Universidade de Aveiro Departamento de Biologia
Ano 2011-2012

VANESSA ISABEL DIAS RIBEIRO DE CARVALHO

Cek1 e Ras1 de *C. albicans*: clonagem, expressão e purificação

C. albicans Cek1 and Ras1: cloning, expression and purification

Universidade de Aveiro Departamento de Biologia
Ano 2011-2012

VANESSA ISABEL DIAS RIBEIRO DE CARVALHO

Cek1 e Ras1 de *C. albicans*: clonagem, expressão e purificação
C. albicans Cek1 and Ras1: cloning, expression and purification

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Aplicada – Ramo de Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Sandra Macedo-Ribeiro, Professor Auxiliar Convidado do Departamento de Biologia da Universidade de Aveiro e sob a co-orientação da Doutora Zsuzsa Sárkány, Investigadora do Instituto de Biologia Molecular e Celular

o júri

presidente

Prof.ª Doutora Maria Adelaide de Pinho Almeida
Professora Auxiliar da Universidade de Aveiro

Doutor Pedro Miguel Dias Castanheira
Investigador do Biocant

Prof.ª Doutora Sandra de Macedo Ribeiro
Professora Auxiliar Convidada da Universidade de Aveiro

Doutora Zsuzsa Sárkány
Investigadora do Instituto de Biologia Molecular e Celular

Agradecimentos

Gostaria de agradecer a todos que, de uma forma ou de outra, contribuíram para que este trabalho fosse realizado.

Agradeço à Professora Doutora Sandra Macedo-Ribeiro pela sua total disponibilidade em aceitar-me como sua aluna de mestrado e integrar-me no seu grupo de trabalho.

Agradeço à Doutora Zsuzsa Sárkány pela sua co-orientação, por todos os conhecimentos transmitidos, pela sua amizade e pelo seu incansável esforço para que este trabalho fosse realizado e concluído com todo o rigor científico. À Doutora Sara Fernandes e Doutor Bruno Almeida pela sua disponibilidade e ajuda na elaboração do trabalho e durante a minha estadia no laboratório.

A todos os restantes membros do laboratório que sempre ajudaram e estiveram sempre disponíveis para que questões práticas do trabalho fossem resolvidas entre estes: Doutor Pedro Barbosa Pereira, Ana Figueiredo, Alexandra Marques, Joana Santos, Ana Marisa Simões, Isabel Abreu e Maria Macedo. Agradeço-te especialmente a ti, Joana F., por sempre estares disposta a ajudar e seres uma excelente colega de trabalho.

Agradeço, também, à Professora Doutora Isabel Abrantes, ao Luís e à Joana por todo o vosso incentivo e ajuda na escolha a fazer quando ingressei neste mestrado.

Quero agradecer à Ivânia, à Aida e à Márcia (apesar de longe) por todo o vosso apoio em prosseguir e por me ajudarem sempre que precisei nos momentos mais difíceis.

E por fim, um agradecimento particular aos meus pais e aos meus irmãos por todo o esforço emocional e financeiro que este período acarretou e por estarem sempre lá a apoiar-me. E a ti meu companheiro e amigo, por todas as horas de ajuda e apoio em tudo o que foi possível.

palavras-chave

Candida albicans, ambiguidade CUG, serina, leucina, Cek1, Ras1.

resumo

Candida albicans é um fungo polimórfico e patogénico que reside de forma comensal nas superfícies mucosas humanas. Este fungo apresenta um código genético ambíguo, uma vez que, o codão universal leucina CUG é predominantemente traduzido como serina (97%), mas também como leucina (3%). A análise de proteínas de *C. albicans* que contêm resíduos CUG em importantes posições funcionais, revela que a ambiguidade do codão modela a função da proteína e poderá ter um papel determinante nas vias de sinalização associadas a mudanças morfológicas e patogénicas.

Com o presente estudo pretende-se investigar o efeito da leucina e da serina nas posições CUG (ambiguidade CUG) na estrutura e função de duas proteínas chave nas vias de sinalização de *C. albicans*, a Ras1 (GTPase) e a Cek1 (Cinase). Estas regulam a transcrição dos genes associados com mudanças morfológicas, patológicas e, por outro lado, contêm resíduos CUG numa posição estritamente conservada e com relevo funcional.

Neste contexto foi possível clonar com sucesso genes sintéticos para os centros ativos da Ras1 e Cek1 (variantes de serina e de leucina para o codão CUG) em vetores que apresentam diferentes caudas de solubilidade (MBP, NusA, Trx, ZTag2 e Gb1). Foram desenvolvidos protocolos de alta expressão bacteriana e de purificação para os domínios ativos Ras1 (ligado à Gb1) e Cek1 (ligado à MBP). A análise dos resultados de purificação analítica e de "Dynamic Light Scaterring" demonstraram que as proteínas recombinantes se apresentam na forma monomérica.

Ensaios de cristalização estão a ser realizados esperando-se determinar as estruturas tridimensionais das proteínas por cristalografia de raio-X. As estruturas da Cek1 e Ras1 com leucina e serina nas posições CUG, conjuntamente com uma análise meticulosa da sua estabilidade e função *in vitro*, irão fornecer informações importantes sobre o papel estratégico da ambiguidade natural do codão.

keywords *Candida albicans*, CUG ambiguity, serine, leucine, Cek1, Ras1.

abstract

The polymorphic fungal pathogen *Candida albicans* has an ambiguous genetic code, as the universal leucine CUG codon is predominantly translated as serine (97%) but also as leucine (3%). Analysis of the rare *C. albicans* proteins containing CUG-encoded residues in functionally relevant positions reveals that codon ambiguity shapes protein function and might have a pivotal role in signaling cascades associated with morphological changes and pathogenesis.

The present study investigates the effect of leucine or serine at CUG positions (CUG ambiguity) in the structure and function of two key effectors of signaling cascades in *C. albicans*, Ras1 (GTPase) and Cek1 (protein kinase), which regulate the transcription of genes associated with morphological changes and pathogenesis. These two proteins contain a CUG residue in a strictly conserved and functionally relevant position.

Synthetic genes coding for the active domains of Ras1 and Cek1 (serine and leucine variants for the CUG codon) were successfully cloned into expression vectors carrying different solubility partners (MBP, NusA, Trx, ZTag2 and Gb1). Furthermore, using an incomplete factorial approach, high level bacterial expression and purification protocols for the active domains of Ras1 (in fusion with Gb1) and Cek1 (in fusion with MBP) were developed. Analytical size exclusion chromatography (SEC) and dynamic light scattering (DLS) results indicate that both recombinant proteins are monomeric.

Crystallization trials must be done aiming for the determination of their three-dimensional structures by X-ray crystallography. The structures of Ras1 and Cek1 with serine or leucine at CUG positions, together with a thorough analysis of their stability and function *in vitro*, will provide valuable insights into a possible strategic role of natural codon ambiguity.

Contents

Abbreviations

A	Adenine
APS	Ammonium PerSulfate
C	Citosine
C. Albicans	*Candida albicans*
Cek1_Leu	Cek1 Leucine variant for CUG codon
Cek1_Ser	Cek1 Serine variant for CUG codon
CV	Column Volumes
D	Diffusion coefficient
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediamine Tetraacetic Acid
EIF4E	Translation initiation factor 4E
EtBr	Ethidium Bromide
G	Guanine
GTP	Guanosine triphosphate
HIV	Human immunodeficiency virus
IF	Incomplete factorial
IMAC	Immobilized affinity chromatography
IPTG	Isopropyl-β-D-thigalactoside
LB	Luria - Bertani Medium
Leu	Leucine
LeuRS	Leucyl-tRNA synthase
MAPK	Mitogen Active Protein Kinase
MI	Morphological Index
mRNA	messenger Ribonucleic Acid
PCR	Polymerase Chain Reaction
PEG	Polyethyleneglycol
P_i	Inorganic phosphate
PI	Polydispersity Index
PMSF	Phenylmethylsulfonyl fluoride
Ras1_Leu	Ras1 Leucine variant for CUG codon

Ras1_Ser	Ras1 Serine variant for CUG codon
R_H	Hydrodynamic radius
RNA	Ribonucleic Acid
S. cerevisiae	*Saccharomyces cerevisiae*
SAP	Shrimp Alkaline Phosphatase
SDM	Site Directed Mutagenesis
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size-exclusion chromatography
Ser	Serine
SerRS	Seryl-tRNA synthase
TAE	Tris-Acetate-EDTA
TEMED	Tetramethylethylenediamine
$tRNA_{CAG}^{Leu}$	Leucine-tRNA
$tRNA_{CAG}^{Ser}$	Serine-tRNA
U	Uracile
V_0	Void volume of the column
V_e	Elution volume
V_t	Total bed volume

1. Introduction

1.1. *Candida albicans*: a human pathogen

The Yeast *Candida albicans* (*C. albicans*) is the most common fungus present in the human skin and mucosal (1), and, although commensal, it also causes superficial infections and life-threatening systemic disease (2). With the increase of immunocompromised individuals due to HIV infection, organ transplantation, and application of chemotherapy, invasive candidiasis has become a serious public health problem in the recent decades (2). This extremely serious medical condition has a mortality rate of about 40% (1,3), and about 90% of the invasive infections are caused by five species of *Candida*: *albicans*, *glabrata*, *parapsilosis*, *tropicalis* and *krusei* (3).

The pathogenic nature of *C. albicans* is attributed to: its multiple surface structures that mediate adhesion to epithelial cells; the capacity to switch its phenotype by regulating gene expression; the ability to introduce hydrolytic enzymes that damage host cells; the capacity to produce biofilms resistant to the antifungal drugs; metabolic and stress adaptation during infection and the capacity to switch between yeast, hyphae and pseudohyphal forms (1). When growth begins, elongated cells are produced and pseudohyphal/hyphal can be distinguished by their general sizes and shapes, and according to their morphological index (MI). The MI measures the relative lengths and septal diameters of the cells (3,4). Other morphological differences that can be mentioned are that hyphal cells are thinner than pseudohyphal cells and have parallel walls without constrictions at the septum site, whereas pseudohyphal cells are bigger and form constrictions between elongated sprout (5) (Figure 1).

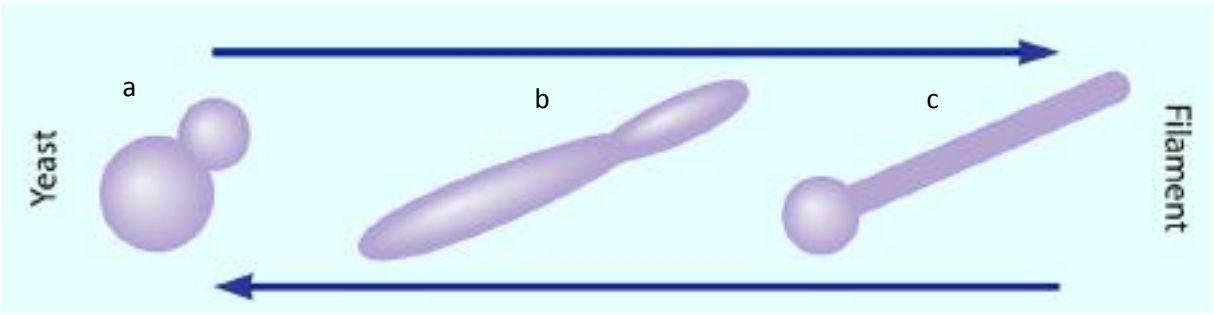


Figure 1- Distinct morphological forms of *Candida albicans* between yeast (a), pseudohyphal (b) and hyphal (c). Figure adapted from Shapiro et al., 2011.

At the base of this morphological types are differences at the molecular level, namely in cell cycle regulation (for review Shapiro et al., 2011). As stated to above, changes in *C. albicans* morphological state are important for the virulence of this fungus. Even though some studies

indicate that both yeast and filamentous states are related with virulence (3,6), filamentous forms are generally accepted as being responsible for tissue invasion and a more deep infection, whereas yeasts are regarded as important for the first state of dissemination and the subsequent infection processes (7). To induce disease, yeast cells adhere to both oral and intestinal epithelia, quickly change its forms to hyphal growth (1) and invade endothelial cells leading to tissue damage. However, the fungus is normally benign and it is well tolerated by the immune system. This morphological switching is regulated by the environmental conditions that the fungus encounters in the medium of invasion and is influenced by the microbiological flora present in the host (1). The host specific immune response to the invasion is mediated by macrophages (8).

1.2. The CUG ambiguity in *C. albicans*

The discovery of the genetic code lead Crick to postulate different rules for the transmission of genetic information, one of them was that the genetic code was frozen during time (9). For this reason the first discoveries about how the genetic code was changing were not well accepted by the scientific community that described them as nature aberrations (10). However, later studies lead to discover that not all changes had a negative impact in nature and that the manipulation of the genetic code rules could have a positive impact (11–16). Many microorganisms such as phages, bacteria or yeast are known to suffer modifications in their phenotypes as a consequence of induced-genetic code ambiguity resulting in a high adaptation potential (14).

In this context, a unique genetic code was described in several species of *Candida* and *Debaryomyces* genera (CTG clade) in which the leucine CUG codon is ambiguously decoded as serine (ser) and leucine (leu), being this sense-to-sense reassignment a characteristic of this two genera, in which canonical leucine-identity of nuclear CUG codons was replaced to serine. This ambiguity has been proven to be a consequence of a mutant serine-tRNA ($tRNA_{CAG}^{Ser}$) which has an anticodon (5'-CAG-3') that matches with the CUG codon from mRNA sequence (17,18). This $tRNA_{CAG}^{Ser}$ gene has evolved from $tRNA_{CGA}^{Ser}$ gene through insertion of an adenosine (A) in the middle position of CGA anticodon which created the CAG anticodon sequence coupled to the mutant $tRNA_{CAG}^{Ser}$ as represented in Figure 2 (12,19).

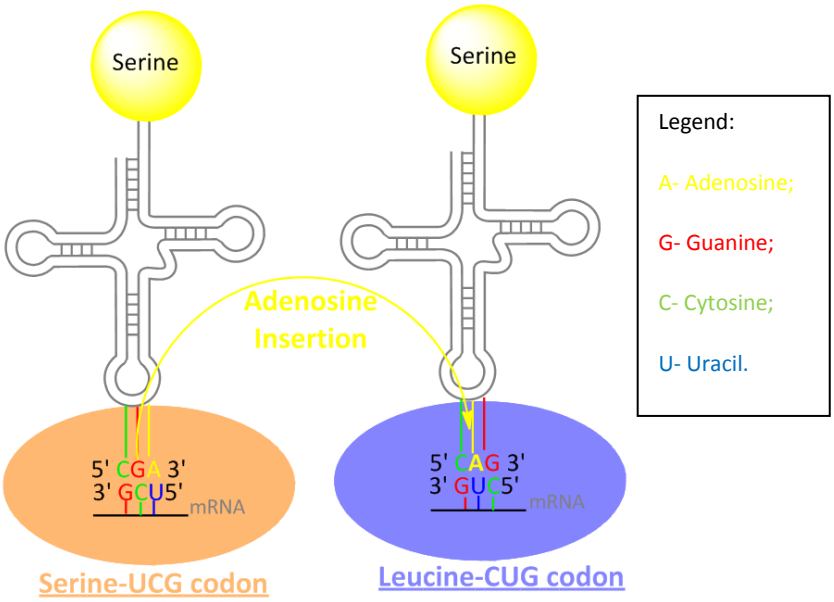


Figure 2- Scheme of the adenosine insertion in the anticodon of $tRNA_{CGA}^{Ser}$ leading to a mutant $tRNA_{CAG}^{Ser}$.

Consequently, in the yeast ancestral CUG codons were ambiguously decoded as serine by mutant $tRNA_{CAG}^{Ser}$ and as leucine by $tRNA_{CAG}^{Leu}$, a tRNA which was later lost during *Candida spp* evolution (12). This created a unique contradiction to the genetic code rules since the same codon was translated into two different amino acids.

The CUG identity alteration from leucine to serine was initiated 272±25 million years ago evolving gradually, until our days that CUG codon is ambiguously decoded as serine and as leucine (12). This is explained by the fact that the novel $tRNA_{CAG}^{Ser}$ has suffered mutations during evolution that affected the recognition and leucylation by the Leucyl-tRNA synthase (LeuRS). Accordingly, the $tRNA_{CAG}^{Ser}$ is recognized by the Seryl-tRNA synthase (SerRS) and LeuRS giving rise to two distinct tRNA species, namely Ser-$tRNA_{CAG}^{Ser}$ and Leu-$tRNA_{CAG}^{Ser}$. These tRNAs participate in protein synthesis making CUG ambiguously decoded (16) (Figure 3). However, SerRS is the main charging enzyme of $tRNA_{CAG}^{Ser}$ and the LeuRS is a poor competitor (16) (Figure 3).

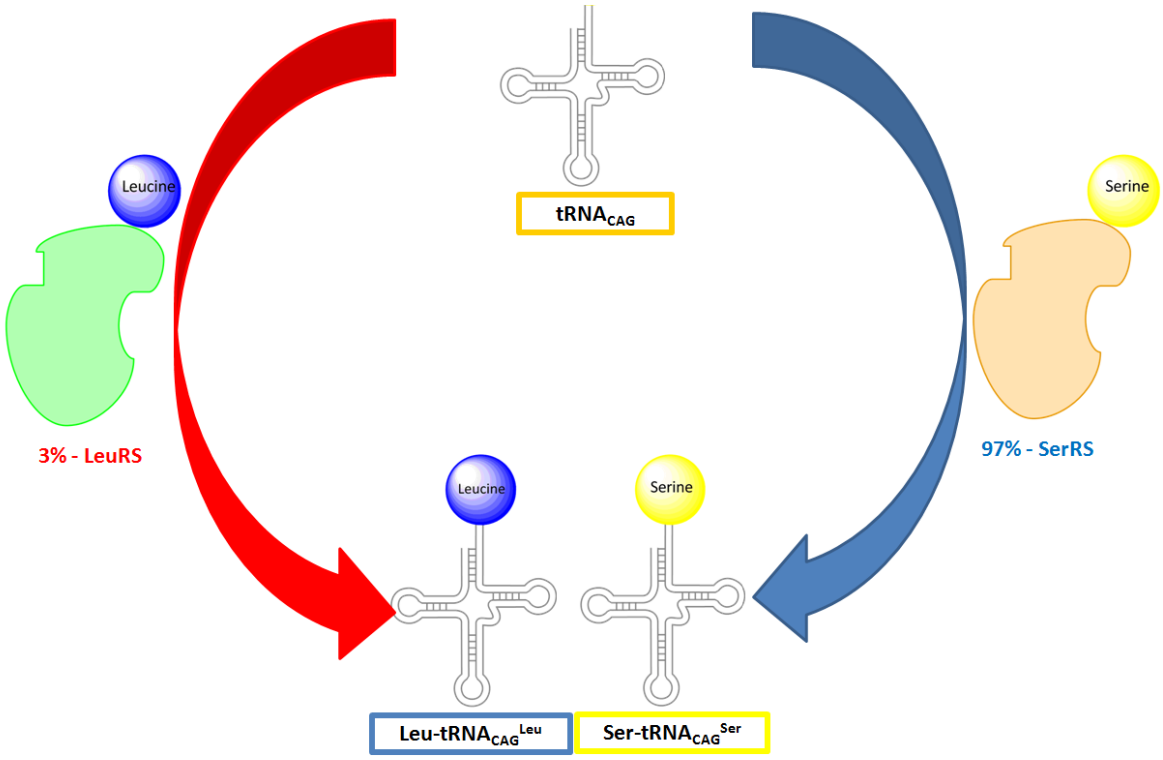


Figure 3- Schematic representation of the competition between the SerRS and the LeuRS for charging the $tRNA_{CAG}$.

In vitro serine and leucine are incorporated into the proteome with different levels of efficiency: 97% of $tRNA_{CAG}^{Ser}$ is charged with serine and only 3% is charged with leucine (20). It has been previously demonstrated that leucine misincorporation can be increased up to 28% in *C. albicans* without visible effects on the growth rate but with an impressive impact in cell morphology (16). In agreement with these findings, the analysis of CUG-containing genes in *C. albicans* showed that in most of the genes the CUG encoded residues is neutral and in general it does not affect the protein folding and consequently its function (21).

Although seemingly tolerated, those surface Leu-for-Ser replacements will hardly occur without subtle structural and functional consequences. In agreement, functional differences were recently reported in *C. albicans* eukaryotic translation initiation factor 4E (EIF4E) upon incorporation of serine or leucine within a non-conserved site at the protein surface (22). Moreover, considering that surface residues are often mediating macromolecular interactions it can be expected that CUG ambiguity in *C. albicans* will have an impact on protein activity and regulation. In the SerRS of *C. albicans*, the amino acids codified by CUG codon are located at the protein surface and

although it does not affect the overall structure, it has functional consequences and the isoform with CUG-encoded leucine is more active (23).

It was previously shown that the partial reversion of CUG decoding from serine to leucine in *C. albicans* up-regulated the expression of genes involved in cell adhesion and hyphal growth as well as increased the secretion of proteases and phospholipases (13,16) features associated with virulence and infection (24). In addition, a small number of proteins associated with morphogenesis and virulence in *Candida species*, were recently identified, particularly proteins involved in Mitogen Active Protein Kinase (MAPK) Pathway (23) as Ras1 and Cek1. These two C albicans proteins are crucial in signal transduction pathways associated with morphological switching and virulence, and moreover contain CUG encoded residues in strictly conserved sites where the replacement of Ser-to-Leu should have a bigger functional impact. Supporting this idea, increased leucine misincorporation up-regulated the expression of an adhesin (13), which is positively controlled by Ras1-dependent signalling cascades (25-26).

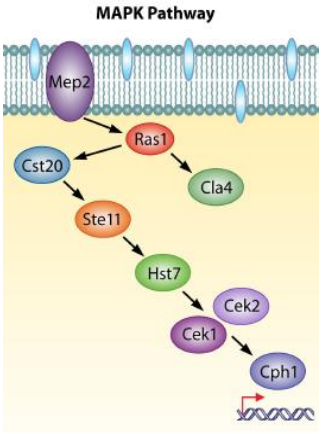


Figure 4- Key cellular signaling cascade regulating the morphogenesis and the vegetative growth in *Candida albicans*. Numerous signaling pathways regulate *C. albicans* morphogenesis being one of them the Mitogen Active Protein Kinase (MAPK) Pathway. Schematic representation taken from the article Shapiro et al., 2011.

1.3. Ras1 and Cek1 proteins from *C. albicans*

Ras1 is a guanine nucleotide binding protein (GTPase) which acts downstream of Mep2, a transmembrane ammonium permease that regulates yeast-to-hypha morphogenesis in response to nitrogen starvation. The small GTPase, is as a sensor of environmental stress signals which activates two signalling cascades the (cAMP)-dependent protein kinase pathway and the MAP kinase-dependent pathway. The activation of cAMP and MAPK pathways culminates in morphological changes and disruption of genes encoding proteins, resulting in reduced virulence (25–27). Another member of the MAPK signal pathway is Cek1, an ERK-family protein kinase. Cek1 is mainly involved in: wild-type yeast-hyphal switching, mating efficiency and virulence (28). The Cst20p-Hst7p-Cek1p-Cph1p MAPK pathway (Figure 4) is the regulator of mating, and invasive hyphal growth under nutrition starvation (28–30).

Cek1 and Ras1 were identified in *C. albicans* as being crucial to the signal transduction mechanism of morphological switching and virulence (25,28,31) and contain CUG-encoded residues. The three-dimensional models of Ras1 and Cek1 have been analyzed in detail using homology models (23). Ras1 GTPase has CUG encoded residue in a conserved Ser/Thr position and Cek1 protein kinase has a CUG-encoded residue in a strictly conserved Ser position (23) (Figure 5 and Figure 6).

Figure 5- Amino acid sequence alignment of Cek1 from different *Candida species (C. albicans, C. tropicalis, C. parapsilosis, C. guilliermondii, C. lusitaniae, C. glabrata*) and *S. cerevisiae*. The CUG encoded residues are boxed in green except for *C. albicans*, which is boxed in blue. Strictly identical residues are shown in red background and conserved residues are shown in red type.

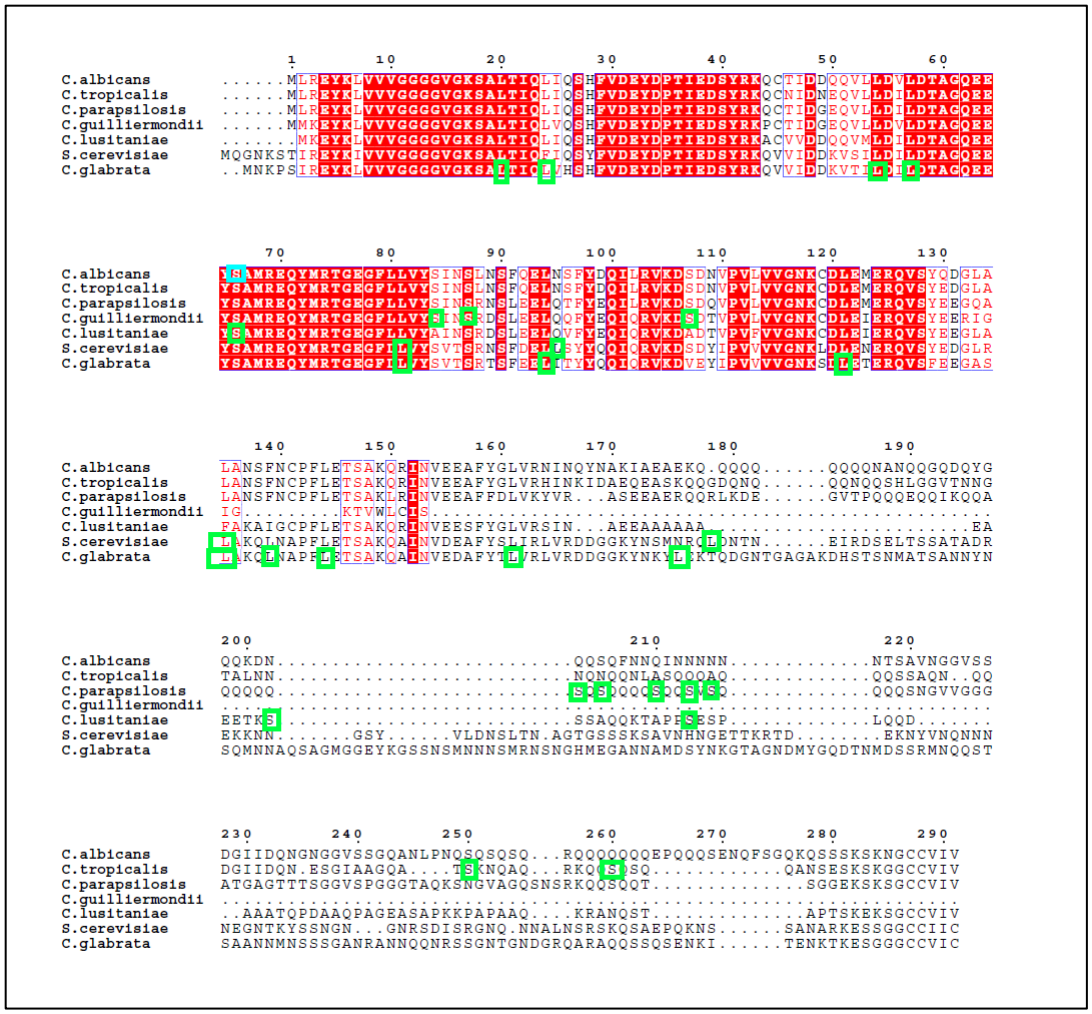


Figure 6- Amino acid sequence alignment of Ras1 from different *Candida species (C. albicans, C. tropicalis, C. parapsilosis, C. guilliermondii, C. lusitaniae, C. glabrata*) and *S. cerevisiae*. The CUG encoded residues are boxed in green except for *C. albicans*, which is boxed in blue. Strictly identical residues are shown in red background and conserved residues are shown in red type.

In *C. albicans* Ras1 GTPase, the CUG-encoded Ser66 is part of the conserved active site whose flexibility is crucial for cycling between active and inactive states and for binding Ras1 effectors and regulators (32,33) (Figure 7). Mutations in this region, locking Ras1 in the active GTP-bound conformation, have been indentified in human tumors (33) and shown to generate transient activated phenotypes in *S. cerevisiae* (34). Likewise, in the *C. albicans* Cek1 protein kinase model (23), the conserved CUG-encoded Ser199 is facing the ATP-binding-pocket, and hence leucine incorporation at this site will likely interfere with its catalytic activity (Figure 7). *C. albicans* Ras1 and Cek1 proteins are plausible early molecular effectors of morphological switching upon

increased leucine misincorporation (e.g., upon pH or oxidative stress) (16). In this work we purpose to develop protocols for the expression and purification of Cek1 and Ras1 active domains in high yields, with the purpose of determining the structural and functional consequences of serine to leucine exchange (CUG-encoded residue) in Ras1 and Cek1 of *C. albicans*.

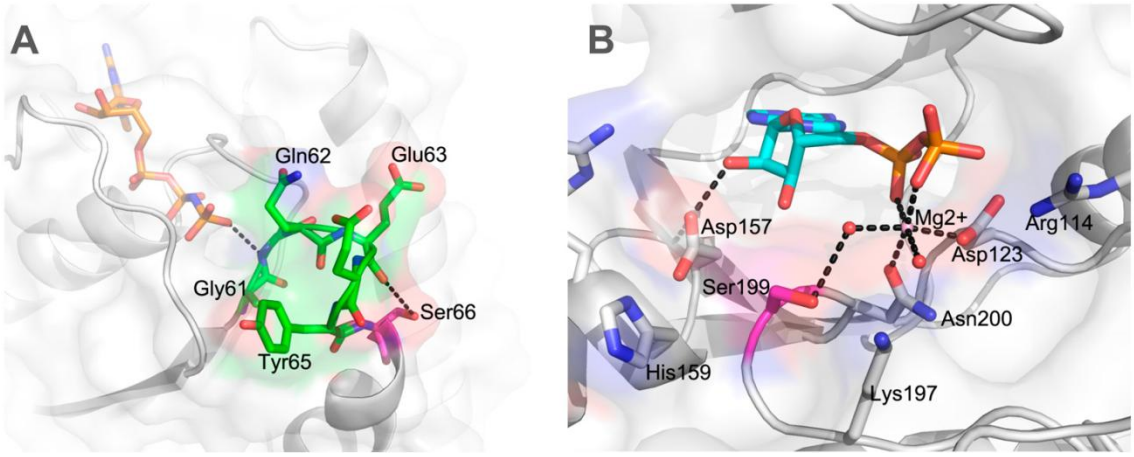


Figure 7- Functional impact of serine/leucine incorporation in *C. albicans* Ras1 and Cek1 proteins. (A) Homology model of *C. albicans* Ras1 GTPase based on the human Ras structure 3K8Y (35). The CUG-encoded Ser66 has carbon atoms represented in magenta and the active site bound GNP (no hydrolyzable GTP) is shown in orange. In human Ras, the incorporation of leucine at position 66 will likely interfere with GTP hydrolysis. (B) Homology model of *C. albicans* Cek1 protein kinase based on the yeast Fus3p structure 2B9F (36). The CUG-encoded Ser 199 (in magenta) is inside ATP-binding pocket (carbon atoms of bound ADP molecule in blue). Introduction of a leucine at this position is expected to compromise the Cek1 active site. Residues represented as sticks are color-coded (nitrogen in blue, oxygen in red phosphate in orange, carbon in gray unless stated otherwise), hydrogen bonds are represented by dashed black lines. Figure and legend adapted from article R. Rocha, Pereira, et al., 2011.

2. Objectives

So far is known that Ras1 (GTPase) and Cek1 (protein kinase), which are crucial in the signal transduction mechanism of morphological switching and virulence in *C. albicans*, contain only one CUG residue in a strictly conserved and functionally relevant position. Hence the CUG ambiguity might regulate the function of Ras1 and Cek1. To confirm this possible strategic role of CUG ambiguous decoding in Ras1 and Cek1, comparative structural and functional *in vitro* studies are essential. These studies require high amount of pure soluble proteins.

The main goals of the present work were to clone, express and purify the active domains of Ras1 and Cek1, containing either serine or leucine at the CUG encoded position. We aimed to design and develop bacterial expression and purification protocols, in order to get high amount of pure soluble recombinant proteins. This is crucial for further structural and functional studies in order to determine consequences of serine or leucine incorporation within the CUG-encoded residue in *C. albicans* Ras1 and Cek1.

3. Materials and methods

3.1. Cloning

Synthetic genes of Ras1_Ser (serine variant for CUG codon) and Cek1_Leu (leucine variant for CUG codon position) were purchased (Eurofins MWG Operon) with the codon optimized for *Escherichia coli* in a standard vector with restriction sites NcoI and Acc65I for 5' and 3' end respectively (see section 6, Figure 36 and Figure 37).

3.1.1. Polymerase Chain Reaction

The gene fragments for the active domains of Ras1_Ser and Cek1_Leu were amplified from the standard vectors by Polymerase Chain Reaction (PCR). Two primers were designed for each PCR reaction. The forward primer (extends in PCR from the start codon towards the stop codon) contained an NcoI restriction site; and the reverse primer (extends from the stop codon towards the start codon) included an Acc65I restriction site (see Table 1). Using these primers in the PCR reactions, the gene fragments (for the active domains of Cek1 and Ras1) were amplified with the restriction sites NcoI and Acc65I for 5' and 3' end respectively.

Table 1 – Primer sequences to amplify gene fragments for the active domains of Ras1_Ser and Cek1_Leu. Restriction site for NcoI and Acc65I are presented in blue and green respectively.

Primer	Sequence (5´→3´)	GC (%)	Length (nt)
Cek1_for	CCGGCCATGGGCAGTCCACGTCAGGTCAGC	70	30
Cek1_rev	CCGGGGTACCTTATTTCATGATTTCTTCATACAG	41.1	34
Ras1_for	CCGGCCATGGCGCTGCGTGAATACAAAC	60.7	28
Ras1_rev	CCGGGGTACCTTAGTTGATGTTGCGCACCAG	58	31

The amplification reactions were performed with Phusion High-Fidelity DNA Polymerase (Thermo Scientific) in a temperature cycler MyCyclerTM Thermal Cycler (Bio-Rad). The cycling conditions are shown in Table 2 -Table 3.

Table 2- PCR conditions for the amplification of gene fragment for the active domain of Cek1_Leu.

Reaction mixture		**Cycling conditions**		
Reagent	**Final Concentration**	**Temperature (°C)**	**Time (min)**	**Number cycles**
10 ng DNA template (pBluescript/CEK1)	0.2 ng ul^{-1}	98 °C	2 min	1
5x Phusion HF Buffer (Thermo Scientific)	1 x	98 °C	10 sec	30
10 µmol Primer forward (Sigma)	0.3 µmol	72 °C	50 sec	
10 µmol Primer reverse (Sigma)	0.3 µmol	72 °C	10 min	
40 mM/each dNTPs (Bioron)	0.8 mM	4°C	∞	
5 U$µl^{-1}$ Phusion High-Fidelity DNA Polymerase (Thermo Scientific)	0.05 U $µl^{-1}$			

Table 3– PCR conditions for the amplification of gene fragment for the active domain of Ras1_Ser.

Reaction mixture		**Cycling conditions**		
Reagent	**Final Concentration**	**Temperature (°C)**	**Time (min)**	**Number cycles**
10 ng DNA template (Pcr2.1/RAS1)	0.2 ng ul^{-1}	98 °C	2 min	1
5x Phusion HF Buffer (Thermo Scientific)	1 x	98 °C	10 sec	30
10 µmol Primer forward (Sigma)	0.3 µmol	72 °C	45 sec	
10 µmol Primer reverse (Sigma)	0.3 µmol	72 °C	10 min	
40 mM/each dNTPs (Bioron)	0.8 mM	4°C	∞	
5 U$µl^{-1}$ Phusion High-Fidelity DNA Polymerase (Thermo Scientific)	0.05 U $µl^{-1}$			

3.1.2. DNA electrophoresis, agarose gel

After performing the PCR reactions a small aliquot (1µl) from reaction mixtures was analyzed on a 1%(w/v) agarose gel. For routine work , 1%(w/v) agarose gels were prepared by dissolving the agarose in TAE buffer (40 mM Tris, 20 mM acetic acid, 1mM EDTA) followed by boiling in a microwave oven until completely melted. Thereafter, the agarose solution was cooled to approximately 60 °C and 2% (v/v) of Ethidium Bromide (EtBr, AppliChem) was added followed by pouring onto a gel tray. Before loading, the DNA samples were mixed with 6X Loading Dye buffer (Fermentas) allowing the visual tracking of DNA migration during electrophoresis. The DNA electrophoresis was performed at 100 volts for 1 hour in a submerged horizontal electrophoresis system (Mini-sub cell GT, Bio-Rad). The electrophosed DNA was analyzed by exposing the gel to the UV light (Gel Doc instrument, Quantity One software, Bio-Rad).

3.1.3. DNA purification from PCR reaction

After confirming the good quality of PCR reaction, the DNA fragments/inserts (for the active domains of Cek1_Leu and Ras1_Ser) were purified from PCR mixture by QIAquick® PCR Purification Kit (Qiagen) according to the manufacturer instructions. The purified DNA was quantified by measuring the absorbance at 260 nm (A_{260}) (NanoDrop ND-1000 Spectrophotometer Thermo Scientific). The A_{260}/A_{280} ratio was used to estimate DNA purity which is expected between 1.8 and 2.0 (37).

3.1.4. Double digestion of inserts

The purified PCR products (gene fragments for active domain of Cek1_Leu and Ras1_Ser) were digested with NcoI and Acc65I restriction enzymes (Fermentas). The double digestions with these two restriction enzymes were performed in 2X Tango buffer using two fold excess of Acc65I. Digestion time of both PCR products (inserts) was 2 hours at 37 °C. After the double digestion, the two restriction enzymes (NcoI and Acc65I) were inactivated at 65 °C for 10 min. Thereafter the double digested inserts were purified by QIAquick® PCR Purification Kit (Qiagen) and DNA was quantified as described in 3.1.3 section.

3.1.5. Expression vectors

Five different pET derived (pETGB, pETZ2, pETNusA, pETMBP and pETTrx) expression vectors were used to clone the gene fragments for the active domain of Cek1_Leu and Ras1_Ser. The main difference between these five pET vectors it is in their solubility tags (see Figure 8). In all five pET constructions the insert will be separated by a TEV cleavage site from an N-terminal His_6tag and solubility tag (depending on the vector).

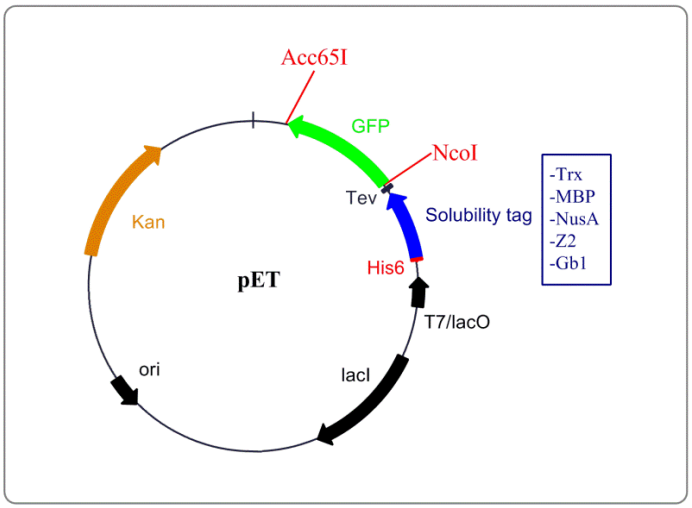


Figure 8 – Schematic representation of five pET vectors with their different solubility tags.

All five vectors were double digested with NcoI and Acc65I under the same conditions described in section 3.1.4 except, the digestion took longer (4 h), due to the bigger size of circular vectors. After digestion the restriction enzymes were inactivated (60 °C, 20 min) and the sample was treated with Shrimp Alkaline Phosphatase (SAP) to avoid recircularization of the cloning vectors. As a consequence of the double digestion the DNA sequence for GFP was cleaved from the pET expression vectors. Thereafter the linearised double digested expression vectors were isolated with agarose gel electrophoresis (DNA fragment was excised under the UV light) and purified using the QIAquick® Gel Extraction Kit (Qiagen). The purified DNA was quantified as described in 3.1.3 section.

3.1.6. Ligation into expression vectors

The two pure double digested inserts (active domains of Cek1_Leu and Ras1_Ser, hereinafter referred as Cek1_Leu and Ras1_Ser) were ligated in the all five pure, double digested expression vectors (see section 3.1.4). Ligations were performed with 1 u T4 DNA ligase (Fermentas) using a molar ratio of 1:4 vector to insert in a total reaction volume of 10 µl, overnight at 20 °C, in 1x T4 DNA ligase buffer (50 mM Tris-HCl, pH 7.5; 10 mM $MgCl_2$, 1 mM ATP, 10 mM Dithiothreitol, Fermentas). In order to convert molar ratios to mass ratios the equation (1) was used:

$$\text{ng of insert} = \frac{\text{ng vector} \times \text{kb size insert}}{\text{kb size of vector}} \times \frac{\text{molar ratio of insert}}{\text{vector}} \qquad (1)$$

3.1.7. Isolation and purification of plasmid DNA

After performing the ligation reaction, 50 µl of *E. coli* DH5α competent cells were transformed with 5µl of ligation mixture as described in section 3.1.9. The positive transformants were selected by plating onto Luria-Bertani (LB) agar plates, [1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1.5% (w/v) agar] supplemented with 50 µg/ml kanamycin (Kan; Sigma) and grown overnight at 37 °C. In order to check if the expression vector contained the insert (Cek1_Leu or Ras1_Ser) five colonies were grown overnight at 37 °C in liquid LB medium with selective antibiotic ($V_{culture}$ = 10 ml) and the plasmid DNA was purified with NZYMiniprep Kit® (NZYTech) according to the manufacturer´s instructions, doubling the volumes of buffers A1, A2 and A3. The purified plasmid DNA was quantified (see section 3.1.3) and tested for the insert by double digestion with NcoI and Acc65I restriction enzyme. Alternatively the positive transformants, after ligation, were tested for the insert by performing a colony PCR. In this method 5µl overnight cultures (of each five colonies, as described above) were used as DNA templates in PCRs. In these PCRs the inserts were amplified with same primers and in the same conditions as were used to amplify gene fragments for the active domains of Ras1_Ser and Cek1_Leu (see section 3.1.1,Table 2 and Table 3). The purified plasmid DNA samples were sent to Eurofins MWG Operon (Germany) for sequencing.

3.1.8. Site directed mutagenesis (SDM)

The Ras1_Leu (leucine variant for CUG codon) and Cek1_Ser (serine variant for CUG codon) were prepared by QuikChange site directed mutagenesis. This method allows site-specific mutation in any double-stranded plasmids. The pET expression plasmids cloned with Ras1_Ser or Cek1_Leu (section 3.1.6) were used as template in the mutagenesis reaction. For both SDM (Cek1 and Ras1) two primers were designed using software provided by Stragene Genomics® (https://www.genomics.agilent.com/homepage.aspx). These primers are two complementary DNA oligonucleotides that contain the desired mutation, flanked by unmodified nucleotide sequence (Table 4). With Ras1_Leu mutagenic primers, the CUG codon was mutated to leucine-UUG in order to create the Leu67 variant of Ras1_Leu. Furthermore, with Cek1_Ser mutagenic primers the CUG codon was mutated to serine UCG to obtain the Ser146 variant of Cek1_Ser. The SDM reactions were performed with Pfu Turbo DNA polymerase (Stratagene) for high fidelity and robust amplification of long DNA templates with a temperature cycler (MyCyclerTM Thermal Cycler, Bio-Rad), using the cycling PCR conditions described in Table 5. The product was then treated with 10 Units of Dpn I endonuclease (Fermentas) for 1h at 37 °C. Dpn I is specific for methylated and hemimethylated DNA (target sequence 5'-Gm^{6}ATC-3') and used to digest the parental DNA template and to select for mutation-containing newly synthesized DNA. DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to Dpn I digestion. After selection of mutated DNA plasmid, 10 µl of the SDM reaction were used to transform 50 µl of *E. coli* DH5α competent cells as described in section 3.1.9. The positive transformants were selected by plating onto LB-agar plates supplemented with 50 µg/ml Kan and growing overnight at 37 °C. DNA plasmid isolation and purification was performed on two isolated colonies as described in section 3.1.7. The pure DNA plasmid samples were sent for sequencing.

Table 4- Primers used to create Ras1_Leu and Cek1_Ser.

Primer	Sequence (5´→3´)	Tm (º C)	Length (nt)
Ras1_67Leu_for	ACAGCTGGTCAAGAAGAGTATTTGGCAATGCGCGA	79.27	35
Ras1_67Leu_rev	TCGCGCATTGCCAAATACTCTTCTTGACCAGCTGT	79.27	35
Cek1_146Ser_for	CTGAAGCCCTCCAACTTGTCGCTGAACTCCAATT	78	34
Cek1_146Ser_rev	AATTGGAGTTCAGCGACAAGTTGGAGGGCTTCAG	78	34

Table 5- Reaction mixture (left) and cycling conditions (right) for SDM of Ras1 and Cek1 active gene fragment.

Reaction mixture		**Cycling conditions**		
Reagent	**Final Concentration**	**Temperature (°C)**	**Time (min)**	**Number cycles**
5 ng µl^{-1} DNA plasmid	0.1 ng µl^{-1}	95 °C	30 sec	1
10x Reaction Buffer (Stragene)	1x	95 °C	30 sec	16
125 ng µl^{-1} primer forward (MWG)	2.5 ng µl^{-1}	55 °C	1 min	
125 ng µl^{-1} primer reverse (Sigma)	2.5 ng µl^{-1}	68 °C	10 min*	
10 mM/each dNTPs (Sigma)	1.28 mM	68 °C	10 min	
2.5 Uµl^{-1} Pfu Turbo DNA polymerase (Stratagene)	0.05 U µl^{-1}	4°C	∞	

3.1.9. Preparation of competent cells and E. coli transformation

In order to induce reproducible and highly efficient transformation of bacterial cells with expression vector, competent cells were prepared. For this purpose, a fresh single colony was inoculated into 50 ml of LB medium (with antibiotic selection, if necessary) which grew overnight at 37 °C with vigorous shaking [180 rpm, in a IKA® KS 4000 I-Control (IKA®) incubator shaker]. In the following day a 2 ml aliquot was used to inoculate 200 ml of LB medium (pre-mixed with $MgCl_2$ to a final concentration of 20 mM) and the cells grew until the optical density at 600 nm (A_{600}) reached the value of 0.4 – 0.5. The cell growth was stopped by incubating the culture on ice for 15 min, the cells were pelleted by gentle centrifugation at 3000 rpm (Hettich Universal 320R) for 5 min at 4 °C, and the pellet was carefully resuspended in 25 ml of ice-cold $CaCl_2$ (100 mM) and then incubated on ice for 20 min. The cells were gently centrifuged again, then resuspended in 2ml of ice-cold $CaCl_2$ (100 mM) and 600 µl glycerol (100%). This resuspension was distributed into eppendorfs pre-chilled in ice-bath and immediately flash-frozen in liquid nitrogen. Competent cells were stored at -80 °C. Table 6 contains the *E.coli* strains (competent cells) used for optimization of Cek1 and Ras1 expression.

Table 6- *E. coli* strains used for optimization of Cek1 and Ras1 expression.

Strain	Description/Application	Antibiotic resistance	Supplier
BL21Star™(DE3)	Lacks a functional RNAase enzyme resulting in higher mRNAs stability.	----------	Stratagene
BL21(DE3)pLysS	Carries a T7 RNA polymerase inhibitor to suppress the basal recombinant protein expression, prior to induction.	Chloramphenicol 34 µg ml^{-1}	Stratagene
Rosetta(DE3)	A BL21(DE3) derivative but provides six extra tRNAs decoding the rare codon AGG, AGA, AUA, CUA,CCC and GGA.	Chloramphenicol 34 µg ml^{-1}	Novagen
Origami2(DE3)	Have mutations in both the thioredoxin reductase and glutathione reductase genes, which greatly enhances disulfide bond formation in the cytoplasm.	Kanamycin 50 µg ml^{-1}	Stratagene

Competent bacterial cells were transformed using the standard heat-shock method. 50µl of these cells were incubated with 10-100 ng of pET expression vectors (see section 3.1.5) on ice during 30 min. The heat shock was performed at 42 °C for 45 sec, followed by incubation on ice for 2 min. Thereafter 600 µl of LB medium were added to each transformation reaction and the cells were allowed to recover at 37 °C for 60 min with shaking (170 rpm). Cells were pelleted by centrifugation (600 *g* for 1 min) at room temperature and 500 µl of the supernatant were discarded. The cells were resuspended in the remaining LB medium and plated on LB-agar plate containing the selected antibiotic. The plates were incubated at 37 °C overnight. (grown overnight at 37 °C until 15 hours). A negative control was performed in parallel, without plasmid DNA addition.

3.2. Protein expression and purification

3.2.1. Trx_Cek1_Leu expression tests

Expression of soluble Cek1_Leu in fusion with Trx tag (pETTrx/Trx_Cek1_Leu construct) was first assessed at analytical scale (50 mL cultures) using an incomplete factorial (IF) approach (38) combining three *E. coli* strains (BL21Star(DE3), Origami2(DE3) and Rosetta(DE3), three temperatures (20, 25 and 30 °C) and two medium (LB and 2YT). Using SAmBA software (http://igs-server.cnrs-mrs.fr/samba/) 9 IF combinations (see Table 7) can be obtained out of the 18 combinations of the corresponding full factorial (3 *E. coli* strains x 3 exp. temperature x 2 medium). In all 9 experiment (see Table 7) the protein expression was induced by adding Isopropyl-β-D-thigalactoside (IPTG) to a final concentration of 0.9 mM and incubating for 3-4 hours or overnight depending on temperature, 25/30 °C and 20 °C respectively. After protein expression the cultures were (9 x 50 ml) harvested at 5000 rpm, 4 °C for 10 min (Hettich Universal 320R centrifuge). The supernatant was discarded and the pellets were resuspended in 2 ml lysis buffer (40 mM Tris-HCl, 200 mM NaCl, 5 mM 2-mercaptoethanol, pH7.5). Cells were lysed by sonication in a Sonoplus HD2200 200 W sonicator (Bandelin Electronic) for 3 x 3 seconds 10 % duty cycle and 10 % output conditions. Then lysates (9 x 2ml) were centrifuged at 13000 rpm (Eppendorf 5414), for 15 min at 4 °C. The supernatants, the soluble fractions, were removed in new tubes and the pellets, the insoluble fractions were resuspended in 2 ml (9 x 2ml) lysis buffer. Then both soluble and insoluble fractions were analyzed by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). The condition yielding the higher levels of soluble recombinant protein was selected for experimental scale-up.

Table 7- IF approach conditions used to test Trx_Cek1_Leu expression.

Combination	Competent cell	Exp. Temp (°C)	Medium
1	Origami2(DE3)	25	LB
2	BL21Star™(DE3)	25	2YT
3	Origami2(DE3)	30	2YT
4	Rosetta(DE3)	30	2YT
5	BL21Star™(DE3)	20	2YT
6	Origami2(DE3)	20	2YT
7	Rosetta(DE3)	20	LB
8	Rosetta(DE3)	25	LB
9	BL21Star™(DE3)	30	LB

3.2.2. High-scale expression of Trx_Cek1_Leu

Trx_Cek1_Leu was overexpressed in *Escherichia coli* strain Rosetta(DE3) (see Table 6). Competent cells previously transformed with pETTrx/Cek1_Leu plasmid were plated onto LB-agar plates containing 50 µg/ml kanamycin and 34 µg/ml chloramphenicol and incubated overnight at 37 °C. A colony was picked and transferred to 50 ml of LB medium containing 50 µg/ml kanamycin and 34 µg/ml chloramphenicol and incubated overnight at 37 °C. In the following day, 700 µl aliquots were used to inoculate 4 X 500 ml portions of LB medium (in 2000 ml Erlenmeyer flasks) containing 50 µg/ml kanamycin and 34 µg/ml chloramphenicol and cells were grown at 37 °C until A_{600} reached 0.7- 0.9. Then, temperature was decreased to 20 °C and Trx_Cek1_Leu expression was induced by adding IPTG to a final concentration of 0.9 mM. Incubation continued at 20 °C for 18 hours (overnight). Cells were harvested by centrifugation for 25 minutes at 4000 rpm (Beckman JLA-8.1000 rotor, Avanti J-26 XPI). The bacterial pellets were resuspended in 30 ml lysis buffer [40 mM Tris-HCl, 500 mM NaCl, 5 %

glycerol (v/v) 5 mM 2-mercaptoethanol, pH7.5] supplemented with 100 µg/ml lysozyme (Sigma) and stored at -20 °C.

3.2.3. Purification of Trx_Cek1_Leu

Immobilized metal affinity chromatography

Ressuspended bacterial pellets were thawed and phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 1 mM. Cells were lysed by sonication in a Sonoplus HD2200 200 W sonicator (Bandelin Electronic) for 3 min with a 50 % duty cycle. Lysates were centrifuged for 45 minutes at 15000 rpm , using a Beckman JA-25.50 rotor. The supernatant was filtered through a 0.2 µm filter and then applied to a 5 ml HisTrap HP column (GE Healthcare) previously equilibrated with binding buffer (20 mM imidazole , 500 mM NaCl, 5% glycerol, 5 mM 2-mercaptoethanol, 40 mM Tris-HCl, pH 7.5). The chromatography system consisted of an Econo gradient pump with a SV-5 buffer select valve (Bio-Rad), a UV-vis detector model 112 (Gilson) and a Kipp & Zonen chart recorder. Alternatively an AKTAprime Plus (GE Healthcare) chromatography system was used. Column was washed with 10 column volumes of binding buffer. Elution was induced by a step gradient of 5 %, 15%, 40% and 100 % elution buffer (500 mM imidazole, 500 mM NaCl, 5% glycerol, 5 mM 2-mercaptoethanol, 40 mM Tris-HCl pH 7.5). Fractions were analysed by SDS-PAGE (12.5 % gel) to assess purity. Thereafter fractions containing His-tagged Trx_Cek1_Leu were combined and concentrated (to a maximum volume of 5 ml) by Amicon Ultra 15 centrifugal concentrator with a 10 kDa cut-off (Millipore Corporation). Subsequently, this sample was further purified by size-exclusion chromatography.

Preparative size-exclusion chromatography (SEC)

A Hiprep 26/60 Sephacryl S-100 (GE Healthcare) column was adapted to AKTAprime Plus (GE Healthcare) or BioLogic Duo Flow system (Bio-Rad). The sample was filtered through a 0.2 µm filter before injection. An isocratic elution in gel filtration buffer [300mM NaCl, 5% (v/v) glycerol, 1mM EDTA, 1mM DTT, 40 mM Tris-HCl, pH 7.5] was monitored by optical density at 280 nm. Collected fractions were analyzed by SDS-PAGE (12.5 % gel) to assess purity. Fractions containing the target protein were pooled and concentrated using Amicon Ultra 15 centrifugal concentrator with a 10 kDa cut-off (Millipore Corporation) to a final concentration of 10 mg/ml. During concentration the buffer was changed to protein storage buffer [150mM NaCl,

5% (v/v) glycerol, 1mM EDTA, 1mM DTT, 20 mM Tris-HCl, pH 7.5]. Concentration of Trx_Cek1_Leu was determined by absorption at 280 nm using a theoretical molar extinction coefficient of 56310 $M^{-1}cm^{-1}$.

3.2.4. Expression test for MBP_Cek1_Leu

Expression of Cek1_Leu in fusion with MBP tag was tested at analytical scale (50 ml) at two different expression temperature (20 °C, and 25 °C) in *Escherichia coli* strain BL21Star™(DE3). The expression test experiments and analyzes of the results were performed as described for Trx_Cek1_Leu (section 3.2.1).

3.2.5. High-scale expression of MBP_Cek1_Leu

The overexpression of MBP_Cek1_Leu in *Escherichia coli* strain BL21Star™(DE3) was performed at 20 °C using the same protocol as described for Trx_Cek1_Leu expression (section 3.2.2) using just one antibiotic: 50 µg/ml kanamycin.

3.2.6. Purification of MBP_Cek1_Leu

Purification of MBP_Cek1_Leu was performed using the same protocol as described for Trx_Cek1_Leu (section 3.2.3) except that a Hiprep 26/60 Sephacryl S-200 (GE Healthcare) column was used for size-exclusion chromatography. The pure MBP_Cek1_Leu was concentrated to a final concentration of 14 mg/ml. Concentration of MBP_Cek1_Leu was determined by absorption at 280 nm using a theoretical molar extinction coefficient of 110045 $M^{-1}cm^{-1}$.

3.2.7. Expression test for Gb1_Ras1_Ser

Expression of Ras1_Ser in fusion with Gb1 tag was assessed at analytical scale (50 ml) at two different expression temperatures (20 °C, and 25 °C) in *E. coli* strain BL21(DE3)pLysS. The expression tests were performed as described in section 3.2.1.

3.2.8. High-scale expression of Gb1_Ras1_Ser

Overexpression of Ras1_Ser in fusion with Gb1 solubility tag (pETGb1) was performed in BL21(DE3)pLysS cells at 25 °C for 4h in presence of 0.9 mM IPTG as described in section 3.2.2. The lysis buffer of Gb1_Ras_Ser contained 5mM of $MgCl_2$ [40 mM Tris-HCl, 5 mM $MgCl_2$ 500 mM NaCl, 5 % glycerol (v/v) 5 mM 2-mercaptoethanol, pH7.5].

3.2.9. Purification of Gb1_Ras1_Ser

Purification of Gb1_Ras1_Ser was performed using the same protocol as described for Trx_Cek1_Leu (section 3.2.3) except that all buffers used for purification and storage of Gb1-Ras1_Ser contained 5 mM $MgCl_2$. Concentration of Gb1_Ras1_Ser was determined by absorption at 280 nm using a theoretical molar extinction coefficient of 24995 $M^{-1}cm^{-1}$.

3.2.10. SDS-PAGE analysis

The Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) is the method used to separate proteins according to their size. Samples from expression and purification were analysed by SDS-PAGE. The SDS-polyacrylamide gels were composed by two layers. The stacking gel, with approximately 1 cm, was composed by 40% of acrylamide, 0,625 M Tris-HCl pH 6.8, 20% SDS, 10% APS and 0,1% TEMED. The separating gel was composed by 12.5% of acrylamide, 1.5 M Tris-HCl pH 8.8, 40% of acrylamide, 20% SDS, 10% APS and 1% TEMED. To analyze samples resulting from protein expression or purification, SDS-PAGE Loading Buffer Dye (2x) was added; to each sample a final concentration of 200 mM DTT was added; samples were boiled for 5 minutes and then were loaded on the stacking gel. The gel was stained with BluePage™ protein staining solution (Fermentas) according to manufacturer instructions. Gel analysis was performed with a densitometer (GS-800, Bio-Rad) and analyzed with the QuantityOne software (Bio-Rad).

3.3. TEV protease cleavage assay

Purified MBP_Cek1_Leu was digested by in-house expressed and purified recombinant TEV protease (200 mM NaCl; 10% glycerol; 2 mM EDTA; 10 mM DTT; 25 mM NaH_2PO_4/Na_2HPO_4 pH 8.0) in order to cleave the MBP tag together with histidine tag. The ratio of TEV protease to fusion protein (MBP_Cek1_Leu) to achieve complete digestion may vary. Therefore pilot experiments were performed with small amount of fusion protein (1.4- 2µg) using several enzyme/substrate molar ratio (1:2.5; 1:5; 1:10; 1:20). The cleavage was performed at room temperature, 2h; in storage buffer of MBP_Cek1_Leu [150mM NaCl, 5% (v/v) glycerol, 1mM EDTA, 1mM DTT, 20 mM Tris-HCl, pH 7.5].

3.4. Analytical Size-exclusion Chromatography

Analytical size-exclusion (gel filtration) chromatography was used for the determination of the molecular weight (oligomerization state) of proteins (MBP_Cek1_Leu, Gb1_Ras1_Ser). The technique is based on the ability of gel filtration media to separate molecules according to size. For analytical size-exclusion chromatography a Superpose 12 10/300 column (GE Healthcare) was assembled in an ÄKTA purifier 10 system (GE Healthcare) and equilibrated with buffer: 150mM NaCl, 5% (v/v) glycerol, 5 mM $MgCl_2$, 1mM EDTA, 1mM DTT, 20 mM Tris-HCl, pH 7.5. Blue dextran 2000 was used for determination of the void volume and ribonuclease A (MW: 13.7 kDa), chymotrypsinogen A (25 kDa), ovalbumin (43 kDa), albumin (67 kDa) and aldolase (158 kDa) were used for column calibration (150 µg of each standard protein). 140 µg MBP_Cek1_Leu and 200 µg Gb1_Ras1_Ser samples were loaded into the column. The calibration curve was prepared by measuring the elution volumes of the standard proteins, calculating their corresponding K_{av} values and plotting them versus the logarithm of their respective molecular weight. The K_{av} parameter was determined according to the equation (2):

$$K_{av} = \frac{(V_e - V_0)}{(V_t - V_0)} \quad (2)$$

where V_e represents the elution volume,V_0 the void volume of the column, and V_t the total bed volume.

3.5. Dynamic Light Scattering (DLS) analysis

This method provides the ability to measure size characteristics of proteins in a liquid medium. Molecules in solution undergo constant diffusion, which is a result of thermal fluctuations in the medium and is referred to as Brownian motion. When a coherent light beam hits a spherical molecule in solution, there is a shift in the wavelength of the incoming light. DLS measures time-dependent fluctuations in the scattering intensity arising from particles undergoing random Brownian motion. They are measured across very short time intervals to produce a correlation curve, from which the particle diffusion coefficient and subsequently the particle size can be obtained. For monodisperse samples, consisting of a single particle size group the correlation curve can be fit to a single exponential form or Cumulant fit (39). The hydrodynamic radius (R_H) is then calculated from the diffusion coefficient (*D*) using the Stokes-Einstein equation (3), where k is the Boltzmann constant, T is the temperature, η is the medium viscosity (39).

$$R_H = \frac{kT}{6\pi\eta D} \quad (3)$$

By definition, the radius measured by DLS is the radius of a hypothetical hard sphere that diffuses with the same speed as the particle under examination. The obtained hydrodynamic radius is an average value, weighted by particle scattering intensity. The size distribution obtained by DLS is a plot of the relative intensity of light scattered by particles in various size classes and is therefore known as an intensity size distribution. If the plot shows a substantial tail, or more than one peak, then Mie theory can make use of the input parameter of sample refractive index to convert the intensity distribution to a volume distribution (40). This will then give a more realistic view of the importance of the tail or second peak (41). The polydispersity is the relative standard deviation of a sample and is used to describe the width of the particle size distribution. Polydispersity can be defined in the following terms: Polydispersity Index (*PdI*) = Relative variance, Polydispersity (*Pd*) = Standard deviation (σ) or % Polydispersity (%*Pd*) = Coefficient of variation = (*PdI*)1⁄2 × 100 (also called the relative polydispersity). A sample is considered monodisperse if the polydispersity is less than 20%, it is medium dispersed if this value is in the range of 20–30%, and it is polydispersed for values above 30%. Molecular size measurements were carried out in a Zeta sizer Nano Zs DLS system (Malvern Instruments) (42) MBP_CEK1 and Gb1_Ras1 samples with concentrations of 14 and 20 mg/ml respectively in their storage buffer were centrifuged at 100,000 rpm for 40 min at 4

°C in an AirfugeTM air-driven ultracentrifuge (Beckman Coulter). Then in a 45-µl DTS 2112 cuvette, three independent measurements were obtained at 20 °C for each sample. All data were then analyzed using DTS (nano) 6.32 software.

3.6. Crystallization screenings

Crystallization trials were performed with the purified Gb1_Ras1_Ser protein (20 mg/mL) and MBP_Cek1_Leu (14.5 mg/ml) by vapor diffusion techniques. Different crystallization conditions were screened using the following crystallization kits: Morpheus HT96 and PACT *premier*™ (Molecular Dimensions), screening variables such as pH, anions, and cations in combination with the precipitant (poliethyleneglycol - PEG). Sitting drops were employed and the droplets contained 1 µl of precipitant solution plus 1 µl of protein solution. The reservoirs contained 350 µl of precipitant solution. Trays were incubated at 20 °C. The crystallization trays were checked after 30 minutes, 24 hours of incubation and thereafter every two days during one week and every week during one month.

4. Results and discussion

4.1. Cloning

The full synthetic genes for Cek1_Leu and Ras1_Ser (see Figure 9A andFigure 10A) were purchased with the codon optimized for *E. coli* in standard vectors: pBluesscript IISK (+) and pCR2.1 respectively (see section 6, Figure 36 and Figure 37. The gene fragments for active domains of Cek1_Leu and Ras1_Ser (see Figure 9B andFigure 10B were amplified by PCR including the restriction sites NcoI and Acc65I for 5' and 3' end respectively. Figure 11 shows the results of these amplifications: gene fragments for active domain of Cek1_Leu and Ras1_Ser.

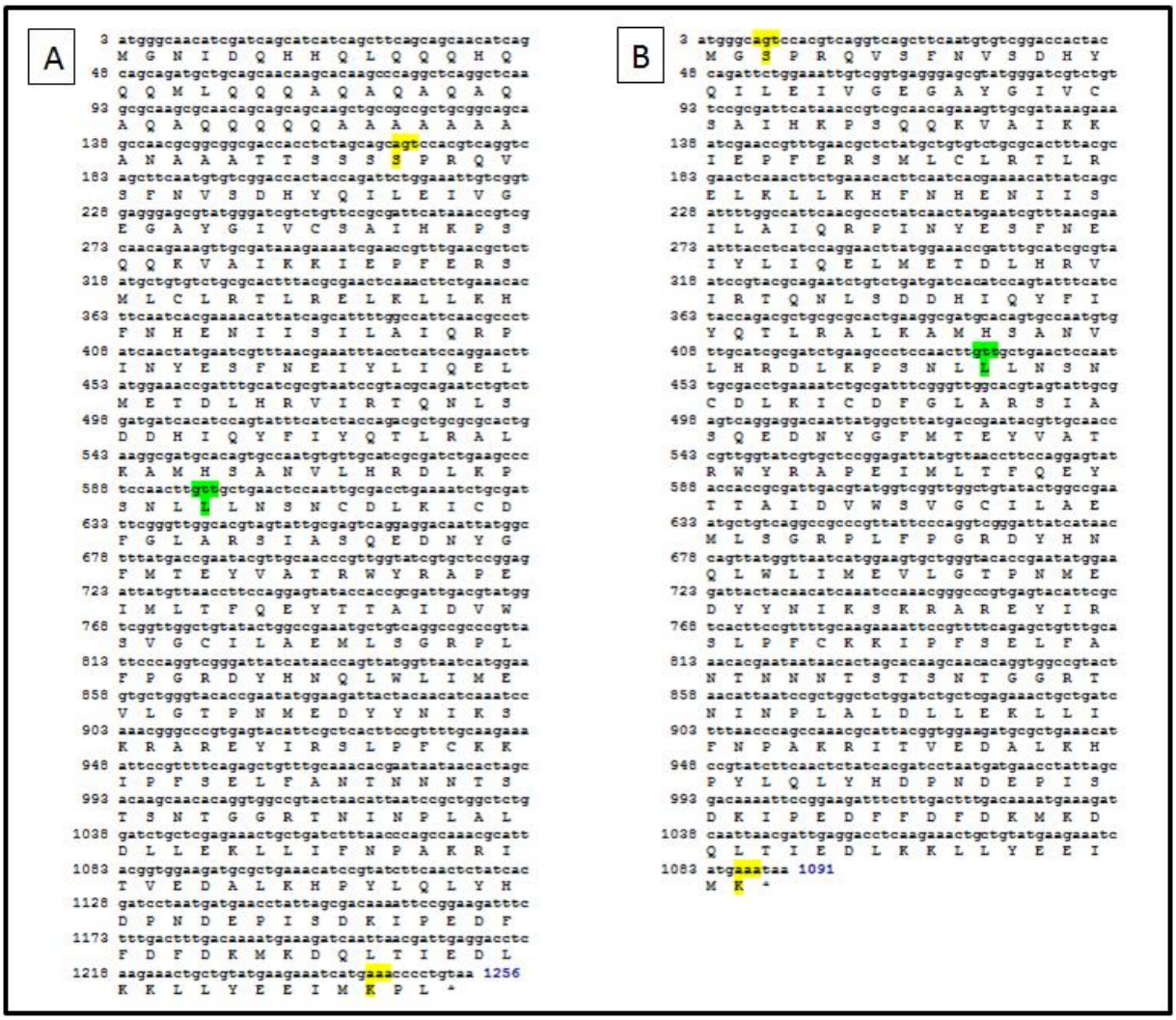


Figure 9- *Candida albicans* Cek1_Leu cDNA: (A) full and (B) amplified fragment for active domain. The CUG encoded residue (Leu variant in the synthetic gene) is highlighted in green at positions: 199 in the full and 146 in the amplified active domain fragment. The first and last residues of the active domain fragment are highlighted in yellow.

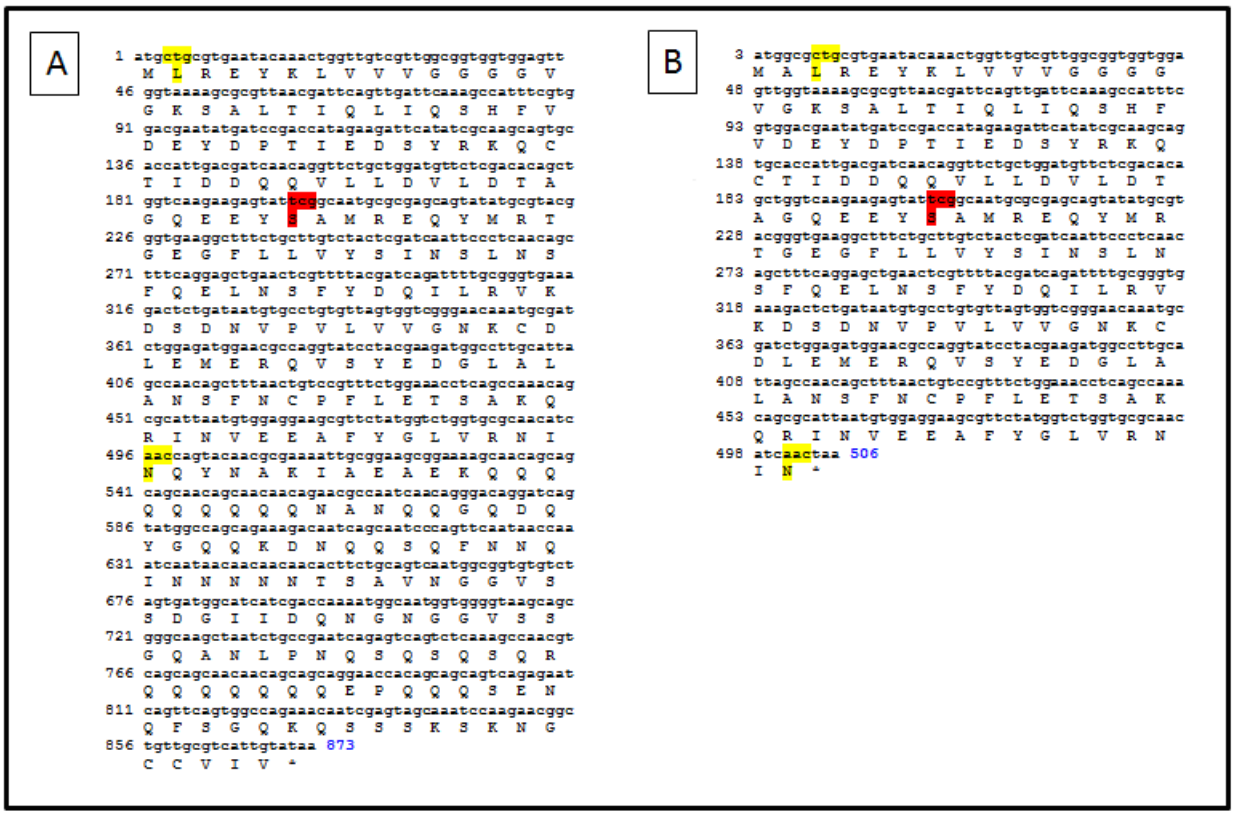


Figure 10- *Candida albicans* Ras1_Ser cDNA: (A) full and (B) amplified fragment for active domain. The CUG encoded residue (Ser variant in the synthetic gene) is highlighted in red at positions: 66 in the full and 67 in the amplified active domain fragment. The first and last residues of the active domain fragment are highlighted in yellow.

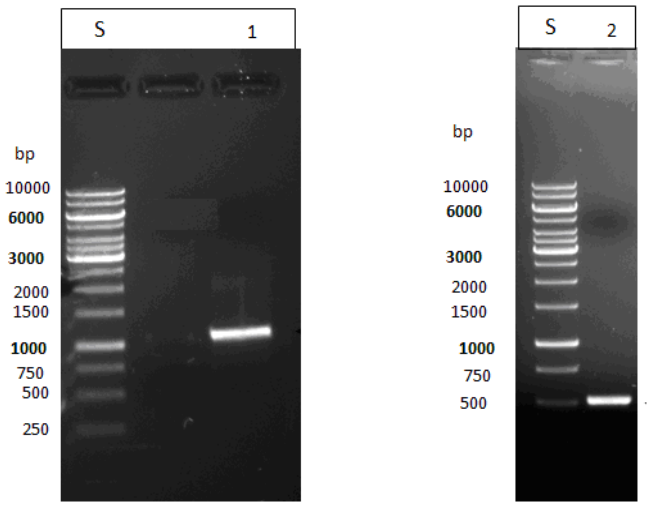


Figure 11- Purified PCR products: lane 1-gene fragment for active domain of Cek1_Leu (1091 bp), lane 2- gene fragment for active domain of Ras1_Ser (506 bp).

The purified PCR products (gene fragments for active domain of Cek1_Leu and Ras1_Ser hereinafter referred as Cek1_Leu and Ras1_Ser) were double digested with NcoI and Acc65I restriction enzymes. These inserts (Cek1_Leu and Ras 1_Ser) were cloned into five pET (pETMBP, pETTrx, pETGb1, pETZ2 and pETNusA) expression vectors at the same NcoI and ACC65I sites. Therefore the vectors were prepared for ligation by double digestion with NcoI and Acc65I restriction enzymes. As a consequence of this double digestion the DNA sequence for GFP was cleaved from the pET expression vectors (see Figure 12). The double digested expression vectors were isolated and purified from agarose gel (see Figure 13). These pure linearised vectors were used to ligate the Cek1_Leu and Ras1_Ser inserts (prepared as described above).

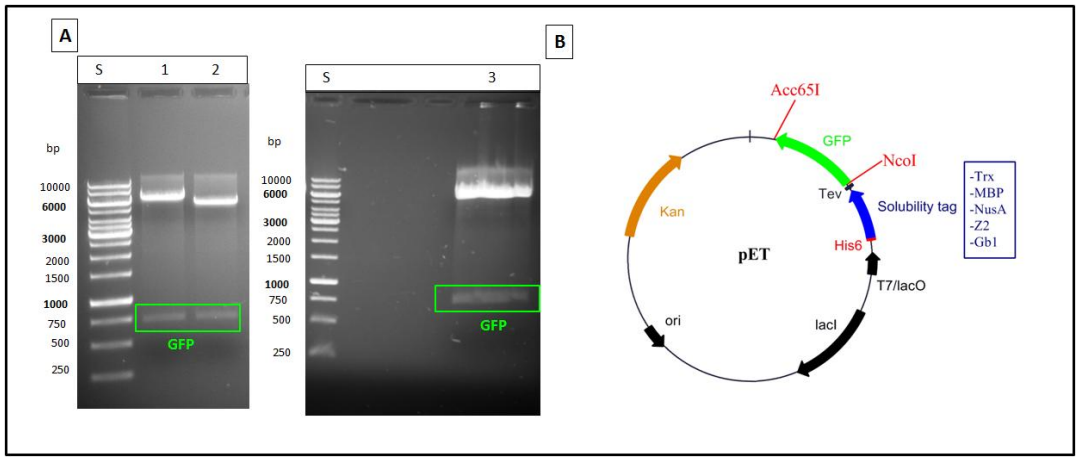


Figure 12- Expression vectors: (**A**) double digestion (NcoI/Acc65I) of: lane 1- pETMBP (l6446 bp), lane 2- pETTrx (5666 bp) and lane 3-pETGb1 (5510 bp); (**B**) schematic representation of plasmids showing the GFP position between NcoI and Acc65I restriction sites.

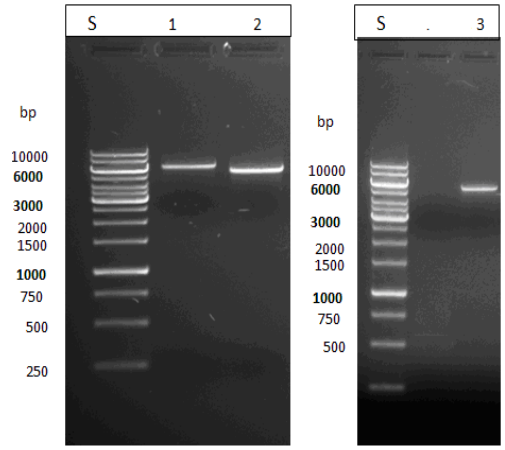


Figure 13- Purified double digested vectors: lane 1- pETMBP, lane 2- pETTrx and lane 3- pETGb1.

After ligation the constructs were transformed and the positive transformants were selected by plating onto LB-agar plates supplemented with 50 µg/ml Kan and grown overnight at 37 °C (see Figure 14).

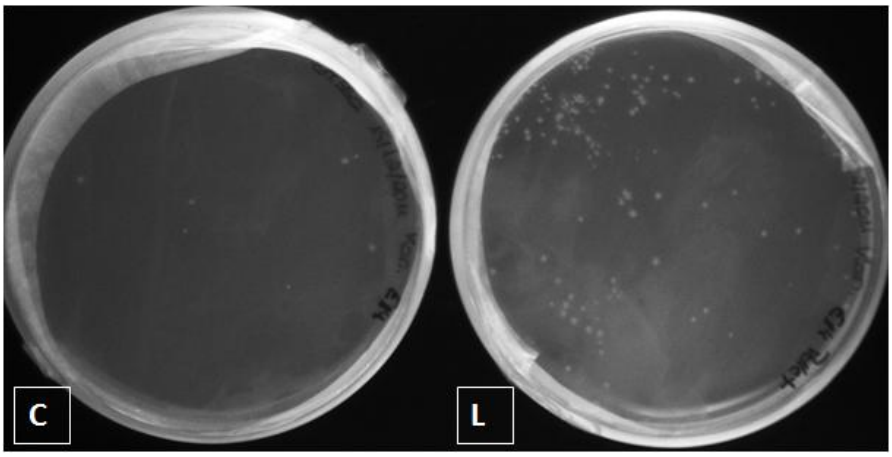


Figure 14- Example for transformation results after ligation: C- control: transformation with ligation mixture containing only the linearised vector (no insert added) L- transformation with ligation mixture containing the linearised pETMBP and Cek1_Leu (insert) at a molar ratio of 1:4.

The plasmid DNA was isolated from the positive transformants and tested for the insert by double digestion with NcoI and Acc65I restriction enzymes (see Figure 15). As is shown in Figure 15 the double digestion separated the insert from the respect expression vector, confirming the success of the ligation reaction.

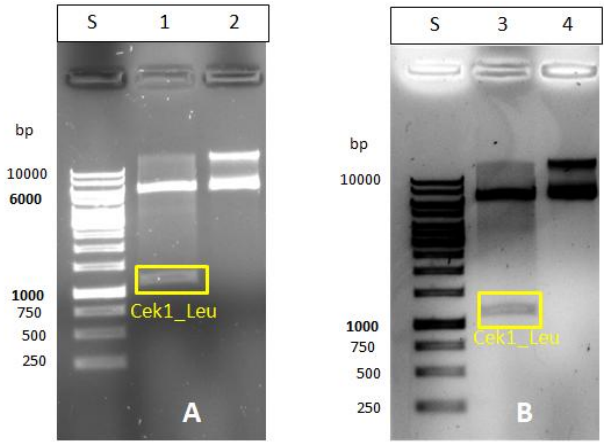


Figure 15- Restriction analyses (NcoI/Acc65I) of positive transformants after ligation of Cek1_Leu in pETTrx (A: lane1- double digested, lane 2- no digestion) and pETMBP (B: lane 3- double digested, lane 4- no digestion.

Alternatively, the positive transformants, after ligation, were tested for the insert by performing a colony PCR (see Figure 16). Using this method the insert was amplified by PCR directly from bacterial colonies.

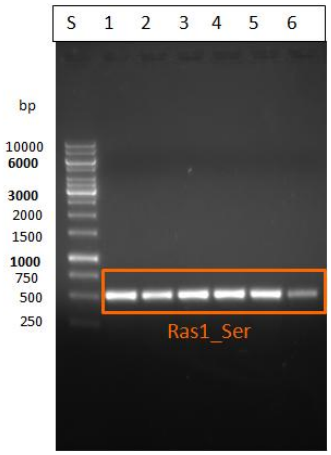


Figure 16- Colony PCR after ligation of Ras1_Ser in pETGb1 vector. Ras1_Ser was successfully amplified by PCR from six different colonies.

Finally the tested DNA plasmid samples were sent to Eurofins MWG Operon (Germany) for sequencing. The results were analyzed by performing alignments using ClustalW tool (http://www.ebi.ac.uk/Tools/clustalw2/index.html) (43).

All of the five expression vectors (pETMBP, pETTrx, pETGb1, pETZ2 and pETNusA) were successfully cloned with Cek1_Leu and Ras1_Ser. Although, the cloning results presented above refer for cloning of Cek1_Leu and Ras1_Ser in two (pETMBP and pETTrx) or one (pETGb1) vectors respectively. These three constructs: pETTrx/Cek1_Leu, pETMBP/Cek1_ Leu and pETGb1/Ras1_Ser were used to express the proteins in large scale.

The Cek1_Ser (serine variant for CUG codon) and Ras1_Leu (leucine variant for CUG codon) variants were prepared by site directed mutagenesis (SDM). The pETMBP/Cek1_Leu and pETGb1/Ras1_Ser constructs were used as templates in mutagenesis reaction. Using these templates with mutagenic primers (see section 3.1.8) the SDM resulted in pETMBP/Cek1_Ser and pETGb1_Ras1_Leu constructs. The mutations were confirmed by sequencing results, which were aligned with the template DNA fragment (see Figure 16 -Figure 17).

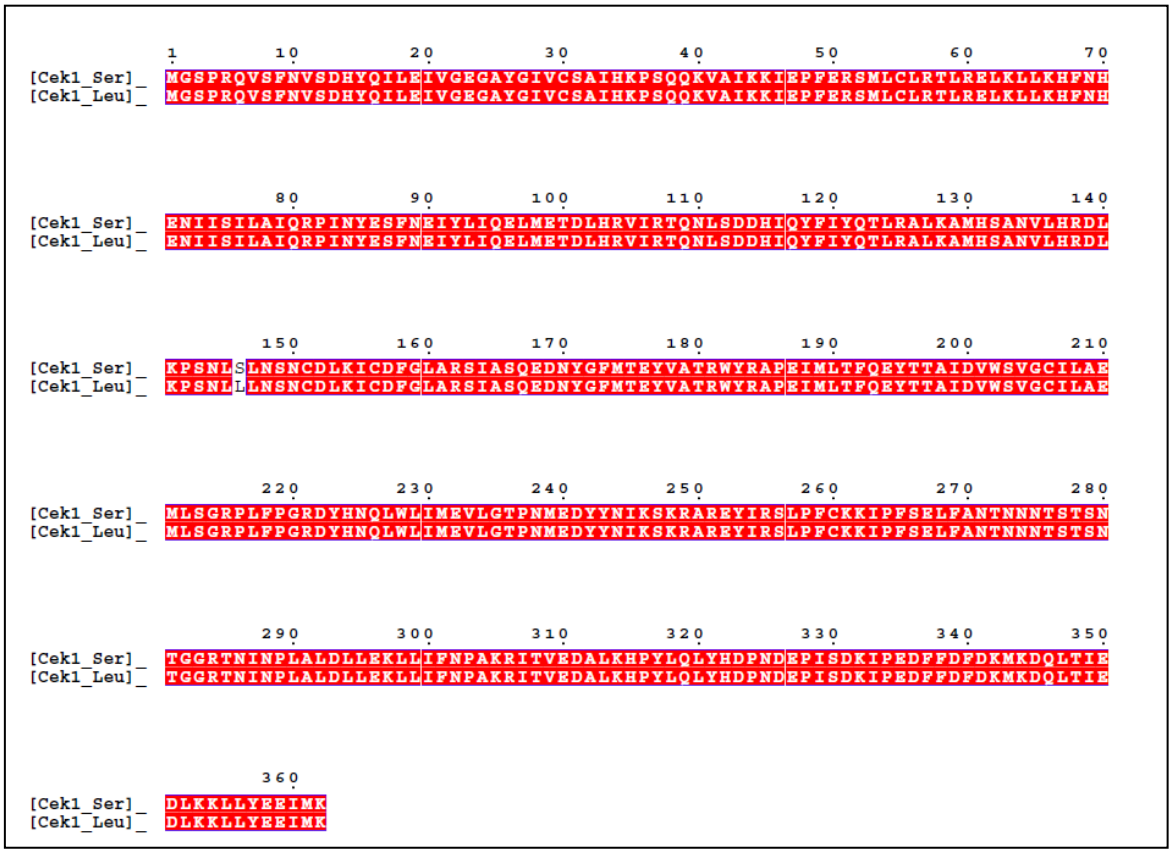


Figure 17- Alignment performed between Cek1_Leu and the product of the SDM (Cek1_Ser) showing to have a serine for the leucine in the position 146. The proteins were aligned with clustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and the image was prepared with ESPRIPT (http://espript.ibcp.fr/ESPript/ESPript/).

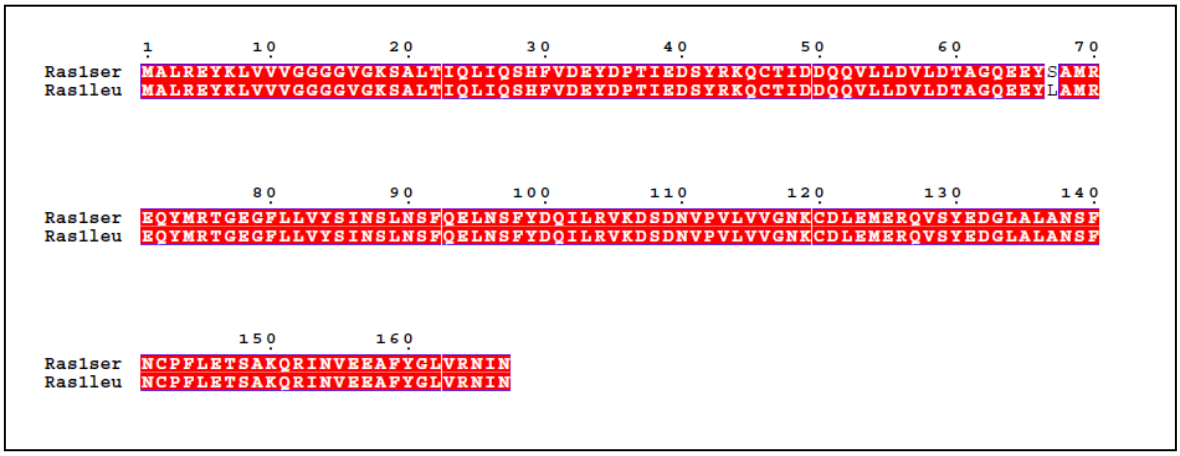


Figure 18- Alignment performed between Ras1_Ser and the product of the SDM (Ras1_Leu) showing to have a leucine for the serine in the position 67. The proteins were aligned with clustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and the image was prepared with ESPRIPT (http://espript.ibcp.fr/ESPript/ESPript/).

4.2. Protein expression and purification

4.2.1. Cek1_Leu

Expression tests were performed with following constructs: pETZ2/Cek1_Leu, pETGb1/Cek1_Leu, pETTrx/Cek1_Leu and pETMBP/Cek1_Leu. The results showed that Cek1_Leu in fusion with Z2 and Gb1 tags were expressed almost exclusively in inclusion bodies, while with Trx and MBP tags the soluble fraction of fusion proteins were observed. Expression tests with pETNusA/Cek1_Leu construct have not been carried out yet.

4.2.1.1. Trx_Cek1_Leu expression test

Expression of soluble Cek1_Leu was assessed in fusion with Trx tag (pETTrx/Cek1_Leu construct) at analytical scale (50 mL cultures) using an incomplete factorial (IF) approach combining three *E. coli* strains [BL21Star(DE3), Origami2(DE3) and Rosetta(DE3)], three temperatures (20 °C, 25 °C and 30 °C) and two expression media (LB and 2YT). The expression patterns of 9 experiments (9 IF combinations, see Table 7 in section 3.2.2) were analysed by polyacrylamide gel electrophoresis. The results showed that at 25 °C and 30 °C Trx_Cek1_Leu was expressed almost exclusively in inclusion bodies (see Figure 19) while at 20 °C, although in small quantity, the expression of soluble form was observed in Rosetta(DE3) and

BL21Star(DE3) cells (see Figure 20). Furthermore the results also showed that the medium type (LB and 2YT) did not affect the expression of soluble Trx_Cek1_Leu (MW 54.79 kDa).

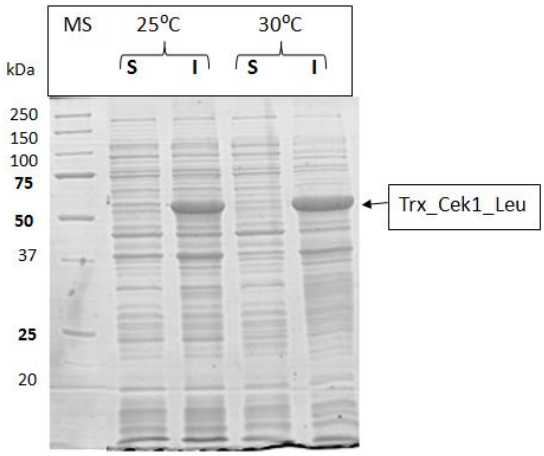


Figure 19- 12.5 % SDS-PAGE analysis: Trx_Cek1 expression pattern in Rosetta(DE3) cells at 25 °C and 30 °C. S- refers to the soluble and I- to the insoluble fractions. Similar expression patterns were obtained in BL21Star(DE3) and Origami2(DE3) cells.

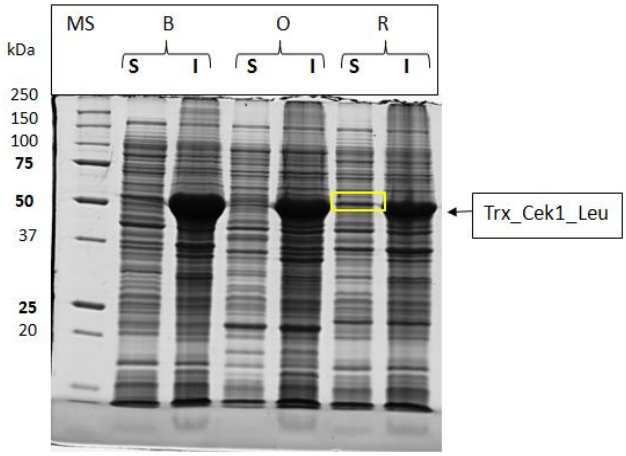


Figure 20-12.5 % SDS-PAGE analysis: Trx_Cek1 expression pattern in B- BL21Star(DE3), O- Origami2(DE3) and R- Rosetta(DE3) 20 °C. S- refers to the soluble and I- to the insoluble fractions.

4.2.1.2. Trx_Cek1_Leu high-scale expression and purification

The expression test patterns described above showed that the highest soluble Trx_Cek1_Leu (MW 54.79 kDa) expression was obtained at 20˚C in *E. coli* Rosetta(DE3), when the cells were collected 18 hours after induction. These expression conditions were selected for large scale production of Trx _Cek1_Leu. The N-terminal His6-tagged Trx_Cek1_Leu was overexpressed in 2000 ml of cell culture. After cell lysis, the soluble fraction was purified in two chromatography steps: first an ion immobilized affinity chromatography (IMAC), which was followed by a preparative size exclusion chromatography (SEC).

For the first purification step (IMAC), a column containing nickel (HisTrap HP column), which is known to bind histidine with high affinity, was used. To separate the His-tagged Trx_Cek1_Leu from other proteins, a step gradient of imidazole was used. A gel analysis of fractions collected during the chromatography is shown Figure 21. The load fraction corresponds to the sample that was loaded into the HisTrap column. The flow-through contains proteins, which do not show any binding to nickel. The 44 mM imidazole fraction (5% elution buffer) represents the proteins with low affinity to nickel. The Trx_Cek1_Leu was mostly eluted in the 92 mM imidazole fractions (15 % elution buffer, see Figure 21).

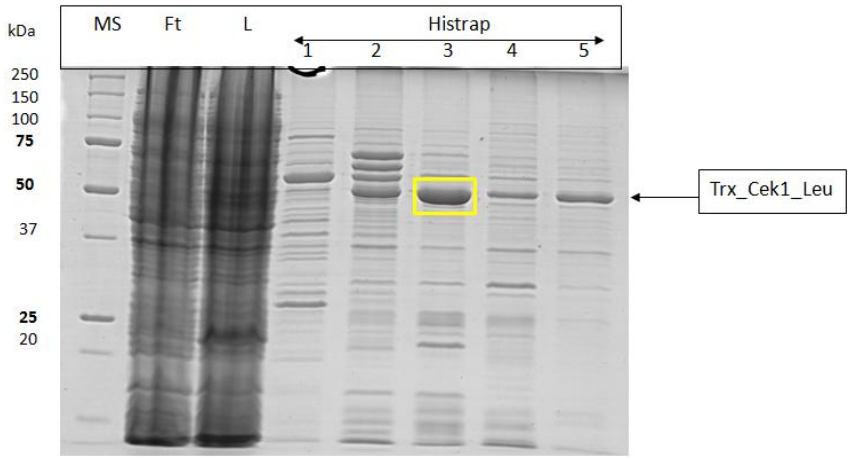


Figure 21- 12.5 % SDS-PAGE analysis of different fractions collected during immobilized metal affinity chromatography. In the first lane, Precision Plus Protein™ Standards molecular mass marker (Bio-Rad). "Ft" refers to the flow-through and "L" to the load. Fractions eluted with: lane1- 44 mM imidazole; lane 2 and 3 - 92 mM Imidazole; lane 4- 140 mM Imidazole and lane 5- 212 mM imidazole.

Fractions containing the most pure and concentrated Trx_Cek1_Leu (lane 3 in the Figure 21) were further purified by preparative size exclusion chromatography using a Hiprep 26/60 Sephacryl S-100 column. Figure 22A shows a typical elution profile for this SEC chromatography step. The first peak in this chromatogram corresponds to aggregates of Trx_Cek1_Leu (eluted in void volume of this column) and the second peak corresponds to monomeric Trx_Cek1_Leu (tested by analytical SEC). Fractions of this second peek were pooled, concentrated and analyzed on 12.5% SDS-PAGE (see Figure 22B). The SDS-PAGE gel showed that, after second purification step, the Trx_Cek1_Leu sample still contains some impurities. Furthermore the amount of this "pure" protein is quite low. From 2000 ml of bacterial broth, 2.2 mg of pure Trx_Cek1_Leu was obtained. Therefore, expression and purification of Cek1_Leu were further tested in fusion with another solubility tag (with MBP tag, see described below).

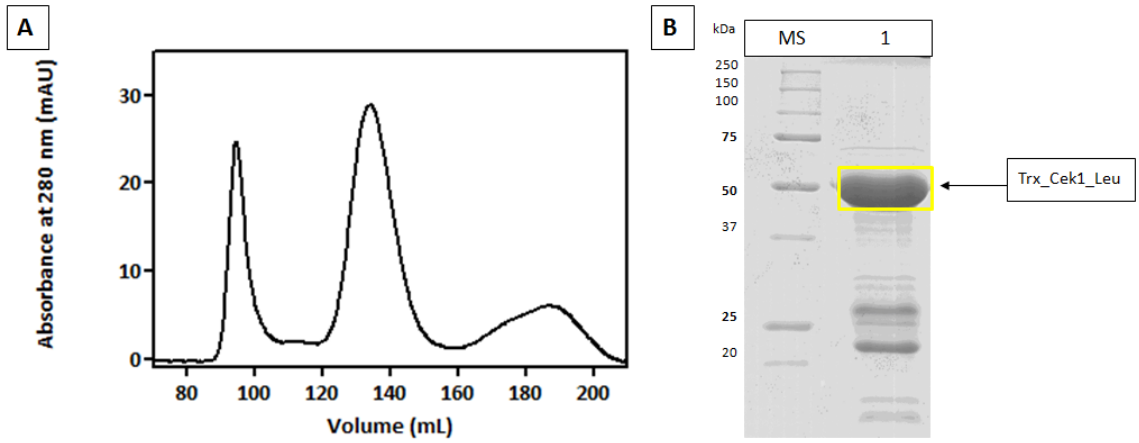


Figure 22- (**A**) SEC Chromatogram and (**B**) 12.5 % SDS-PAGE analysis of concentrated Trx_Cek1_Leu, in the first lane: Precision Plus ProteinTM Standards molecular mass marker (Bio-Rad)

4.2.1.3. MBP_Cek1_Leu high-scale expression and purification

Expression of Cek1_Leu in fusion with MBP tag (pETMBP/Cek1_Leu) was first tested at analytical scale (50 ml) at two different expression temperature (20 °C, and 25 °C) in *E. coli* strain BL21Star^{TM}(DE3). The large scale expression of MBP_Cek1_Leu (84.98 kDa) was performed at 20 °C for 18 hours (see Figure 23).

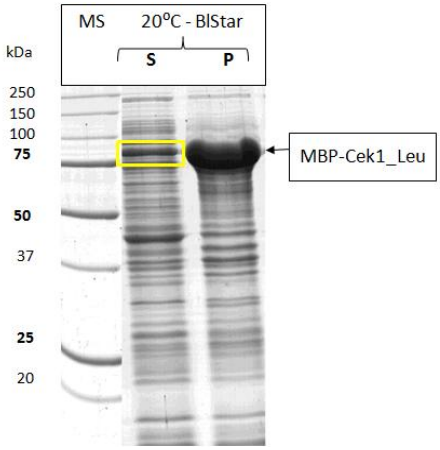


Figure 23- 12.5 % SDS-PAGE analysis: MBP_Cek1_Leu expression pattern in BL21Star™(DE3) cells at 20 °C. S- refers to the soluble and I- to the insoluble fractions

Purification of MBP_Cek1_Leu was performed using the same protocol as for purification of Trx_Cek1_Leu (described above) except that a Hiprep 26/60 Sephacryl S-200 (GE Healthcare) column was used for preparative SEC chromatography. In the first purification step MBP_Cek1_Leu was eluted from HisTrap column in all: 92 mM, 140 mM, 212 mM, 308 mM and 500 mM imidazole fractions (see Figure 24). Therefore a small (50 µl) sample from all these fractions was loaded to an analytical size exclusion column (Superose 12 10/300). The obtained analytical SEC chromatograms confirmed the MBP_Cek1_Leu eluted in 92 mM imidazole fractions contained less aggregates. Thus this fraction (lane 2 from Figure 24) was used to perform the second purification step, the preparative size exclusion chromatography.

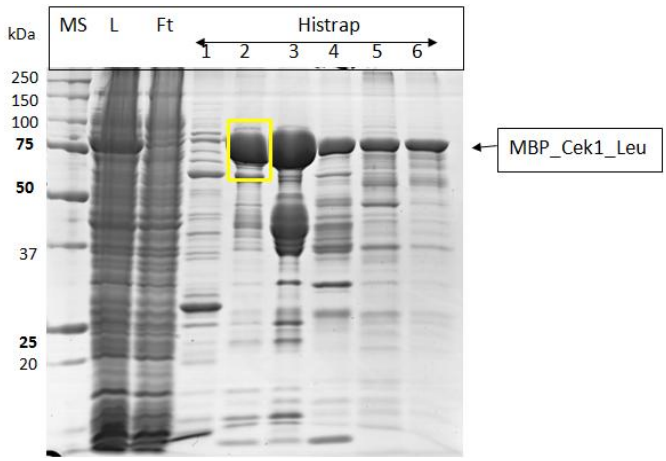


Figure 24- 12.5 % SDS-PAGE analysis of different fractions collected during immobilized metal affinity chromatography. In the first lane, Precision Plus Protein™ Standards molecular mass marker (Bio-Rad). "Ft" refers to the flow-through and "L" to the load. Fractions eluted with: 1- 44 mM, 2- 92 Mm, 3- 140 mM, 4- 212 mM, 5- 308 mM and 6- 500 mM imidazole.

The elution profile of this preparative SEC showed that, some of the protein was eluted in the void volume (first small peak of the chromatogram see Figure 25). However the significant amount of the protein eluted at a volume compatible with monomeric MBP_Cek1_Leu (second peak of the SEC chromatogram see Figure 25). The oligomerization state of the protein was determined by analytical SEC (see section 4.4.1).The fractions composing the second peak of preparative SEC chromatogram were pooled and concentrated to a final concentration of 14.5 mg/ml. This expression and purification protocol provided 8.5 mg pure soluble MBP_Cek1_Leu (see Figure 26) from 2000 ml bacterial broth.

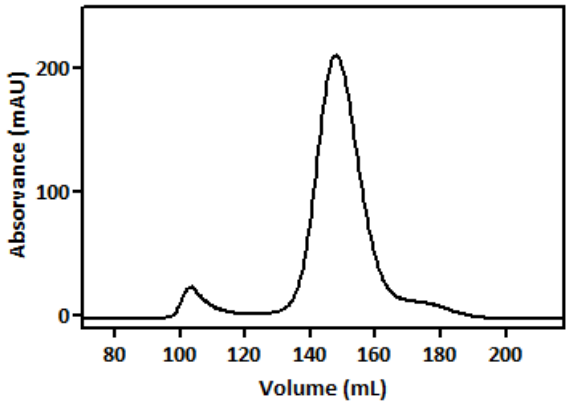


Figure 25- SEC chromatogram. MBP_Cek1_Leu eluted in 92 mM imidazole from HisTrap was applied to a Hiprep 26/60 Sephacryl S-200 column.

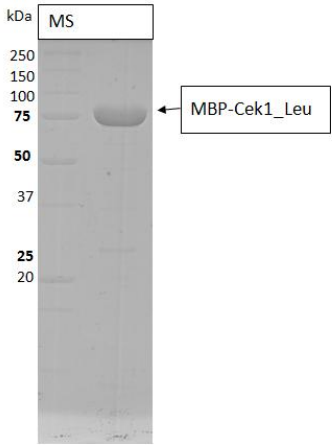


Figure 26- 12.5 % SDS-PAGE analysis of pure MBP_Cek1_Leu

4.2.2. Ras1_Ser

Expression of Ras1_Ser was tested in fusion with Trx (pETTrx/Ras1_Ser), MBP (pETMBP/Ras1_Leu) and Gb1 (pETGb1/Ras1_Ser) solubility tags. The expression tests and purification of Trx_Ras1_Ser and MBP_Ras1_Ser were performed by Zsuzsa Sárkány. The results of these experiments showed that Ras1_Ser in fusion with Trx tag forms soluble aggregates, while the soluble MBP_Ras1_Ser is monomeric (data not shown). As a consequence of these results it was decided to perform further expression tests with a smaller solubility tag as the Gb1. The results showed that Ras1_Ser in fusion with Gb1 was expressed in a significant amount in soluble form (detailed results described below). Expression tests with pETZ2/Ras1_Ser and pETNusA/Ras1_Ser constructs have not been carried out yet.

4.2.2.1. Gb1_Ras1_Ser high-scale expression and purification

Expression of Ras1_Ser in fusion with Gb1 solubility tag (pETGb1/Ras1_Ser construct) was first tested at analytical scale (50 ml) at two different expression temperature (20 °C, and 25 °C) in *E. coli* strain BL21(DE3)pLysS. The large scale overexpression of Gb1_Ras1_Ser (27.780 kDa) was performed at 25 °C for 4h (see Figure 27). Purification of Gb1_Ras_Ser was performed with the same two step purification protocol as used for purification of Trx_Cek1_Leu (see section 4.2.1.3). In the first purification step Gb1_Ras1_Ser was eluted from HisTrap column mainly in two 92 mM imidazole fractions (see Figure 28). These fractions were further purified by a preparative SEC on a Hiprep 26/60 Sephacryl S-200 column. In this purification step the aggregates, eluted in void volume, were separated from the monomeric Gb1_Ras1_Ser (see Figure 29). The oligomerization state of the protein was determined by analytical SEC (see section 4.4.1). The fractions composing the second peak of preparative SEC chromatogram were pooled and concentrated to a final concentration of 20 mg/ml. This expression and purification protocol provided 17 mg pure soluble Gb1_Ras1_Ser (see Figure 30) from 2000 ml bacterial broth.

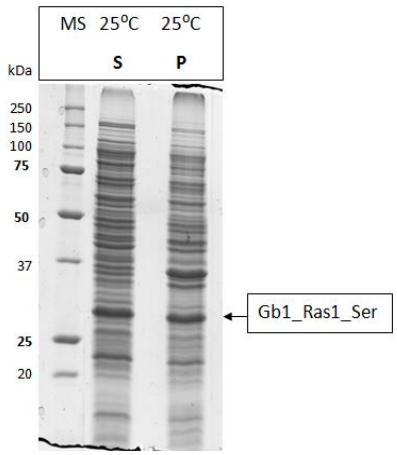


Figure 27- 12.5 % SDS-PAGE: Gb1_Ras1_Ser expression pattern in BL21(DE3)PLysS cells. In the first lane, Precision Plus ProteinTM Standards molecular mass marker (Bio-Rad) (MS). S- refers to the soluble and I- to the insoluble fractions.

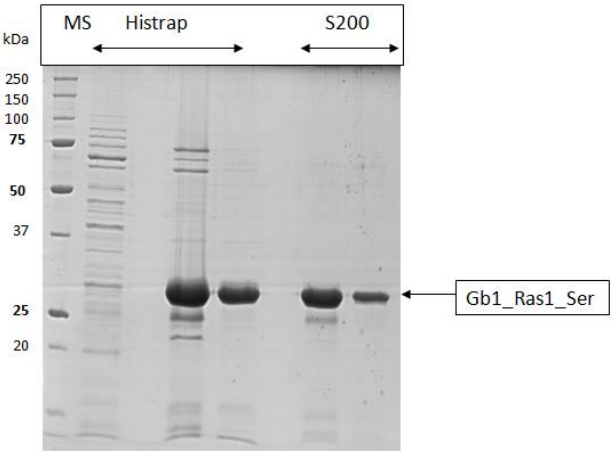


Figure 28- 12.5% SDS-PAGE analysis of different fractions eluted from Histrap and SEC.

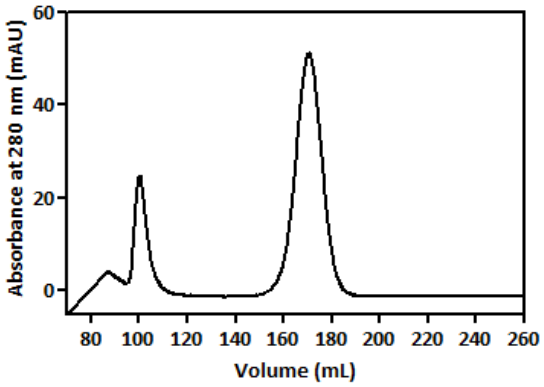


Figure 29- SEC chromatogram. Gb1_Ras1_Ser eluted in 92 mM imidazole from HisTrap was applied to a Hiprep 26/60 Sephacryl S-200 column.

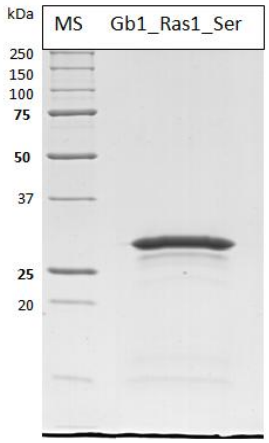


Figure 30- 12.5 % SDS-PAGE analysis of pure Gb1_Ras1_Ser.

4.3. TEV protease cleavage assay

Purified MBP_Cek1_Leu was digested by in-house expressed and purified recombinant TEV protease in order to cleave the MBP tag together with histidine tag. Pilot digestions were performed using several enzyme/substrate molar ratios (1:2.5; 1:5; 1:10; 1:20; see Figure 31) at room temperature for 2h. The results, analyzed in 12.5% SDS-PAGE, show that in all four conditions (different enzyme/substrate molar ratios), the His6_ MBP tag was almost completely cleaved MBP_Cek1_Leu. Therefore for a scale-up of the process a molar ratio of 1:20 (enzyme/ substrat) will be used. The His6_MBP tags and TEV (which possesses itself a His6Tag) will be separated from Cek1_Leu by a nickel affinity chromatography (HisTrap). Cek1_Leu should be collected in the flow-through.

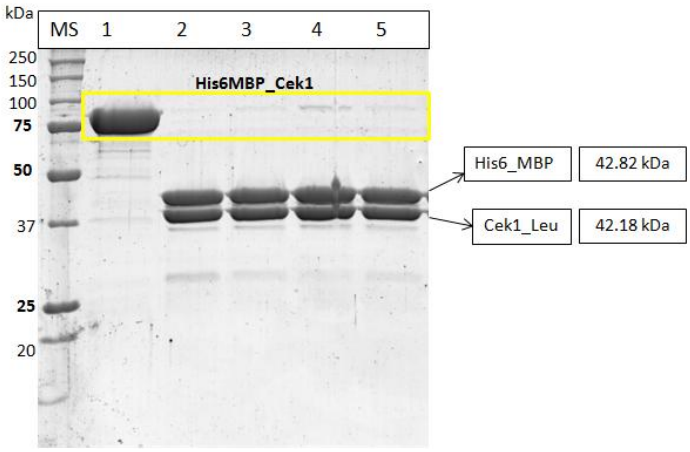


Figure 31- 12.5%SDS-Page analysis of TEV protease cleavage pilot experiments, performed with different enzyme/substrate molar ratios: 1- 0:1 (control); 2- 2- 1:2.5; 3- 1:5 4- 1:10; 5- 1:20.

4.4. Protein homogeneity/oligomerization state

4.4.1. Analytical size-exclusion chromatography (SEC)

The oligomerization state of the pure MBP_Cek1_Leu and Gb1_Ras1_Ser were assessed by analytical size exclusion chromatography. The pure MBP_Cek1_Leu and Gb1_Ras1_Ser were eluted from Superose 12 10/300 column as a single sharp peak with 11.95 and 12.96 ml of elution volumes (V_e) respectively, [see Figure 32A-B]. The superposed SEC chromatograms of standard molecular weight markers are shown in Figure 33A. The calibration curve for determination of molecular weights (see Figure 33B) was prepared as described in section 3.4 (materials and methods). Using this calibration curve (Figure 33B) the molecular weights of

MBP_Cek1_Leu (K_{av}=0.28) and Gb1_Ras1_Ser (K_{av}=0.34) were found to be 89.2 and 31 kDa, respectively. These values are consistent with a monomer of MBP_Cek1_Leu (84.98kDa) and Gb1_Ras1_Ser (27.78) kDa. Thus, the analytical size exclusion chromatography confirmed the monomeric that both fusion proteins are likely to be monomeric in solution (MBP_Cek1_Leu and Gb1_Ras1_Ser).

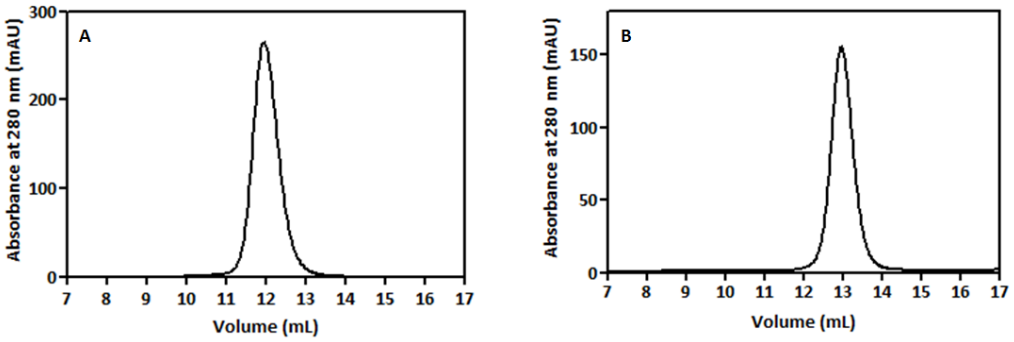


Figure 32- Analytical SEC chromatogram of (A) MBP_Cek1_Leu and (B) Gb1_Ras1_Ser

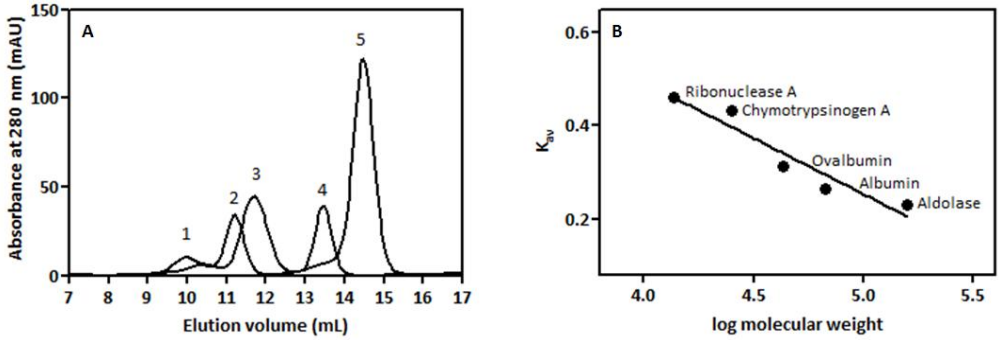


Figure 33- Calibration of Superpose 12 10/300 column. (A) Two analytical SEC chromatograms of standard molecular weight markers are superposed, where 1 refers to aldolase (Molecular weight: 158 kDa; elution volume: 11.08 ml), 2 to albumin (67 kDa; 11.68 ml), 3 to ovalbumin (43 kDa; 12.45 ml), 4 to chymotrypsinogen A (25 kDa; 14.47 ml) and 5 to ribonuclease A (13.7 kDa; 14.96 ml). (B) Calibration curve for determination of molecular weight by analytical SEC.

4.4.2. Dynamic Light Scattering (DLS)

The homogeneity of pure concentrated MBP_Cek1_Leu (14 mg/ml) and Gb1_Ras1_Ser (20 mg/ml) were also assessed by DLS. Since the obtained DLS data (size distributions by intensity) show a substantial tail, the plots of the size distribution by intensity were converted to volume distribution (see Figure 34Figure 35) to get a more accurate interpretation of the data. The DLS

spectrums of MBP_Cek1_Leu and Gb1_Ras1_Ser showed one major population by size (see Figure 34 Figure 35) with a polydispersity (%*Pd*) of 18.8 and 18.5% respectively. As a rule of thumb, samples with % Pd <~20%, are considered to be monodisperse (see section 3.5). The R_H (hydrodynamic radius) values were obtained from plots of size distribution as a function of volume. The MBP_Cek1_Leu and Gb1_Ras1_Ser R_H values were found to be 4.3 and 2.43 nm respectively. These values were used to estimate the molecular weight (assuming a globular protein) using an empirical calibration graph developed by Malvern Instruments available in the Nano software (42). Accordingly, the molecular mass of MBP_Cek1_Leu and Gb1_Ras1_Ser were estimated to be 100 kDa and 26.6 kDa respectively. These values correspond to a monomeric form of MBP_Cek1_Leu (84.98kDa) and Gb1_Ras1_Ser (27.78 kDa). Thus, the DLS results (the pure MBP_Cek1_Leu and Gb1_Ras1_Ser samples are monodispers and monomeric) are consistent with the data obtained from analytical SEC (described above section 4.4.1).

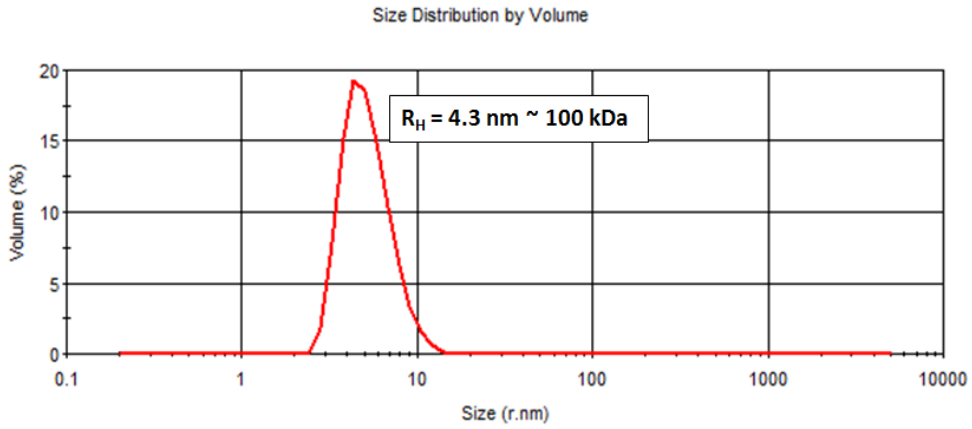


Figure 34- Distribution of hydrodynamic radii by volume from DLS spectrum for MBP_Cek1_Leu.

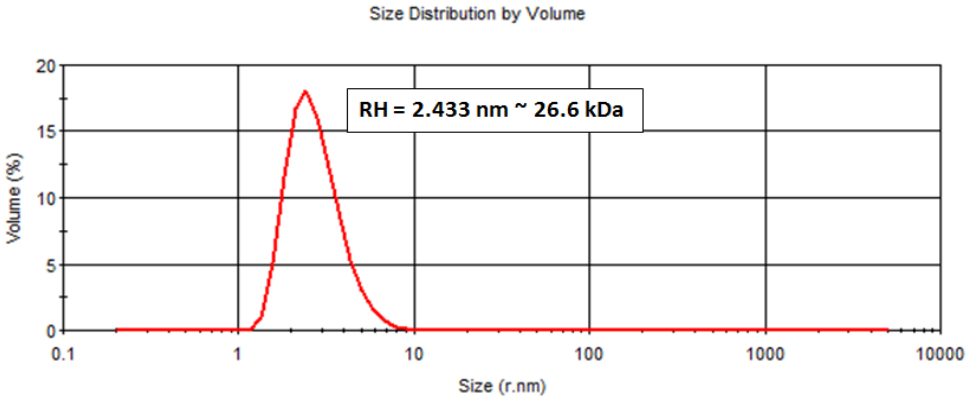


Figure 35- Distribution of hydrodynamic radii by volume from DLS spectrum for Gb1_Ras1_Ser.

4.5. Crystallization screenings

Crystallization trials were performed with the purified Gb1_Ras1_Ser protein (20 mg/mL) and MBP_Cek1_Leu (14.5 mg/ml) by vapor diffusion (sitting drop) techniques using the following crystallization kits: Morpheus HT96 and PACT *premier*™ Crystallization Screen MD1-29 (Molecular Dimensions Limited). In the majority of these crystallization conditions precipitation and phase separation was observed. Further crystallization screening for the preliminary crystallization conditions are going to be done with Ras1_Ser, Ras1_Leu, Cek1_Leu and Cek1_Ser after removal of the solubility tags.

5. Conclusions and future perspectives

In the present work, the catalytic domains of Cek1 (Leu variant for CUG codon) and Ras1 (Ser variant for CUG codon) were successfully cloned into five pET-derived expression vectors with variable 6 His-tagged N-terminus solubility tags (Gb1, NusA, MBP, Trx and Z2) cleavable with TEV protease. Moreover, the Ser variant (for CUG codon) of Cek1 and Leu variant (for CUG codon) of Ras1 were also prepared by site directed mutagenesis.

Using an incomplete factorial approach, high level bacterial expression and purification protocols for the active domains of Cek1 (in fusion with MBP) and Ras1 (in fusion with Gb1) were developed. Furthermore, analytical size exclusion chromatography and dynamic light scattering results indicated that both recombinant fusion proteins are monomeric. To avoid any possible effect of the solubility tags on the crystallization and enzymatic activity of Cek1 and Ras1 (for future studies), the His6_MBP and His6_Gb1 cleavage is being implemented in our expression and purification protocol.

Crystallization screenings for both proteins are currently in progress, aiming the determination of their three-dimensional structures by X-ray crystallography. The structures of Ras1 and Cek1 with Ser or Leu at CUG positions, together with a thorough analysis of their stability and function *in vitro*, will provide valuable insights into a possible strategic role of natural codon ambiguity in signaling pathways associated with morphogenesis and virulence.

6. Appendix

eurofins mwg|operon

Gene Synthesis Quality Assurance Documentation

Order No.: **2351840/Sarkany**

Plasmid Name:	pBluescript II SK(+)-cek1_candida_alb	**Internal Name:**	N112
Gene Name:	cek1_candida_alb	**Gene Size:**	1262bp
Vector Backbone:	pBluescript II SK(+)	**Antibiotic Selection:**	Ampicillin
Cloning:	cloned via SmaI (blunt)	**Quantity:**	5.3µg
Designation:	E. coli XL-10 Gold		
Genotype:	Tet^r *Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB lacI*^{q} *ZΔM15 Tn10* (Tet^r) Amy Cam^r]^a		

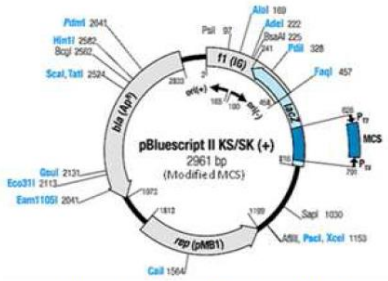
Plasmid Map

PCR Product: cek1_candida_alb **5', 3' Cloning Sites:** NcoI/Acc65I
cloned via SmaI (blunt)

Modified Multiple Cloning Site of pBluescript

```
....ACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGT
      M13 Forward primer
     SfiI                    EcoRV
ACGGCCGTCAAGGCCAAGCTTCCCGGGATATCACGTGAAGCTTGCAAGCTCCAGCTTTTGTTCCC
            HindIII SmaI       Eco72I HindIII

TTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA....
                                  M13 Reverse Primer
```



Please Note:

Verify sequence after each cloning step. All material left at Eurofins MWG Operon will be discarded after 1 month.

Sequence analysis was done via doublestrand DNA sequencing. Sequence congruence was 100%.

Please find your sequence trace files in your personal account on our **ECOM system (order ID. 2351840)**.

The plasmid DNA has been lyophilised. We recommend to dissolve it in 10mM Tris buffer or TE.

Eurofins MWG Operon 28.02.2011

CERTIFIED ISO 9001:2000 ACCORDING TO

René Scheel
Project Manager Gene Synthesis

Sebastian Kubny
Project Assistant Gene Synthesis

Figure 36- Quality Assurance Documentation of the synthetic gene of Cek1_Leu (leucine variant for CUG codon position) purchased from Eurofins MWG Operon.

eurofins mwg|operon

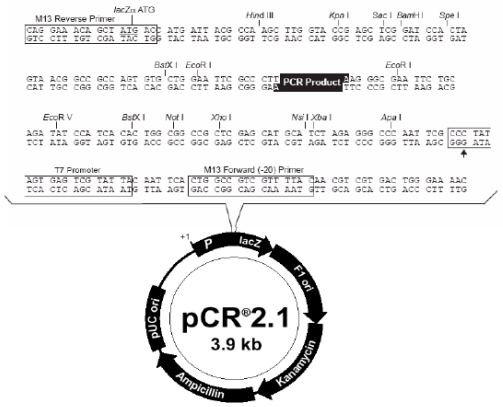
Gene Synthesis Quality Assurance Documentation

Order No.: **2351840/Sarkany**

Plasmid Name:	pCR2.1-ras1_candida_alb	**Internal Name:**	N113-8
Gene Name:	ras1_candida_alb	**Gene Size:**	884bp
Vector Backbone:	pCR2.1	**Antibiotic Selection:**	Ampicillin and Kanamycin
Cloning:	cloned via TOPO-TA	**Quantity:**	1.4µg
Designation:	E. coli TOP10		
Genotype:	F- *mcr* A Δ(*mrr -hsd* RMS-*mcr* BC) Φ 80*lac* ZΔM15 Δ*lac* X74 *rec* A1 *ara* D139 Δ(*ara-leu*)7697 *gal* U *gal* K *rps* L (Str^{R}) *end* A1 *nup* G		

Plasmid Map

PCR Product: ras1_candida_alb **5', 3' Cloning Sites:** NcoI/Acc65I cloned via TOPO-TA



Please Note:

Verify sequence after each cloning step. All material left at Eurofins MWG Operon will be discarded after 1 month.

Sequence analysis was done via doublestrand DNA sequencing. Sequence congruence was 100%.

Please find your sequence trace files in your personal account on our **ECOM system (order ID. 2351840)**.

The plasmid DNA has been lyophilised. We recommend to dissolve it in 10mM Tris buffer or TE.

Eurofins MWG Operon 07.02.2011

CERTIFIED ISO 9001:2000 ACCORDING TO

René Scheel
Project Manager Gene Synthesis

Sebastian Kubny
Project Assistant Gene Synthesis

Figure 37- Quality Assurance Documentation of the synthetic gene of Ras1_Ser (serine variant for CUG codon position) purchased from Eurofins MWG Operon.

7. Bibliography

1. Sudbery PE. Growth of Candida albicans hyphae. Nature reviews. Microbiology. 2011 Aug 16;9(10):737–48.

2. Pfaller MA, Diekema DJ. Epidemiology of invasive candidiasis: a persistent public health problem. Clinical microbiology reviews. 2007 Jan 1; 20(1):133–63.

3. Shapiro RS, Robbins N, Cowen LE. Regulatory circuitry governing fungal development, drug resistance, and disease. Microbiology and molecular biology reviews : MMBR. 2011 Jun; 75(2):213–67.

4. Merson-Davies LA, Odds FC. A morphology index for characterization of cell shape in Candida albicans. Journal of general microbiology. 1989 Nov; 135(11):3143–52.

5. Sudbery P, Gow N, Berman J. The distinct morphogenic states of Candida albicans. Trends in microbiology. 2004 Jul ;12(7):317–24.

6. Bahn Y-S, Staab J, Sundstrom P. Increased high-affinity phosphodiesterase PDE2 gene expression in germ tubes counteracts CAP1-dependent synthesis of cyclic AMP, limits hypha production and promotes virulence of Candida albicans. Molecular Microbiology. 2003 Aug 27; 50(2):391–409.

7. Saville SP, Lazzell AL, Monteagudo C, Lopez-Ribot JL. Engineered control of cell morphology in vivo reveals distinct roles for yeast and filamentous forms of Candida albicans during infection. Eukaryotic cell. 2003 Oct;2(5):1053–60.

8. Moyes DL, Runglall M, Murciano C, Shen C, Nayar D, Thavaraj S, et al. A biphasic innate immune MAPK response discriminates between the yeast and hyphal forms of Candida albicans in epithelial cells. Cell host & microbe. 2010 Sep 16;8(3):225–35.

9. Crick FHC. The origin of the genetic code. Journal of Molecular Biology. 1968 Dec;38(3):367–79.

10. Jukes TH, Osawa S. Evolutionary changes in the genetic code. Comparative biochemistry and physiology. B, Comparative biochemistry. 1993 Nov;106(3):489–94.

11. Santos M a, Cheesman C, Costa V, Moradas-Ferreira P, Tuite MF. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. Molecular microbiology. 1999 Mar;31(3):937–47.

12. Massey SE, Moura G, Almeida R. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. Genome. 2003;13(4):544–57.

13. Miranda I, Rocha R, Santos MC, Mateus DD, Moura GR, Carreto L, et al. A genetic code alteration is a phenotype diversity generator in the human pathogen Candida albicans. PloS one. 2007 Jan;2(10):e996.

14. Moura GR, Carreto LC, Santos MAS. Genetic code ambiguity: an unexpected source of proteome innovation and phenotypic diversity. Current opinion in microbiology. 2009 Dec;12(6):631–7.

15. Grosjean H, de Crécy-Lagard V, Marck C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. FEBS letters. 2010 Jan 21;584(2):252–64.

16. Gomes AC, Miranda I, Silva RM, Moura GR, Thomas B, Akoulitchev A, et al. A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans. Genome biology. 2007 Jan;8(10):R206.

17. Santos M a, Keith G, Tuite MF. Non-standard translational events in Candida albicans mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. The EMBO journal. 1993 Mar;12(2):607–16.

18. Santos M a, Tuite MF. The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. Nucleic acids research. 1995 May 11;23(9):1481–6.

19. Moura GR, Paredes J a, Santos M a S. Development of the genetic code: insights from a fungal codon reassignment. FEBS letters. 2010 Jan 21;584(2):334–41.

20. Suzuki T, Ueda T, Watanabe K. The "polysemous" codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. The EMBO journal. 1997 Mar 3;16(5):1122–34.

21. Rocha R, Barbosa Pereira PJ, Santos M a S, Macedo-Ribeiro S. Purification, crystallization and preliminary X-ray diffraction analysis of the seryl-tRNA synthetase from Candida albicans. Acta crystallographica. Section F, Structural biology and crystallization communications. 2011 Jan 1;67(Pt 1):153–6.

22. Feketová Z, Masek T, Vopálenský V, Pospísek M. Ambiguous decoding of the CUG codon alters the functionality of the Candida albicans translation initiation factor 4E. FEMS yeast research. 2010 Aug 1;10(5):558–69.

23. Rocha R, Pereira PJB, Santos M a S, Macedo-Ribeiro S. Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proceedings of the National Academy of Sciences of the United States of America. 2011 Aug 23;108(34):14091–6.

24. Shapiro RS, Uppuluri P, Zaas AK, Collins C, Senn H, Perfect JR, et al. Hsp90 orchestrates temperature-dependent Candida albicans morphogenesis via Ras1-PKA signaling. Current biology : CB. 2009 Apr 28;19(8):621–9.

25. Leberer E, Harcus D, Dignard D, Johnson L, Ushinsky S, Thomas DY, et al. Ras links cellular morphogenesis to virulence by regulation of the MAP kinase and cAMP signalling pathways in the pathogenic fungus Candida albicans. Molecular microbiology. 2001 Nov;42(3):673–87.

26. Leberer E, Ziegelbauer K, Schmidt A, Harcus D, Dignard D, Ash J, et al. Virulence and hyphal formation of Candida albicans require the Ste20p-like protein kinase CaCla4p. Current biology : CB. 1997 Aug 1;7(8):539–46.

27. Rocha CR, Schröppel K, Harcus D, Marcil A, Dignard D, Taylor BN, et al. Signaling through adenylyl cyclase is essential for hyphal growth and virulence in the pathogenic fungus Candida albicans. Molecular biology of the cell. 2001 Nov;12(11):3631–43.

28. Csank C, Schröppel K, Leberer E, Harcus D, Mohamed O, Meloche S, et al. Roles of the Candida albicans mitogen-activated protein kinase homolog, Cek1p, in hyphal development and systemic candidiasis. Infection and Immunity. 1998;66(6):2713–21.

29. Chen J, Chen J, Lane S, Liu H. A conserved mitogen-activated protein kinase pathway is required for mating in Candida albicans. Molecular Microbiology. 2002 Dec 9;46(5):1335–44.

30. Whiteway M, Dignard D, Thomas DY. Dominant negative selection of heterologous genes: isolation of Candida albicans genes that interfere with Saccharomyces cerevisiae mating factor-induced cell cycle arrest. Proceedings of the National Academy of Sciences of the United States of America. 1992 Oct 15;89(20):9410–4.

31. Lengeler KB, Davidson RC, Souza CD, Harashima T, Shen W-chiang, Wang P, et al. Signal Transduction Cascades Regulating Fungal Development and Virulence. Society. 2000;64(4):746–85.

32. Pai EF, Krengel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. The EMBO journal. 1990 Aug;9(8):2351–9.

33. Buhrman G, Wink G, Mattos C. Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. Structure (London, England : 1993). 2007 Dec;15(12):1618–29.

34. Marshall MS, Gibbs JB, Scolnick EM, Sigal IS. Regulatory function of the Saccharomyces cerevisiae RAS C-terminus. Molecular and cellular biology. 1987 Jul;7(7):2309–15.

35. Buhrman G, Holzapfel G, Fetics S, Mattos C. Allosteric modulation of Ras positions Q61 for a direct role in catalysis. Proceedings of the National Academy of Sciences of the United States of America. 2010 Mar 16;107(11):4931–6.

36. Reményi A, Good MC, Bhattacharyya RP, Lim WA. The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network. Molecular cell. 2005 Dec 22;20(6):951–62.

37. Manchester KL. Value of A260/A280 ratios for measurement of purity of nucleic acids. BioTechniques. 1995 Aug;19(2):208–10.

38. Benoit I, Coutard B, Oubelaid R, Asther M, Bignon C. Expression in Escherichia coli, refolding and crystallization of Aspergillus niger feruloyl esterase A using a serial factorial approach. Protein expression and purification. 2007 Sep;55(1):166–74.

39. Berne, B. J., and Pecora R. Dynamic Light Scattering: With Application to Chemistry, Biology and Physics 3rd Ed. Mineola, NY: 2000.

40. Technical note MRK656-01. 2012;Available from: http://www.malvern.com

41. Zhou, L., and Chen J. Therapeutic Protein Characterization Using Light Scattering Techniques, Technical Note MRK704 01. Worcestershire, U.K: 2010.

42. MAN0317 Issue2. In: Zetasizer Nano Series User Manual.

43. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research. 1994 Nov 11;22(22):4673–80.

Communication at scientific meeting

Carvalho, V.; Sárkány, Z.; Barbosa, PJP; and Macedo-Ribeiro, S. "CUG ambiguity in the structure of Ras1 and Cek1 from *Candida albicans*" I^{3}S Scientific Retreat (2012), Póvoa do Varzim (poster presentation).