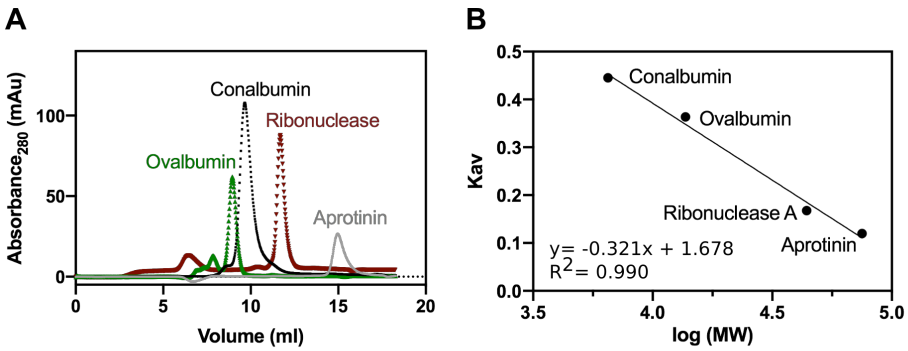


Figure 4.8 Analysis of N-terminal truncated Cst5 constructs identified by screening of the truncation library. (A) 12.5% SDS-PAGE gel analysis of purified Cst5 proteins. (B-D) SEC profile of the Cst5 constructs using a Superdex 75 10/300 GL column.

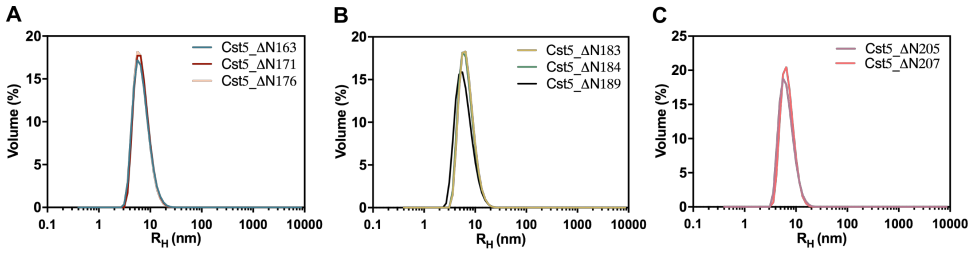


C

Cst5	ΔN163	ΔN176	ΔN183	ΔN184	ΔN189	ΔN205	ΔN211
Estimated MW (kDa)	77.3	79.0	78.3	79.0	64.1	60.3	59.6
Theoretical MW (kDa)	29.7	28.3	27.6	27.5	26.9	25.3	24.6

Figure 4.9 Calibration of analytical SEC. Superdex 75 10/300 GL column. (A) Elution profile of the standard molecular weight (MW) markers: conalbumin (75 kDa; Ve=9.7 ml), ovalbumin (44 kDa, Ve=10.4 ml), ribonuclease A (13.7 kDa, Ve=13.6 ml), aprotinin (6.5 kDa, Ve=15.0 ml) in a Superdex 75 10/300 GL equilibrated in 20 mM HEPES pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 1 mM EDTA, 1 mM DTT. Ve correspond to elution volume. (B) Calibration curve of Superdex 75 10/300 GL. (C) Estimated MW of Cst5 truncation constructs by analytical SEC.

The Cst5 constructs were analyzed by DLS, a technique to assess the homogeneity of proteins that also provides hydrodynamic radius (R_H) values to estimate the molecular weight (Figure 4.10A-C). Data showed that most Cst5 constructs were monodisperse (polydispersity (Pd) ~20%) with an estimated molecular weight higher than expected for globular monomeric proteins (Figure 4.7A, 4.10D). Altogether, these results suggest that in these conditions Cst5 constructs could be homodimers. Although, Cst5_ΔN205 and ΔN207 display multiple visible peaks (distribution by intensity, not shown) with a Pd value around 30%, suggesting the presence of high molecular weight aggregated species.



D

Cst5	ΔN163	ΔN171	ΔN176	ΔN183	ΔN184	ΔN189	ΔN205	ΔN207
Pd (%)	21.5	21.6	22.5	19.1	21.2	18.5	32.2	29.4
R_H (nm)	3.53	3.58	3.50	3.56	3.50	3.19	3.43	3.52
Estimated MW (kDa)	64.6	66.6	63.3	65.6	63.2	50.9	60.2	64.1
Theoretical MW (kDa)	29.7	29.9	28.3	27.6	27.5	26.9	25.3	24.6

Figure 4.10 DLS of N-terminal truncation constructs of Cst5. (A) DLS of Cst5 ΔN163, ΔN171, ΔN176 (B) ΔN183, ΔN184, ΔN189 (C) ΔN205, ΔN207. (D) Data obtained from DLS for Cts5 constructs: Pd, R_H and the estimated MW calculated with Zetasizer software (Zetasizer v7.12, Malvern Instruments).

The melting temperatures of Cst5 constructs were determined by DSF (Figure 4.11). The analysis of melting temperature (Tm) revealed no major differences between Cst5 constructs, and the deletions between amino acids Leu163 and Val207 did not interfere with overall thermal stability of Cst5. The Cst5_ΔN207 presented the lowest Tm, however based on DLS data, this construct was not a homogenous sample, which could interfere with DSF measurements.

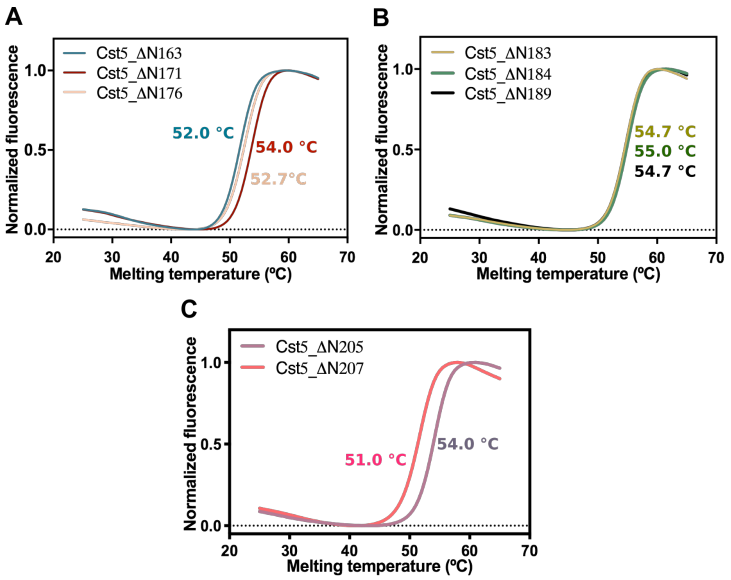


Figure 4.11 Representative thermal denaturation curves measured by DSF. (A) Cst5 ΔN163 (52.0 ˚C), ΔN171 (54.0 ˚C), ΔN176 (52.7 ˚C). (B) ΔN183 (54.7 ˚C), ΔN184 (55.0 ˚C), ΔN189 (54.7 ˚C). (C) ΔN205 (54.0 ˚C), ΔN207 (51.0 ˚C).

4.3.5 Interaction of Cek1 variants with Cst5

The BLI was used to study the molecular interaction between Cek1 and the scaffold protein Cst5 (Figure 4.12). In order to evaluate if the effect of CUG ambiguity could affect this interaction, BLI was tested with both variants of Cek1 active domain (ΔN-Cek1_Ser and ΔN-Cek1_Leu) and Cst5_ΔN184 construct. Although interaction was observed for both Cek1 variants, good quality binding data could only be obtained for the ΔN-Cek1_Ser variant, with an affinity in the low micromolar range (K_D=1.7×10^{-7} ± 1.7 ×10^{-9} M).

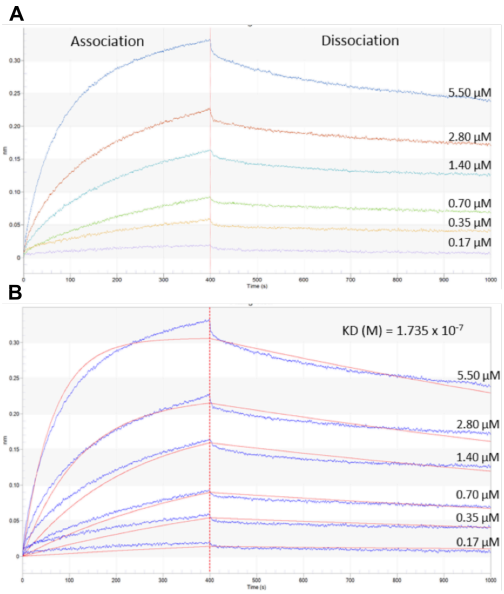


Figure 4.12 Interferometry analysis of the interaction between Cst5 and ΔN-Cek1_Ser protein. (A) The biotinylated Cst5 was bound to streptavidin biosensors and incubated with several dilutions of ΔN-Cek1_Ser 0.17 µM (purple), 0.35 µM (yellow), 0.70 µM (green), 1.40 µM (light blue), 2.80 µM (red), 5.60 µM (dark blue). The real time binding response (nm) is plotted against time for different concentrations of ΔN-Cek1_Ser. (B) Curve fitting for the determination of the kinetic parameters. The data were analyzed using the Data Analysis Software (version 9.0 –Forte Bio) and a 1:1 binding interaction model with global fitting.

4.4 Discussion

The identification of PPI in *C. albicans* is a challenging task, mostly due to the alternative CUG codon translation that hinders the study of large protein interaction networks. We previously demonstrated that the reassignment of CUG codon from leucine to serine in *C. albicans* modulates the stability and enzymatic activity of Cek1 (Fraga *et al.* 2019). Previous data suggested the rewiring of the

protein interaction networks regulating the signal transduction across the MAPK cascade centered on Cek1 (Fraga *et al.* 2019), but the role of ambiguous translation of the CUG codon in finely tuning those PPIs, remains undisclosed. In this regard, we aimed to characterize the cross-talk between Cek1, the scaffold Cst5 and the MAPKK Ste7 and to study the impact of the CUG-encoded for the regulation of this signaling cascade.

Here, we focused on the characterization of the Cst5 scaffold protein that has been demonstrated to be able to assemble multiprotein complexes in the Cek1 pathway (Cote *et al.* 2011, Yi *et al.* 2011). The sequence and structural analysis of Cst5 predicted a RING domain at the Cst5 N-terminal region, with a typical zinc-finger motif (residues 66-137) plus a PH domain (residues 221-324). The library-based ESPRIT method (Mas *et al.* 2017) was successfully applied to the Cst5 target protein, allowing the production of Cst5 soluble fragments. These results showed that Cst5 N-terminal truncations, which contain the PH domain, were highly soluble and selected for further biophysical characterization.

In contrast, all constructs that included the RING domain failed to produce soluble Cst5 fragments. The RING domain is characterized by a cysteine-rich motif (zinc-finger motif), which forms a specific tertiary structure stabilized by the coordination of the metal ion (Deshaies *et al.* 2009, Cote *et al.* 2011). Thus, to overcome the remarkably lower expression yield of the Cst5 constructs containing RING domain, additional zinc enrichment of growth media could be used to improve solubility and maintain a stable and well-folded structure. Moreover, semi-conserved residues within RING domain are also described to be crucial to form the domain's hydrophobic core (Deshaies *et al.* 2009), and structural changes due to the lack of zinc coordination might expose those residues leading to insolubility and aggregation.

Pure and soluble Cst5 domain fragments were obtained covering a window between residue Leu163 and Glu211 to the C-terminal residue Pro 383. Our results showed that Cst5 constructs display differences at the N-terminus position, which does not influence the overall protein thermal stability. These fragments include part of the RING-PH interdomain linker and the PH domain, a region that was described to be involved in the specific interaction with *C. albicans* Ste11 in Y2H system (Cote *et al.* 2011). Moreover, SEC and DLS revealed that Cst5 constructs show an apparent molecular weight consistent with a dimeric state. Although all the Cst5 constructs analyzed lack the RING domain, previous data showed that the second zinc-finger of the Cst5 RING domain has a central role in Cst5

homodimerization (Cote *et al.* 2011). Further experiments will be required to unambiguously confirm the oligomeric state of the purified Cst5 fragments.

As mentioned before, *C. albicans* scaffold protein Cst5 is strikingly different from *S. cerevisiae* Ste5 (Cote *et al.* 2011). The lack of the vWA domain, that in *S. cerevisiae* binds Ste7 and facilitates Fus3 phosphorylation, suggest a distinct regulation in a vWA domain-independent manner of the Cek1 MAPK signaling pathway. Moreover, the presence of several CUG-encoded residues within the Cst5 PH domain, suggest that the identity of the CUG residues might further modulate this interaction network.

Our results revealed that serine variant of the active domain of Cek1 (ΔN-Cek1_Ser) interacts with the Cst5_ΔN184 fragment. A previous study using the Y2H system, where all CUG codons are translated as leucines, mapped the interaction with Cst5 to the N-terminal region of Cek1 (Cote *et al.* 2011). However, this region (residues 1 to 85), enriched in amino acids repeats, is missing in the Cek1 active domain. Further studies are needed to deeply understand the molecular interaction between Cek1 and Cst5, using for example, the full-length Cek1 protein, which contain this N-terminal tail, other Cst5 constructs and serine or leucine variant of these proteins. Preliminary data showed that the leucine variant of the Cek1 active domain (ΔN-Cek1_Leu) also binds to the Cst5_ΔN184 (data not shown), but the experimental data could not be fitted applying the same binding model used for ΔN-Cek1_Ser. Although those results suggesting a distinct behavior of the ΔN-Cek1_Leu and a possible impact of the CUG residue identity on this protein interaction, further experiments will be required to confirm those results. We previously demonstrated that a leucine at the CUG position had a negative impact on the Cek1 auto-phosphorylation, since no phosphorylation at the conserved tyrosine of the T×Y motif was detected in the ΔN-Cek1_Leu variant (Fraga *et al.* 2019). Whether the differences in the interaction between Cek1 and Cst5 result directly from the presence of the leucine residues at the CUG position or from the absence of a phosphorylated tyrosine residue, still needs to be determined.

As mentioned above, future work will be focused on studying the interaction between MAPK Cek1 and other Cst5 constructs as well as with other crucial proteins involved in this cascade, in particular, Hst7. Furthermore, the CUG-encoded residues present in the Cst5 (3 CUGs) and Hst7 (4 CUGs), expressed as serines will be mutated to leucines to evaluate if a serine or leucine insertion at the CUG position could modulate protein interactions.

In summary, we have successfully produced different constructs of *C. albicans* scaffold protein Cst5 and assessed the interaction of Cst5_ΔN184 with the active domain of Cek1 MAPK. The molecular tools and findings presented in this chapter denote the first steps towards the understanding of the protein interaction networks regulating *C. albicans* Cek1 signaling pathway and provide novel information with implications for better understanding *C. albicans* virulence attributes.

4.5 References

Arnold, K., L. Bordoli, *et al.* (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.*Bioinformatics* 22: 195-201.

Bhattacharyya, R. P., A. Remenyi, *et al.* (2006). The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway.*Science* 311: 822-6.

Blackwell, C. and J. D. Brown (2009). The application of tandem-affinity purification to Candida albicans.*Methods Mol Biol* 499: 133-48.

Bond, C. S. and A. W. Schuttelkopf (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments.*Acta Crystallogr D Biol Crystallogr* 65: 510-2.

Boysen, J. H., S. Fanning, *et al.* (2009). Detection of Protein–Protein Interactions Through Vesicle Targeting.*Genetics* 182: 33-39.

Cereija, T. B., S. Alarico, *et al.* (2019). The structural characterization of a glucosylglycerate hydrolase provides insights into the molecular mechanism of mycobacterial recovery from nitrogen starvation.*IUCrJ* 6: 572-585.

Cote, P., T. Sulea, *et al.* (2011). Evolutionary reshaping of fungal mating pathway scaffold proteins.*mBio* 2: e00230-10.

Das, S., R. Bhuyan, *et al.* (2019). Network analysis of hyphae forming proteins inCandida albicansidentifiesimportant proteins responsible for pathovirulence in the organism.*Heliyon* 5.

Deshaies, R. J. and C. A. Joazeiro (2009). RING domain E3 ubiquitin ligases.*Annu Rev Biochem* 78: 399-434.

Fraga, J. S., Z. Sarkany, *et al.* (2019). Genetic code ambiguity modulates the activity of a C. albicans MAP kinase linked to cell wall remodeling.*Biochim Biophys Acta Proteins Proteom* 1867: 654-661.

Garrenton, L. S., S. L. Young, *et al.* (2006). Function of the MAPK scaffold protein, Ste5, requires a cryptic PH domain.*Genes Dev* 20: 1946-58.

Good, M., G. Tang, *et al.* (2009). The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation.*Cell* 136: 1085-97.

Inouye, C., N. Dhillon, *et al.* (1997). Ste5 RING-H2 domain: role in Ste4-promoted oligomerization for yeast pheromone signaling.*Science* 278: 103-6.

Kaneko, A., T. Umeyama, *et al.* (2004). Tandem affinity purification of theCandidaalbicansseptin protein complex.*Yeast* 21: 1025-1033.

Kelley, L. A., S. Mezulis, *et al.* (2015). The Phyre2 web portal for protein modeling, prediction and analysis.*Nat Protoc* 10: 845-58.

Mas, P. J. and D. J. Hart (2017). ESPRIT: A Method for Defining Soluble Expression Constructs in Poorly Understood Gene Sequences.*Methods Mol Biol* 1586: 45-63.

Palzer, S., Y. Bantel, *et al.* (2013). An Expanded Genetic Code in Candida albicans To Study ProteinProtein Interactions In Vivo.*Eukaryot Cell* 12: 816-827.

Ratushny, V. and E. A. Golemis (2008). Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system.*Biotechniques* 44: 655-662.

Remenyi, A., M. C. Good, *et al.* (2005). The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network.*Mol Cell* 20: 951-62.

Schoeters, F., C. A. Munro, *et al.* (2018). A High-Throughput Candida albicans Two-Hybrid System.*mSphere* 3.

Schoeters, F. and P. Van Dijck (2019). Protein-Protein Interactions in Candida albicans.*Frontiers in Microbiology* 10.

Scholz, J., H. Besir, *et al.* (2013). A new method to customize protein expression vectors for fast, efficient and background free parallel cloning.*BMC Biotechnol* 13: 12.

Stynen, B., P. Van Dijck, *et al.* (2010). A CUG codon adapted two-hybrid system for the pathogenic fungus Candida albicans.*Nucleic Acids Res* 38: e184.

Subotic, A., E. Swinnen, *et al.* (2017). A Bimolecular Fluorescence Complementation Tool for Identification of Protein-Protein Interactions in Candida albicans.*G3 (Bethesda)* 7: 3509-3520.

Tsumoto, K., M. Umetsu, *et al.* (2004). Role of arginine in protein refolding, solubilization, and purification.*Biotechnol Prog* 20: 1301-8.

Westermarck, J., J. Ivaska, *et al.* (2013). Identification of protein interactions involved in cellular signaling.*Mol Cell Proteomics* 12: 1752-63.

Yi, S., N. Sahni, *et al.* (2011). Utilization of the mating scaffold protein in the evolution of a new signal transduction pathway for biofilm development.*MBio* 2: e00237-10.

Chapter 5

General discussion

5.1 General discussion

C. albicans is an opportunistic fungal pathogen that is able to adapt and survive in different microenvironments, using its morphological and phenotype switching, as well as the ability to adhere to host tissues. Such adaptation capability is highly coordinated by signaling pathways, such as the mitogen activated protein kinases (MAPKs), which mediate these common virulence features and enable *C. albicans* infections (Alonso-Monge *et al.* 2009). In response to adverse environmental conditions, *C. albicans* ambiguously translates the universal leucine CUG codon predominantly as serine (97%), but also as leucine (3%). *C. albicans* was shown to be well adapted and tolerant to genetic code ambiguity: increased levels of leucine incorporation at CUG position under stress conditions had no visible effect on the growth phenotype but had a significant impact on *C. albicans* morphogenesis, adhesion and hyphal growth, underlining a correlation between mistranslation and virulence (Miranda *et al.* 2007). The Cek1 MAPK pathway, which includes several proteins containing these CUG-encoded residues, is tightly associated with known *C. albicans* virulence traits such as morphogenesis, hyphal formation and cell wall structure (Alonso-Monge *et al.* 2006, Ernst *et al.* 2011). Hence, the main goal of this work was to unveil the structural and functional effects of the CUG residue identity in regulatory proteins of the Cek1-mediated pathway, a key to understand how this unusual codon translation mechanism can modulate the *C. albicans* pathogenicity.

5.1.1 Impact of CUG residue identity in *C. albicans* Cek1

Given the differences in the biochemical properties of the two amino acids, changing the identity of the CUG codon from leucine, a hydrophobic amino acid to a polar serine is expected to affect protein function. We hypothesized that the unique CUG-encoded residue in Cek1 could modulate the enzymatic activity, and consequently the flow of information across this signaling cascade. *C. albicans* Cek1 shares 63% sequence identity to *S. cerevisiae* Fus3, a MAPK of the yeast mating pathway. Based on Fus3 structure, the CUG-encoded residue of Cek1, structurally equivalent to Fus3 Ser141, is located at the ATP binding pocket and stabilizes a magnesium-bound water molecule.

Our results showed that the identity of the CUG residue impacts on the stability and kinase activity of both active domain (ΔN-Cek1) and the full-length Cek1 protein. The insertion of a serine at the CUG position stabilized the protein (higher melting temperature) and increased the enzymatic activity compared to leucine variant. Moreover, the autophosphorylation of Cek1 occurred exclusively in the serine variant of the active domain of Cek1 (ΔN-Cek1_Ser) at the tyrosine residue of the conserved ^{231}TEY^{233} motif, located in the kinase activation loop. In contrast, in the full-length Cek1 protein, the autophosphorylation of the Tyr233 residue was also observed in the leucine variant (full-length Cek1_Leu).

In summary, we showed that leucine insertion at the CUG site interferes with Cek1 autophosphorylation, but the biochemical differences observed *in vitro* between the serine and leucine variants are subtle, although they might have an impact on *C. albicans* virulence. This adds to previous reports showing that ambiguous translation of the CUG codon had an impact on the activity of *C. albicans* proteins (Rocha *et al.* 2001, Zhou *et al.* 2013), but the exact mechanism underlying the variation in autophosphorylation activity of Cek1 serine and leucine variants is still undetermined.

S. cerevisiae Fus3 is activated through a canonical mechanism by the upstream kinases but it was shown to be autophosphorylated during expression in *E. coli*, and the presence of the scaffold protein Ste5 induces Fus3 autophosphorylation on the tyrosine residue of the activation loop (mono-phosphorylated Fus3) (Bhattacharyya *et al.* 2006). The comparison of the unphosphorylated and mono-phosphorylated structures of MAPK Fus3 reveals no major structural differences apart from a more flexible loop in the mono-phosphorylated state, which facilitates the access of substrates to the active site. In the unphosphorylated Fus3, the activation loop is folded into the catalytic cleft, blocking the binding of ATP and substrate peptides (Bhattacharyya *et al.* 2006, Good *et al.* 2009). The data here presented showed that Cek1 autophosphorylation activity was higher when a serine was inserted at the CUG site. Furthermore, current experimental evidence suggests that autophosphorylation occurs in *cis*. This serine at the CUG site is structurally equivalent to the serine that in Fus3 binds a water molecule within the coordination sphere of the active site magnesium ion, that stabilizes the bound nucleotide. In *C. albicans* Cek1, the presence of the serine at the CUG site might position the nucleotide and the tyrosine in the activation loop in a conformation to serve as an intramolecular phosphoacceptor. A similar *cis*-autophosphorylation intramolecular reaction has been demonstrated for Fus3 MAPK

in the presence of the Ste5, which induce the Fus3 autophosphorylation of Tyr182 of the activation loop, increasing Fus3 kinase activity (Bhattacharyya *et al.* 2006).

5.1.2 Cek1 in the context of the MAPK signaling cascade

Structural differences between the scafoold protein Ste5 from *S. cerevisiae* and Cst5 from *C. albicans*, such as the disappearance of the vWA domain from Cst5, suggest distinct regulatory mechanism in *C. albicans* Cek1 signaling pathway (Cote *et al.* 2011). Indeed, the Ste5 vWA domain binds tightly to the MAPKK Ste7 and weakly to the MAPK Fus3 (Good *et al.* 2009). Additionally, Ste7 also interacts with Fus3, trough docking interaction motifs well conserved in *C. albicans* Hst7 (Remenyi *et al.* 2005, Bhattacharyya *et al.* 2006, Good *et al.* 2009). This interaction promotes tethering of Ste5-Fus3, inducing allosteric Fus3 phosphorylation by Ste7 (Bhattacharyya *et al.* 2006). Together, these observations and the lack of vWA domain in *C. albicans* Cst5, indicate an alternative mechanism for MAPK Cek1 phosphorylation/activation and for the overall regulation of the Cek1-mediated pathway.

Although some studies have identified protein interaction networks in *C. albicans*, the alternative CUG codon usage in this human pathogen presents significant technical challenges. The classic yeast-two-hybrid (Y2H) assay in *S. cerevisiae*, identified the interactions between the scaffold protein Cst5 with the MAP kinase module, Ste11, Hst7 and Cek1 (Stynen *et al.* 2010, Cote *et al.* 2011). Moreover, an adapted yeast two-hybrid system in *C. albicans* (C2H) enabled the identification of the Cek1 interaction with upstream kinase Hst7, which was not observed by Y2H assay in *S. cerevisiae*. Since all the proteins, except Ste11, forming the central protein-protein interaction (PPI) network of this cascade have CUG residues, we speculate that CUG residue identity could rewire protein interactions in this signal transduction pathway.

We aimed to study the interaction between Cek1 and the other components of this multi-enzymatic cascade: the upstream MAPKK Hst7 and the scaffold protein Cst5, which also contain CUG residues (Hst7 – 4 CUGs and Cst5 – 3 CUGs). The results revealed that ΔN-Cek1_Ser variant interacted with the scaffold protein Cst5 (Cst5_ΔN184), lacking the N-terminal RING domain. Interestingly, the Cst5_ΔN184 construct contains a region (Fus3 binding domain), with low amino acid sequence conservation, that in *S. cerevisiae* Ste5 binds to Fus3 and stimulates its autophosphorylation on Tyr182 (Bhattacharyya *et al.* 2006). This Ste5-Fus3 binding

domain does not play a significant role in the cascade activation but rather downregulates the signaling information by triggering a feedback phosphorylation of Ste5 by Fus3. Moreover, the interaction between the ΔN-Cek1_Leu and the Cst5_ΔN184 construct showed that the CUG-encoded residue of MAPK Cek1 impacts on Cst5 binding. These results established the framework for future research aiming to delineate the interaction sites between Cek1, Ste7 and Cst5 and to understand the modulatory role of CUG residues identity on these PPIs.

5.1.3 Cek1 N-terminal tail

The N-terminal tail of the MAPK Cek1 (full-length Cek1) is enriched in amino acids repeats, including glutamine, alanine and a combined glutamine/alanine containing segments. These low complexity regions, i.e. regions containing little diversity in their amino acid composition, are generally devoid of regular three-dimensional structure, playing crucial roles in the kinase activity and in regulation of protein interaction networks (Schaefer *et al.* 2012). Some of the fungal regulatory proteins kinases involved in morphogenetic signaling networks that modulate transcription and PPIs are enriched in amino acid repeats, which enable yeast adaptation and survival to diverse environmental conditions (Chavali *et al.* 2017). In *C. albicans,* 30% of the proteins contain homorepeats, and polyglutamine repeats represent the most prevalent homorepeats (Dr. Chavali, personal communication). Although proteins containing amino acid repeats have been studied and characterized intensively in the context of protein aggregation in neurodegenerative diseases (Zoghbi *et al.* 2000), research about the functional role of homorepeats is currently in exponential growth (Chavali *et al.* 2017, Totzeck *et al.* 2017, Mier *et al.* 2018). To explore the functional role of this N-terminal tail, with the characteristics of an intrinsically disordered region, on the structure and function of the MAPK Cek1, full-length Cek1 was produced for biochemical and biophysical studies. We postulated that the N-terminal tail of the Cek1 protein could modulate its activity and ultimately the signal transduction across the Cek1 signaling cascade.

Interestingly, our results demonstrated that the N-terminal tail of Cek1 MAPK positively impacts the thermal stability and autophosphorylation activity of both the serine and leucine variants of the full-length protein. This observation implies a role of this N-terminal flexible tail on Cek1 function with potential impact in signal transduction. The mechanism by which the N-terminal low complexity region could

enhance Cek1 activity and stability remains to be disclosed. The present results showed that the N-terminal tail of Cek1 potentiates autophosphorylation of Tyr233, and partially compensates for the deleterious effects resulting from the incorporation of the leucine residue on the enzymatic activity of Cek1. Hence, these results provide the first insights about the interplay between the CUG codon ambiguity and the flexible N-terminal tail, which are worth further investigation.

The *in vitro* structural and functional studies should be complemented by assays in *C. albicans* cells. The host lab has prepared several engineered *C. albicans* strains expressing serine and leucine variants of both full-length Cek1 and ΔN-Cek1. The preliminary results indicate that no significant differences are observed in the viability and growth rates of the different *C. albicans* strains over-expressing those Cek1 variants. However, expression of the *C. albicans* Cek1 leucine variants enhances their susceptibility to the cell wall disturbing agent calcofluor white. Interestingly, *C. albicans* strains expressing the leucine variants show a decrease in the total abundance of Cek1 protein, as detected by western blot with our Cek1-specific antibody, in agreement with the decreased stability observed with the recombinant leucine variant.

5.1.4 Concluding remarks

Additional studies will be necessary to obtain an integrated view of the role of CUG-residues in *C. albicans* virulence. Complementary assays with *C. albicans* strains will have to be performed to analyze *C. albicans* phenotype and morphological transitions together with host cell recognition and cell wall components in response and adaption to changing environments during infection. Furthermore, the *in vitro* reconstitution of the Cek1 MAPK signaling will be a useful approach to tackle the role of CUG ambiguity in cell signaling associated with *C. albicans* pathogenesis. First, a complete characterization of the recombinant proteins MAPKK Hst7 and the scaffold Cst5 constructs should be done. A combined analysis of the Cek1 enzymatic activity and PPIs between with its molecular partners should constitute the next step to address the correlation between macromolecular complex assembly and Cek1 phosphorylation. Structural studies, combining circular dichroism analysis, small angle X-ray scattering and protein crystallization of isolated proteins and Cek1 complexes with Cst5 and Hst7, should provide key information about the structural flexibility and conformational changes requires for MAPK complex assembly.

In conclusion, this work supports our initial hypothesis that CUG ambiguity impacts in the activity and function of *C. albicans* Cek1. Understanding the structure-function relationship of these pivotal proteins provides a overview of the functional impact of CUG residue identity in the fungal pathogen *C. albicans* and could guide the design of new and more efficient antifungal agents against *C. albicans* infections.

5.2 References

Alonso-Monge, R., E. Roman, *et al.* (2009). Fungi sensing environmental stress.*Clin Microbiol Infect* 15 Suppl 1: 17-9.

Alonso-Monge, R., E. Roman, *et al.* (2006). The MAP kinase signal transduction network in *Candida albicans*.*Microbiology* 152: 905-12.

Bhattacharyya, R. P., A. Remenyi, *et al.* (2006). The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway.*Science* 311: 822-6.

Chavali, S., P. L. Chavali, *et al.* (2017). Constraints and consequences of the emergence of amino acid repeats in eukaryotic proteins.*Nat Struct Mol Biol* 24: 765-777.

Cote, P., T. Sulea, *et al.* (2011). Evolutionary reshaping of fungal mating pathway scaffold proteins.*mBio* 2: e00230-10.

Ernst, J. F. and J. Pla (2011). Signaling the glycoshield: maintenance of the Candida albicans cell wall.*Int J Med Microbiol* 301: 378-83.

Good, M., G. Tang, *et al.* (2009). The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation.*Cell* 136: 1085-97.

Mier, P. and M. A. Andrade-Navarro (2018). Glutamine Codon Usage and polyQ Evolution in Primates Depend on the Q Stretch Length.*Genome Biol Evol* 10: 816-825.

Miranda, I., R. Rocha, *et al.* (2007). A genetic code alteration is a phenotype diversity generator in the human pathogen Candida albicans.*PLoS One* 2: e996.

Remenyi, A., M. C. Good, *et al.* (2005). The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network.*Mol Cell* 20: 951-62.

Rocha, C. R., K. Schroppel, *et al.* (2001). Signaling through adenylyl cyclase is essential for hyphal growth and virulence in the pathogenic fungus Candida albicans.*Mol Biol Cell* 12: 3631-43.

Schaefer, M. H., E. E. Wanker, *et al.* (2012). Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks.*Nucleic Acids Res* 40: 4273-87.

Stynen, B., P. Van Dijck, *et al.* (2010). A CUG codon adapted two-hybrid system for the pathogenic fungus Candida albicans.*Nucleic Acids Res* 38: e184.

Totzeck, F., M. A. Andrade-Navarro, *et al.* (2017). The Protein Structure Context of PolyQ Regions.*PLoS One* 12: e0170801.

Zhou, X. L., Z. P. Fang, *et al.* (2013). Aminoacylation and translational quality control strategy employed by leucyl-tRNA synthetase from a human pathogen with genetic code ambiguity.*Nucleic Acids Res* 41: 9825-38.

Zoghbi, H. Y. and H. T. Orr (2000). Glutamine repeats and neurodegeneration.*Annu Rev Neurosci* 23: 217-47.

Annexes

Author publications related to the topic of the PhD thesis

Fraga J.S., Sárkány Z., Silva A., Correia I., Pereira P.J.B.P., Macedo-Ribeiro S. (2019). Genetic code ambiguity modulates the activity of a *C. albicans* MAP kinase linked to cell wall remodeling. *BBA - Proteins and Proteomics* 1867: 654-661. doi: 10.1016/j.bbapap.2019.02.004

Fraga J., Maranha A, Mendes V, Pereira PJ, Empadinhas N, Macedo-Ribeiro S. (2015). Structure of mycobacterial maltokinase, the missing link in the essential GlgE-pathway. *Scientific Rep* 5:8026. doi:10.1038/srep08026

BBA - Proteins and Proteomics 1867 (2019) 654–661

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Genetic code ambiguity modulates the activity of a *C. albicans* MAP kinase linked to cell wall remodeling

Joana S. Fraga[a,b,c], Zsuzsa Sárkány[a,b,1], Alexandra Silva[a,b], Inês Correia[a,b], Pedro José Barbosa Pereira[a,b], Sandra Macedo-Ribeiro[a,b,*]

[a] IBMC-Instituto de Biologia Molecular e Celular, Universidade do Porto, 4200-135 Porto, Portugal
[b] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal
[c] ICBAS-Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, 4050-313 Porto, Portugal

ARTICLE INFO

Keywords:
Biophysics
Cek1 protein kinase
Autophosphorylation
Signal transduction
Candida albicans
CUG codon translation ambiguity

ABSTRACT

The human fungal pathogen *Candida albicans* ambiguously decodes the universal leucine CUG codon predominantly as serine but also as leucine. *C. albicans* has a high capacity to survive and proliferate in adverse environments but the rate of leucine incorporation fluctuates in response to different stress conditions. *C. albicans* is adapted to tolerate this ambiguous translation through a mechanism that combines drastic decrease in CUG usage and reduction of CUG-encoded residues in conserved positions in the protein sequences. However, in a few proteins, the residues encoded by CUG codons are found in strictly conserved positions, suggesting that this genetic code alteration might have a functional impact. One such example is Cek1, a central signaling protein kinase that contains a single CUG-encoded residue at a conserved position, whose identity might regulate the correct flow of information across the MAPK cascade. Here we show that insertion of a leucine at the CUG-encoded position decreases the stability of Cek1, apparently without major structural alterations. In contrast, incorporation of a serine residue at the CUG position induces the autophosphorylation of the conserved tyrosine residue of the Cek1 ^{231}TEY^{233} motif, and increases its intrinsic kinase activity *in vitro*. These findings show that CUG ambiguity modulates the activity of Cek1, a key kinase directly linked to morphogenesis and virulence in *C. albicans*.

1. Introduction

Over the years it has been repeatedly demonstrated that the genetic code is not immutable and the presence of non-standard translation in evolutionarily distant organisms was frequently observed [1]. Variations in codon assignment, including several sense-to-sense codon alterations, have been identified in numerous mitochondrial genomes [2], which are smaller than nuclear genomes and where genetic code changes are less likely to be harmful [3]. Alterations in the canonical genetic code involving nonsense-to-sense codon modifications, such as the reassignment of standard termination codons to different amino acids occurred frequently in mitochondrial genomes [3], possibly through a mechanism preceded by codon-anticodon mismatches and dual function of stop codons [4]. Reassignment of termination codons to glutamine, cysteine or tryptophan also occurs in nuclear genomes [3], and recent studies showed that in some ciliates termination codons can be ambiguously recognized as either sense or stops in a context-dependent manner [5]. Furthermore, ambiguous meaning of translation stop codons are at the basis of natural expansion of the genetic code, which occurs by insertion of non-canonical amino acids such as selenocysteine at UGA sites in prokaryotes and eukaryotes, and pyrrolysine at UAG codons in *Methanosarcinaceae* methyltransferase coding genes [6]. The reassignment of the "universal" leucine CUG codon to serine or alanine is the only known case of alteration of sense codon meaning in nuclear genomes and has been demonstrated in different yeast species with biotechnological or biomedical relevance. The first evidence for the change in the meaning of the CUG codon was shown in yeast species belonging to the CTG-clade that comprise the most pathogenic *Candida* species, including *C. albicans* [7]. The yeast *Pachysolen tannophilus* translates CUG codons as alanine instead of the canonical leucine, with

Abbreviations: MAPK, mitogen activated protein kinase; MBP, maltose binding protein; PCR, polymerase chain reaction; TEV, tobacco etch virus; Tm, melting temperature; TNP-ATP, 2′,3′-*O*-trinitrophenyl-adenosine-5′-triphosphate

* Corresponding author at: IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, 4200-135 Porto, Portugal.
E-mail address: sribeiro@ibmc.up.pt (S. Macedo-Ribeiro).
[1] Present address: LEPABE – Departamento de Engenharia Química, Faculdade de Engenharia da Universidade do Porto, 4200–465 Porto, Portugal.

https://doi.org/10.1016/j.bbapap.2019.02.004
Received 5 December 2018; Received in revised form 13 February 2019; Accepted 18 February 2019
Available online 20 February 2019
1570-9639/

implications in biotechnological applications [8]. Some yeasts ambiguously decode the CUG codon as leucine and serine through different mechanisms, an intriguing feature considering that accuracy during translation of the genetic code is critical to ensure adequate cell function and mistranslation can negatively impact cell survival [9,10]. However, recent data show that cells are relatively tolerant to errors in protein translation, and that increased mistranslation (*e.g.*, under stress conditions) enhances microbial fitness and modulates host-microbe interactions [11].

C. albicans translates the CUG codon predominantly as serine (97%) with only a small percentage (3%) being translated as leucine [12,13]. In response to environmental and stress conditions, such as low pH or during oxidative stress, the incorporation of leucine increases up to 5% [14]. The levels of leucine insertion at CUG codons can be artificially increased in engineered *C. albicans* strains to between 28% and 98%, resulting in colonies with high morphological diversity, enhanced expression of genes involved in cell adhesion, as well as increased resistance to antifungal agents, oxidative stress and decreased susceptibility to phagocytosis by macrophages [13,15,16]. Those results suggested that proteome expansion resulting from ambiguous CUG translation might result in phenotypic diversity, increase fitness and modulate host-pathogen interactions. Interestingly, increased incorporation of leucine at CUG codons up to 28% did not result in significant changes in growth rates [13,14], demonstrating the high tolerance of *C. albicans* to CUG ambiguity. Indeed, in *C. albicans* and other CTG-clade species the distribution of residues encoded by CUG codons is prevalent in non-conserved and partially exposed regions at the protein surface, where both leucine and serine can be incorporated without major impact on protein structure or function [17]. However, in a few proteins, the residues encoded by CUG codons are located in functionally relevant positions, where a serine or other polar amino acids are generally conserved in homologous proteins. In contrast, in *Ascoideia asiatica* the CUG codon is translated either as leucine or serine at approximately equal proportions in a stochastic manner, and CUG-encoded residues in key structural and functional sites are strictly avoided [18].

Some studies have demonstrated that CUG ambiguity impacts the activity and regulation of *C. albicans* proteins with no major changes on the global protein structure. Examples are *C. albicans* eukaryotic translation initiation factor 4E that contains a CUG-encoded residue located at the protein surface [19], and the seryl- and leucyl-tRNA synthetases that contain a single CUG-encoded residue each at highly conserved amino acid sequence positions [17,20,21]. Previous analysis of the impact of CUG ambiguous translation in the *C. albicans* proteome revealed that CUG codons are over-represented on key signaling pathways, such as the Mitogen Activated Protein Kinase (MAPK) pathway, that respond to various external cues and mediate the regulation of morphogenesis and virulence [17]. In particular, Cek1, a key MAPK, contains one CUG-encoded residue in the nucleotide binding pocket, in a position where a serine is strictly conserved in homologous proteins and ambiguity is predicted to have functional impact [17,22]. Cek1 is activated by sequential phosphorylation by upstream MAPK cascade activators (Cst20-Ste11-Ste7) in response to extracellular stimuli, leading to the phosphorylation of the transcription factor Cph1 [23]. This kinase is involved in distinct essential mechanisms of *C. albicans*: invasive hyphal growth [24–26], biofilm formation [27], cell wall integrity, host recognition [28,29], and response to oxidative and osmotic stress [30], and therefore it is a key target to explore CUG ambiguity-mediated morphological changes.

To assess the impact of CUG ambiguity on Cek1 kinase, two recombinant variants of the active domain of Cek1, having a serine or a leucine residue at the CUG-encoded position (Cek1_Ser or Cek1_Leu) were expressed and purified for structural and functional studies. Biochemical and biophysical data show that although the incorporation of leucine at this position does not induce Cek1 unfolding, it reduces the protein's thermal stability and decreases its enzymatic activity. In addition, we show for the first time that only the Cek1_Ser variant from *C. albicans* is autophosphorylated *in vitro* at the tyrosine residue of the conserved threonine-glutamate-tyrosine (^{231}TEY^{233}) motif within the kinase activation loop.

2. Materials and methods

2.1. Cloning, expression and purification of Cek1

A synthetic gene encoding full-length *Candida albicans* Cek1 (serine variant for CUG codon at position 199) with codon usage optimized for expression in *Escherichia coli* was obtained from a commercial supplier (Eurofins Scientific). The active domain of Cek1 (amino acids 58–421) was amplified by PCR and cloned into expression vector pETMBP (EMBL dabatase) in frame with a N-terminal hexahistidine tag, the maltose binding protein (MBP) solubility tag and a tobacco etch virus (TEV) protease recognition site. This tagged form of the active domain of Cek1 is herein named Cek1. The Cek1_Ser (serine variant – UCC replacing the CUG codon) expression plasmid was used as template to produce a plasmid expressing Cek1_Leu (leucine variant – UUA replacing the CUG codon) by site-directed mutagenesis (QuikChange, Agilent Technologies) (Table S1). To prepare the Cek1 phosphorylation-null mutants, Thr231 and Tyr233 of the conserved ^{231}TEY^{233} phosphorylation motif were mutated to valine and phenylalanine (Cek1^{TV}, Cek1^{YF}, Cek1$^{TV/YF}$), respectively. Furthermore, using the same strategy a phosphomimetic mutant where Thr231 was mutated to aspartate (Cek1^{TD}) and the kinase-dead mutant, with Lys100 replaced by arginine (Cek1^{KR}) were produced (Table S1). All residue numbers refer to the full-length amino acid sequence of *C. albicans* Cek1.

Cek1 variants (Cek1_Ser and Cek1_Leu) and Cek1 mutants were overexpressed in *E. coli* strain BL21 Star™ (DE3) (Stratagene). Following transformation, *E. coli* cells were grown in Luria-Bertani medium at 37 °C until an OD_{600} of 0.6–0.8. At this point, the temperature was decreased to 20 °C and protein expression was induced with addition of 0.9 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and allowed to proceed overnight. Cells were harvested by centrifugation at 4 °C and 4000*g* for 30 min, resuspended in lysis buffer (40 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 5 mM 2-β-mercaptoethanol) supplemented with 100 μg/ml lysozyme and stored at −20 °C. Prior to lysis by sonication (SONOPULS HD2200, Bandelin Electronic), phenylmethylsulfonyl fluoride was added to the cell suspension. The crude cell extract was clarified by centrifugation at 4 °C and 35,000g for 45 min and loaded onto a HisTrap HP column (5 ml, GE Healthcare) pre-equilibrated in buffer A (40 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 5% (v/v) glycerol, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol). Bound proteins were eluted with buffer A supplemented with 75 mM imidazole. Cek1-containing fractions were pooled and further purified by size exclusion chromatography on a HiPrep 26/60 Sephacryl S-100 column (GE Healthcare) equilibrated with protein buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 5 mM $MgCl_2$, 1 mM DTT). Fractions from the preparative size exclusion chromatography were analyzed by SDS-PAGE and those containing pure Cek1 were pooled and concentrated on centrifugal ultrafiltration devices (10 kDa cutoff; Merck Millipore). Final protein concentration was estimated by measuring the absorbance of the samples at 280 nm using the theoretical extinction coefficient for each variant (Table S2) prior to storing the samples at −80 °C. Additionally, the purity and oligomeric state of the purified proteins were analyzed by analytical size exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) using protein buffer as mobile phase. The average yields for the purified Cek1_Ser and Cek1_Leu variants were 5 and 1 mg per liter of *E. coli* culture, respectively.

In order to remove the fusion tags (His_6 and MBP tags), purified Cek1 was digested with TEV protease overnight at 4 °C in protein buffer. The buffer of the sample was then exchanged to 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% (v/v) glycerol using a HiPrep 26/10 desalting

column (GE Healthcare), prior to a second immobilized metal affinity chromatography step using a HisTrap HP column equilibrated with buffer A. Untagged Cek1 present in the HisTrap column flow through was further purified on a Superdex 75 10/300 GL column equilibrated with protein buffer and eluted protein fractions containing pure Cek1, as assessed by SDS-PAGE, were pooled, concentrated and stored at −80 °C until used.

2.2. Dynamic light scattering

Molecular size measurements were carried out in a Zetasizer Nano ZS DLS system (Malvern Instruments) using a DTS 2112 cuvette. Cek1 solution was diluted in protein buffer to 1 mg/ml and three independent measurements were made for each sample at 20 °C. The intensity size distribution obtained by dynamic light scattering is a plot of relative intensity of light scattered by particles in different size classes. However, when there is more than one peak in this plot, the intensity size distribution must be converted to a volume size distribution for a more realistic view of the data [31], as was the case for Cek1 protein samples. All data were analyzed using dynamic light scattering (nano) software (Malvern Instruments) to determine the polydispersity (Pd) (relative standard deviation) and the mean hydrodynamic radius (R_H) values from the volume distributions, used to estimate the molecular weight (assuming a globular protein) from a built-in empirical calibration graph.

2.3. Circular dichroism

The overall secondary structure of Cek1 variants (Cek1_Ser and Cek1_Leu) and of selected mutants was assessed by far-UV circular dichroism using a J815 circular dichroism spectrometer (JASCO). Measurements of samples at 0.1 mg/ml were performed at 20 °C with 1 nm bandwidth, 1 s response, 100 nm/min scanning speed and 6 accumulations. The circular dichroism spectra (190–260 nm) were analyzed using the JASCO software (JASCO) to determine the secondary structure content of the samples.

2.4. Differential scanning fluorimetry

The thermal stability of the Cek1 variants (0.3 mg/ml) was determined by following SYPRO Orange (Invitrogen) fluorescence using an iCycler iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad). The melting curves (excitation/emission, 470 nm/570 nm) were obtained by increasing the temperature from 25 to 85 °C in 0.5 °C steps with 30 s hold time. Data were analyzed with CFX Manager software (Bio-Rad), which calculates the melting temperature (Tm) from the inflection point of the melting curve. The thermal denaturation assay was performed in triplicate for each protein and data represent the average of three independent experiments with standard error. A statistical analysis was performed using one-way ANOVA (Tukey multicomparison test) $p < 0.05$ with Prism7 (GraphPad Software).

2.5. Cek1 phosphorylation in vitro

The immunoblot analysis of Cek1 proteins was performed using polyclonal anti-Cek1 serum produced in-house by immunizing rabbits with the untagged active domain of Cek1. Purified proteins were separated on 12% (w/v) polyacrylamide SDS-PAGE gels and electrotransferred onto nitrocellulose membranes. The membranes were blocked with 5% (w/v) BSA in TBS-Tween (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20) for 1 h at room temperature (RT), incubated overnight at 4 °C with rabbit anti-Cek1 serum (1:5000), followed by an incubation with anti-rabbit IgG HRP-conjugated antibody (1:10000) for 1 h at RT. Cek1 samples were detected with ECL Plus reagent (GE Healthcare).

The phosphorylation state of recombinant Cek1 proteins was evaluated by immunoblot analysis using a phospho-specific antibody. After blocking, membranes were incubated with rabbit anti-phospho p44/42 MAPK antibody (1:2000, Cell Signaling Technology) overnight at 4 °C, and then with anti-rabbit IgG HRP-conjugated antibody (1:10000) for 1 h at RT. Phosphorylated samples were detected with ECL Plus reagent.

Cek1 autophosphorylation activity was probed by mixing Cek1_Ser or the phosphomimetic mutant $Cek1^{TD}$_Ser with the untagged catalytically inactive mutant (untagged $Cek1^{KR}$_Ser, used as substrate, 2:1 M ratio). The reaction at RT was initiated by addition of 1 mM ATP and the phosphorylation status of the catalytically inactive mutant at different time points was evaluated by immunoblot with rabbit anti-phospho p44/42 MAPK antibody and anti-Cek1 serum, as described above.

2.6. Cek1 activity assay

The activity of purified Cek1 was monitored by measuring the production of ADP using a chemiluminescent ADP detection assay (ADP-Glo kit, Promega). The reactions were performed in 384-well plates, kept in the dark, using 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 0.1 mg/ml BSA as assay buffer. The standard curves of ATP to ADP conversion were determined using different ATP/ADP concentrations according to the manufacturer's protocol. The reactions were started by the addition of 1 mM ATP to 10 μM Cek1, and stopped by addition of ADP-Glo™ Reagent after 1 h incubation. Measurements were performed using an endpoint luminescence protocol in a Synergy 2 plate reader (BioTek).

3. Results

3.1. The C. albicans Cek1_Ser variant is more stable than Cek1_Leu

In order to evaluate the structural impact of serine or leucine insertion at the CUG-encoded position of Cek1 (Fig. 1a), a thorough biophysical characterization was performed. Analytical size exclusion chromatography showed no differences on the elution profile of the two Cek1 variants, with elution volumes compatible with monomeric proteins (85 kDa; Fig. 1b). In accordance, dynamic light scattering measurements indicated that both Cek1 variants were monodisperse, with hydrodynamic radii consistent with a monomeric organization (Fig. 1c). Additionally, the presence of leucine or serine at the CUG-encoded position had no major impact on the fold of the Cek1 kinase, since no major structural differences were observed by circular dichroism spectroscopy (Fig. 1d). Indeed, the circular dichroism spectra revealed no significant differences in secondary structure content between the two forms (Cek1_Ser: 32% α-helix, 43% β-sheet, 1.6% β-turn; Cek1_Leu: 34% α-helix, 41% β-sheet, 1.6% β-turn), indicating a preservation of the overall structure. Interestingly, the incorporation of serine at the CUG-encoded position lead to an increase (+3 °C) in the Tm of Cek1, as determined by differential scanning fluorimetry (Fig. 1e).

The stability of the two Cek1 variants was unaltered by the addition of nucleotides, since their Tm remained unchanged upon incubation with ATP (Table S3). Altogether, these results indicate that the presence of a leucine residue at the single CUG-encoded position did not induce protein misfolding but significantly reduced the thermal stability of Cek1, which was not due to differences in the phosphorylation states of the two variants, as shown below.

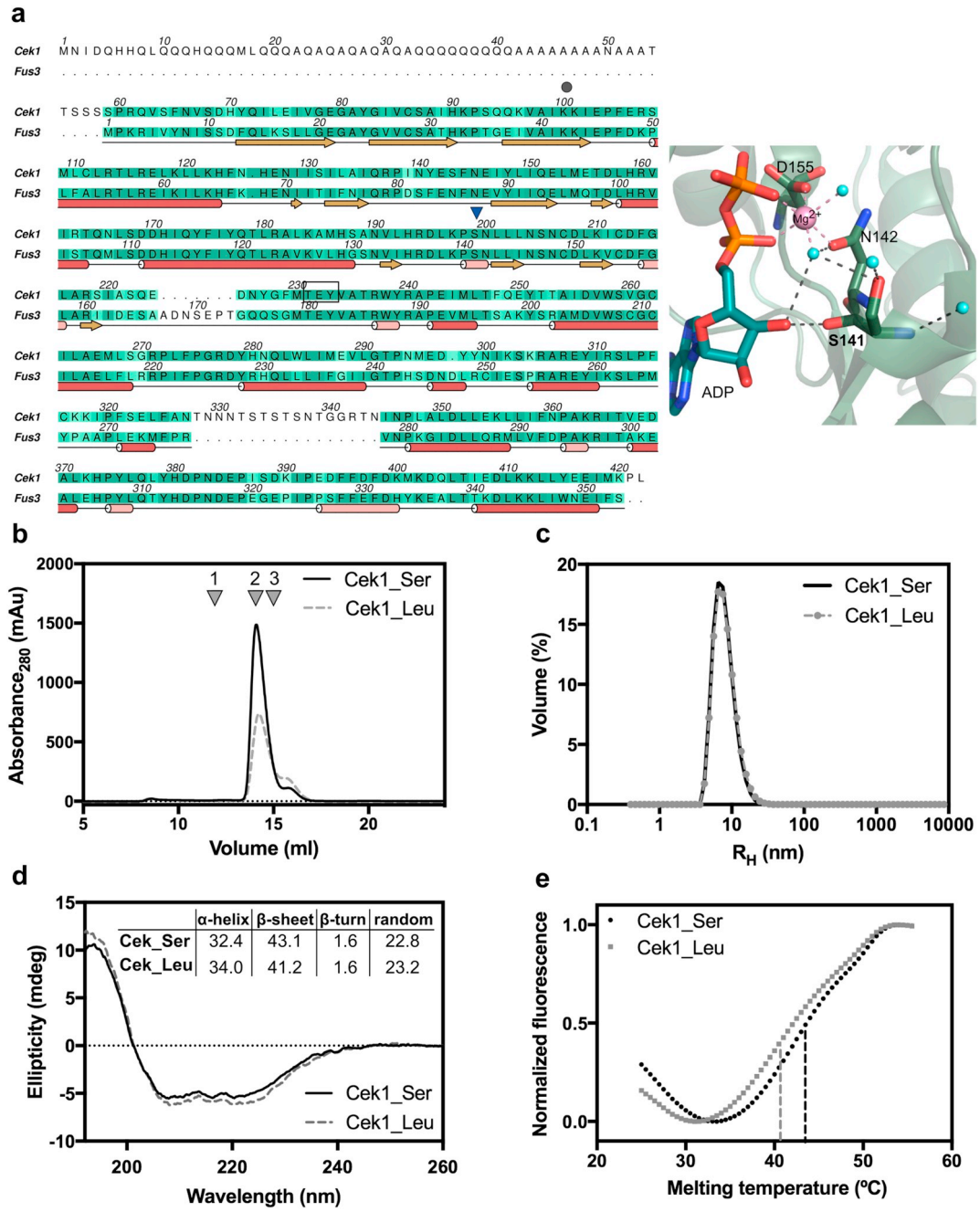


Fig. 1. Biophysical characterization of *Candida albicans* Cek1 variants. (a) Amino acid sequence alignment of *C. albicans* Cek1 and of *S. cerevisiae* orthologue Fus3. The secondary structure elements of *S. cerevisiae* Fus3 (PDB accession code 2B9F) are represented below the corresponding amino acid sequence (red cylinders, α-helices; pink cylinders, 3_{10} helices; yellow arrows, β-sheets). Important residues are marked: CUG-encoded residue (Ser199; blue triangle), catalytic residue (Lys100; black circle) and ^{231}TEY^{233} motif (black box). Detailed view of the ATP-binding pocket of *S. cerevisiae* Fus3: Ser141 (shown as sticks) of Fus3 is structurally equivalent to the *C. albicans* CUG-emcoded residue 199. Hydrogen bonds are represented as dashed lines. The magnesium ion is represented as a pink sphere and blue spheres represent ordered water molecules. (b) Size exclusion chromatography of Cek1_Ser (solid line) and Cek1_Leu (dashed line). Both variants elute as monomers (85 kDa). Downward triangles indicate the elution volume of protein standards: 1 - Aldolase (158 kDa), 2 – Conalbumin (75 kDa) and 3 – Ovalbumin (43 kDa). (c) The two Cek1 variants are monodisperse (Pd = 26.8% Cek1_Ser; Pd = 29.3% Cek1_Leu) and monomeric (R_H Cek1_Ser = 4.05 nm; R_H Cek1_Leu = 4.18 nm, in good agreement with the estimated molecular mass of Cek1), as assessed by dynamic light scattering. (d) Far-UV circular dichroism spectra of Cek1_Ser and Cek1_Leu revealing no significant differences in overall secondary structure. (e) Representative thermal denaturation measured by differential scanning fluorimetry for both variants of Cek1. The normalized fluorescence intensity is plotted as a function of the temperature (°C). The temperature at the inflection point of the unfolding transition defines the Tm value for each protein (Tm Cek1_Ser = 43.1 ± 0.7 °C; Tm Cek1_Leu = 40.5 ± 0.3 °C).

3.2. Autophosphorylation at the conserved ^{231}TEY^{233} motif of Cek1 requires a serine at the CUG-encoded position

MAP kinases, including the *Saccharomyces cerevisiae* MAPK Fus3, are often found to be autophosphorylated when expressed in *E. coli* [32]. The phosphorylation status of recombinant Cek1 variants was assessed using anti-phospho p44/42, a specific antibody that recognizes the phosphorylation of both threonine (Thr231) and tyrosine (Tyr233) residues of the ^{231}TEY^{233} motif [33]. Immunoblot analysis of both Cek1 variants with anti-phospho p44/42 revealed that only recombinant

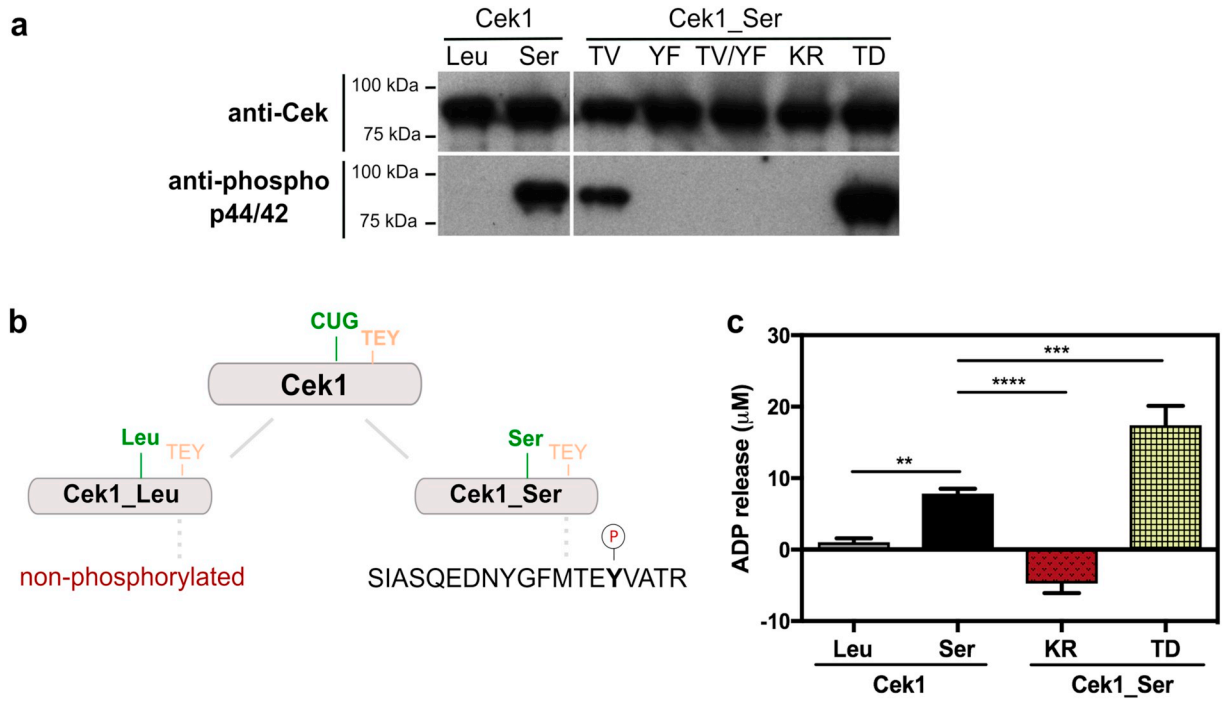


Fig. 2. Cek1_Ser is autophosphorylated at Tyr233 and displays higher catalytic activity than Cek1_Leu. (a) Immunoblot analysis of Cek1_Leu, Cek1_Ser and Cek1_Ser mutants (Cek1^{TV}_Ser, Cek1^{YF}_Ser, Cek1$^{TV/TF}$_Ser, Cek1^{KR}_Ser, Cek1^{TD}_Ser) probed with rabbit anti-Cek serum (upper panel) and rabbit anti-phospho p44/42 (bottom panel). Autophosphorylation was detected in Cek1_Ser and Cek1^{TV}_Ser, indicating that autophosphorylation occurs on Tyr233. (b) Schematic representation of phosphorylation results of Cek1 variants obtained by immunoblot analysis combined with mass spectrometry: phosphorylation was not detected in Cek1_Leu, while in Cek1_Ser a phosphorylated peptide was identified, indicating that autophosphorylation occurred on Tyr233. (c) Activity assay by quantification of ADP release of Cek1 variants, inactive (Cek1^{KR}_Ser, negative control) and phosphomimetic (Cek1^{TD}_Ser, positive control) mutants. Cek1_Ser released significantly more ADP compared to Cek1_Leu, while Cek1^{TD}_Ser was the most active. Data are the average of three determinations and error bars represent standard deviation. Statistics were analyzed using one-way ANOVA (Tukey multicomparison test, $p > .05$ non-significant (ns), $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$).

Cek1_Ser was autophosphorylated *in vitro*, and no phosphorylation was detected in the Cek1_Leu variant using this approach (Fig. 2a). The autophosphorylation activity of Cek1_Ser was further explored using two single mutants (Cek1^{TV}_Ser; Cek1^{YF}_Ser) and a double mutant (Cek1$^{TV/YF}$_Ser) of the putative phosphorylation sites (Thr231 mutated to valine and Tyr233 mutated to phenylalanine), as well as a kinase-dead mutant (Cek1^{KR}_Ser, Lys100 mutated to arginine) and the putative phosphomimetic variant (Cek1^{TD}_Ser, Thr231 mutated to aspartate). As for the Cek1_Ser and Cek1_Leu variants, these specific mutations had no effect on the fold and overall secondary structure content of the enzyme, as assessed by circular dichroism (Fig. S1). Mutation of Thr231 to valine (Cek1^{TV}_Ser) at the ^{231}TEY^{233} motif had no impact on the phosphorylation signal of Cek1_Ser (Fig. 2a). In contrast, replacement of Tyr233 by phenylalanine (Cek1^{YF}_Ser) completely abolished the autophosphorylation of recombinant Cek1_Ser, indicating that Cek1_Ser autophosphorylation occurs on Tyr233 (Fig. 2a). Therefore, when Thr231 and Tyr233 were mutated to valine and phenylalanine (Cek1$^{TV/YF}$_Ser), respectively, recognition by anti-phospho p44/42 was also completely abolished. As previously described, the substitution of the catalytic lysine residue by an arginine decreased the kinase activity by preventing transfer of the phosphate group from the nucleotide phosphate to the acceptor protein [34]. Thus, mutation of the catalytic lysine to arginine (Cek1^{KR}_Ser) abolished the autophosphorylation signal, while the putative phosphomimetic mutant (Cek1^{TD}_Ser) displayed an enhanced autophosphorylation signal in the immunoblot assay, as expected (Fig. 2a). Analysis of the thermal stability of the Cek1^{YF}_Ser and Cek1$^{TV/YF}$_Ser mutants (Table S4), which are not autophosphorylated, showed no significant differences in the melting temperatures compared to wild-type Cek1_Ser, underscoring that differences in the melting temperatures of Cek1_Leu and Cek1_Ser do not result from Cek1_Ser autophosphorylation.

To further confirm that Cek1_Ser was autophosphorylated on Tyr233 of the ^{231}TEY^{233} motif, as suggested by the immunoblot data, both variants of Cek1 were analyzed by mass spectrometry. Protein samples were analyzed as-purified and after incubation with 1 mM ATP. Interestingly, using this approach we were not able to detect phosphorylated peptides in the as-purified Cek1 variants. However, upon incubation with ATP, a peptide (SIASQEDNYGFMTEpYVATR) was identified in the Cek1_Ser variant sample, displaying a mass compatible with a single phosphorylation (Fig. 2b and S2a). Although the experimental procedure used is not quantitative, the measured signal for the phosphorylated peptide was approximately 20-fold lower than for the non-phosphorylated peptide (Fig. 2b and S2a). This result further supported the immunoblot data for the Cek1_Ser mutants, suggesting that the autophosphorylation of Cek1_Ser occurred at Tyr233. In agreement with the immunoblot results, no phosphorylated peptide could be detected by mass spectrometry on the Cek1_Leu variant sample (Fig. S2b), suggesting that a leucine residue at the CUG-encoded position had a negative impact on the autophosphorylation activity of Cek1. In summary, recombinant Cek1_Ser expressed in *E. coli* was readily autophosphorylated at Tyr233 of the ^{231}TEY^{233} motif, suggesting that the presence of a serine at the CUG-encoded position was required for protein autophosphorylation and may be crucial for regulating the kinase activity.

3.3. Codon decoding ambiguity affects Cek1 protein activity

In contrast to Cek1_Leu, Cek1_Ser is autophosphorylated *in vitro*. To investigate the influence of phosphorylation on the enzymatic activity of Cek1 and unveil the role of CUG ambiguity on its function, the catalytic activity of Cek1 variants was measured. Using a luminescent ADP-detection assay, Cek1_Ser displayed a 10-fold higher activity than

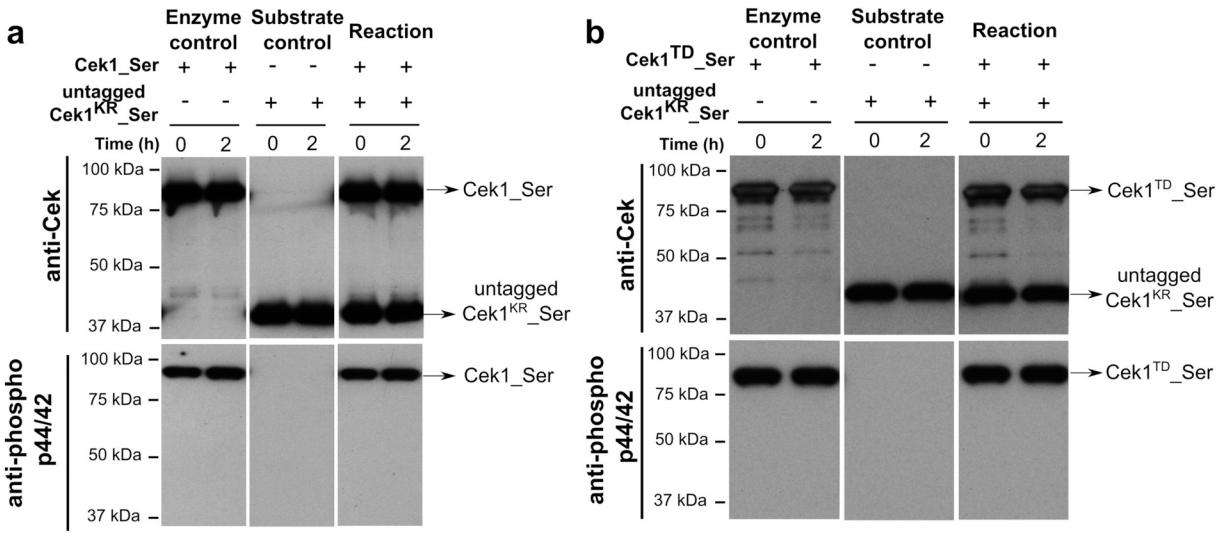


Fig. 3. The autophosporylation of Cek1 is an intramolecular event. Transphosphorylation assay of the inactive mutant of Cek1_Ser (untagged Cek1^{KR}_Ser, substrate, 42 kDa) by (a) Cek1_Ser variant (85 kDa) and (b) Cek1^{TD}_Ser phosphomimetic mutant (85 kDa), probed with rabbit anti-Cek serum (upper panel) and rabbit anti-phospho p44/42 (bottom panel). Phosphorylation of untagged Cek1^{KR}_Ser was not observed in the reaction, only the autophosphorylation of the two enzymes was detected.

Cek1_Leu (Fig. 2c). As expected, the Cek^{KR}_Ser mutant displayed no catalytic activity, while the activity of mutant Cek1^{TD}_Ser was significantly increased (~2-fold; Fig. 2c). Considering that the CUG-encoded residue is located in the ATP binding pocket, we reasoned that the differences in the autophosphorylation activity observed for the two Cek1 variants could result from dissimilarities in ATP binding affinity. With this aim, we used TNP-ATP, a fluorescent ATP analog that binds to the ATP binding pocket of protein kinases [35], to assess the affinity of both variants to the nucleotide. This assay showed that TNP-ATP bound similarly to both Cek1 variants (5.7 μM for Cek1_Ser and 8 μM for Cek1_Leu) (Fig. S3) and that differences in enzyme activity did not seem to be correlated with alterations in ATP binding affinity. However, the fluorophore bound to the nucleotide might mask small differences in ATP binding. Overall, these results demonstrated that Cek1_Ser had intrinsic autophosphorylation activity and was more active than Cek1_Leu *in vitro*. Indeed, the Cek1^{TD}_Ser mutant, mimicking the phosphorylation of the ^{231}TEY^{233} motif threonine, in the activation loop, displayed significantly enhanced catalytic activity.

3.4. Autophosphorylation of Cek1 does not occur in trans

In order to explore if the autophosphorylation of Cek1_Ser occurred through a cis- (intramolecular) or trans-autophosphorylation (intermolecular) mechanism [36], the inactive kinase-dead mutant (Cek1^{KR}_Ser) was used as substrate of Cek1_Ser or of its phosphomimetic mutant, Cek1^{TD}_Ser. Although both Cek1_Ser (Fig. 3a) and Cek1^{TD}_Ser (Fig. 3b) were autophosphorylated, neither was able to phosphorylate the substrate, suggesting that Cek1 kinase is unable to phosphorylate its inactive form, supporting the cis-autophosphorylation mechanism.

4. Discussion

C. albicans is a human commensal that, under specific conditions, is able to cause disease that ranges from superficial mucosal to life-threatening systemic infections [37]. This opportunistic human pathogen has the remarkable ability to quickly adapt to varied host niches and threats elicited by host immune defenses. Hence, in response to multiple stress factors and/or damage induced by antifungal drugs, the activation of highly coordinated signaling pathways generates the diversity required for survival and virulence [38]. Interestingly, many of the enzymes involved in signal transduction cascades in *C. albicans* are encoded by genes that contain CTG codons, stochastically translated into serine (97%) and leucine (3%) residues often located in highly conserved sites [17,22]. One of these enzymes is the MAPK Cek1 (with one CUG-encoded residue) that is central for *C. albicans* morphogenesis, cell wall biosynthesis and host interaction [28,29]. Accordingly, we hypothesized that the fluctuations in serine/leucine incorporation levels induced by different stress conditions might impact the function of Cek1 with downstream consequences on *C. albicans* virulence traits. To assess the role of the identity of the CUG-encoded residue in Cek1 function, we characterized the structural and functional features of the Cek1_Ser and Cek1_Leu variants of the kinase active domain. Cek1 displays high amino acid sequence identity (63%) with the structurally characterized Fus3 [22,39], the MAPK that mediates the mating pheromone response in *S. cerevisiae*. The unique conserved CUG-encoded residue in the nucleotide binding pocket of Cek1 (Ser199) is structurally equivalent to a serine residue in Fus3 that stabilizes a water molecule from the coordination sphere of the magnesium ion that interacts with the phosphate moiety of the nucleotide (Fig. 1a) [17,22]. Consequently, the presence of a leucine residue at this position is expected to interfere with the nucleotide binding and/or phosphate transfer activity of Cek1. Our results showed that although the ambiguity at the CUG-encoded position does not affect the overall secondary structure of Cek1, incorporation of leucine at position 199 decreases considerably the thermal stability of the kinase. This difference likely results from the expected loss of one or two hydrogen bonds upon replacement of the polar serine residue by the hydrophobic leucine [40,41]. Although the CUG-encoded residue faces the ATP binding pocket, binding to TNP-ATP did not seem to be altered by the identity of the amino acid at position 199.

Autophosphorylation is a widespread feature among protein kinases and is also observed in Fus3 [32], a protein that is only fully activated *via* a kinase cascade that results in dual phosphorylation on the tyrosine and threonine residues of the activation loop TEY motif [42]. Interestingly, following expression in *E. coli*, only Cek1_Ser is recognized by the anti-phospho p44/42 antibody. Our results showed that Cek1_Ser autophosphorylation occurs on the conserved Tyr233 of the ^{231}TEY^{233} motif within the activation loop. In agreement, *in vitro* activity assays showed that the catalytic activity of Cek1_Ser is significantly higher than that of Cek1_Leu. Although our mass spectrometry data suggested the occurrence of autophosphorylation on Tyr233 upon incubation with

ATP, we cannot exclude the possibility that the *in vitro* kinase activity assays, based on the quantification of ADP release, report on the combined effects of Cek1 autophosphorylation and intrinsic ATPase activity, as previously described for some protein kinases [43,44]. Further, our autophosphorylation results using the kinase dead mutant Cek^{KR}_Ser as substrate, suggested that Cek1 autophosphorylation did not occur *via* a trans-autophosphorylation mechanism, and likely involves an intramolecular reaction as proposed for *S. cerevisiae* Fus3 [32]. In yeast MAPKs, enzyme activation requires an order-to-disorder transition, meaning that in the inactive state part of the activation loop, including the phosporylatable tyrosine, is positioned at the substrate binding pocket and obstructs access of substrate peptides to the active site [32,39] [45],. Therefore, kinase activity is inhibited *via* a pseudo-substrate mechanism, prior to full activation by the upstream kinase, which phosphorylates both residues in the TEY motif resulting in its displacement from the catalytic site. In the resting state, positioning of the tyrosine side chain close to the kinase catalytic cleft, highly conserved in *C. albicans* Cek1, may explain the identification of a small subpopulation of monophosphorylated protein in recombinant Cek1. Therefore, in our Cek1 samples the enzyme catalytic site might be predominantly inaccessible to substrate peptides or proteins, explaining the inability of Cek1_Ser or $Cek1^{TD}$_Ser to phosphorylate the generic substrate myelin basic protein [32] and the synthetic Cph1 peptide [23] (data not shown).

Although our current data do not allow to infer the precise autophosphorylation mechanism of Cek1 and why this activity is compromised in the Cek1_Leu variant, we can speculate that the presence of a leucine at the CUG-encoded position impacts negatively the phosphoryl transfer activity of Cek1 either by directly diminishing the catalytic activity or by modifying the dynamics of the activation loop and therefore altering the residence time of the phosphoryl-accepting tyrosine at the kinase catalytic site. Cek1 forms part of a highly regulated multienzyme assembly, whose activation is triggered in response to different external signals [46]. In *S. cerevisiae*, the scaffold protein Ste5 enhances Fus3 autophosphorylation on the tyrosine residue of the TEY motif [32] and, together with the phosphatase Msg5, maintains the monophosphorylated form of Fus3, which has an inhibitory role on the signaling pathway [33]. Interestingly, the allosteric Fus3 autophosphorylation mechanism mediated by bipartite binding of the scaffold protein Ste5 is likely not conserved in Cek1, since this region is structurally distinct in Cst5, the *C. albicans* MAPK scaffold protein [46]. In *C. albicans*, CUG translational ambiguity might play a role in rewiring the collective behavior of these macromolecular assemblies with consequences for the flow of information across these signaling cascades.

5. Conclusion

In this study, we evaluate the effect of CUG translational ambiguity on the MAP kinase Cek1, a key protein of the signaling pathway related to morphogenesis and virulence in *C. albicans*. Biochemical and biophysical data revealed that Cek1_Ser variant induces an increase on the protein thermal stability compared to leucine variant, without any changes on the overall secondary structure. Moreover, only the Cek1_Ser variant was autophosphorylated *in vitro* at the tyrosine residue of the ^{231}TEY^{233} motif of the activation loop. Accordingly, incorporation of a serine residue at the CUG position seems to induce a more active protein with possible implications on the Cek1 function. Further studies will be required to understand the molecular details of CUG codon ambiguous translation in the context of the Cek1 full-length protein and for *C. albicans* host recognition and pathogenicity.

CRediT author statement

Joana S. Fraga: methodology, formal analysis, investigation, writing - original draft preparation. **Zsuzsa Sarkany**: conceptualization, methodology, writing- reviewing and editing. **Alexandra Silva:** formal analysis, writing- reviewing and editing. **Inês Correia**: formal analysis, writing- reviewing and editing. **Pedro José Barbosa Pereira:** validation, resources, writing- reviewing and editing. **Sandra Macedo-Ribeiro:** conceptualization, validation, resources, supervision, writing - original draft, reviewing and editing, funding acquisition.

Acknowledgements

We thank Hugo Osório for performing mass spectrometry measurements at the i3S Proteomics Scientific Platform (part of the Portuguese Mass Spectrometry Network, integrated in the National Roadmap of Research Infrastructures of Strategic Relevance (ROTEIRO/0028/2013; LISBOA-01-0145-FEDER-022125)). We acknowledge the Biochemical and Biophysical Technologies Scientific Platform of i3S for support with biophysical characterization.

Funding

This work was funded by (i) FEDER-Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020-Operational Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through FCT/MCTES-Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior (Portugal) in the framework of project POCI-01-0145-FEDER-007274 (Institute for Research and Innovation in Health Sciences), and by (ii) FEDER through Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement in the framework of Projects Norte-01-0145-FEDER-000008 and Norte-01-0145-FEDER-000012. J.S.F. is the recipient of PhD fellowship SFRH/BD/94403/2013 from FCT/MCTES.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbapap.2019.02.004.

References

[1] R.D. Knight, S.J. Freeland, L.F. Landweber, Rewiring the keyboard: evolvability of the genetic code, Nat. Rev. Genet. 2 (2001) 49–58.
[2] R.D. Knight, L.F. Landweber, M. Yarus, How mitochondria redefine the code, J. Mol. Evol. 53 (2001) 299–313.
[3] A.R. Bezerra, A.R. Guimaraes, M.A.S. Santos, Non-standard genetic codes define new concepts for protein engineering, Life (Basel) 5 (2015) 1610–1628.
[4] H. Seligmann, Alignment-based and alignment-free methods converge with experimental data on amino acids coded by stop codons at split between nuclear and mitochondrial genetic codes, Biosystems 167 (2018) 33–46.
[5] E.C. Swart, V. Serra, G. Petroni, M. Nowacki, Genetic codes with no dedicated stop codon: context-dependent translation termination, Cell 166 (2016) 691–702.
[6] A. Ambrogelly, S. Palioura, D. Söll, Natural expansion of the genetic code, Nat. Chem. Biol. 3 (2007) 29–35.
[7] M.A.S. Santos, G. Keith, M.F. Tuite, Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5′-CAG-3′ (leucine) anticodon, EMBO J. 12 (1993) 607–616.
[8] S. Muhlhausen, P. Findeisen, U. Plessmann, H. Urlaub, M. Kollmar, A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes, Genome Res. 26 (2016) 945–955.
[9] M. Santos, P.M. Pereira, A.S. Varanda, J. Carvalho, M. Azevedo, D.D. Mateus, N. Mendes, P. Oliveira, F. Trindade, M.T. Pinto, R. Bordeira-Carrico, F. Carneiro, R. Vitorino, C. Oliveira, M.A.S. Santos, Codon misreading tRNAs promote tumor growth in mice, RNA Biol. 15 (2018) 773–786.
[10] M. Kapur, S.L. Ackerman, mRNA translation gone awry: translation fidelity and neurological disease, Trends Genet. 34 (2018) 218–231.
[11] C.R. Evans, Y. Fan, K. Weiss, J. Ling, Errors during gene expression: single-cell heterogeneity, stress resistance, and microbe-host interactions, mBio 9 (2018) (e01018–01018).
[12] M.A.S. Santos, M.F. Tuite, The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*, Nucleic Acids Res. 23 (1995) 1481–1486.
[13] I. Miranda, R. Rocha, M.C. Santos, D.D. Mateus, G.R. Moura, L. Carreto, M.A.S. Santos, A genetic code alteration is a phenotype diversity generator in the human pathogen *Candida albicans*, PLoS ONE 2 (2007) e996.
[14] A.C. Gomes, I. Miranda, R.M. Silva, G.R. Moura, B. Thomas, A. Akoulitchev, M.A.S. Santos, A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*, Genome Biol. 8 (2007) R206.

[15] I. Miranda, A. Silva-Dias, R. Rocha, R. Teixeira-Santos, C. Coelho, T. Gonçalves, M.A.S. Santos, C. Pina-Vaz, N.V. Solis, S.G. Filler, A.G. Rodrigues, *Candida albicans* CUG mistranslation is a mechanism to create cell surface variation, MBio 4 (2013) (e00285–00213).
[16] A.R. Bezerra, J. Simoes, W. Lee, J. Rung, T. Weil, I.G. Gut, M. Gut, M. Bayes, L. Rizzetto, D. Cavalieri, G. Giovannini, S. Bozza, L. Romani, M. Kapushesky, G.R. Moura, M.A.S. Santos, Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 11079–11084.
[17] R. Rocha, P.J.B. Pereira, M.A.S. Santos, S. Macedo-Ribeiro, Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 14091–14096.
[18] S. Muhlhausen, H.D. Schmitt, K.T. Pan, U. Plessmann, H. Urlaub, L.D. Hurst, M. Kollmar, Endogenous stochastic decoding of the CUG codon by competing Ser- and Leu-tRNAs in Ascoidea asiatica, Curr. Biol. 28 (2018) 2046–2057 (e2045).
[19] Z. Feketova, T. Masek, V. Vopalensky, M. Pospisek, Ambiguous decoding of the CUG codon alters the functionality of the *Candida albicans* translation initiation factor 4E, FEMS Yeast Res. 10 (2010) 558–569.
[20] X.L. Zhou, Z.P. Fang, Z.R. Ruan, M. Wang, R.J. Liu, M. Tan, F.M. Anella, E.D. Wang, Aminoacylation and translational quality control strategy employed by leucyl-tRNA synthetase from a human pathogen with genetic code ambiguity, Nucleic Acids Res. 41 (2013) 9825–9838.
[21] Q.Q. Ji, Z.P. Fang, Q. Ye, Z.R. Ruan, X.L. Zhou, E.D. Wang, C-terminal domain of Leucyl-tRNA synthetase from pathogenic *Candida albicans* recognizes both tRNASer and tRNALeu, J. Biol. Chem. 291 (2016) 3613–3625.
[22] Z. Sárkány, A. Silva, P.J.B. Pereira, S. Macedo-Ribeiro, Ser or Leu: structural snapshots of mistranslation in *Candida albicans*, Front. Mol. Biosci. 1 (2014) 1–14.
[23] P. Maiti, P. Ghorai, S. Ghosh, M. Kamthan, R.K. Tyagi, A. Datta, Mapping of functional domains and characterization of the transcription factor Cph1 that mediate morphogenesis in *Candida albicans*, Fungal Genet. Biol. 83 (2015) 45–57.
[24] E. Roman, R. Alonso-Monge, Q. Gong, D. Li, R. Calderone, J. Pla, The Cek1 MAPK is a short-lived protein regulated by quorum sensing in the fungal pathogen *Candida albicans*, FEMS Yeast Res. 9 (2009) 942–955.
[25] N.A. Gow, F.L. van de Veerdonk, A.J. Brown, M.G. Netea, *Candida albicans* morphogenesis and host defence: discriminating invasion from colonization, Nat. Rev. Microbiol. 10 (2012) 112–122.
[26] C. Csank, K. Schroppel, E. Leberer, D. Harcus, O. Mohamed, S. Meloche, D.Y. Thomas, M. Whiteway, Roles of the *Candida albicans* mitogen-activated protein kinase homolog, Cek1p, in hyphal development and systemic candidiasis, Infect. Immun. 66 (1998) 2713–2721.
[27] S. Yi, N. Sahni, K.J. Daniels, K.L. Lu, T. Srikantha, G. Huang, A.M. Garnaas, D.R. Soll, Alternative mating type configurations (a/alpha versus a/a or alpha/alpha) of *Candida albicans* result in alternative biofilms regulated by different pathways, PLoS Biol. 9 (2011) e1001117.
[28] M. Galan-Diez, D.M. Arana, D. Serrano-Gomez, L. Kremer, J.M. Casasnovas, M. Ortega, A. Cuesta-Dominguez, A.L. Corbi, J. Pla, E. Fernandez-Ruiz, *Candida albicans* beta-glucan exposure is controlled by the fungal CEK1-mediated mitogen-activated protein kinase pathway that modulates immune responses triggered through dectin-1, Infect. Immun. 78 (2010) 1426–1436.
[29] C. Herrero-de-Dios, R. Alonso-Monge, J. Pla, The lack of upstream elements of the Cek1 and Hog1 mediated pathways leads to a synthetic lethal phenotype upon osmotic stress in *Candida albicans*, Fungal Genet. Biol. 69 (2014) 31–42.
[30] R.A. Monge, E. Roman, C. Nombela, J. Pla, The MAP kinase signal transduction network in *Candida albicans*, Microbiology 152 (2006) 905–912.
[31] J. Stetefeld, S.A. McKenna, T.R. Patel, Dynamic light scattering: a practical guide and applications in biomedical sciences, Biophys. Rev. 8 (2016) 409–427.
[32] R.P. Bhattacharyya, A. Remenyi, M.C. Good, C.J. Bashor, A.M. Falick, W.A. Lim, The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway, Science 311 (2006) 822–826.
[33] M.J. Nagiec, P.C. McCarter, J.B. Kelley, G. Dixit, T.C. Elston, H.G. Dohlman, Signal inhibition by a dynamically regulated pool of monophosphorylated MAPK, Mol. Biol. Cell 26 (2015) 3359–3371.
[34] A.C. Carrera, K. Alexandrov, T.M. Roberts, The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring ATP, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 442–446.
[35] L.E.W. LaConte, S. Srivastava, K. Mukherjee, Probing protein kinase-ATP interactions using a fluorescent ATP analog, Methods Mol. Biol. 1647 (2017) 171–183.
[36] J. Beenstock, N. Mooshayef, D. Engelberg, How do protein kinases take a selfie (autophosphorylate)? Trends Biochem. Sci. 41 (2016) 938–953.
[37] P.E. Sudbery, Growth of *Candida albicans* hyphae, Nat. Rev. Microbiol. 9 (2011) 737–748.
[38] R.S. Shapiro, N. Robbins, L.E. Cowen, Regulatory circuitry governing fungal development, drug resistance, and disease, Microbiol. Mol. Biol. Rev. 75 (2011) 213–267.
[39] A. Remenyi, M.C. Good, R.P. Bhattacharyya, W.A. Lim, The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network, Mol. Cell 20 (2005) 951–962.
[40] C.N. Pace, H. Fu, K. Lee Fryar, J. Landua, S.R. Trevino, D. Schell, R.L. Thurlkill, S. Imura, J.M. Scholtz, K. Gajiwala, J. Sevcik, L. Urbanikova, J.K. Myers, K. Takano, E.J. Hebert, B.A. Shirley, G.R. Grimsley, Contribution of hydrogen bonds to protein stability, Protein Sci. 23 (2014) 652–661.
[41] D.C. Rees, A.D. Robertson, Some thermodynamic implications for the thermostability of proteins, Protein Sci. 10 (2001) 1187–1194.
[42] A. Gartner, K. Nasmyth, G. Ammerer, Signal transduction in Saccharomyces cerevisiae requires tyrosine and threonine phosphorylation of FUS3 and KSS1, Genes Dev. 6 (1992) 1280–1292.
[43] S. Ahmad, M.A. Hughes, G.L. Johnson, J.E. Scott, Development and validation of a high-throughput intrinsic ATPase activity assay for the discovery of MEKK2 inhibitors, J. Biomol. Screen. 18 (2013) 388–399.
[44] C.M. Rominger, M.D. Schaber, J. Yang, R.R. Gontarek, K.L. Weaver, T. Broderick, L. Carter, R.A. Copeland, E.W. May, An intrinsic ATPase activity of phospho-MEK-1 uncoupled from downstream ERK phosphorylation, Arch. Biochem. Biophys. 464 (2007) 130–137.
[45] P.A. Lochhead, Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation, Sci. Signal. 2 (2009) (pe4).
[46] P. Cote, T. Sulea, D. Dignard, C. Wu, M. Whiteway, Evolutionary reshaping of fungal mating pathway scaffold proteins, mBio 2 (2011) (e00230–00210).

SUPPPLEMENTARY DATA

Genetic code ambiguity modulates the activity of a *C. albicans* MAP kinase linked to cell wall remodeling

Joana S. Fraga [1,2,3], Zsuzsa Sárkány [1,2#], Alexandra Silva [1,2], Inês Correia [1,2], Pedro José Barbosa Pereira [1,2], Sandra Macedo-Ribeiro [1,2,*]

[1] IBMC-Instituto de Biologia Molecular e Celular, Universidade do Porto, 4200-135 Porto, Portugal
[2] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal
[3] ICBAS-Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, 4050-313 Porto, Portugal

[#]Present address: LEPABE – Departamento de Engenharia Química, Faculdade de Engenharia da Universidade do Porto, 4200-465 Porto, Portugal

[*]To whom correspondence should be addressed: Sandra Macedo-Ribeiro, IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, 4200-135 Porto, Portugal; sribeiro@ibmc.up.pt; Tel. (+351) 226 074 953

Table S1. Oligonucleotides used to amplify the active domain of Cek1 and for site-directed mutagenesis.

Oligonucleotide	Sequence (5´ to 3´)	Use
Cek1_For	CCGGCCATGGGCAGTCCACGTCAGGTCAGC	gene amplification
Cek1_Rev	CCGGGGTACCTTATTTCATGATTTCTTCATACAG	gene amplification
Cek1_Leu_For	TGGAGTTCAGCAACAAGTTTAAGGGCTTCAGAT CGCGATG	mutagenesis
Cek1_Leu_Rev	CATCGCGATCTGAAGCCCTTAAACTTGTTGCTGA ACTCCA	mutagenesis
Cek1_TV_For	GAGGACAATTATGGCTTTATGGTCGAATACGTTG CAACCCGTTG	mutagenesis
Cek1_TV_Rev	CAACGGGTTGCAACGTATTCGACCATAAAGCCA TAATTGTCCTC	mutagenesis
Cek1_YF_For	CAATTATGGCTTTATGACCGAATTCGTTGCAACC CG	mutagenesis
Cek1_YF_Rev	CGGGTTGCAACGAATTCGGTCATAAAGCCATAA TTG	mutagenesis
Cek1_TV/YF_For	GGAGGACAATTATGGCTTTATGGTCGAATTCGTT GCAACCCGTTGGTA	mutagenesis
Cek1_TV/YF_Rev	TACCAACGGGTTGCAACGAATTCGACCATAAAG CCATAATTGTCCTCC	mutagenesis
Cek1_KR_For	GCAACAGAAAGTTGCGATAAGGAAAATCGAACC GTTTGAAC	mutagenesis
Cek1_KR_Rev	GTTCAAACGGTTCGATTTTCCTTATCGCAACTTT CTGTTGC	mutagenesis
Cek1_TD_For	CAACGGGTTGCAACGTATTCGTCCATAAAGCCAT AATTGTCCTC	mutagenesis
Cek1_TD_Rev	GAGGACAATTATGGCTTTATGGACGAATACGTT GCAACCCGTTG	mutagenesis

Table S2. Extinction coefficient of Cek1 variants and Cek1_Ser mutants based on amino acid sequence.

Protein	Extinction coefficient (M^{-1} cm^{-1})
Cek1_Ser	109,670
Cek1_Leu	109,670
Cek1^{TV}_Ser	109,670
Cek1^{YF}_Ser	108,180
Cek1$^{TV/YF}$_Ser	108,180
Cek1^{KR}_Ser	109,670
Cek1^{TD}_Ser	109,670

Table S3. Thermal stability of Cek1_Ser and Cek1_Leu variants incubated with ATP. Data shown are mean ± standard error of the mean (n=3). Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ for a 95% confidence interval comparing each variant, Cek1_Ser or Cek1_Leu with ATP ($p>0.05$ non-significant (ns), *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

Protein	Melting temperature (°C) (mean ± SD)	Significance (adjusted *p* value)
Cek1_Ser	43.1 ± 0.7	ns (0.9999)
Cek1_Ser with ATP	43.2 ± 0.8	
Cek1_Leu	40.5 ± 0.3	ns (0.9999)
Cek1_Leu with ATP	39.9 ± 0.9	

Table S4. Thermal stability of Cek1_Ser and Cek1_Leu variants and Cek1_Ser mutants. Data shown are mean ± standard error of the mean (n=3). Statistics were analyzed using one-way ANOVA (Tukey multicomparison test) $p<0.05$ for a 95% confidence interval comparing all protein with Cek1_Ser ($p>0.05$ non-significant (ns), *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

Protein	Melting temperature (°C) (mean ± SD)	Significance (adjusted *p* value)
Cek1_Ser	43.1 ± 0.7	-
Cek1_Leu	40.5 ± 0.3	** (0.007)
Cek1^{TV}_Ser	43.0 ± 0.7	ns (0.9999)
Cek1^{YF}_Ser	43.3 ± 0.5	ns (0.9999)
Cek1$^{TV/YF}$_Ser	44.4 ± 0.6	ns (0.9999)
Cek1^{KR}_Ser	47.2 ± 0.6	**** (0.0001)
Cek1^{TD}_Ser	46.4 ± 0.2	** (0.0014)

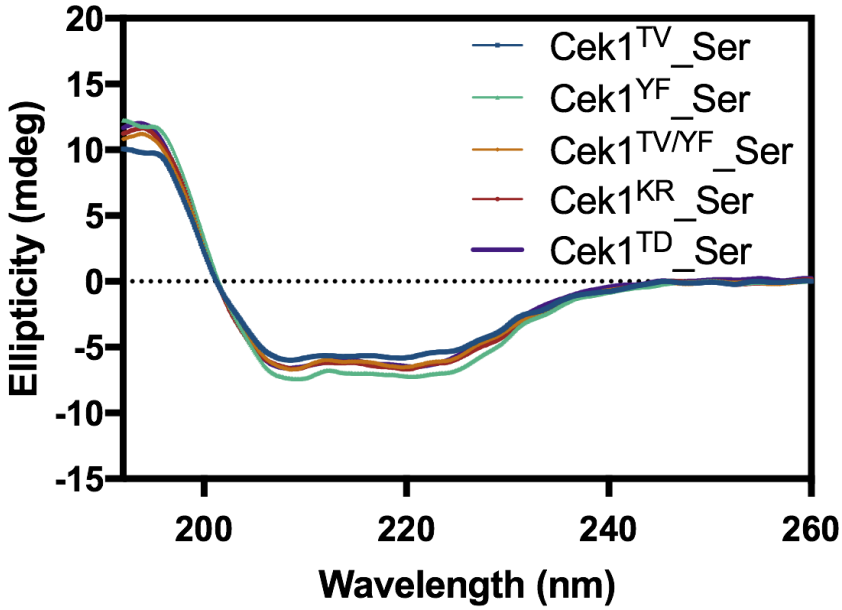


Figure S1. Far-UV CD spectra of Cek1_Ser mutants (Cek1^{TV}_Ser, Cek1^{YF}_Ser, Cek1$^{TV/YF}$_Ser, Cek1^{KR}_Ser, Cek1^{TD}_Ser). The mutations did not induce any major differences on the secondary structure content of the enzyme.

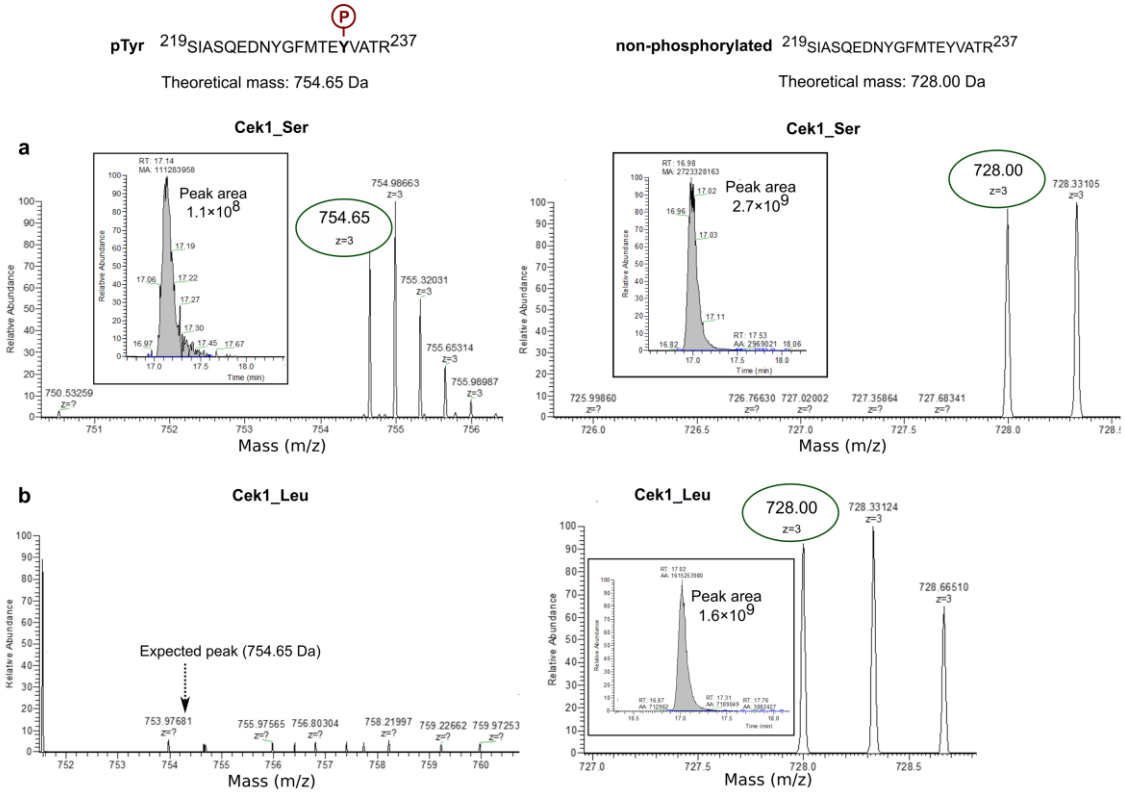


Figure S2. LC-MS/MS analysis of the Cek1 variants. Representative extracted ion chromatogram with the peak area and the respective m/z for the peptide of interest (SIASQEDNYGFMTEYVATR). The integrated peak areas (insets) of the phosphorylated peptide and of its equivalent non-phosphorylated version were used to calculate an apparent percentage of phosphorylation of Cek1. An arrow marks the position of the peptide ion corresponding to the expected mass of a single phosphopeptide on the tyrosine residue, whose theoretical mass (754.65220), as well as that of its non-phosphorylated counterpart (727.99676), was calculated using the online server ProteinProspector (MS-isotope). (*a*) MS spectrum for the Cek1_Ser variant: Left panel - phosphopeptide (SIASQEDNYGFMTEpYVATR, peak area 1.1×10^8); Right panel: the equivalent non-phosphorylated peptide (peak area 2.7×10^9). Comparative analysis of the peak areas corresponding to phosphorylated and non-phosphorylated peptide fragments, suggested that only a small percentage of Cek1_Ser is autophosphorylated on tyrosine 233. *(b)* MS analysis of the Cek1_Leu variant revealed absence of the phosphorylated peptide (Left panel), while the non-phosphorylated peptide was detectable (Right panel).

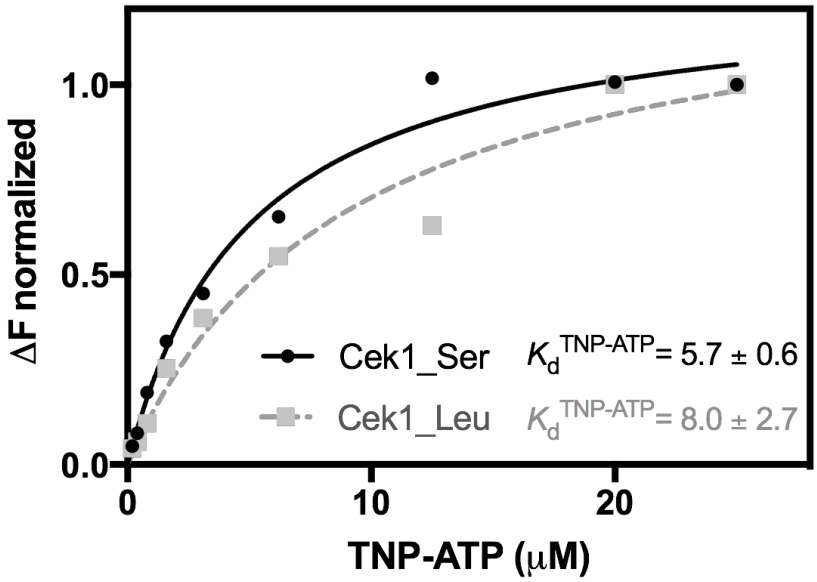


Figure S3. Binding of TNP-ATP to Cek1 variants. Both Cek1 variants bind the nucleotide with comparable affinity. The data are representative of three independent experiments.

Supplementary materials and methods

Mass spectrometry

Protein identification and quantitation was performed by nanoLC-MS/MS, using an equipment composed by an Ultimate 3000 liquid chromatography system coupled to a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Samples were loaded onto a trapping cartridge (Acclaim PepMap C18, Thermo Scientific) and after 3 min loading the trap column was switched in-line to a EASY-Spray column (PepMap RSLC, C18, Thermo Scientific). Separation was achieved using a formic acid (FA: 0.1% (v/v)) and acetonitrile (AC: 80% (v/v)) gradient (10 min (2.5% to 10% AC), 180 min (10% to 35% AC), 30 min (35% to 99% AC) and 30 min (hold 99% AC)). Subsequently, the column was equilibrated with 2.5% AC for 40 min. Data acquisition was controlled with Xcalibur 4.0 and Tune 2.8 software (Thermo Scientific). The mass spectrometer was operated in data-dependent (dd) positive acquisition mode alternating between a full scan (m/z 380-1580) and subsequent HCD MS/MS of the 10 most intense peaks from full scan (normalized collision energy of 27%). The ESI spray voltage was 1.9 kV. The raw data were processed using Proteome Discoverer 2.2.0.388 software (Thermo Scientific) and searched against the SwissProt/UniProt database. The Sequest HT search engine was used to identify tryptic peptides. The ion mass tolerance was 10 ppm for precursor ions and 0.02 Da for fragment ions. The maximum allowed number of missing cleavage sites was set to 2. Cysteine carbamidomethylation was defined as constant modification. Methionine oxidation, phosphorylation (S, T and Y) and protein N-terminus acetylation were defined as variable modifications. Peptide confidence was set to high. The processing node Percolator was enabled with the following settings: maximum delta Cn 0.05; decoy database search target FDR 1%, validation was based on q-value.

Fluorescence measurements

The nucleotide binding assay was performed with the fluorescent analog of ATP, TNP-ATP (Jena Bioscience). The fluorescence emission scanning spectra were obtained with a Fluoromax-4 spectrofluorometer (HORIBA) in protein buffer at 20°C using a quartz cuvette (10x10 mm). Samples were excited at 410 nm and emission spectra scanned from 500-600 nm, with excitation and emission slits set at 6 nm. For fluorescence titrations, Cek1 variants (5 μM) were mixed with TNP-ATP (0-25 μM) and emission was recorded at 547 nm wavelength. The fluorescence of TNP-ATP was subtracted from the total fluorescence (enzyme plus TNP-ATP) to yield the specific fluorescence enhancement and data were normalized to the maximum fluorescence and analyzed using the nonlinear curve fitting to determine TNP-ATP binding affinity (K_d $^{TNP-ATP}$) with Prism7 (GraphPad Software).