
**COLD-ADAPTED BACTERIA
RESPONSE TO TEMPERATURE
CHANGES: FROM PHYSIOLOGY TO
EXPLOITATION OF THEIR
BIOTECHNOLOGICAL POTENTIAL**

Marzia Calvanese



Dottorato in Biotecnologie – XXXV ciclo

Università di Napoli Federico II

Dottorato in Biotecnologie – XXXV ciclo

Università di Napoli Federico II



**COLD-ADAPTED BACTERIA
RESPONSE TO TEMPERATURE
CHANGES: FROM PHYSIOLOGY TO
EXPLOITATION OF THEIR
BIOTECHNOLOGICAL POTENTIAL**

Marzia Calvanese

Dottorando: Marzia Calvanese

Relatore: Ermenegilda Parrilli

Correlatore: Maria Luisa Tutino

Coordinatore: Prof. Marco Moracci

Settore Scientifico Disciplinare: CHIM11

Alla mia Famiglia, il mio porto sicuro

Index	
Riassunto	1
Summary	4
Introduction	5
1 Cold adaptation strategies of the Antarctic bacteria to temperature changes	5
2 <i>Pseudoalteromonas haloplanktis</i> TAC125 as a model for adaptation to environmental fluctuations	7
2.1 Studies on the acclimation strategies adopted by psychrophilic bacterium	8
3 <i>Pseudoalteromonas haloplanktis</i> TAC125 as unconventional host for difficult-to-express protein production	9
Chapter 1: Combined analysis of the metabolome and transcriptome to explore adaptative mechanisms in <i>Ph</i>TAC125 in response to temperature changes	15
Metabolic Robustness to Growth Temperature of a Cold-Adapted Marine Bacterium	16
1.2 In-depth extracellular metabolic analysis in <i>Pseudoalteromonas haloplanktis</i> TAC125 towards the medium optimization	48
1.2.1 Introduction	48
1.2.2 Material and Methods	49
1.2.2.1 Strains and growth conditions	49
1.2.2.2 RNA Sequencing.....	49
1.2.2.3 Extracellular metabolomics and iron detection.....	49
1.2.3 Results.....	50
1.2.3.1 Differences in the extracellular metabolite composition at 15°C and 0°C The 0 and 15°C transcriptomes of <i>Ph</i> TAC125	50
1.2.3.2 Transcriptomic analysis reveals differentially expressed genes associated to lactate and 2-oxoglutarate	56
1.2.4 Discussion	57
Chapter 2: Improvement of <i>Pseudoalteromonas haloplanktis</i> TAC125 as host for recombinant protein production: development of high-copy number plasmids	63
Improvement of <i>Pseudoalteromonas haloplanktis</i> TAC125 as host for recombinant protein production: development of high-copy number plasmids	64
2.1 Introduction	65
2.2 Material and methods	66

2.2.1 Bacterial strains, media, and plasmids.....	66
2.2.2 Plasmid vector preparation	67
2.2.3 Construction of the plasmid library	67
2.2.4 Electroporation of KrPL	67
2.2.5 FACS sorting.....	68
2.2.6 Selection of mutants with higher plasmid copy number and fluorescence assays.....	69
2.2.7 SDS-PAGE.....	69
2.2.8 Plasmid Copy Number determination via qPCR	69
2.2.9 Plasmid loss frequency assays (Plasmid stability assay).....	70
2.2.10 Sequence analysis	71
2.2.11 Deletions of DNA sequences in the replication origin	71
2.2.12 RNA isolation and RNA-Seq data analysis of pMtBL	71
2.3 Results.....	72
2.3.1 Set up of a library of psychrophilic plasmids containing randomly mutated pMtBL replication origins	72
2.3.2 FACS sorting of psychrophilic cells harboring the randomly mutated plasmid library	72
2.3.3 Analysis of FACS-sorted psychrophilic populations and molecular characterization of some selected clones.	74
2.3.4 Sequence analysis of the isolated mutant plasmids and deletion mutations affecting the replication sequence	75
4 Discussion	77
Chapter 3: Study of condensation in the Antarctic bacterium <i>Pseudoalteromonas haloplanktis</i> TAC125 producing a recombinant intrinsically disordered protein, the case of <i>hCDKL5</i>	87
3.1 Introduction.....	88
3.1.1 The intrinsically disordered proteins (IDPs) and their recruitment into protein condensates	88
3.1.2 The study of the human cyclin-dependent kinase-like 5 (<i>hCDKL5</i>) containing an intrinsically disordered region (IDR)	89
3.1.3 An integrated approach of plasmid design and strain engineering finalized to improve the <i>hCDKL5</i> production in <i>Pseudoalteromonas</i> <i>haloplanktis</i> TAC125.....	90
3.2 Materials and Method	92
3.2.1 Bacterial strains, media, and plasmids	92
3.2.3 Construction of the fICDKL5 mutants	92
3.2.3 Growth conditions	92

3.2.4 Imaging KrPL <i>LacY</i> ⁺ cells	93
3.2.5 Analysis of the recombinant proteins production	93
3.3 Results and Discussion	95
3.3.1 Design and construction of <i>h</i> CDKL5 mutants	95
3.3.2 Cell imaging experiments <i>in vivo</i> of mCherry-flCDKL5 and mutants expressed in KrPL <i>LacY</i> ⁺ cells.....	97
3.4 Conclusion	101
Communications and Publications	105
Appendix	108

Riassunto

Introduzione

Alcuni habitat della Terra sono poco colonizzati dai microorganismi a causa delle basse temperature, le quali influenzano negativamente la fisiologia e la funzionalità cellulare. Pertanto, alcuni microrganismi, noti come psicofili, hanno evolutivamente sviluppato una serie di strategie adattive tali da consentire la loro proliferazione in questi ambienti estremi. Inoltre, l'avanzamento nelle tecnologie "omiche" ha consentito di identificare i meccanismi adattativi sviluppati in risposta alle basse temperature. In particolare, un elevato numero di geni codificanti principalmente per proteine trascrizionali e traduzionali sono stati identificati dall'analisi dei genomi dei batteri psicofili, indicando che la trascrizione, la traduzione, e il *fold*ing proteico sono dei processi limitanti a basse temperature. Ulteriori studi hanno riportato che le basse temperature riducono la velocità di reazione; pertanto, gli psicofili producono enzimi adattati al freddo caratterizzati da bassa termostabilità, elevata flessibilità strutturale e attività specifica per mantenere velocità catalitiche adeguate. Infine, le analisi dei trascritti di alcuni microorganismi hanno dimostrato che l'esposizione a basse temperature induce una variazione della composizione della membrana, a causa della sua rigidità a basse temperature. Gli studi sull'adattamento ambientale dei microorganismi adattati al freddo sono diventati cruciali negli ultimi decenni a causa degli effetti dei cambiamenti climatici nell'ambiente antartico. Studi dimostrano che questi effetti influenzeranno negativamente la composizione e la funzione della comunità microbica. Pertanto, è necessario approfondire le nostre conoscenze sui processi evolutivi per poter comprendere i meccanismi molecolari dei microrganismi psicofili influenzati dal riscaldamento globale e per incorporare queste informazioni in un sistema predittivo. *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*) è uno dei batteri psicofili maggiormente studiati. Negli ultimi anni numerosi studi sono stati condotti in questo batterio consentendo a una migliore comprensione della plasticità genomica e dei processi metabolici che contribuiscono all'adattamento di *PhTAC125* alle basse temperature. Tuttavia, sono necessarie ulteriori ricerche per valutare se questo batterio psicofilo possa adattarsi ai cambiamenti ambientali causati dal riscaldamento globale, evitando così la sua estinzione. Inoltre, *PhTAC125* costituisce non solo un modello in ambito ecologico, ma possiede anche un enorme potenziale come ospite non convenzionale per la produzione di proteine ricombinanti. Infatti, *PhTAC125* rappresenta una valida alternativa ai sistemi convenzionali (*E. coli*, *B. subtilis*), grazie alle sue caratteristiche fisiologiche come una rapida crescita in un ampio intervallo di temperature ed un'efficiente sintesi proteica. Inoltre, una serie di significativi miglioramenti come l'ottimizzazione di un terreno di coltura, lo sviluppo di efficienti sistemi di espressione e tecniche di mutagenesi nel genoma batterico hanno determinato la produzione di proteine eucariotiche complesse in *PhTAC125*.

Capitolo 1: Analisi di metabolomica e trascrittomica per determinare i meccanismi di adattamento in *PhTAC125* a due diverse temperature

In questo capitolo, le strategie di adattamento al freddo in *PhTAC125* a 15 °C e 0°C sono state valutate integrando i dati ottenuti dall'analisi di trascrittomica e metabolomica. L'analisi dei trascritti ha mostrato che i geni coinvolti nel metabolismo energetico, nella sintesi degli acidi grassi e nel metabolismo degli aminoacidi sono maggiormente espressi a 0 °C rispetto a 15 °C. L'analisi metabolica, invece, non ha evidenziato variazioni significative nei profili dei metaboliti prodotti alle due temperature. Questa robustezza metabolica è dovuta alla capacità del batterio psicrofilo di modulare l'espressione di numerosi geni coinvolti nel metabolismo centrale di *PhTAC125*. Un altro obiettivo di questo studio è stato l'ottimizzazione del mezzo GG, identificando potenziali fattori che possano limitare la crescita di *PhTAC125*; per questo motivo l'analisi di metabolica extracellulare è stata ripetuta analizzando un maggior numero di campioni raccolti durante le diverse fasi di crescita a 15 °C e 0 °C. In questa analisi, nuovi metaboliti sono stati rilevati, ma tuttavia solo i profili del lattato e del α -chetoglutarato hanno mostrato cambiamenti significativi alle due temperature. Inoltre, i risultati ottenuti dalla determinazione di ioni ferrici nel terreno di coltura indicherebbero una condizione di ferro limitante durante la fase esponenziale a 15 °C. L'aumento di solfato di ferro nel terreno di coltura a 15 °C ha determinato una riduzione significativa dei livelli extracellulari di lattato e α -chetoglutarato, ma tuttavia, i parametri cinetici non sono migliorati. Infine, i risultati di metabolomica sono stati integrati con quelli della trascrittomica, rivelando una maggiore espressione dei geni coinvolti nell'assimilazione delle fonti di azoto e nel catabolismo del lattato, invece una riduzione dei livelli dei trascritti coinvolge i geni codificanti gli enzimi del ciclo dell'acido tricarbossilico (TCA) a 15 °C. Complessivamente, questi risultati suggeriscono che le fonti di carbonio, azoto e ferro potrebbero essere sbilanciate durante la crescita di *PhTAC125* a 15 °C.

Capitolo 2: Ottimizzazione di *Pseudoalteromonas haloplanktis* TAC125 come ospite per la produzione di proteine ricombinanti: sviluppo di plasmidi ad alto numero di copie

Negli ultimi anni, efficienti sistemi di espressione sono stati sviluppati in *PhTAC125*, consentendo la produzione in forma solubile e attiva di proteine considerate difficili da esprimere in ospiti convenzionali. Nonostante ciò, altri fattori sono stati rilevati che limitino l'efficienza di produzione di proteine ricombinate, come ad esempio il basso numero di copie dei vettori psicrofili. A tale scopo, l'origine di replicazione (OriR) è stata mutata in modo casuale; tali sequenze sono state clonate in un vettore di espressione psicrofilo contenente il gene che codifica per una proteina fluorescente (GFP) e trasferite in *PhTAC125*. Successivamente, le cellule indotte sono state sottoposte a diversi cicli di selezione mediante FACS. Questa strategia ha consentito di isolare e caratterizzare l'origine di replicazione di sette cloni con un numero di copie superiore a quella iniziale. Inoltre, uno dei cloni selezionati ha permesso di ottenere un miglioramento di due ordini di grandezza nel numero di copie del

plasmide. Questi risultati hanno consentito di ottimizzare *PhTAC125* come sistema non convenzionale per la produzione di proteine ricombinanti.

Capitolo 3: Studio di separazione di fase liquido-liquido nel batterio antartico *Pseudoalteromonas haloplanktis* TAC125 che produce una proteina umana ricombinante intrinsecamente disordinata, *hCDKL5*

Recentemente, è stato dimostrato che *PhTAC125* è l'unico microorganismo procariotico in grado di produrre la proteina umana, *hCDKL5*.

hCDKL5 è una serina/treonina chinasi che ad eccezione del dominio catalitico all' N-terminale, è caratterizzata dalla presenza di ampie regioni non strutturate. Le mutazioni nella sequenza di questa chinasi determinano lo sviluppo di una grave patologia dello sviluppo neurologico, denominata disturbo da deficienza di CDKL5 (CDD). Lo sviluppo di nuovi plasmidi psicrofili e i miglioramenti apportati al batterio hanno consentito di superare alcuni fattori critici nella produzione ricombinate di *hCDKL5* aprendo la strada ad una potenziale terapia proteica sostitutiva. Tuttavia, la produzione eterologa di *hCDKL5* in *PhTAC125* evidenzia ancora alcuni limiti causati dall' insolubilità del prodotto proteico. Una possibile ipotesi è che l'apparente insolubilità di *hCDKL5* potrebbe essere dovuta alla formazione di condensati. Le proprietà di condensazione di questa proteina sono state studiate *in vivo* osservando la produzione ricombinante in *PhTAC125* di *hCDKL5* fusa all'N-terminale con una proteina fluorescente. Inoltre, sono stati costruiti diversi mutanti di *hCDKL5* progettati per determinare il contributo di ciascun dominio alla condensazione della proteina. I risultati hanno dimostrato che *hCDKL5* forma condensati intracellulari nel batterio psicrofilo e che una regione del dominio catalitico potrebbe essere responsabile della loro formazione. Ulteriori studi saranno necessari per comprendere in dettaglio la correlazione tra la struttura di *hCDKL5* e la formazione di condensati.

Summary

Psychrophilic organisms possess several evolutionary adaptations which allow them to thrive at low temperatures. However, the information on the acclimation properties of these microorganisms remains very limited. In this study, the cold-adaptations of a model microorganism, *PhTAC125*, were assessed by integrating transcriptomic and metabolomic data at 15 °C and 0 °C. These analyses highlighted the metabolic robustness of the *PhTAC125* due to its capability to modulate the expression of numerous metabolic genes, producing similar profiles of metabolites involved in the *PhTAC125* central pathways at two temperatures. Furthermore, the extracellular metabolic analysis allowed us to identify potential nutrient limitations in the growth medium, named GG. Indeed, a deeper extracellular metabolic analysis and the determination of the extracellular ferric ion concentration revealed that the significant changes in the lactate and 2-oxoglutarate profiles could be an adaptive response of the bacterium in dealing with Fe-limitation. In addition, the integration of transcriptomic and metabolic data suggested that the carbon, nitrogen, and iron sources could be unbalanced during the *PhTAC125* growth at 15 °C. The next studies will focus on setting up an optimized growth medium starting from these considerations. *PhTAC125* is also considered one of the most interesting unconventional hosts for difficult-to-express protein production. In the last years, numerous studies have focused on the development of an efficient gene expression technology in *PhTAC125*, however, other bottlenecks are limiting the overall efficiency of the recombinant production system, as an example is the low copy number of the psychrophilic vectors. For this purpose, the replication origin (OriR) was mutated randomly, and the collection of OriR-sequences was cloned into a psychrophilic expression vector containing the gene coding for a fluorescent protein (GFP). The library was transferred into the *PhTAC125* and subjected to several rounds of selection by FACS, allowing the isolation and characterization of seven clones at higher PCN than the wild-type one, among which one of the selected clones allowed to achieve a two-order of magnitude enhancement in plasmid copy number. The improvements developed in the last years allowed *PhTAC125* to be the only prokaryotic host able to produce the full length of Human Cyclin-dependent kinase-like 5 (*hCDKL5*), which is characterized mainly by a long unstructured region, except for its N-terminal catalytic domain. Despite the successful recombinant production of *hCDKL5* in *PhTAC125*, the human protein expression still highlights some limiting factors related to its solubility. Some peculiar features make us think that the apparent insolubility of *hCDKL5* was not associated with the formation of inclusion bodies (never observed in *PhTAC125*) but with the formation of liquid-like condensates. The condensation properties of this protein were investigated *in vivo* during the recombinant expression of a fluorescent protein fused to the N-terminal of *hCDKL5* into *PhTAC125*. Furthermore, several rationally designed *hCDKL5* mutants were constructed to uncover the contribution of each domain to its condensation. The findings demonstrated that the *hCDKL5* forms intracellular clusters in the psychrophilic bacterium, and a region of the catalytic domain drives the protein assembly into condensates.

Introduction

1 Cold adaptation strategies of the Antarctic bacteria to temperature changes

The majority of the Earth's biosphere is permanently exposed to temperatures below 5°C and covered with ice¹. These extreme temperatures and often unfavorable conditions make these habitats less favorable for microbial communities. Indeed, cold temperatures negatively influence the structural and functional properties of cellular components, such as cell integrity, solute diffusion rates, membrane fluidity, enzyme kinetics, protein folding, rates of transcription, translation, and cell division¹⁻³. However, microorganisms known as psychrophiles possess several evolutionary adaptations in order to maintain vital cellular functions in a wide array of temperatures (-20 °C - 25 °C)^{1,4-6}. The generation of novel sequencing technologies enabled the sequencing of several genomes isolated from psychrophilic bacteria, producing a large amount of data that could provide information concerning cold adaptation⁵. Key actors of adaptive mechanisms in cold-adapted bacteria are represented in Figure 1. Furthermore, the genomic analysis of psychrophilic organisms has been combined with other "omic" technologies, such as transcriptomics, proteomics, and metabolomics, allowing differential expression analysis of genes, proteins, and metabolites, respectively, under environmental stress⁷. From the genome analysis of psychrophilic microorganisms isolated from different ecological niches, it was possible to identify similar genomic features for the adaptation to cold stress. In particular, the comparison of *Rhodococcus* and *Polaromonas* psychrophilic genomes indicated that a high number of genes mainly encode for transcriptional and translational proteins (RNA-polymerase, tRNAs, rRNAs, chaperones). It is possible to assume that protein synthesis may be a rate-limiting step for growth and metabolic activity at low temperatures; therefore, a high capacity for translation may result in vital maintenance of the cellular processes⁸. In addition, certain genes essential for the cold-temperature growth were found to be in multiple copies in these genomes (catalase, the iron uptake regulator fur, sigma-70 polymerase), suggesting that genomic redundancy could be a cold-adaptive strategy⁸. In addition, the analysis of other psychrophilic genomes, like *Shewanella spp.*⁹, *Psychrobacter arcticus* 273-4¹⁰, and *Alteromonas sp.* SN2¹¹ showed the presence of a large number of elements that contribute to genome plasticity, such as plasmids and transposons. Recombination and transposition may serve as an adaptive strategy to increase genetic diversity and the potential for survival and growth in cold-temperature environments⁸. The genomic and transcriptomic analyses in some cold-adapted microorganisms (*Shewanella livingstonensis* Ac10¹², *Sphingopyxis alaskensis*¹³, *Psychrobacter arcticus* 273-4^{10,14}) highlighted the up-regulation of RNA helicase enzymes, which can unwind the secondary DNA and RNA structures for efficient translation in the cold¹⁵, and the overproduction of proteins involved in folding assistance, such as molecular chaperones and cold-shock proteins¹⁶. The overexpression of these proteins could be a strategy to overcome challenges related to the reduction of the activity

of transcriptional and translational enzymes, and the increment of DNA/RNA secondary structure stability under low temperatures⁷. Further studies reported that cold temperatures reduce the biochemical reaction rates; therefore, psychrophiles produce cold-adapted enzymes characterized by low thermostability, high structural flexibility, and specific activity in order to maintain adequate catalytic rates². In addition, the proteomic analysis highlighted that the amino acid composition of psychrophilic enzymes is mainly characterized by a high content of asparagine, methionine, and glycine, which increase protein flexibility and lower the formation of hydrogen bonding and salt bridges¹⁷⁻¹⁹. Transcriptomic analyses in *Pseudomonas putida* KT2440²⁰ and *Shewanella oneidensis* MR-1²¹ have shown that exposure to cold temperatures induces a rapid up-regulation of genes involved in membrane biogenesis, such as fatty acid and LPS biosynthesis, peptidoglycan biosynthesis, and outer membrane proteins. Indeed, it is well known that the cell membrane is in a rigid form in freezing temperatures, thus inactivating the function of certain transmembrane proteins²². Psychrophile membrane adaptations include increased unsaturated fatty acids in membrane phospholipids and reduced size and charge of the lipid headgroups^{5,23,24}.

Studies on environmental adaptation and survival strategies of cold-adapted microorganisms have become crucial in the last few decades since the Antarctic environment is especially sensitive to global climate change, as indicated by the retreat of glaciers and ice sheets, ocean warming, acidification, deoxygenation, and changes in nutrient availability^{25,26}. Furthermore, the ocean absorbs around ~90% of the heat generated by greenhouse gas emissions, predicted to increase the average sea surface temperature between 1 and 4 °C over the next decades²⁷. These effects will negatively affect microbial community composition and function, species distributions, and the interactions among organisms.²⁸⁻³¹. In particular, the changes in biological processes will also alter the microbial food web through cellular growth efficiency, carbon cycling, and energy fluxes³². Unfortunately, it is challenging to predict how the functions and composition of microbial ecosystems will be altered in the next decades. Therefore, a profound knowledge of evolutionary processes is necessary to understand the molecular mechanisms of psychrophilic microorganisms under global warming and to incorporate this information into a predictive model on the effects of temperature changes.

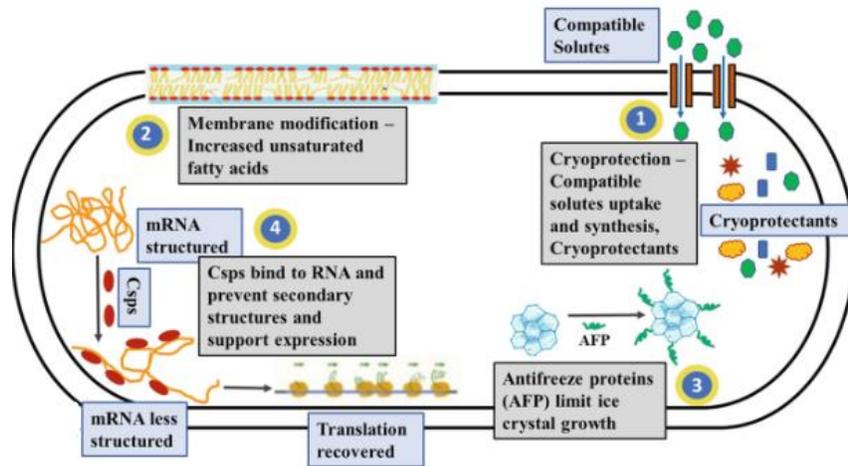


Figure 1 Mechanisms for cold adaptation. Available information regarding cold-adapted bacteria in cold environments shows the development of adaptive mechanisms involved in the modulation of the membrane composition, production of anti-freeze proteins, chaperones, cold-shock proteins, and cold-adapted enzymes. The Figure was adapted by Casanueva, A. *et al.*⁶.

2 *Pseudoalteromonas haloplanktis* TAC125 as a model for adaptation to environmental fluctuations

One of the most studied bacteria adapted to cold environments is *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*), an Antarctic marine γ -proteobacterium isolated from Antarctic coastal seawater³³. *PhTAC125*, recently renamed *Pseudoalteromonas translucida* TAC125, is the first Antarctic bacterium whose genome was fully sequenced and annotated, allowing the identification of two chromosomes and a small cryptic plasmid, pMtBL³³. Following the development of new sequencing technologies, the *PhTAC125* genome was resequenced, revealing the presence of a large plasmid, pMEGA, which was undiscovered until now³⁴. Among the most interesting features of pMEGA is the presence of a putative error-prone polymerase regulated through the SOS response and other genes that might play a crucial role in *PhTAC125* survival in cold-adapted environments and for its adaptation to environmental changes³⁴. *PhTAC125* is able to grow in a wide range of temperatures (from -2.5 °C to 25 °C) thanks to the development of a vast array of favorable physiological features that enables it to survive and adapt to these extreme environments. The accumulation of numerous data for *PhTAC125* in the last few years, including its complete genome sequence^{33,34}, and its proteome³⁵, has led to a better understanding of acclimation mechanisms under extreme temperatures. These analyses performed at 4 °C have demonstrated that protein synthesis and protein folding are the main upregulated functions in *PhTAC125*, suggesting that both cellular processes are limiting factors for bacterial development in cold environments³⁵. In particular, the Antarctic bacterium overproduces chaperone proteins, such as ribosome-bound trigger factors and heat shock proteins³⁵. The cellular function of these chaperones is to assist the folding of proteins and prevent misfolding, interacting with newly polypeptide chains on the ribosome. The efficient translation of *PhTAC125* in the cold is also due to the

overexpression of RNA helicase enzymes, which may help to unwind the mRNA secondary structures, and Peptidyl-prolyl Isomerase enzymes, which accelerates the proline isomerization process and the folding of membrane proteins. Furthermore, the *PhTAC125* genome contains a high number of rRNA clusters (about 9) and tRNA genes (about 106) to allow a correct translation avoiding ribosome stalling processes³³.

2.1 Studies on the acclimation strategies adopted by psychrophilic bacterium

The study on *PhTAC125* metabolism at different temperatures was investigated by the genome-scale metabolic model reconstruction³⁶. In this study, metabolic response caused by changes in gene expression at 4 °C and 18 °C suggests that amino acid degradation and fatty acid metabolism appear to be essential for metabolic adaptation. In particular, fatty acid metabolism modifies the membrane lipid composition in order to maintain membrane fluidity and function of membrane proteins at low temperatures. On the other hand, amino acid degradation could be used as an alternative carbon and energy source during nutrient limitation³⁶. Successively, the regulation of the carbon sources assimilation in *PhTAC125* was investigated using transcriptomics and intracellular ¹H-¹³C NMR³⁷. These analyses showed that *PhTAC125* can use simultaneously two strategies (diauxic and co-utilization) for nutrient uptake when the bacterium was grown in the defined medium embedding 19 different amino acids. Furthermore, it is demonstrated that the bacterium uses different amino acids as a carbon source depending on the phase of cell growth, whereas the diauxic shifts occur when the preferred sources are exhausted. Finally, these findings were embedded in a theoretical model indicating that this metabolic phenotype is compatible with a tight gene regulation that allows the modulation of assimilatory pathways³⁷. This behavior might suggest that *PhTAC125* evolved plastic strategies for carbon source assimilation in response to poor and variable nutrient availability in marine environments. Another essential aspect of the ocean food web is the exchange of dissolved organic matter from phototrophic organisms to bacteria³⁸. Recently, these interactions were studied by setting up a co-culture of *Phaeodactylum tricornutum* diatom and *PhTAC125* bacterium³⁹. Daly and coworkers demonstrated the carbon fixed by phytoplankton can be used as carbon and energy sources for bacterial cells by diatom-dead cells or diatom-released compounds. Overall, these studies helped to elucidate the genome plasticity and metabolic processes contributing to *PhTAC125* cold adaptation. At the same time, further research is necessary to reliably predict how psychrophilic bacteria can respond to global warming, avoiding their extinction. Relatively to this aspect, Toll-Riera and her coworkers have examined the potential evolutionary adaptation of the psychrophilic bacterium, *PhTAC125*⁴⁰. In this study, it was demonstrated that *PhTAC125* can be adapted until 30 °C and not beyond. This adaptation has involved different mutations in the *PhTAC125* genome, whose most prevalent occur in the Lon protease. Another interesting observation was the reduction of the copy number of chromosome 2. These

evolutionary adaptations allow for improvement in the fitness of the bacterium, increasing the degradation of misfolded proteins and reducing the levels of protein expression of chromosome 2⁴⁰.

Although advances in the understanding of the mechanisms underlying cold adaptation have emerged from the analyses of the genome, and proteome of *PhTAC125*, there remains much more to be uncovered. Furthermore, *PhTAC125* constitutes not just a considerable model of study in the ecology field but also holds enormous potential as an unconventional host for recombinant protein production.

3 *Pseudoalteromonas haloplanktis* TAC125 as unconventional host for difficult-to-express protein production

During the past few decades, considerable progress has been made in microbial platform engineering, improving the productivity and yields of recombinant protein. Advances in bioprocessing techniques have provided to achieve higher cell densities through growth media optimization⁴¹, bioreactor design, and development of fermentation processes⁴². Furthermore, the yields of the heterologous proteins have been further enhanced by directly engineering host cells⁴³. The combination of systems biology models⁴⁴, omics datasets, and metabolic engineering^{45,46} have provided a more comprehensive knowledge about the cellular processes influencing protein quantity and quality. In addition, the recent innovations in cell engineering, including the use of RNAi, ribozyme engineering, and gene-editing tools like CRISPR/Cas9, zinc finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs) have enabled to edit individual host cell genes to fine-tune cell physiology and the recombinant protein production⁴⁷. All advancements in this area have increased and facilitated the production of high-level protein production in several organisms, including bacteria, yeast, mammalian cells, yeast, and insect cells⁴⁸⁻⁵⁰. Although the availability of a variety of expression hosts, *E. coli* is the preferred microorganism for protein production at laboratory and industrial scales due to its inexpensive and fast high-density growth, availability of several vectors and host strains, well-known genetics and physiology, and good productivity⁵¹. Despite all these advantages, the production of some recombinant proteins results in heavily proteolyzed, insoluble, and nonfunctional forms⁵². These phenomena can be attributed to the metabolic burden caused by the elevated production of heterologous proteins, lacking proper post-translational modifications, and the generation of inclusion bodies⁵³. Although various strategies of genetic engineering and synthetic biology have been applied to overcome these limitations, it is necessary to consider the physiological limitations of the cell system, and for this reason, a valid alternative could be the use of unconventional hosts with a high capability for protein production. In this regard, *PhTAC125* appears as a promising biological system for the manufacture of high-value proteins, thanks to its physiological features: fast growth in a wide range of temperatures and efficient protein synthesis^{33,35}. Furthermore, the Antarctic

bacterium was the first polar bacterium for which an efficient gene expression technology was developed⁵⁴. In particular, either constitutive or inducible systems^{55–59} were developed exploiting the autonomous replication sequence derived from one of its endogenous plasmids, pMtBL⁵⁵. Moreover, the analysis of the main metabolic pathways and the screening of the growth performances in the presence of different amino acids allowed the formulation of a synthetic medium, GG, guaranteeing the recombinant protein production at a subzero temperature⁵⁸. So, the implementation of psychrophilic systems and the formulation of synthetic media based on bacterial metabolism^{36,58} allowed the production of difficult-to-express proteins in an active and soluble form, whose expression is extremely complex in conventional hosts (*E. coli*, *B. Subtilis*)^{58,60–62}. Furthermore, advances in metabolic engineering also provided a tool for the modulation of critical metabolic pathways in competition with recombinant protein production, aiding the uncoupling of microbial growth from the product formation⁶³. Although these advancements have greatly facilitated recombinant DNA studies in this Antarctic bacterium, significant drawbacks still limit its potential exploitation in this application. Therefore, the focus of future studies will be based on the improvement of the recombinant production yields, the translational efficiency, and the expression plasmid copy number.

References

1. Rodrigues, D. F. & Tiedje, J. M. Coping with our cold planet. *Appl. Environ. Microbiol.* 74, 1677–1686 (2008).
2. D'Amico, S., Collins, T., Marx, J. C., Feller, G. & Gerday, C. Psychrophilic microorganisms: Challenges for life. *EMBO Rep.* 7, 385–389 (2006).
3. Piette, F. *et al.* Life in the cold: A proteomic study of cold-repressed proteins in the antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Appl. Environ. Microbiol.* 77, 3881–3883 (2011).
4. Mikucki, J. A. & Priscu, J. C. Bacterial diversity associated with blood falls, a subglacial outflow from the Taylor Glacier, Antarctica. *Appl. Environ. Microbiol.* 73, 4029–4039 (2007).
5. Casanueva, A., Tuffin, M., Cary, C. & Cowan, D. A. Molecular adaptations to psychrophily: The impact of 'omic' technologies. *Trends Microbiol.* 18, 374–381 (2010).
6. Margesin, R. & Miteva, V. Diversity and ecology of psychrophilic microorganisms. *Res. Microbiol.* 162, 346–361 (2011).
7. De Maayer, P., Anderson, D., Cary, C. & Cowan, D. A. Some like it cold: Understanding the survival strategies of psychrophiles. *EMBO Rep.* 15, 508–517 (2014).
8. Raymond-Bouchard, I., Tremblay, J., Altshuler, I., Greer, C. W. & Whyte, L. G. Comparative Transcriptomics of Cold Growth and Adaptive Features of a Eury- and Steno-Psychrophile. *Front. Microbiol.* 9, 1–14 (2018).

9. Zhao, J. S., Deng, Y., Manno, D. & Hawari, J. *Shewanella* spp. genomic evolution for a cold marine lifestyle and in-situ explosive biodegradation. *PLoS One* 5, (2010).
10. Ayala-Del-Río, H. L. *et al.* The genome sequence of psychrobacter arcticus 273-4, a psychoactive siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. *Appl. Environ. Microbiol.* 76, 2304–2312 (2010).
11. Math, R. K. *et al.* Comparative genomics reveals adaptation by *Alteromonas* sp. SN2 to marine tidal-flat conditions: Cold tolerance and aromatic hydrocarbon metabolism. *PLoS One* 7, (2012).
12. Kawamoto, J. *et al.* Eicosapentaenoic acid plays a beneficial role in membrane organization and cell division of a cold-adapted bacterium, shewanella livingstonensis Ac10. *J. Bacteriol.* 191, 632–640 (2009).
13. Ting, L. *et al.* Cold adaptation in the marine bacterium, *Sphingopyxis alaskensis*, assessed using quantitative proteomics. *Environ. Microbiol.* 12, 2658–2676 (2010).
14. Bergholz, P. W., Bakermans, C. & Tiedje, J. M. *Psychrobacter arcticus* 273-4 Uses Resource Efficiency and Molecular Motion Adaptations for Subzero Temperature Growth. *J. Bacteriol.* 191, 2340–2352 (2009).
15. Cartier, E. A. *et al.* A biochemical and functional protein complex involving dopamine synthesis and transport into synaptic vesicles. *J. Biol. Chem.* 285, 1957–1966 (2010).
16. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* 16, 574–581 (2009).
17. Huston, A. L., Haeggström, J. Z. & Feller, G. Cold adaptation of enzymes: Structural, kinetic and microcalorimetric characterizations of an aminopeptidase from the Arctic psychrophile *Colwellia psychrerythraea* and of human leukotriene A4 hydrolase. *Biochim. Biophys. Acta - Proteins Proteomics* 1784, 1865–1872 (2008).
18. Michaux, C. *et al.* Crystal structure of a cold-adapted class C β -lactamase. *FEBS J.* 275, 1687–1697 (2008).
19. Siddiqui, K. S. & Cavicchioli, R. Improved thermal stability and activity in the cold-adapted lipase B from *Candida antarctica* following chemical modification with oxidized polysaccharides. *Extremophiles* 9, 471–476 (2005).
20. Frank, S. *et al.* Functional genomics of the initial phase of cold adaptation of *Pseudomonas putida* KT2440. *FEMS Microbiol. Lett.* 318, 47–54 (2011).
21. Gao, H., Wang, X., Yang, Z. K., Palzkill, T. & Zhou, J. Probing regulon of ArcA in *Shewanella oneidensis* MR-1 by integrated genomic analyses. *BMC Genomics* 9, (2008).
22. Durack, J., Ross, T. & Bowman, J. P. Characterisation of the Transcriptomes of Genetically Diverse *Listeria monocytogenes* Exposed to Hyperosmotic and Low Temperature Conditions Reveal Global Stress-Adaptation Mechanisms. *PLoS One* 8, 1–15 (2013).
23. Deming, J. W. Psychrophiles and polar regions. *Curr. Opin. Microbiol.* 5, 301–309 (2002).

24. Guan, Z., Tian, B., Perfumo, A. & Goldfine, H. The polar lipids of *Clostridium psychrophilum*, an anaerobic psychrophile. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1831, 1108–1112 (2013).
25. Brierley, A. S. & Kingsford, M. J. Impacts of Climate Change on Marine Organisms and Ecosystems. *Curr. Biol.* 19, R602–R614 (2009).
26. Henson, S. A. *et al.* Rapid emergence of climate change in environmental drivers of marine ecosystems. *Nat. Commun.* 8, 1–9 (2017).
27. Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A. & Trenberth, K. E. Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nat. Clim. Chang.* 1, 360–364 (2011).
28. Philippart, C. J. M. *et al.* Impacts of climate change on European marine ecosystems: Observations, expectations and indicators. *J. Exp. Mar. Bio. Ecol.* 400, 52–69 (2011).
29. Burrows, M. T. *et al.* Ocean community warming responses explained by thermal affinities and temperature gradients. *Nat. Clim. Chang.* 9, 959–963 (2019).
30. Mieszkowska, N., Burrows, M. T., Hawkins, S. J. & Sugden, H. Impacts of Pervasive Climate Change and Extreme Events on Rocky Intertidal Communities: Evidence From Long-Term Data. *Front. Mar. Sci.* 8, (2021).
31. Viitasalo, M. & Bonsdorff, E. Global climate change and the Baltic Sea ecosystem: direct and indirect effects on species, communities and ecosystem functioning. *Earth Syst. Dyn.* 13, 711–747 (2022).
32. Mocali, S. *et al.* Ecology of cold environments: New insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. *Sci. Rep.* 7, 1–13 (2017).
33. Medigue, C. *et al.* Coping with cold: The genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res.* 15, 1325–1335 (2005).
34. Qi, W. *et al.* New insights on *Pseudoalteromonas haloplanktis* TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies. *Sci. Rep.* 9, 1–14 (2019).
35. Piette, F. *et al.* Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Mol. Microbiol.* 76, 120–32 (2010).
36. Fondi, M. *et al.* Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis*TAC125. *Environ. Microbiol.* 17, 751–766 (2015).
37. Perrin, E. *et al.* Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nat. Commun.* 11, 1–16 (2020).
38. Lønborg, C., Carreira, C., Jickells, T. & Álvarez-Salgado, X. A. Impacts of Global Change on Ocean Dissolved Organic Carbon (DOC) Cycling. *Front. Mar. Sci.* 7, 1–24 (2020).
39. Daly, G., Perrin, E., Viti, C., Fondi, M. & Adessi, A. Scaling down the microbial loop: data-driven modelling of growth interactions in a diatom–bacterium co-culture. *Environ. Microbiol. Rep.* 13, 945–954 (2021).

40. Toll-Riera, M., Olombrada, M., Castro-Giner, F. & Wagner, A. A limit on the evolutionary rescue of an Antarctic bacterium from rising temperatures. *Sci. Adv.* 8, 1–11 (2022).
41. Wurm, F. M. & Hacker, D. First CHO genome. *Nat. Biotechnol.* 29, 718–720 (2011).
42. Porter, A. J., Racher, A. J., Preziosi, R. & Dickson, A. J. Strategies for Selecting Recombinant CHO Cell Lines for cGMP Manufacturing : Improving the Efficiency of Cell Line Generation. (2010) doi:10.1002/btpr.443.
43. Carinhas, N., Oliveira, R., Alves, P. M., Carrondo, M. J. T. & Teixeira, A. P. Systems biotechnology of animal cells: the road to prediction. *Trends Biotechnol.* 30, 377–385 (2012).
44. Galleguillos, S. N. *et al.* What can mathematical modelling say about CHO metabolism and protein glycosylation ? *Comput. Struct. Biotechnol. J.* 15, 212–221 (2017).
45. Lewis, A. M., Abu-absi, N. R., Borys, M. C. & Li, Z. J. The Use of ‘ Omics Technology to Rationally Improve Industrial Mammalian Cell Line Performance. doi:10.1002/bit.25673.
46. Kildegaard, H. F., Baycin-hizal, D., Lewis, N. E. & Betenbaugh, M. J. The emerging CHO systems biology era : harnessing the ‘ omics revolution for biotechnology. *Curr. Opin. Biotechnol.* 24, 1102–1107 (2013).
47. Lalonde, M. & Durocher, Y. Therapeutic glycoprotein production in mammalian cells. *J. Biotechnol.* 251, 128–140 (2017).
48. Mckenzie, E. A. & Abbott, W. M. Expression of recombinant proteins in insect and mammalian cells. *Methods* 147, 40–49 (2023).
49. Owczarek, B. & Gerszberg, A. A Brief Reminder of Systems of Production and Chromatography-Based Recovery of Recombinant Protein Biopharmaceuticals. 2019, (2019).
50. Puetz, J. & Wurm, F. M. Recombinant Proteins for Industrial versus Pharmaceutical Purposes : A Review of Process and Pricing. (2019).
51. Ramadan, H. A. I., Saini, K. S., Baeshen, N. A. & Redwan, E. M. Production of Biopharmaceuticals in *E. coli* : Current Scenario and Future Perspectives. 25, 953–962 (2015).
52. Hossain, Z. & Afroz, H. Prospects and applications of nanobiotechnology : a medical perspective. 1–8 (2012).
53. Tripathi, N. K. & Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front. Bioeng. Biotechnol.* 7, (2019).
54. Parrilli, E., De Vizio, D., Cirulli, C. & Tutino, M. L. Development of an improved *Pseudoalteromonas haloplanktis* TAC125 strain for recombinant protein secretion at low temperature. *Microb. Cell Fact.* 7, 1–10 (2008).
55. Tutino, M. L. *et al.* A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles* 5, 257–264 (2001).
56. Duilio, A. *et al.* Promoters from a cold-adapted bacterium: definition of a consensus motif and molecular characterization of UP regulative elements. *Extremophiles* 8, 125–32 (2004).

57. Papa, R., Rippa, V., Sannia, G., Marino, G. & Duilio, A. An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125. *J. Biotechnol.* 127, 199–210 (2007).
58. Sannino, F. *et al.* A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl. Microbiol. Biotechnol.* 101, 725–734 (2017).
59. Colarusso, A., Lauro, C., Calvanese, M., Parrilli, E. & Tutino, M. L. Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a cell factory: Iptg-inducible plasmid construction and strain engineering. *Microorganisms* 8, 1–24 (2020).
60. Vigentini, I., Merico, A., Tutino, M. L., Compagno, C. & Marino, G. Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J. Biotechnol.* 127, 141–50 (2006).
61. Dragosits, M. *et al.* Influence of growth temperature on the production of antibody Fab fragments in different microbes: a host comparative analysis. *Biotechnol. Prog.* 27, 38–46 (2011).
62. Unzueta, U. *et al.* Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A. *Appl. Microbiol. Biotechnol.* 99, 5863–5874 (2015).
63. Fondi, M. *et al.* Modelling hCDKL5 heterologous expression in bacteria. *Metabolites* 11, (2021).

Chapter 1: Combined analysis of the metabolome and transcriptome to explore adaptive mechanisms in *PhTAC125* in response to temperature changes

Psychrophilic organisms possess several evolutionary adaptations which allow them to thrive at low temperatures. However, the information on the regulatory networks and metabolic pathways of these microorganisms remains very limited. In this chapter, the cold-adaptive strategies in *PhTAC125* at 15 °C and 0 °C were assessed by integrating transcriptomic and metabolomic data. The transcriptomic analysis showed that the bacterium increased the abundance of transcripts involved in energy metabolism, fatty acid synthesis, and amino acid metabolism at 0 °C compared to 15 °C. These mechanisms appear to be an important adaptive strategy for *PhTAC125* survival at cold temperatures. On the other hand, the metabolic analyses highlighted no significant changes in metabolite profiles at the two temperatures. This metabolic robustness is due to the capability of the *PhTAC125* to modulate the expression of numerous metabolic genes, producing similar profiles of metabolites involved in the *PhTAC125* central pathways.

Another main aim of this study was the optimization of the GG medium identifying potential limitations to *PhTAC125* growth; for this reason, the extracellular metabolic analysis was repeated, analyzing a major number of samples during different phases of the bacterial growth at 15 °C and 0 °C. In this analysis, new metabolites were revealed, and just the lactate and 2-oxoglutarate profiles showed significant changes at two temperatures. In addition, the determination of the extracellular ferric ion concentration would indicate an iron-restricted condition during the exponential growth phase of the Antarctic bacterium at 15 °C. Successively, *PhTAC125* was grown with higher iron sulfate concentrations at 15 °C, revealing a reduction of the lactate and 2-oxoglutarate extracellular levels; however, the kinetics parameters didn't improve. Understanding of metabolomic achievements was enhanced by the integration of transcriptomic data, revealing the up-regulation of genes involved in the assimilation of nitrogen sources, the lactate catabolism, and the reduction of the tricarboxylic acid cycle enzymes at 15 °C. Overall, these findings suggest that the carbon, nitrogen, and iron sources could be unbalanced during the *PhTAC125* growth at 15 °C. The next studies will focus on setting up an optimized growth medium starting from these considerations.



Metabolic Robustness to Growth Temperature of a Cold-Adapted Marine Bacterium

 Christopher Riccardi,^a  Marzia Calvanese,^b  Veronica Ghini,^{c,d}  Tania Alonso-Vásquez,^a  Elena Perrin,^a  Paola Turano,^c
 Giorgio Giurato,^{e,f}  Alessandro Weisz,^{e,f}  Ermenegilda Parrilli,^b  Maria Luisa Tutino,^b  Marco Fondi^{a,g}

^aDepartment of Biology, University of Florence, Florence, Italy

^bDipartimento di Scienze Chimiche, University of Naples, Napoli, Italy

^cCenter of Magnetic Resonance (CERM), University of Florence, Florence, Italy

^dConsorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Florence, Italy

^eDepartment of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Baronissi, Italy

^fGenome Research Center for Health-GRGS, Baronissi, Italy

^gCentro interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC), University of Florence, Italy

Christopher Riccardi, Marzia Calvanese, and Veronica Ghini contributed equally to this study. Author order was determined alphabetically based on their first name.

ABSTRACT Microbial communities experience continuous environmental changes, with temperature fluctuations being the most impacting. This is particularly important considering the ongoing global warming but also in the “simpler” context of seasonal variability of sea-surface temperature. Understanding how microorganisms react at the cellular level can improve our understanding of their possible adaptations to a changing environment. In this work, we investigated the mechanisms through which metabolic homeostasis is maintained in a cold-adapted marine bacterium during growth at temperatures that differ widely (15 and 0°C). We have quantified its intracellular and extracellular central metabolomes together with changes occurring at the transcriptomic level in the same growth conditions. This information was then used to contextualize a genome-scale metabolic reconstruction, and to provide a systemic understanding of cellular adaptation to growth at 2 different temperatures. Our findings indicate a strong metabolic robustness at the level of the main central metabolites, counteracted by a relatively deep transcriptomic reprogramming that includes changes in gene expression of hundreds of

Editor Rachel Mackelprang, California State University, Northridge

Copyright © 2023 Riccardi et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Marco Fondi, marco.fondi@unifi.it.

The authors declare no conflict of interest. Received 12 November 2022
Accepted 26 January 2023

metabolic genes. We interpret this as a transcriptomic buffering of cellular metabolism, able to produce overlapping metabolic phenotypes, despite the wide temperature gap. Moreover, we show that metabolic adaptation seems to be mostly played at the level of few key intermediates (e.g., phosphoenolpyruvate) and in the cross talk between the main central metabolic pathways. Overall, our findings reveal a complex interplay at gene expression level that contributes to the robustness/resilience of core metabolism, also promoting the leveraging of state-of-the-art multi-disciplinary approaches to fully comprehend molecular adaptations to environmental fluctuations.

IMPORTANCE This manuscript addresses a central and broad interest topic in environmental microbiology, i.e., the effect of growth temperature on microbial cell physiology. We investigated if and how metabolic homeostasis is maintained in a cold-adapted bacterium during growth at temperatures that differ widely and that match measured changes on the field. Our integrative approach revealed an extraordinary robustness of the central metabolome to growth temperature. However, this was counteracted by deep changes at the transcriptional level, and especially in the metabolic part of the transcriptome. This conflictual scenario was interpreted as a transcriptomic buffering of cellular metabolism, and was investigated using genome-scale metabolic modeling. Overall, our findings reveal a complex interplay at gene expression level that contributes to the robustness/resilience of core metabolism, also promoting the use of state-of-the-art multi-disciplinary approaches to fully comprehend molecular adaptations to environmental fluctuations.

KEYWORDS cold-adaptation, genome-scale modeling, metabolomics, transcriptomics

Microorganisms are able to colonize virtually every environmental niche on Earth (1). They have adapted for millions of years prospering under conditions such extreme as water boiling or freezing points, high radiation, acidic or alkaline pH values, heavy metal pollution, and high salinity (2). Growth temperature, in particular, is one of the environmental parameters that mostly impact the physiology of microorganisms and that is thought to have played a key role in their adaptation, selection, and diversification (3). Given the geological history of our planet, it is

reasonable to think that adaptation to changing temperatures has independently occurred many times in evolution; consequently, there exists a vast array of molecular strategies for this purpose, disseminated in the microbial kingdom (4–6). Their characterization is key in this phase of Earth's life, as global change is imposing rapid/dramatic modifications in basic environmental parameters (including temperature) that, in turn, will solicitate the activation of such temperature-adaptation related pathways in microbial communities. Understanding which genes get activated or which compounds get secreted in the environment following an increase of water temperature will help us model and predict the scenarios of microbial communities in a changing environment. Without necessarily invoking global change, the need to rewire cellular networks in response to temperature shifts is likely a common feature in natural microbial communities. The temperature in the Southern Ocean, for example, is anywhere from -2 to 10°C, since Antarctic water temperature fluctuation responds to the seasonal advance and retreat of sea ice (7). Therefore, marine microorganisms are exposed to seasonal oscillations in temperature. Upper-ocean microbes can experience a higher variability in sea surface temperature and their working temperatures exceed the *in situ* Eulerian temperature range by up to 10°C (8). Recent findings demonstrate how even upper-ocean microbes experience along-trajectory temperature variability up to 10°C greater than seasonal fluctuations as a result of large-scale climate variability, indicating a remarkable thermal tolerance by the drifting microbial populations in fluctuating marine environments (8, 9). From a cellular viewpoint, previous studies of cold- and heat-adapted microbes have revealed a variety of molecular adaptations that allow their activity and survival under extreme conditions. These initially involve the change in the expression of specific gene sets but, ultimately, the regulatory changes imposed by temperature increase/decrease are mostly implemented at the metabolic level, since this represents, in the words of Prof. Oliver Fiehn, “the ultimate response of biological systems to genetic or environmental changes” (10). These variegated adaptation strategies observed in temperature-stressed microorganisms indicate that: (i) each microorganism may follow a peculiar route to maintain cellular homeostasis when facing temperature fluctuations (11–13), and (ii) there exists intense cross talk between regulatory and metabolic

networks in the response thereof (14). Here, we specifically investigate the molecular mechanisms through which metabolic homeostasis is maintained in a marine bacterium that (in its natural settings) is known to experience such broad seasonal temperature fluctuations. Indeed, rather than studying the consequences of a cold shock event (a circumstance that rarely occurs in nature), we here focus on the comparison of the main cellular networks in cells growing at 2 different temperatures. This was performed through the integration of metabolomic and transcriptomic data unitedly with genome-scale metabolic modeling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*). This bacterium (15) has been isolated from an Antarctic coastal seawater sample collected in the vicinity of the French Antarctic station Dumont d'Urville, Terre Adélie (66° 40' S; 140° 01' E) and has received much attention in the last decade due to the interest in characterizing its cold-adaptation and nutritional adaptation strategies, as well as its biotechnological potential (16–19). We show that different growth temperatures induce broad transcriptional changes that involve genes of many key metabolic pathways. This transcriptional rewiring, however, is scarcely reflected at the level of the core metabolism, as most key central metabolites show overlapping trends at the 2 tested temperatures. The obtained -omics data were used to compute the flux distributions sustaining growth at low and high temperature, and this provided a mechanistic understanding of the possible adaptation strategies to temperature fluctuations.

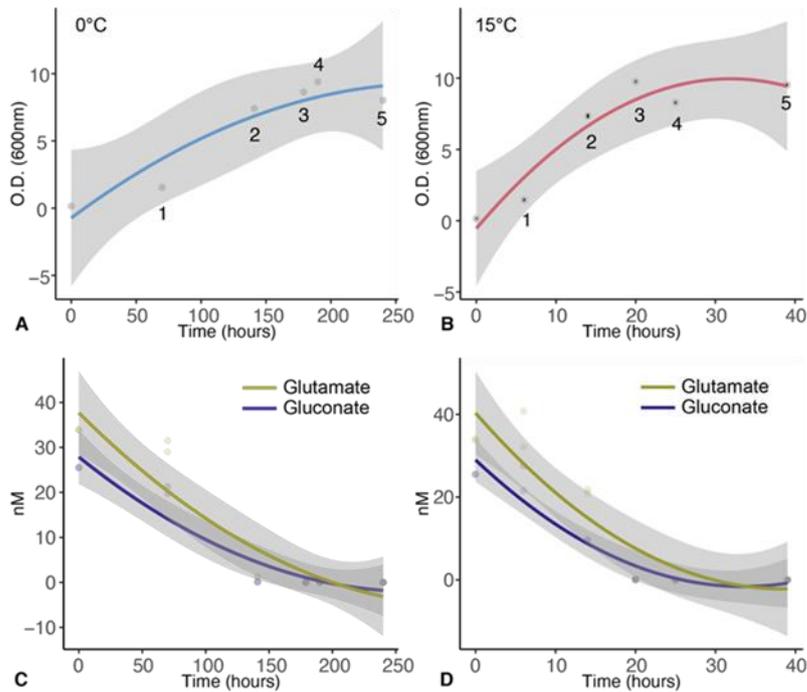


FIG 1 (A) Growth curve of *PhTAC125* at 0°C. Numbers indicate the sampling points for metabolomic and transcriptomic experiments. (B) Growth curve of *PhTAC125* at 15°C. Numbers indicate the sampling points for the metabolomic experiments. The transcriptome of *PhTAC125* growing cells was sampled at time point 1. (C) Glutamate and gluconate uptake at 0°C. (D) Glutamate and gluconate uptake at 15°C.

RESULTS

We cultivated *PhTAC125* cells in a bioreactor and sampled the 0 and 15°C growth curves at 5 different time points (Fig. 1A and B) that overall resembled the same physiological conditions for the 2 experiments. For the 0°C growth, we sampled the following time points: 70, 141, 179, 190, and 240 h. The 15°C curve was sampled at 6, 14, 20, 25, and 39 h. The average growth rate of the 2 cultures varied, being 0.11 (standard deviation (s.d.) 0.011) h⁻¹ and 0.016 (s.d. 6.4e⁻⁰⁵) h⁻¹ at 15° and 0°C, respectively. The same calculation was repeated and limited to the exponential phases of the two cultures, yielding 0.27 (s.d. 0.006) h⁻¹ and 0.027 (s.d. 0.0009) h⁻¹ at 15° and 0°C, respectively. The samples obtained were used for intracellular and extracellular metabolites quantification through nuclear magnetic resonance (NMR) (see Materials and Methods). For 2 of these time points (Fig. 1A and B, labeled as “1”), we also performed quantified gene expression levels using RNA-Seq. *PhTAC125* metabolome is qualitatively and quantitatively robust to temperature shift. Overall, we assigned and analyzed the concentration of 34 intracellular metabolites in the 2 growth curves. These metabolites

represented key intermediates of central metabolic pathways, such as the TCA cycle, amino acids biosynthesis, glycolysis, Pentose Phosphate Pathway (PPP), and nucleic acids biosynthesis. Our NMR metabolomic approach was not able to discriminate between the oxidized and reduced forms of NAD and NADP, so we generally refer to NADX and NADPX in the rest of the manuscript. First, we monitored the overall trends of their intracellular concentrations with respect to the beginning of the growth experiment (Fig. 2A) by computing the Pearson product moment for each metabolite across the five time points and the relative statistical support (Spearman correlation, P value, 0.05 (Fig. 2B)).

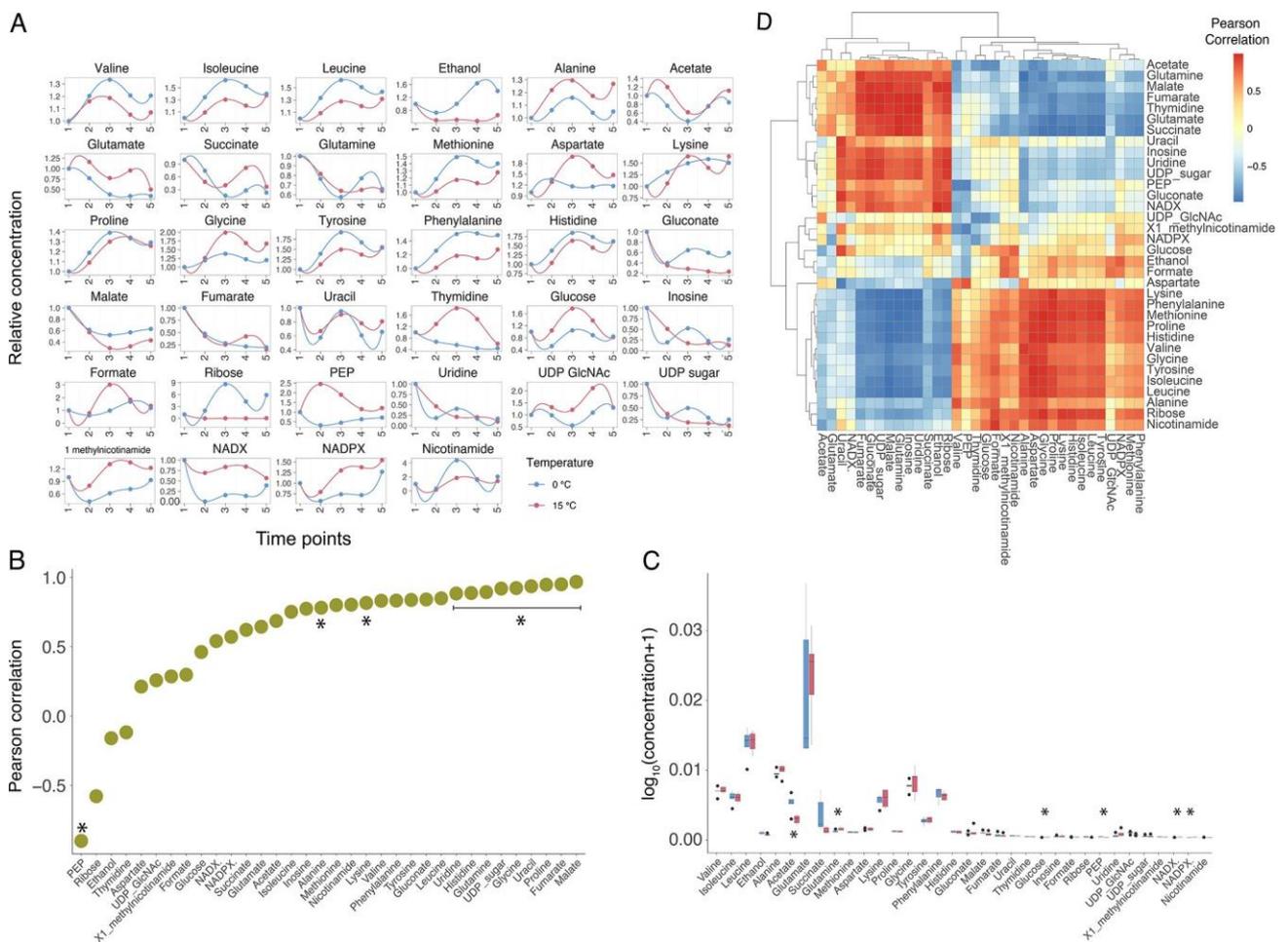


FIG 2 (A) Normalized intracellular metabolites concentration across the 5 time points. (B) Correlation of each intracellular metabolite at 0 and 15°C (asterisks indicate Spearman correlation P value, 0.05). (C) Comparison of the concentration of each intracellular metabolite at 0 and 15°C (asterisks indicate statistically significant correlations, i.e., P value, 0.05). (D) All-against-all correlations between intracellular metabolites at 0 and 15°C

Strikingly, for about 90% of the analyzed metabolites, we found a

positive correlation between the variation of their concentrations during the 2 separate growth experiments (at 0 and 15°C). These positive values ranged from 0.21 in the case of aspartate to 0.96 in the case of malate. Ethanol and thymidine displayed a Pearson product moment correlation (PPM) quite close to zero (-0.16 and -0.1, respectively), thus showing no correlation in the two experiments. Finally, the concentrations of PEP and ribose were negatively correlated in the 2 growth experiments, with PPM of -0.9 and -0.57, respectively. The negative correlation of PEP was supported statistically (Spearman correlation, P value = 0.03692). To qualitatively assess the change in concentration of each metabolite against each other, we computed the all-against-all correlation among the metabolites identified in our study. The results of this analysis are reported in Fig. 2D. We observed 2 main blocks of metabolites: 1 embedding amino acids (with the exception of glutamate and glutamine), and 1 mostly including intermediates of the main central metabolic pathways (TCA, PPP, nucleic acids biosynthesis). Each of these clusters was characterized by a strong (and statistically supported) correlation among the representatives of the same cluster, and an equally relevant anti-correlation with the members of the other cluster. Overall, the intracellular concentration of amino acids was shown to increase over time, whereas the level of, for example, TCA and PPP intermediates was shown to decrease throughout the growth experiment.

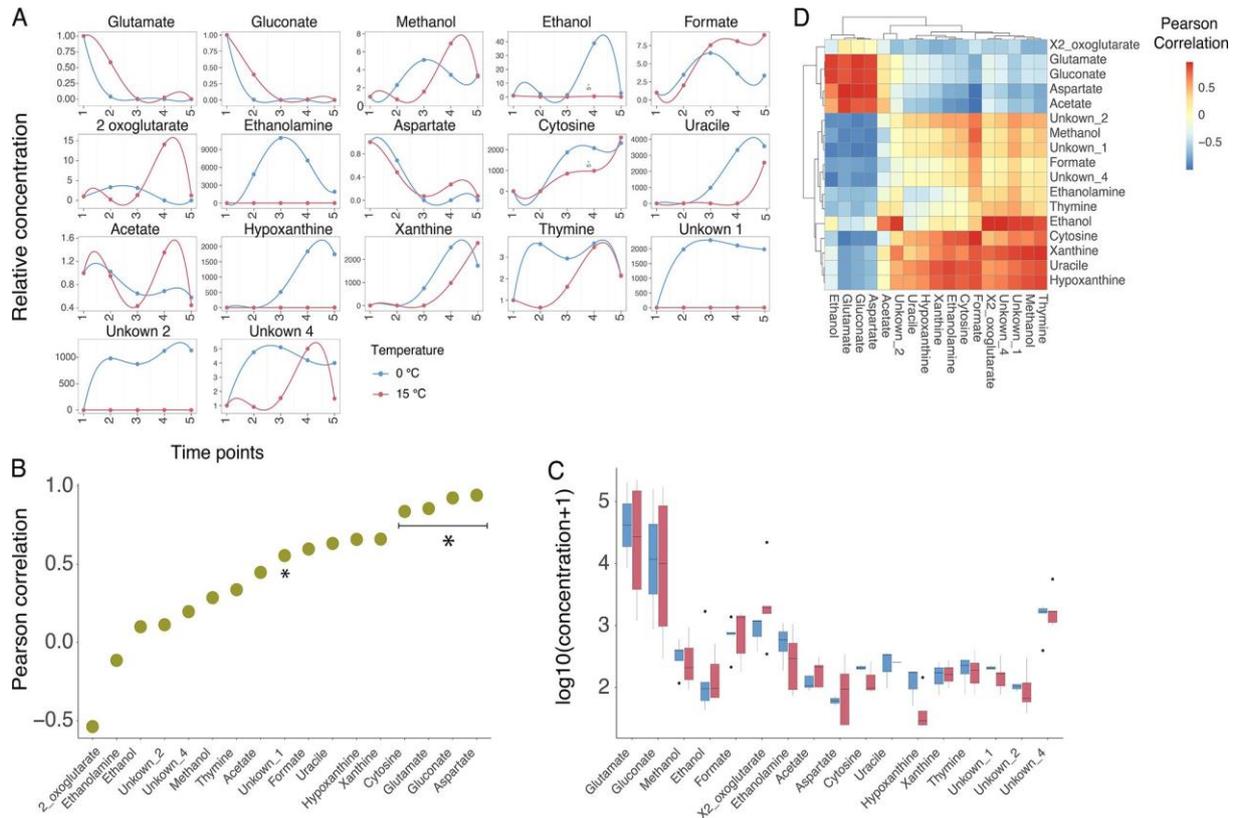


FIG 3 (A) Normalized extracellular metabolites concentration across the 5 time points. (B) Correlation of each intracellular metabolite at 0 and 15°C (asterisks indicate Spearman correlation P value, 0.05) (C) Comparison of the concentration (in a.u.) of each extracellular metabolite at 0 and 15°C (asterisks indicate statistically significant correlations, i.e., P value, 0.05). (D) All-against-all correlations between extracellular metabolites at 0 and 15°C.

To assess whether the growth temperatures had a role in determining the observed trends in metabolite concentrations, we computed an all-against-all correlation for each metabolite at 0° and 15°C separately, and then evaluated whether these 2 correlation matrices differed significantly. The t -test performed supported no significant differences between the means of these 2 comparisons (P values = 0.74 and 0.89, respectively) suggesting that, despite the large difference in growth temperatures, the overall dynamics of intracellular core metabolites were maintained, pointing to an apparent structural metabolic robustness to growth temperature. This robustness was also conserved quantitatively, as the average concentration of each metabolite was maintained similar across the 2 different growth experiments (Fig. 2C). Indeed, except for 6 metabolites out of 34 (acetate, glutamine, glucose, PEP, NAD⁺, and NADP⁺, t test, P value, 0.05, but P value. 0.05 after correction for multiple testing with the Bonferroni method), no other average metabolite concentration showed a significant

difference at 0 and 15°C (Fig. 2C and Fig. S1). A similar scenario was observed for the pool of (17) quantified extracellular metabolites (Fig. 3A). Consistently with growth conditions, glutamate and gluconate concentrations decreased over time until their exhaustion from the growth medium (Fig. 3A), a trend that (i) indicated the almost simultaneous consumption of these 2 nutrients and (ii) was observed at both 0 and 15°C. Out of the 17 extracellular metabolites that were quantified in our experiments, 16 showed a positive correlation between their trends at 0 and 15°C (Fig. 3B). In particular, 10 of them displayed a PPM above 0.5. The only metabolite whose trend differed in the 2 experiments was 2-oxoglutarate (PPM = 20.4). This situation was mirrored at the quantitative level (Fig. 3C). Indeed, none of the extracellular metabolites showed a statistically significant difference between 0° and 15°C growth experiments.

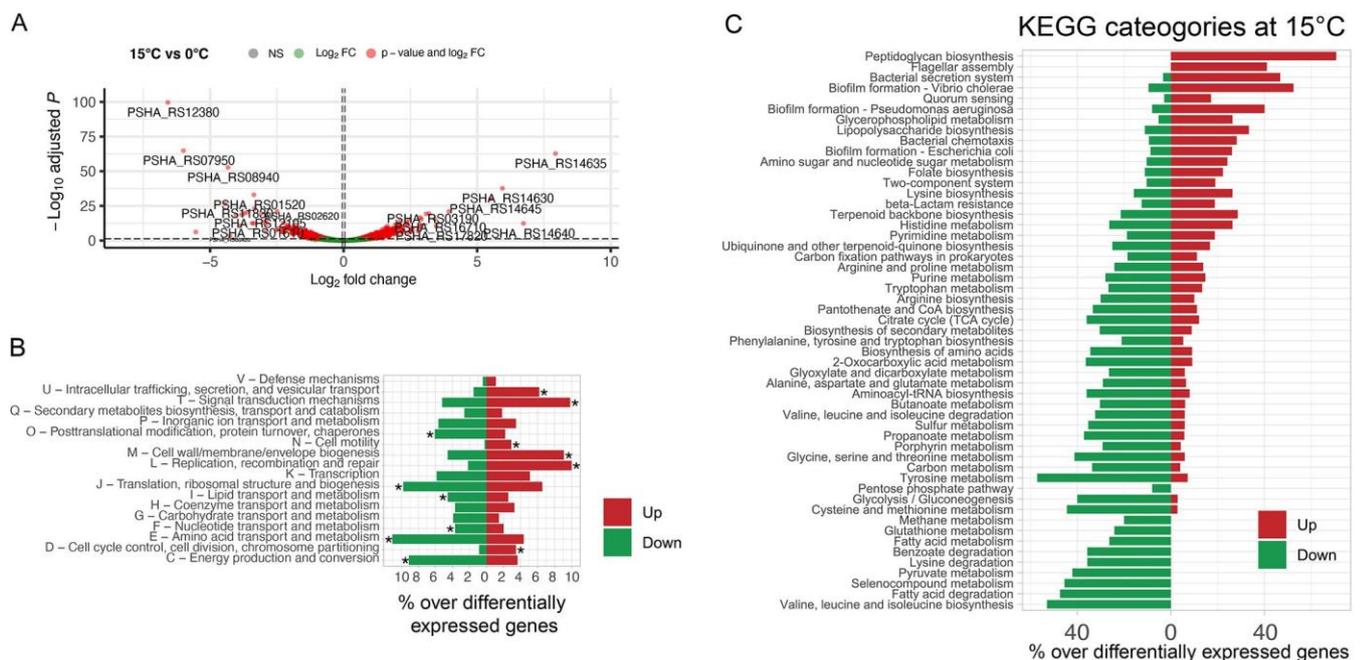


FIG 4 (A) Volcano plot of up- and downregulated genes. (B) COG categories of up- and downregulated genes (asterisks indicate significantly enriched functional categories). (C) Percentage of differentially expressed genes over the total gene of a subset of *PhTAC125* central metabolic pathways. Differentially expressed genes are represented in green (downregulated) or red (upregulated).

Finally, metabolomic data allowed us to evaluate the uptake kinetics of glutamate and gluconate in the tested conditions (Fig. 1C and D), revealing that the 2 carbon sources were taken up almost simultaneously in the 2 experiments, and were depleted/exhausted around T4. The 0 and 15°C transcriptomes of

PhTAC125. We next characterized the transcriptomes of *PhTAC125* at 0 and 15°C. We sampled both curves during exponential growth (Fig. 1A and B, sample points “1”), and sequenced the transcriptome using RNA-Seq technology (see Materials and Methods section). The main features of transcriptomic data are shown in Table S1 and Fig. S2 and S3. Differentially expressed genes (DEGs) were identified using a \log_2 fold change of 1 (or -1) and an adjusted *P* value of 0.05 as thresholds. Overall, we identified 607 differentially expressed genes in the comparison between growth at 15° and 0° C, with an almost identical amount of up- and downregulated genes at 15°C (304 and 303 genes, respectively). The outliers of up- and downregulated genes are shown in Fig. 4A and described in Table 1. Remarkably, out of 607 DEGs, 359 were metabolic protein-coding genes (roughly 59%). Here we first describe specific features of outlier DEGs, and then focus our attention on broader functional categories that embedded the highest fraction of differentially expressed genes. Among the top 5 downregulated genes, 2 lacked a clear functional annotation (PSHA_RS12380 and PSHA_RS08940), and further experiments are required to understand their role in the adaptation to growth at warmer temperatures. Two of them (PSHA_RS07950 and PSHA_RS06965) were annotated as TonB-dependent receptors, suggesting their involvement in the uptake and transport of large substrates, possibly siderophores complexes and/or vitamins. The remaining gene (PSHA_RS11830) is annotated as NAD-dependent succinate-semialdehyde dehydrogenase, responsible for the conversion of succinate-semialdehyde to succinate. The downregulation of this gene is in line with the lower intracellular concentration of succinate at 15°C, as discussed later in the text. Strikingly, among the over-expressed genes during growth at 15°C, we found a prevalence of cold shock related proteins (4 out of 5) and a RNase R encoding gene (PSHA_RS14645). Although no functional studies have been done on cold shock responses and cold shock proteins (CSPs) from psychrophilic bacteria, similarities with the CSPs that are produced in mesophiles have been observed. In particular, the over-expressed CSPs in *PhTAC125* share an amino acid sequence identity ranging between 63.8% and 65.6% with the mesophilic *cspE* of *E. coli*. Furthermore, these proteins contain highly conserved RNA-binding motifs, RNP1 (K-G-F-G-F-I) and RNP2 (V-F-V-H-F) (20, 21), indicated by a black box in Fig. S4.

TABLE 1 The list of the top-five down- and upregulated genes and their functions

Locus tag	log ₂ FC	P _{adj} value	Annotation
PSHA_RS12380	26.58	2.43e ⁻¹⁰⁰	Hypothetical protein
PSHA_RS07950	25.99	1.52e ⁻⁶⁵	TonB-dependent receptor
PSHA_RS06965	25.53	6.72e ⁻⁰⁷	TonB-dependent receptor
PSHA_RS11830	24.44	2.64e ⁻²⁷	NAD-dependent succinate-semialdehyde dehydrogenase
PSHA_RS08940	-4.33	1.85e ⁻⁵³	PA2169 family four-helix-bundle
PSHA_RS14635	7.93	1.99e ⁻⁶³	Cold-shock protein
PSHA_RS14640	6.73	4.02e ⁻¹³	Cold-shock protein
PSHA_RS14630	5.94	2.03e ⁻³⁸	Cold-shock protein
PSHA_RS14645	5.48	9.73e ⁻³¹	Ribonuclease R
PSHA_RS16715	3.94	1.25e ⁻²¹	Cold-shock protein

PSHA_RS14630, PSHA_RS14635, PSHA_RS14640 and PSHA_RS16715 contain 1 highly conserved nucleic acid-binding domain, called cold shock domain (CSD, <http://pfam.xfam.org/family/PF00313>), which is annotated as RNA chaperone/anti-terminator. Moreover, the NCBI identifies both genes PSHA_RS14635 and PSHA_RS14640 as encoding for the same protein (WP_011329604.1), while the others encode for proteins with a different identifier but virtually same sequence. Cold shock proteins would normally counteract the deleterious effects of temperature drop, enabling the cells to grow at low temperatures (22). Jiang et al. proposed the role of *cspA* as an RNA chaperone capable of melting the RNA secondary structure (23), thereby enhancing translation of mRNAs at low temperatures. However, counterintuitively with their given name, not all members of the CSP family are cold-inducible, and their expression is activated upon different stresses (24). For this reason, CSPs might be required for bacterial adaptation to environmental changes. The fifth upregulated protein, RNase R, belongs to the RNR family. In *Escherichia coli*, the RNase R consists of a central nuclease domain, 2 cold shock (CSD) domains near the N-terminal region of the protein, an S1 domain and a highly basic C-terminal region (25). Cairrão et al. showed that the *mnr* gene is co-transcribed with flanking genes as an operon induced under cold shock in *E. coli* (locus tag b4179) (26). There is an important analogy here since, although *PhTAC125*'s operon map isn't available, its top 4-upregulated genes are mapped consecutively on the genome, with coding sense on the same strand (1), spaced by an average distance of 216 nucleotides. The homology to a variety of domains involved in stress response, together with our observation of a strong upregulation during growth at 15°C, may

indicate that *PhTAC125* strives to mediate the elimination of detrimental secondary structures and a temperature rise promotes the expression of enzymes that are required for the correct processing of rRNA precursors. Extending the analysis to a broader functional level, we identified significant differences in 12 COG categories. Specifically, intracellular trafficking, secretion and vesicular transport, signal transduction mechanisms, cell motility, cell wall/membrane/envelope biogenesis, replication, recombination and repair and cell cycle control, cell division, and chromosome partitioning functional categories were found to be over-represented among upregulated genes during growth at 15°C. This matches the results from recent works on the characterization of the evolutionary pathways responsible for thermal adaptation, and the overall notion that one of the main effects of temperature increases on cell physiology consists in the disruption of membrane integrity caused by increased fluidity (27, 28). Overexpression of cell wall and membrane biogenesis-related genes might help overcome this feature associated with growth at higher temperatures. Conversely, genes involved in post-translational modification, protein turnover, chaperones, translation, ribosomal structure and biogenesis, lipids, nucleic acids and amino acids transport and metabolism, as well as energy production functional categories, were over-represented among downregulated genes during growth at 15°C (Fig. 4B). We believe that this analysis depicts the system-level adaptation of bacterial life to diverse growth temperatures, and is in line both with the physiological features observed in this work and with previously obtained data on cold/warm adaptation. Indeed, the significant over-expression of genes broadly related to cell replication (COG categories L, M, and D) is in line with the higher growth rate observed at 15°C in respect to 0°C, over the entire growth period (Fig. 1 and B), and in the specific time interval where RNA was sampled. Further, the over-expression of cell motility at warmer temperatures is in perfect agreement with previous assays on *PhTAC125* motility that had shown a reduced swimming capability of this strain at 0°C (18). Surprisingly, the upregulation of cell growth related processes is counteracted by a general downregulation of cellular metabolism. Key processes of *PhTAC125* core metabolism such as amino acid, lipid, and nucleotide metabolism together with energy production/conversion, were found to be downregulated (Fig. 4C). As mentioned previously, among 607 DEGs, 359 (59.14%) were metabolic genes. Zooming in at the level of the single pathways (Fig. 4C) revealed that the

downregulation affected most of the key central pathways of *PhTAC125* metabolic network (including, for example, glycolysis, TCA cycle, PPP, and amino acids biosynthesis genes). In these pathways, the number of downregulated genes strongly outpaced that of upregulated genes (Fig. 4C). Thus, despite consistent intracellular and extracellular metabolic profiles between the 2 tested growth temperatures, the underlying expression of metabolic genes showed an opposite trend, with a remarkable number of DEGs in the 2 growth conditions. For example, the intracellular pools of key TCA intermediates (fumarate and malate) (Fig. 2C) were shown to be similar between 0 and 15°C, despite more than 50% of the TCA genes being differentially expressed in the 2 conditions (Fig. 4C). Similarly, the intracellular concentration of amino acids was shown to be comparable between the 2 conditions (Fig. 2C), whereas the level of expression of their corresponding biosynthetic genes was shown to be significantly different (Fig. 4C). Accordingly, we hypothesize that the transcriptional network of *PhTAC125* provides a buffering mechanism through which metabolic homeostasis is maintained, at least at the level of the central metabolism. In the next section we will combine metabolomic and transcriptomic data with a genome-scale metabolic reconstruction of *PhTAC125* to unravel the intimate mechanisms through which this balance is achieved. Genome-scale modeling of growth temperature adaptation. A genome-scale metabolic reconstruction exists for the strain *PhTAC125* (16). We, thus, exploited this resource to get a mechanistic interpretation of metabolic homeostasis. First, we checked whether the model was able to represent the experimentally determined phenotypes. We thus constrained the *PhTAC125* genome-scale reconstruction with uptake rates of glutamate and gluconate measured at 0 and 15° degrees, and ran an FBA simulation to predict the cellular growth rates. Average uptake rates of glutamate and gluconate were respectively computed between T1 and T3 for both growth curves (that is after 14 and 141 h for 15 and 0°C, see Materials and Methods), and resulted in 0.10 and 0.08 mmol/gCDW*h⁻¹ at 0°C and 0.47 and 0.62 mmol/gCDW*h⁻¹ at 15°C. As shown in Fig. 5A, at 0°C the results of this simulation (0.021 h⁻¹) were in line with the measured growth rates (0.027 h⁻¹). Conversely, at 15°C a discrepancy was observed between the experimental growth rate and the one predicted *in silico*, 0.27 versus 0.13 h⁻¹, respectively. Although there could be many possible explanations

for this, we hypothesize that constraining the model only based on the uptake rates of the provided nutrients may not be enough to correctly represent the metabolic phenotype of the cells exposed to different temperatures. For this reason, and to provide a more realistic picture of how metabolic homeostasis is maintained at the level of the central metabolism, we combined metabolomic and transcriptomic data with the genome-scale modeling of *PhTAC125* metabolism. We decided to focus on T1 (beginning of the exponential growth phase) as this time point should better resemble the cellular physiological state in which FBA assumptions hold the most (i.e., metabolic steady state).

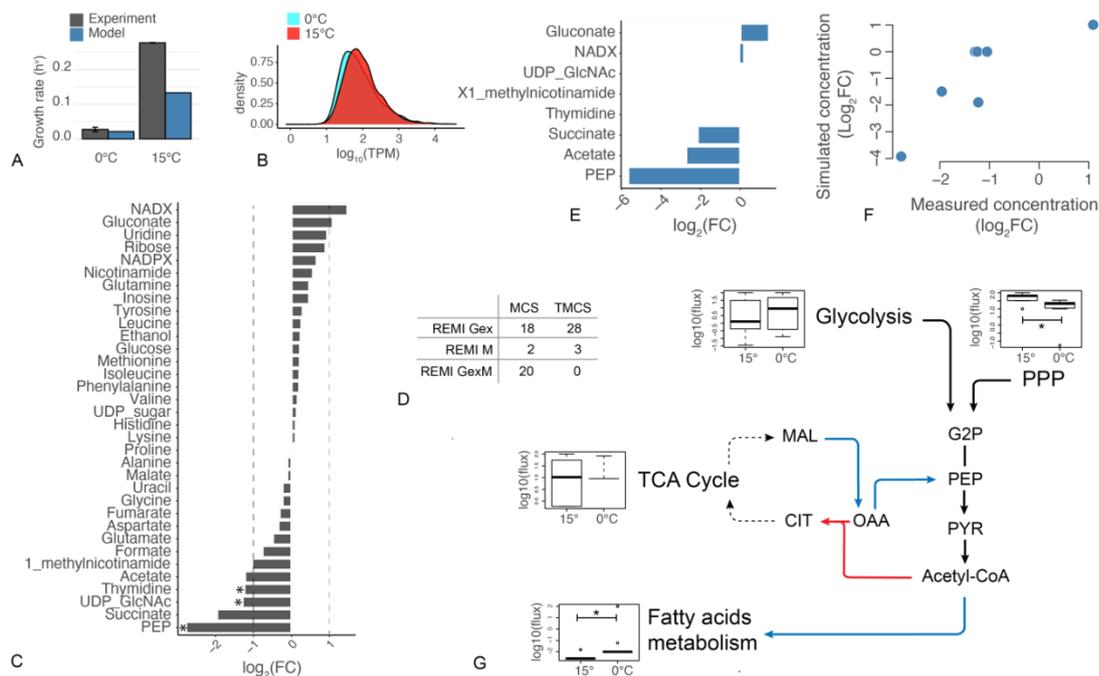


FIG 5 (A) Comparison between experimental and simulated growth rates. (B) Distribution of TPM values of gene expression at T1. (C) log₂FC of metabolites intracellular concentration at T1 during growth at 15°C versus growth at 0°C. (D) REMI output in terms of Maximal and Theoretical Consistency Scores following transcriptomic and metabolomic data integration. (E) Genome-scale modeling prediction of metabolites log₂FC concentration showing a log₂FC > 1 during the metabolomic experiment. (F) Correlation between log₂FC of simulated (y axis) versus measured (x axis) internal metabolites concentration. (G) Working model of key metabolic adjustments at the 2 different temperatures. Blue and red arrows indicate fluxes predicted to increase during growth at 0° and 15°C, respectively. Boxplots represent the average flux values for each pathway. Asterisks indicate a significant (i.e., *P* value, 0.05) Kolmogorov-Smirnov statistical test.

Then, we computed the log₂ fold change (log₂FC) of central intracellular metabolites in the 2 conditions (Fig. 5C). Most (76%) of the metabolites showed a | log₂FC | lower than 1, confirming an overall robustness of the central metabolism to growth

temperature. Only 8 metabolites displayed a $|\log_2FC|$ lower/greater than 1 in the contrast between 15° and 0° metabolomic data, and for only 3 of them we obtained a statistical support (P value, 0.05) (Fig. 5C). More specifically, NADX and gluconate were the 2 metabolites showing a $\log_2FC > 1$ (thus being more abundant at 15°) whereas PEP, succinate, UDP-GlcNAc, thymidine, acetate and X1- methyl-nicotinamide displayed a $\log_2FC, -1$ (thus being more abundant at 0°). We then constrained the model using all available experimental data obtained in this work using REMI (29) (see Materials and Methods). Briefly, this approach allows translating gene expression and metabolite abundance data resulting from a “perturbation” experiment into differential flux distributions among the 2 resulting conditions. In our case, we used data (metabolomes and transcriptomes sampled at the same time point) from 15° and 0° growths to analyze the systemic metabolic alteration(s) occurring in this pair of conditions. FC and TPM values (Fig. 5B and C) were used to constrain the model with metabolic and expression data, respectively. Further, the model was also constrained by setting the boundaries of uptake reactions to represent the actual medium used for the experiment (Schatz medium with glutamate and gluconate as the sole carbon sources, see Materials and Methods). The outcome of data integration into the metabolic reconstruction is summarized by REMI through the computation of the theoretical maximum consistency score (TMCS) and the maximum consistency score (MCS). The first indicates the number of available omics data (for metabolites and reactions), whereas the latter represents the number of those constraints that are consistent with fluxes, and could be integrated into REMI models. As a result, the MCS is always equal to or smaller than the TMCS. In other words, MCS is the largest fraction of available data (metabolomics and transcriptomics) that could be incorporated into an FBA model from a given set of constraints (the abundance of metabolites and transcripts), while ensuring that the model still achieves the required metabolic functionalities and remains feasible. TMCS indicates the number of genes and metabolites with available relative abundance values that can potentially (either because above the specified threshold or their actual inclusion in the metabolic reconstruction) be integrated into the model. The proportion between TMCS and MCS obtained in this work is comparable to

that from other studies where REMI was used for the same purposes (29) (Fig. 5D). Out of data integration and FBA simulations, we obtained 2 distinct flux distributions, i.e., the flux distribution resembling growth at 0°C and the 1 theoretically accounting for the growth at 15°C. The differences between the 2 will represent the most likely metabolic alterations in response to growth temperature and, consequently, will highlight those pathways/reactions that contribute the most to maintaining the observed metabolic homeostasis. First, we checked whether our simulations were accounting for the actual differences between 0° and 15°C intracellular metabolomes. We, thus, computed a matching coefficient (Simple Matching Coefficient (SMC)) between the predicted and measured \log_2FC of each of the 34 internal metabolites, accounting for how many times the model correctly predicted the increase (or decrease) of its internal concentration. Overall, we found an SMC of 71% between simulated and measured metabolic data, revealing that the model is capable of accounting for most of the central metabolome rewiring in respect to growth temperature. Then, we focused on those metabolites that showed a marked change between growth at 15 and 0°C (i.e., for which $|\log_2FC| > 1$) (Fig. 5C). Figure 5F shows that, except for 3 metabolites (UDP-GlcNAc, Thymidine, and 1-methylnicotinamide) for which the model does not predict any difference between the 2 growth conditions, for the remaining 5 metabolites the model correctly predicts a higher (NAD⁺ and gluconate) or lower (acetate, PEP and succinate) internal production at 15°. Overall, we found a significant, positive correlation of 0.72 between the measured and the predicted FC of internal concentrations of the 8 metabolites for which $|\log_2FC| > 1$ (Fig. 5F) (Spearman correlation, P value = 0.04). Next, we examined to what extent the central, interconnected pathways showed significantly altered flux distributions. While no clear signal could be identified for glycolysis and TCA cycle (mirroring what we observed with DEGs analysis), both PPP and fatty acids metabolism showed significant differences in their fluxes between 0° and 15° growth simulations. As for fatty acids metabolism, fluxes representative of growth at 0°C were significantly higher (Kolmogorov-Smirnov test, P value = 8.96×10^{-10}) than those resembling metabolism at 15°C (Fig. 5F). The

opposite was observed for PPP simulated fluxes which displayed, on average, significantly lower values at 0° vs 15°C (Kolmogorov-Smirnov test, P value = 0.0030). A sustained activity of PPP at 15° is in line with the increased internal concentration of gluconate and ribose (2 key PPP intermediates) at this temperature (Fig. 5C). Similarly, an increase in fatty acids biosynthesis (and fatty acids metabolism in general) is in line with previous simulations and experiments concerning the involvement of this process in the adaptation to growth at (relatively) low temperatures (16). We then focused on the analysis of simulated metabolic fluxes around the metabolite that showed the highest degree of variation, PEP. We asked which metabolic rewiring could lead to an increased production of this metabolite at 0°C. The analysis of fluxes revealed an increased activity at 0°C with respect to 15°C of both the reaction converting L-malate to 2-oxaloacetate ([S]-malate:NAD1 oxidoreductase) and the reaction converting 2-oxaloacetate to PEP (phosphate:oxaloacetate carboxy-lyase). Also, ATP:pyruvate, H₂O phosphotransferase reaction displayed an increase in flux at 0°C. Overall, this would allow the partial redirection of key TCA cycle intermediates to the production of PEP. At the same time, we also recorded an increased flux at 0°C in the reaction redirecting acetyl-CoA to fatty acids biosynthesis. Conversely, reactions leading to the production of citrate from OAA and Acetyl-CoA resulted to be less active at 0°C. Overall, our experimentally constrained simulations seem to suggest a working model for the metabolic adaptation to growth at different temperatures. Considering growth at 0°C, an increased level of intracellular PEP observed experimentally might be functional to its conversion to acetyl-CoA and its consequent tunneling into fatty acids metabolism (Fig. 5G), and might be the outcome of fluxes redirection from the TCA cycle to PEP production, rather than from glycolysis or PPP. Consistent with this idea is the overall increase of key TCA intermediates (Fig. 5A) measured during growth at 0°C with respect to growth at 15°C that suggests an overall increased activity of this central pathway in the 0°C growth condition.

DISCUSSION

In this work, we have studied how a cold-adapted bacterium rewires

its central metabolism when growing at 2 distinct temperatures that overall resemble a seasonal shift. This was done by characterizing a pool of 34 intracellular and 17 extracellular central metabolites during 5 different time points of its growth curves (i.e., at 0°C and 15°C) and by evaluating gene expression during the initial stages of its exponential phases. To our surprise, the top 5 upregulated genes at 15°C contain a cold shock domain, which is usually recruited to counteract the deleterious effects of temperature drop. A possible explanation for the apparent paradox where CspA is activated, not only following cold stress, but also under non-stress and other stress conditions which entail a downregulation of bulk gene expression and protein synthesis is presented in (30). CSPs have been found in almost all types of bacteria and are mainly induced after a rapid temperature downshift to regulate the adaptation to cold stress but are also present under normal conditions to regulate other biological functions (31). For example, in *E. coli*, only 4 (*cspA*, *cspB*, *cspG*, and *cspI*) of the 9 CSP genes are cold-induced (32). Two of them, *cspE* and *cspC*, are constitutively expressed at physiological temperatures, and act as ‘housekeeping RNA chaperones’ to modulate the global gene expression (33, 34). Furthermore, these CSPs are also involved in the transcription anti-termination mechanism, which is based upon preventing the formation of secondary structures on the nascent mRNA (35–37). In *PhTAC125*, four cold shock-like proteins *cspE* are upregulated during growth at 15°C. This indicates that these CSP proteins may not play a role in cold-adaptation in *PhTAC125* but, analogously to *E. coli*, they may act as chaperones by destabilizing secondary structures in target RNA at high temperature so that the single-stranded state of target RNA is maintained. This may then enable efficient transcription and translation. From a broader perspective, the comparison between the 0° and 15°C transcriptomes produced more than 600 differentially expressed genes, assigned to a dozen of different functional categories. On the other hand, the intracellular concentration of nearly 90% of the analyzed metabolites correlates positively in the 2 conditions. We interpret this as the capability of *PhTAC125* regulatory network to buffer the temperature shift to produce strikingly similar metabolic phenotypes, despite the temperature gap. This is also reflected in the conservation of the overall metabolic network structure (i.e., the presence of correlated clusters of metabolites) (Fig. 2C and Fig.

3C). Remarkably, this conservation involved metabolites from many different pathways, such as amino acids metabolism, TCA cycle, PPP, and glycolysis, suggesting that such robustness is propagated at the level of the entire metabolism and is not restricted to specific pathways. Studies on the metabolic response to growth at different temperatures in other microorganisms have generally shown larger variability among intracellular pools of metabolites (38), and/or the tendency to activate genetic mechanisms that shut down metabolism under longer-term high temperature stress (39). This is not the case for *PhTAC125* that, through a global transcriptional buffering, maintains very consistent trends of its intracellular and extracellular metabolomes. This mirrors, for example, the response of *E. coli* to genetic and environmental perturbations and its capability to maintain metabolite levels stable, reflecting the rerouting of fluxes in the metabolic network (40). A few metabolites stood out in this conservation of metabolite pools, including PEP and ribose (showing opposite trends at 0 and 15°C), and ethanol and thymidine (showing no correlation between the 2 experiments). Also, NAD⁺ and NADP⁺ levels differed in the 2 conditions, as shown in Fig. 2A. To unravel this complex interplay and produce a mechanistic interpretation of this adaptation, -omics data (together with data on carbon sources uptake rates) were used to constrain a genome-scale metabolic reconstruction and derive the most-likely metabolic phenotypes at these 2 temperatures. This approach provided evidence that most of the metabolic adaptation is probably played at the level of the phosphoenolpyruvate (PEP)–pyruvate-oxaloacetate node and specifically involves the increase of TCA fluxes and their redirection to PEP production. Consistent with this idea, is the overall reduction of key TCA intermediates at 15°C (Fig. 5A) and the observation that, upon growth on non-glycolytic compounds (e.g., glutamate and gluconate) the cycle intermediates malate or oxaloacetate must be converted to pyruvate and PEP for the synthesis of glycolytic intermediates. In our case, however, the availability of an increased pool of PEP would rather allow its conversion to acetyl-CoA, and its consequent tunneling into fatty acids metabolism (Fig. 5G). Also the overexpression of genes involved in valine, leucine, and isoleucine degradation (Fig. 4C) at 0°C can be associated with the higher activity of fatty acids metabolism, as the degradation of these 2 amino acids leads to the production of acetyl-CoA (41). Consistently with this observation, our model

predicts an increased activity of fatty acids metabolism at 0° (Fig. 5G). The importance of fatty acids metabolism in growth at low temperatures is largely known. For example, the fluidity of the rigidified membrane (42) imposed by low temperatures can be restored through the modulation of the (i) saturated and unsaturated fatty acids, (ii) fatty acid chain length, and (iii) the proportion of *cis* to *trans* fatty acids thus, ultimately, through the modulation of fatty acids metabolism (as observed in this work). This also validates and extends previous findings on general cold-adaptation strategies (16, 43, 44), pinpointing the role of specific metabolic reactions in this process. We also found an increased PPP activity at 15° with respect to 0°C (Fig. 5G). Increased PPP activity signifies increased NADPH production, and is linked with many biosynthetic processes. Interestingly, high activity of the PPP pathway has been described as an adaptive mechanism (in soil bacteria) to temperature stress (45), in particular by linking the requirement of sugar units for biofilm formation to the overall activity in this metabolic pathway. We know for a fact from the work of (18) that *PhTAC125* displays strong propensity to form more biofilm at 15°C versus 0°C when grown in the same minimal medium used in this work. Also, we found a remarkable fraction of overexpressed biofilm-related genes (Fig. 4C), reinforcing the idea of a temperature-mediated switch to this kind of lifestyle. Thus, we infer that a similar mechanism is at play in this marine bacterium where the redirection of fluxes in PPP at 15°C would be the metabolic basis for the formation of cellular aggregates at higher temperatures. Moreover, the increased PPP activity can respond to other cell requirements linked to the DNA replication recombination and repair, since the ribose produced by PPP is necessary for these cellular activities that result more active at 15°C (Table S2), as suggested by transcriptomic analysis (Fig. 5B). Very recently, Macarena Toll-Riera et al. published (28) a work in which *PhTAC125* was evolved at increasingly higher temperatures to study the evolutionary potential of upper thermal tolerance, and characterize the genomic basis of temperature adaptation. Interestingly, the authors found almost no metabolic genes among those that apparently provided an increased fitness at higher temperatures. Indeed, clones that were selected for their improved growth at higher temperatures (up to 30°C) had mainly

mutations in proteases-encoding genes or in genes involved in chromosome copy number reduction, energy production and conversion and cell wall biosynthesis. Thus, apparently, the metabolism of *PhTAC125* was not the primary target of mutations that led to an increased fitness in the tested conditions. This, in turn, may indicate the presence of an already optimized and plastic metabolism that allows growth in a broad range of temperatures, in line with the metabolic robustness that we have characterized in this work. Conclusions. The aim of our experiments was to evaluate the effect of growth temperatures on cellular homeostasis, particularly at the metabolic level. We have shown that a cold-adapted marine bacterium expresses 2 very similar metabolic phenotypes in response to 2 widely different temperatures, by adjusting the expression of key enzymes and fine-tuning the intracellular concentration of key intermediates. This illuminated on the possible molecular mechanisms that marine microbes may use to adapt to broad (seasonal) temperature changes. In particular, we showed that specific changes at the level of regulatory circuits can buffer such variations, and maintain the underlying metabolic network robust to temperature fluctuations. In the future, it will be interesting to investigate whether similar mechanisms are at play in natural microbial assemblages.

MATERIALS AND METHODS

Strains and growth conditions. *PhTAC125* (46) cells were grown in a 1.5 L GG medium (47) in a Stirred Tank Reactor 3 L fermenter (Applikon) connected to an eZ2 Bio Controller (Applikon) at 2 different temperatures (0°C and 15°C). This medium contains glutamate and gluconate as the only carbon sources. The bioreactor was equipped with the standard pH-, pO₂-, level- and temperature sensors for the bioprocess monitoring. For the growths of the *PhTAC125* bacterium, the pre-culture was centrifuged (6000 × *g*, 20 min, 4°C); the cells were washed twice with fresh medium, and then used to inoculate the bioreactor with a starting OD₆₀₀ of 0.2, in aerobic conditions (50% and 30% dissolved oxygen at 15°C and 0°C, respectively), in stirring (500 rpm at 15°C and 250 rpm at 0°C). The bacterial culture was carried out at 15°C for 40 h or at 0°C for 240 h. Each culture condition was repeated three times. Cell growth was monitored, measuring the OD₆₀₀ about every 2 h

in the experiments at 15°C, and every 8 h at 0°C. Three different measurements were performed at each time point for each biological replicate. For intra- and extracellular metabolites analysis, the samples were taken in triplicate at 5 different time points during the growth in GG medium at 2 different temperatures (70 h, 141 h, 179 h, 190 h, and 240 h at 0°C, and 6.5 h, 14 h, 20 h, 25 h, and 39 h at 15°C). For the analysis of extracellular metabolites, aliquots (1 mL) of cell cultures were harvested during the growth and centrifuged for 15 min at 13,000g at 4°C; the supernatant was recovered, filtered (Filtropur 0.2 mm, SARSTED AG & Co. KG) and stored at 280°C. The analysis of intracellular metabolites was performed on 60 OD₆₀₀ pellets recovered during the growth by centrifuging for 20 min at 6,000 rpm at 4°C.

Transcriptomics. For the RNA-Seq experiment, 1 OD₆₀₀ pellets were recovered during the exponential growth phase (~1/1.5 OD/mL) at 15°C and 0°C by centrifugation (10 min, 13,000g, 4°C). Cell pellets were washed in RNase-free PBS three times and stored at 280°C. Total RNA was isolated from the cells using the Directzol RNA Kit (Zymo Research) following the manufacturer's instructions. Contaminating genomic DNA was then removed through treatment with RNase-free DNase I (Roche).

RNA sequencing. RNA concentration was measured using Quant-IT RNA assay kit-high sensitivity and a Qubit Fluorometer (Life Technologies), and its quality and integrity assessed with the Agilent 4200 TapeStation System (Agilent Technologies). Indexed libraries were prepared starting from 400 ng of total RNA according to Universal Prokaryotic RNA-Seq Library Prep kit (Tecan). Final libraries were sequenced at a concentration of 1.7 pM/lane on the NextSeq 500 platform (Illumina Inc) in paired mode 2 × 75bp. For each experimental condition, 3 biological replicates were prepared.

RNA-Seq data analysis. The 6 samples were assessed for base call quality and adapter content using fastp (48), allowing down to a mean quality threshold of 20 (i.e., probability of incorrect base call of 1 in 100) and minimum read length of 40 nucleotides. A median of 98.9% of the reads passed quality check, indicating that sequencing was carried out pristinely, and our data were biologically reliable. Salmon (49) index was built on *PhTAC125*'s transcriptome free of tRNA and rRNA sequences, providing the entire genomic sequence as background decoy to account for possible underlying DNA contamination. CDS and genomic FASTA

were retrieved on the NCBI, using assembly accession GCF_000026085.1. Moreover, CDS and genomic files were concatenated with data from pMEGA plasmid (NZ_MN400773.1) and pMtBL plasmid (NZ_AJ224742.1). The transcriptome quantification step was performed using the `- validateMappings` flag, to ensure a sensitive selective alignment of the sequencing reads. The resulting mapping rates ranged between 56.8% and 66.21% (average of 62.47%), whereas roughly half of the reads represented ribosomal DNA sequences (discarded for downstream analyses). The integration of transcript-level abundance estimates from Salmon with the data analysis pipeline was performed using R package `tximport` (R Core Team, 2022, <https://www.R-project.org/>, R version 4.0.3). The testing of changes in the overall transcriptional output was handled through the statistical engine in R package `DESeq2`, with an expected proportion of false positives set at 5%. The volcano scatterplot showing statistical significance versus magnitude of change for 607 differentially expressed genes was produced using the library `EnhancedVolcano` from the R Bioconductor 3.14 suite.

Metabolomics. ^1H NMR-based metabolomic analyses were performed on cell lysates and growth media to monitor the intracellular metabolites, and the uptake and release of the extracellular metabolites, respectively, by measuring their concentration levels in samples collected at different time points during the cell growth. Medium samples were prepared in 5.00 mm NMR tubes by mixing 60 mL of a potassium phosphate buffer (1.5 M K_2HPO_4 , 100% (vol/vol) $^2\text{H}_2\text{O}$, 10 mM sodium trimethylsilyl [2,2,3,3- $^2\text{H}_4$]propionate (TMSP), pH 7.4), and 540 mL of each growth medium. Cell lysate samples were prepared in 5.00 mm NMR tubes by mixing 60 mL of $^2\text{H}_2\text{O}$ and 540 mL of samples. All the NMR spectra were recorded using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency, and equipped with a 5 mm PATXI ^1H - ^{13}C - ^{15}N and ^2H -decoupling probe including a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). A BTO 2000 thermocouple served for temperature stabilization at the level of approximately 0.1 K at the sample. Before measurement, samples were kept for 5 min inside the NMR probe head, for temperature equilibration at 300 K. ^1H NMR spectra were acquired with water peak suppression and a

standard NOESY pulse sequence using 128 scans, 65536 data points, a spectral width of 12019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mixing time of 0.1 s. The raw data were multiplied by a 0.3 Hz exponential line broadening before applying Fourier transformation. Transformed spectra were automatically corrected for phase and baseline distortions. All the spectra were then calibrated to the reference signal of TMSP at δ 0.00 ppm using TopSpin 3.5 (Bruker BioSpin srl). The metabolites, whose peaks in the spectra were well resolved, were assigned, and their levels analyzed using a dedicated R script developed in-house. In total, 34 and 17 metabolites were identified and quantified in the cell lysate and in the growth medium spectra, respectively. The assignment was performed using an internal ^1H NMR spectral library of pure organic compounds (BBIORFCODE, Bruker BioSpin), stored reference NMR spectra of metabolites, and spiking experiments. Matching between new NMR data and databases was performed using the Assure NMR software (Bruker BioSpin). The relative concentrations of the various metabolites were calculated by integrating the corresponding signals in defined spectral ranges, using in-house developed R 3.0.2 scripts. Similarly, downstream data analysis was performed using R. Raw data, post-Raw data, and post-processing codes are made available at <https://github.com/combogenomics/MetRob015>. Uptake rates of gluconate and glutamate at the 2 different temperatures were computed as the ratio between the average growth rate between T1 and T3 (μ) and the biomass yield (λ). The former was computed as:

$$\mu = \frac{\log OD_{T3} - \log OD_{T1}}{T3 - T1}$$

The latter was computed according to the following relationship:

$$\lambda = \frac{\text{Biomass } \left(\frac{g}{L}\right)}{\text{Consumed C source (mol)}}$$

Biomass was obtained from OD values as described in (50), using 0.74 as a scaling factor for the growth at 15°C and 0.66 for the growth at 0°C.

Genome-scale metabolic modeling. The genome-scale metabolic

reconstruction used in this work is the one recently used in (51). All the simulations were run in MATLAB 2019a, using the COBRA toolbox (52) version 2.7.4. Metabolomic and transcriptomic data were integrated using REMI method (29), providing TPM gene expression and relative concentration values for metabolites at time point T1. Both for metabolites and gene expression, we selected the top 3% as upregulated and the bottom 3% as downregulated (REMI default is 5%). Consistently with the original REMI publication, we used the 2-fold change as the cutoff threshold to identify the significant gene expression and metabolite changes. All the other parameters were left as default. The predicted internal concentration of each metabolite was computed using the *computeFluxSplits* function implemented in the COBRA toolbox. This function computes the relative contributions of fluxes to the net production and consumption of a specific set of metabolites included in the model. Statistical tests on flux distributions were calculated using R. To estimate the difference in the activity of the main central metabolic pathways at 0° and 15°C, we computed the predicted flux of each reaction in these pathways (TCA cycle, glycolysis, PPP, and fatty acids metabolism) at the 2 different temperatures, and averaged this number of reactions by the total number of reactions included in that pathway. From this latter analysis, we excluded those reactions that (i) had a flux equal to zero (were inactive) in both conditions and (ii) had a different sign (i.e., changed direction) at the 2 temperatures.

Data availability. The authors confirm that the data supporting the findings of this study are available within the article and/or its supplemental materials. Collection of sequence data produced in this study is available under the BioProject PRJNA886636.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.

FIG S2, PDF file, 0.1 MB.

FIG S3, PDF file, 0.04 MB.

FIG S4, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.01 MB.

TABLE S2, PDF file, 0.02 MB.

ACKNOWLEDGMENTS

We thank Dr. Assunta Sellitto for technical assistance for RNA-Seq libraries preparation.

P.T. and V.G. acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and, specifically, the CERM/CIRMMP Italy Centre.

This study was supported by Regione Campania, Progetto GENOMAeSALUTE (POR CAMPANIA FESR 2014/2020, azione 1.5; CUP: B41C17000080007), by PNRA (Programma Nazionale di Ricerche in Antartide) grant PNRA18_00075 and PNRA18_00335, and by a PRIN-MUR (RESEARCH PROJECTS OF RELEVANT NATIONAL INTEREST– 2020 Call) Project: 20208LLXEJ.

REFERENCES

1. Rothschild LJ, Mancinelli RL. 2001. Life in extreme environments. *Nature* 409:1092–1101. <https://doi.org/10.1038/35059215>.
2. Shu W-S, Huang L-N. 2022. Microbial diversity in extreme environments. *Nat Rev Microbiol* 20:219–235. <https://doi.org/10.1038/s41579-021-00648-y>.
3. Hegedus S, Csonka J. 2010. *Astrobiology: physical origin, biological evolution, and spatial distribution*, p 27–57. Nova Science Publishers, New York, NY.
4. Bennett AF, Lenski RE. 1997. Phenotypic and evolutionary adaptation of a model bacterial system to stressful thermal environments, p 135–154. *In* Bijlsma R, Loeschcke V (ed), *Environmental Stress, Adaptation and Evolution*. Birkhäuser Basel, Basel, Switzerland.
5. Saarinen K, Laakso J, Lindström L, Ketola T. 2018. Adaptation to fluctuations in temperature by nine species of bacteria. *Ecol Evol* 8:2901–2910. <https://doi.org/10.1002/ece3.3823>.
6. Walworth NG, Zakem EJ, Dunne JP, Collins S, Levine NM. 2020. Microbial evolutionary strategies in a dynamic ocean. *Proc Natl Acad Sci U S A* 117:5943–5948. <https://doi.org/10.1073/pnas.1919332117>.
7. Auger M, Morrow R, Kestenare E, Sallée J-B, Cowley R. 2021. Southern Ocean *in-situ* temperature trends over 25 years emerge from interannual variability. *Nat Commun* 12:514. <https://doi.org/10.1038/s41467-020-20781-1>.
8. Doblin MA, van Sebille E. 2016. Drift in ocean currents impacts intergenerational microbial exposure to temperature. *Proc Natl Acad Sci U S A* 113:5700–5705. <https://doi.org/10.1073/pnas.1521093113>.
9. Kim H, Ducklow HW. 2016. A decadal (2002–2014) analysis for dynamics of heterotrophic bacteria in an Antarctic coastal ecosystem: variability and physical and biogeochemical forcings. *Front Mar Sci* 3. <https://doi.org/10.3389/fmars.2016.00214>.
10. Fiehn O. 2002. Metabolomics the link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171. <https://doi.org/10.1023/A:1013713905833>.
11. Lu H, Ulanov AV, Nobu M, Liu W-T. 2016. Global metabolomic responses of *Nitrosomonas europaea* 19718 to cold stress and altered ammonia feeding patterns. *Appl Microbiol Biotechnol* 100:1843–1852. <https://doi.org/10.1007/s00253-015-7095-y>.
12. Czajka JJ, Abernathy MH, Benites VT, Baidoo EEK, Deming JW, Tang YJ. 2018. Model metabolic strategy for heterotrophic bacteria in the cold ocean based on *Colwellia psychrerythraea* 34H. *Proc Natl Acad Sci U S A* 115:12507–12512. <https://doi.org/10.1073/pnas.1807804115>.

13. Kurdrid P, Phuengcharoen P, Senachak J, Saree S, Hongsthong A. 2020. Revealing the key point of the temperature stress response of *Arthrospira platensis* C1 at the interconnection of C- and N- metabolism by proteome analyses and PPI networking. *BMC Mol and Cell Biol* 21:43. <https://doi.org/10.1186/s12860-020-00285-y>.
14. Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L. 2010. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol* 6:364. <https://doi.org/10.1038/msb.2010.18>.
15. Birolo L, Tutino ML, Fontanella B, Gerday C, Mainolfi K, Pascarella S, Sannia G, Vinci F, Marino G. 2000. Aspartate aminotransferase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125. *Eur J Biochem* 267:2790–2802 <https://doi.org/10.1046/j.1432-1327.2000.01299.x>.
16. Fondi M, Maida I, Perrin E, Meller A, Mocali S, Parrilli E, Tutino ML, Liò P, Fani R. 2015. Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125: Modelling of *P. haloplanktis* TAC125 metabolism. *Environ Microbiol* 17:751–766. <https://doi.org/10.1111/1462-2920.12513>.
17. Mocali S, Chiellini C, Fabiani A, Decuzzi S, de Pascale D, Parrilli E, Tutino ML, Perrin E, Bosi E, Fondi M, Lo Giudice A, Fani R. 2017. Ecology of cold environments: new insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. *Sci Rep* 7:839. <https://doi.org/10.1038/s41598-017-00876-4>.
18. Ricciardelli A, Casillo A, Vergara A, Balasco N, Corsaro MM, Tutino ML, Parrilli E. 2019. Environmental conditions shape the biofilm of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Microbiol Res* 218: 66–75. <https://doi.org/10.1016/j.micres.2018.09.010>.
19. Parrilli E, Tedesco P, Fondi M, Tutino ML, Lo Giudice A, de Pascale D, Fani R. 2021. The art of adapting to extreme environments: the model system *Pseudoalteromonas*. *Phys Life Rev* 36:137–161. <https://doi.org/10.1016/j.plrev.2019.04.003>.
20. Landsman D. 1992. RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res* 20:2861–2864. <https://doi.org/10.1093/nar/20.11.2861>.
21. Yamanaka K, Fang L, Inouye M. 1998. The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Mol Microbiol* 27:247–255. <https://doi.org/10.1046/j.1365-2958.1998.00683.x>.
22. Keto-Timonen R, Hietala N, Palonen E, Hakakorpi A, Lindström M, Korkeala H. 2016. Cold shock proteins: a minireview with special emphasis on Csp-family of enteropathogenic *Yersinia*. *Front Microbiol* 7:1151. <https://doi.org/10.3389/fmicb.2016.01151>.
23. Jiang W, Hou Y, Inouye M. 1997. CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J Biol Chem* 272:196–202. <https://doi.org/10.1074/jbc.272.1.196>.
24. Caballero CJ, Menendez-Gil P, Catalan-Moreno A, Vergara-Irigaray M, García B, Segura V, Irurzun N, Villanueva M, Ruiz de Los Mozos I, Solano C, Lasa I, Toledo-Arana A. 2018. The regulon of the RNA chaperone CspA and its auto-regulation in *Staphylococcus aureus*. *Nucleic Acids Res* 46: 1345–1361. <https://doi.org/10.1093/nar/gkx1284>.
25. Phadtare S, Severinov K. 2010. RNA remodeling and gene regulation by cold shock proteins. *RNA Biol* 7:788–795.

- <https://doi.org/10.4161/rna.7.6.13482>.
26. Cairrão F, Cruz A, Mori H, Arraiano CM. 2003. Cold shock induction of RNase R and its role in the maturation of the quality control mediator SsrA/tmRNA. *Mol Microbiol* 50:1349–1360. <https://doi.org/10.1046/j.1365-2958.2003.03766.x>.
 27. Angilletta MJ, Jr. 2009. Thermal adaptation. Oxford University Press, England.
 28. Toll-Riera M, Olombrada M, Castro-Giner F, Wagner A. 2022. A limit on the evolutionary rescue of an Antarctic bacterium from rising temperatures. *Sci Adv* 8:eabk3511. <https://doi.org/10.1126/sciadv.abk3511>.
 29. Pandey V, Hadadi N, Hatzimanikatis V. 2019. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Comput Biol* 15:e1007036. <https://doi.org/10.1371/journal.pcbi.1007036>.
 30. Brandi A, Pon CL. 2012. Expression of *Escherichia coli* cspA during early exponential growth at 37°C. *Gene* 492:382–388. <https://doi.org/10.1016/j.gene.2011.10.047>.
 31. Horn G, Hofweber R, Kremer W, Kalbitzer HR. 2007. Structure and function of bacterial cold shock proteins. *Cell Mol Life Sci* 64:1457–1470. <https://doi.org/10.1007/s00018-007-6388-4>.
 32. Xia B, Ke H, Inouye M. 2001. Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in *Escherichia coli*: S1 domain complements cold-shock protein deficiency. *Mol Microbiol* 40:179–188. <https://doi.org/10.1046/j.1365-2958.2001.02372.x>.
 33. Phadtare S, Inouye M. 2001. Role of CspC and CspE in regulation of expression of RpoS and UspA, the stress response proteins in *Escherichia coli*. *J Bacteriol* 183:1205–1214. <https://doi.org/10.1128/JB.183.4.1205-1214.2001>.
 34. Phadtare S, Tadigotla V, Shin W-H, Sengupta A, Severinov K. 2006. Analysis of *Escherichia coli* global gene expression profiles in response to over-expression and deletion of CspC and CspE. *J Bacteriol* 188:2521–2527. <https://doi.org/10.1128/JB.188.7.2521-2527.2006>.
 35. Bae W, Xia B, Inouye M, Severinov K. 2000. *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. *Proc Natl Acad Sci U S A* 97:7784–7789. <https://doi.org/10.1073/pnas.97.14.7784>.
 36. Phadtare S, Inouye M, Severinov K. 2002. The nucleic acid melting activity of *Escherichia coli* CspE is critical for transcription antitermination and cold acclimation of cells. *J Biol Chem* 277:7239–7245. <https://doi.org/10.1074/jbc.M111496200>.
 37. Phadtare S, Severinov K. 2005. Nucleic acid melting by *Escherichia coli* CspE. *Nucleic Acids Res* 33:5583–5590. <https://doi.org/10.1093/nar/gki859>.
 38. Guan Y, Yin D, Du X, Ye X. 2018. Metabolomics approach used for understanding temperature-related pectinase activity in *Bacillus licheniformis* DY2. *FEMS Microbiology Lett* 365. <https://doi.org/10.1093/femsle/fny255>.
 39. Kim S, Kim Y, Suh DH, Lee CH, Yoo SM, Lee SY, Yoon SH. 2020. Heat-responsive and time-resolved transcriptome and metabolome analyses of *Escherichia coli* uncover thermo-tolerant mechanisms. *Sci Rep* 10:17715. <https://doi.org/10.1038/s41598-020-74606-8>.
 40. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa

- T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M. 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316:593–597. <https://doi.org/10.1126/science.1132067>.
41. Massey LK, Sokatch JR, Conrad RS. 1976. Branched-chain amino acid catabolism in bacteria. *Bacteriol Rev* 40:42–54. <https://doi.org/10.1128/br.40.1.42-54.1976>.
 42. Los DA, Murata N. 2004. Membrane fluidity and its roles in the perception of environmental signals. *Biochim Biophys Acta* 1666:142–157. <https://doi.org/10.1016/j.bbamem.2004.08.002>.
 43. Graumann PL, Marahiel MA. 1999. Cold shock response in *Bacillus subtilis*. *J Mol Microbiol Biotechnol* 1:203–209.
 44. Ting L, Williams TJ, Cowley MJ, Lauro FM, Guilhaus M, Raftery MJ, Cavicchioli R. 2010. Cold adaptation in the marine bacterium, *Sphingopyxis alaskensis*, assessed using quantitative proteomics. *Environ Microbiol* 12:2658–2676. <https://doi.org/10.1111/j.1462-2920.2010.02235.x>.
 45. Bore EK, Apostel C, Halicki S, Kuzyakov Y, Dippold MA. 2017. Microbial metabolism in soil at subzero temperatures: adaptation mechanisms revealed by position-specific ¹³C labeling. *Front Microbiol* 8:946. <https://doi.org/10.3389/fmicb.2017.00946>.
 46. Médigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN, Cheung F, Cruveiller S, D'Amico S, Duilio A, Fang G, Feller G, Ho C, Mangenot S, Marino G, Nilsson J, Parrilli E, Rocha EPC, Rouy Z, Sekowska A, Tutino ML, Vallenet D, von Heijne G, Danchin A. 2005. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* 15:1325–1335. <https://doi.org/10.1101/gr.4126905>.
 47. Sannino F, Giuliani M, Salvatore U, Apuzzo GA, de Pascale D, Fani R, Fondi M, Marino G, Tutino ML, Parrilli E. 2017. A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl Microbiol Biotechnol* 101:725–734. <https://doi.org/10.1007/s00253-016-7942-5>.
 48. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
 49. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419. <https://doi.org/10.1038/nmeth.4197>.
 50. Giuliani M, Parrilli E, Ferrer P, Baumann K, Marino G, Tutino ML. 2011. Process optimization for recombinant protein production in the psychrophilic bacterium *Pseudoalteromonas haloplanktis*. *Process Biochemistry* 46:953–959. <https://doi.org/10.1016/j.procbio.2011.01.011>.
 51. Fondi M, Gonzi S, Dziurzynski M, Turano P, Ghini V, Calvanese M, Colarusso A, Lauro C, Parrilli E, Tutino ML. 2021. Modelling hCDKL5 heterologous expression in bacteria. *Metabolites* 11:491. <https://doi.org/10.3390/metabo11080491>.
 52. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnúsdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A,

Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, et al. 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc 14:639–702. <https://doi.org/10.1038/s41596-018-0098-2>.

Supplementary materials

Accession Number	Number of Reads	Temperature
SRX17780709	18,376,681	15°C
SRX17780711	25,089,797	15°C
SRX17780712	31,526,393	15°C
SRX17780713	17,586,523	0°C
SRX17780714	23,037,026	0°C
SRX17780710	29,173,509	0°C

Table S1. Main features of the transcriptomic dataset BioProject PRJNA886636. Reads numbers refer to the raw, unfiltered files.

Locus tag	log ₂ FC	Adj. p-value	Annotation
PSHAa0090	1.1108113	0.00000925	xerC site-specific recombinase
PSHAa1218	1.8251366	1.52E-13	recN DNA repair protein recN
PSHAa2356	1.0226152	0.0000609	topB DNA topoisomerase III
PSHAa2361	1.2276516	0.0000048	DNA topoisomerase III (N terminal part)
PSHAa2873	1.1514558	0.000014	lexA transcriptional repressor for SOS response
PSHA_RS17095	1.382628818	0.00000669	xni putative exonuclease IX

Table S2. Expression data for DNA-replication related-genes

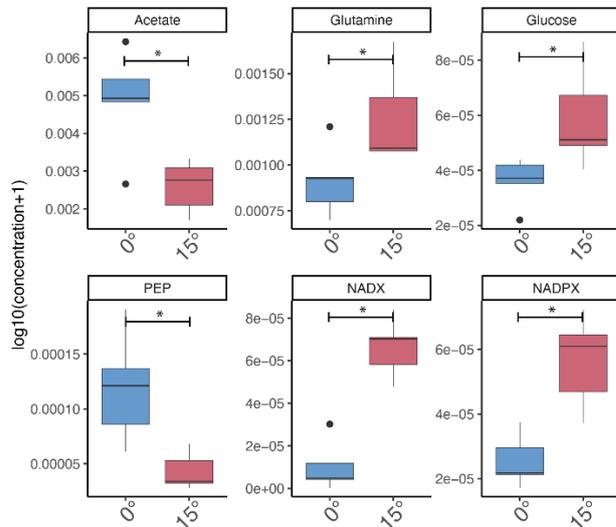


Figure S1. Boxplots of intracellular metabolites that showed a statistically significant difference concentration across all the time points analyzed.

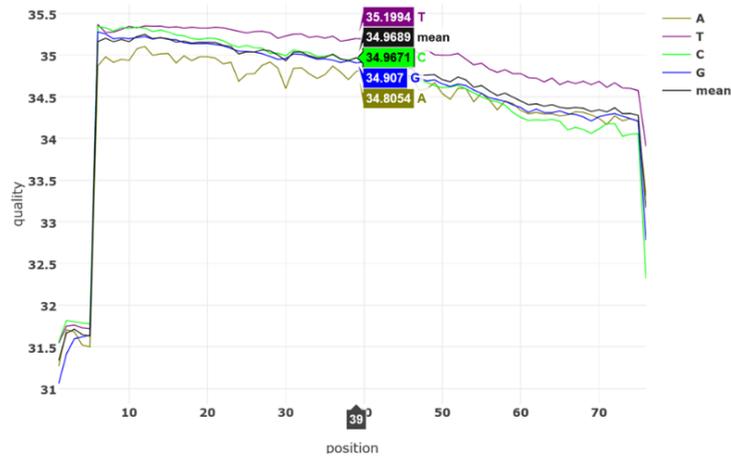


Figure S2. Quality check report performed using fastp (Shifu Chen et al., 2018). Position 39 is being displayed as a sample representative of the quality at read mid-length.

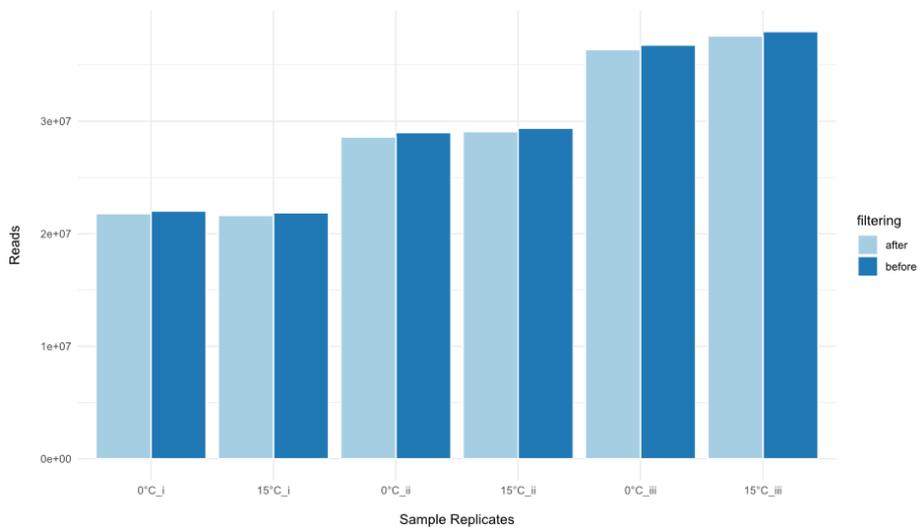


Figure S3. Sample replicates reads count before and after QC with fastp.

		RPN1		RPN2		
Ph RS16715	-MSTTTGSKVWFNEA	KGFGEI	EQESG-PI	VFAHFS	SAITSDGFKTLAEGQRVQFTVTQGQK	58
Ph RS14630	MSNTTTGTVKFFNEA	KGFGEI	EQESG-AD	VFAHFS	SAISGDGFKTLAEGQRVQFTVTQGQK	59
Ph RS14635	MSNTTTGSKVWFNEA	KGFGEI	EQESG-AD	VFAHFS	SAIVSDGFKTLAEGQRVQFTVTQGQK	59
Ph RS14640	MSNTTTGSKVWFNEA	KGFGEI	EQESG-AD	VFAHFS	SAIVSDGFKTLAEGQRVQFTVTQGQK	59
Eco cspE	-MSKIKGNVWFNES	KGFGEI	TPEDGSKD	VFVHFS	SAIQTNDFKTLAEGQRVEFEITNGAK	59
	. . * ** : **	*** **	. *	***_****	:*:* * ** *. * : . * *	
Ph RS16715	GPQAENIVCI		68			
Ph RS14630	GPQAENIVCI		69			
Ph RS14635	GPQAENIVCI		69			
Ph RS14640	GPQAENIVCI		69			
Eco cspE	GPSAANVIAL		69			
	** * * : . :					

Figure S4. Sequence alignment of four psychrophilic CSPs with the mesophilic CspE protein. Gaps indicated by hyphens were introduced to improve alignment, and identical amino acids are indicated with asterisks. The RNA-binding motifs RPN1 and RPN2 are boxed in black.

1.2 In-depth extracellular metabolic analysis in *Pseudoalteromonas haloplanktis* TAC125 towards the medium optimization

1.2.1 Introduction

Microorganisms produce and secrete many primary and secondary metabolites during their growth in response to different environmental signals, providing important information about the changes in microbial metabolism¹. Various environmental factors can affect the uptake and secretion of metabolites, such as temperature, pH, concentration of nutrients, dissolved oxygen, and other parameters². Much progress has been made in the study of temperature effects on microbial metabolism, but many gaps in knowledge remain, especially for the marine microbial communities in cold Antarctic regions. To this end, Riccardi and coworkers investigated by transcriptomic and metabolic analysis the cold-adaptation mechanisms in *PhTAC125* during its growth at two different temperatures (Riccardi *et. al* 2022 paragraph 1.1). The intracellular and extracellular metabolites were analyzed and quantified at 5 different time points during the growth in GG medium at 15 and 0 °C, moreover, in the same growth conditions a gene differential expression analysis, limited to a single point time, at the beginning of the exponential growth phase was performed. The findings reveal a significant transcriptomic perturbation that includes changes in the expression of more than 600 genes, maintaining similar trends of intracellular and extracellular metabolites involved in the *PhTAC125* central pathways at the two different temperatures. This section aims to analyze in deeper the extracellular metabolite profiles at two temperatures taking into account metabolites overlooked during the first metabolic analysis of *PhTAC125*. This could allow the identification of potential limits of GG³ as a medium for *PhTAC125* cell growth helping to set up a new growth medium to improve the growth performance of the bacterium at different temperatures. In this contest not only the impact of the carbon and nitrogen sources were evaluated but also the impact of iron in the *PhTAC125* growth was assessed. Indeed, although carbon and nitrogen sources are principal components of organic material and their availability in the environment is necessary for the growth and survival of microorganisms, iron can enjoy a status of notable importance for several key biological processes, including amino acid synthesis, oxygen transport, respiration, the citric acid cycle, and DNA biosynthesis⁴. However, iron acquisition presents a challenge for the majority of microorganisms due to its extremely low solubility under aerobic conditions limiting availability⁴. Specifically, marine bacteria typically require micromolar levels of total iron for growth, yet the iron concentration in the surface waters of the oceans is only 0.01-2 nM⁵. Microorganisms have, therefore, had to adapt to maintain sufficiently high levels of this micronutrient. The most basic acclimation mechanisms include iron accumulation in ferritin- or bacterioferritin-based iron reserve complexes⁶ and the secretion of high-affinity iron-binding siderophores⁷. When iron is bound to siderophores, membrane transporters transport the iron into the microorganism⁸. Additionally, iron scarcity can result in metabolic shifts toward enzymes that do not require iron; this has occurred in some marine bacteria that have adapted to the ocean's iron scarcity by reducing the expression of many enzymes of the tricarboxylic acid (TCA) cycle⁹, and using flavodoxin-based respiratory-chain enzymes instead of iron-containing

ferredoxin-based enzymes¹⁰. Therefore, iron deprivation results in decreased TCA enzyme activities and subsequent NADH formation and ATP production. To cope with the low efficiency of the electron transport chain under Fe-limitation, a further acclimation strategy used by heterotrophic marine bacteria is the upregulation of glyoxylate shunt⁹. Serving as an alternative pathway within the TCA cycle, the glyoxylate shunt allows the production of oxaloacetate, which can serve as a precursor for gluconeogenesis, and bypasses the NADH-producing steps which are required within the electron transport chain in the production of ATP¹¹.

1.2.2 Materials and Methods

1.2.2.1 Strains and growth conditions

*PhTAC125*¹² cells were grown in a 1.5 L GG medium³ in a Stirred Tank Reactor 3 L fermenter (Applikon) connected to an eZ2 Bio Controller (Applikon) at two different temperatures (0 °C – 15 °C). The bioreactor was equipped with the standard pH-, pO₂-, level- and temperature sensors for the bioprocess monitoring. The original formulation contained 5 mg/L of ferrous sulfate (FeSO₄·7H₂O), while a further variation with 20mg/L of FeSO₄·7H₂O has been tested for the Fe-replete condition. For the growths of the *PhTAC125* bacterium, the pre-culture was centrifuged (6000 rpm, 20 min, 4 °C); the cells were washed twice with fresh medium, and then used to inoculate the bioreactor with a starting OD₆₀₀ of 0.2, in aerobic conditions (50% and 30% dissolved oxygen at 15 °C and 0 °C, respectively), in stirring (500 rpm at 15 °C and 250 rpm at 0 °C). Cell growth was monitored, measuring the OD₆₀₀ about every two hours in the experiments at 15 °C, and every eight hours at 0 °C. The bacterial culture was carried out at 15 °C for 160 h or at 0 °C for 240 h. The cell growth was monitored and recorded at intervals of 1h. Three different measurements were performed at each time point for each biological replicate. For the analysis of extracellular metabolites, aliquots (1 mL) of cell cultures were harvested during the growth and centrifugated for 15 min at, 13,000 rpm, at 4 °C; the supernatant was recovered, filtered (Filtropur 0.2 µm, SARSTED AG & Co. KG) and stored at -80 °C.

1.2.2.2 RNA Sequencing

The experimental details on the RNA-Seq technique are reported in the section *Materials and Methods* of paragraphs 1.1 “Metabolic robustness to growth temperature of a cold-adapted marine bacterium” in chapter 1.

1.2.2.3 Extracellular metabolomics and iron detection

¹H NMR-based metabolomic analysis were performed on cell media to monitor the uptake and release of the metabolites by measuring their concentration levels in samples collected at different time points during the cell growth. Sample preparation, spectral acquisition and processing were performed according to procedures developed at CERM^{15–20}. Samples were prepared in 5.00 mm NMR tubes by mixing 60 µL of a potassium phosphate buffer (1.5 M K₂HPO₄, 100% (v/v)

H₂O₂, 10 mM sodium trimethylsilyl [2,2,3,3-²H₄] propionate (TMSP), pH 7.4) and 530 µL of growing medium. The NMR spectra were recorded using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm PATXI ¹H-¹³C-¹⁵N and ²H-decoupling probe including a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). A BTO 2000 thermocouple served for temperature stabilization at the level of approximately 0.1 K at the sample. Before measurement, samples were kept for 5 minutes inside the NMR probe head, for temperature equilibration at 300 K. ¹H NMR spectra were acquired with water peak suppression and a standard NOESY pulse sequence using 128 scans, 65536 data points, a spectral width of 12019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s and a mixing time of 0.1 s. The raw data were multiplied by a 0.3 Hz exponential line broadening before applying Fourier transformation. Transformed spectra were automatically corrected for phase and baseline distortions. All the spectra were then calibrated to the reference signal of TMSP at δ 0.00 ppm using TopSpin 3.5 (Bruker BioSpin srl). The metabolites, whose peaks in the spectra were well resolved, were assigned and their levels analyzed using a dedicated R script developed in-house. We identified and quantified 18 metabolites in the growing medium spectra. The assignment was performed using an internal ¹H NMR spectral library of pure organic compounds (BBIORFCODE, Bruker BioSpin), stored reference NMR spectra of metabolites and spiking experiments. Matching between new NMR data and databases was performed using the Assure NMR software (Bruker BioSpin). The relative concentrations of the various metabolites were calculated by integrating the corresponding signals in defined spectral range, using in house developed R 3.0.2 scripts. Similarly, downstream data analysis was performed using R. Raw data and post-processing codes are made available at <https://github.com/combogenomics/MetRob015>.

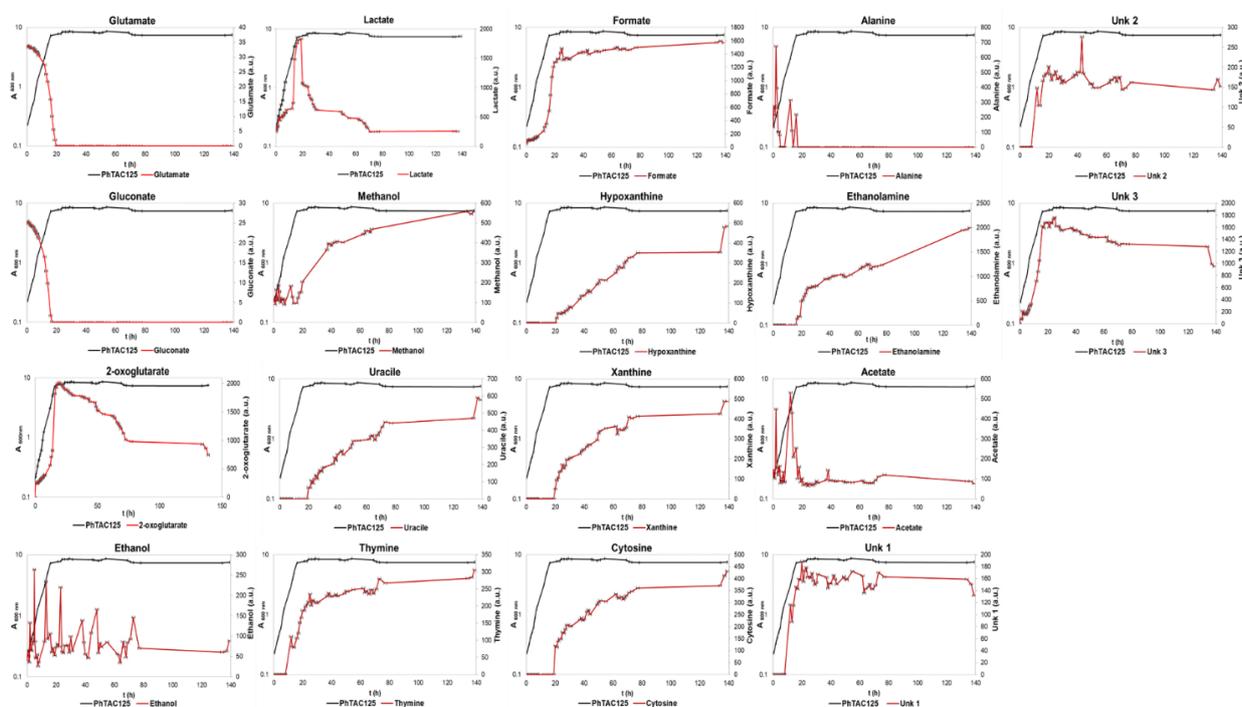
The ferric ion concentration was determined using inductively coupled plasma atomic-emission spectrometry (ICP-AES). For the quantification of the iron concentration with ICP-AES, a calibration line was generated. The wavelength used for iron determination was 238.204 nm but an extra wavelength was used (239.562 nm) to control the interferences from other elements at the standard wavelength.

1.2.3 Results

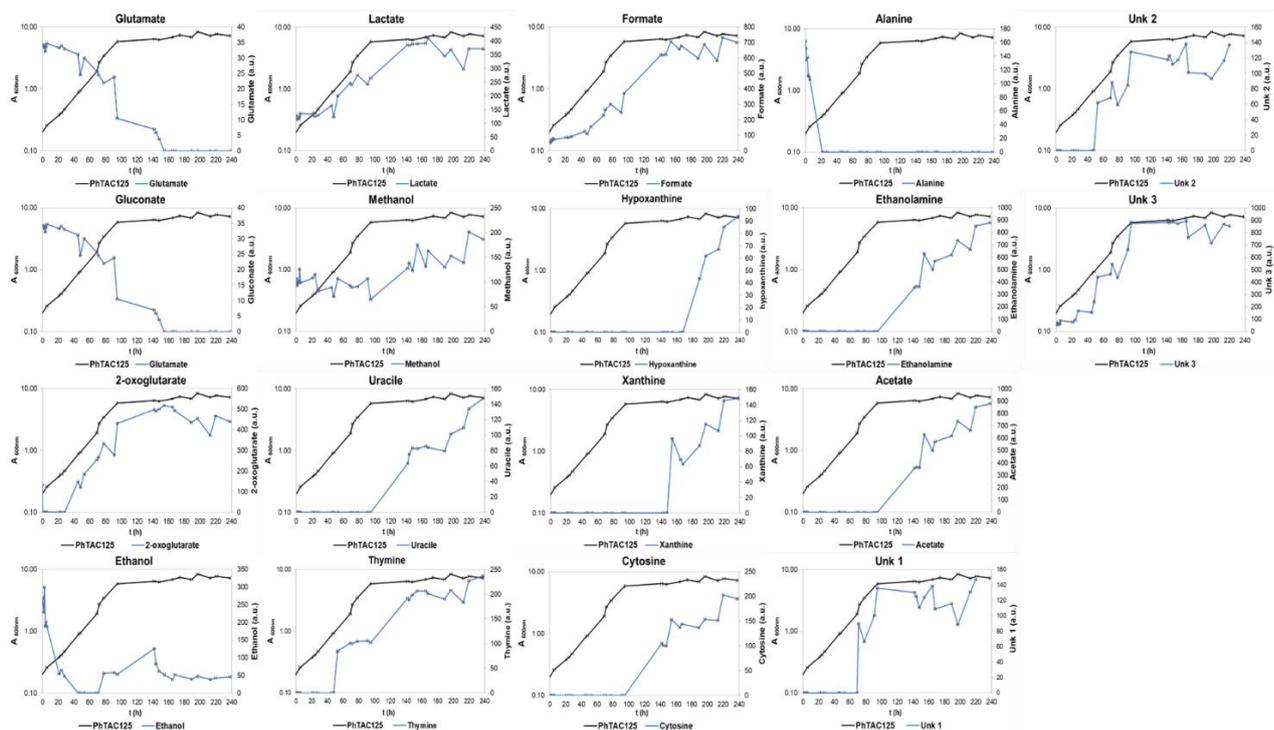
1.2.3.1 Differences in the extracellular metabolite composition at 15°C and 0°C The 0 and 15°C transcriptomes of *PhTAC125*

In this section, an in-depth study on the changes in the levels of extracellular metabolites of *PhTAC125* at two different temperatures was carried out. To this end, 30 samples were collected at regular intervals during the growths of *PhTAC125* in a bioreactor at 15 °C and 0 °C and analyzed by ¹H-NMR. The obtained data shed light on the capability of *PhTAC125* to produce similar metabolic phenotypes at 15 °C and 0°C, shown in Figures 3.1.1 and 3.1.2, respectively. Indeed, two of the 18 extracellular metabolites detected showed

different abundance levels and secretion profiles at two temperatures (Figure 3.1.3). In particular, lactate and 2-oxoglutarate are the metabolites that displayed the largest variation. At 0 °C, the lactate was already detected at 1h, whereas the 2-oxoglutarate after 28h of growth (Figure 3.1.3 A-B). These metabolites continued to accumulate in the medium until the end of the growth. At 15 °C, an extracellular accumulation of these metabolites was observed from the beginning of the growth, with a significant decrease after 19h (Figure 3.1.3 C-D). This reduction occurs as cells deplete their carbon sources (gluconate and glutamate at 17h and 19h, respectively) and begin to rely on their ability to use for environmental lactate and 2-oxoglutarate (Figure 3.1.3 A - B). As mentioned above, the cellular Fe:C ratio for several marine bacteria involved changes in the regulation of the metabolic pathways, including the TCA cycle²¹. To understand whether the profiles of 2-oxoglutarate and lactate in *PhTAC125* are linked to the remodeling of metabolic pathways in response to carbon and iron-deficient conditions, it is necessary to determine the extracellular levels of iron.



3.1.1 Growth curve and extracellular metabolite profiles of *PhTAC125* at 15 °C. The growth curve of *PhTAC125* in the GG medium at 15 °C is shown as a black line. The optical density (600 nm) was measured until 140 hours. The time-resolved extracellular metabolite concentrations (a.u.) were measured using ¹H NMR spectroscopy during the *PhTAC125* growth in the same growth conditions, indicated with red lines. Error bars represent the standard deviation between duplicate samples.



3.1.2 Growth curve and extracellular metabolite profiles of *PhTAC125* at 0 °C. The growth curve of *PhTAC125* in the GG medium at 0°C is shown as a black line. The optical density (600 nm) was measured until 240 hours. The time-resolved extracellular metabolite concentrations (a.u.) were measured using ^1H NMR spectroscopy during the *PhTAC125* growth in the same growth conditions, indicated with blue lines. Error bars represent the standard deviation between duplicate samples.

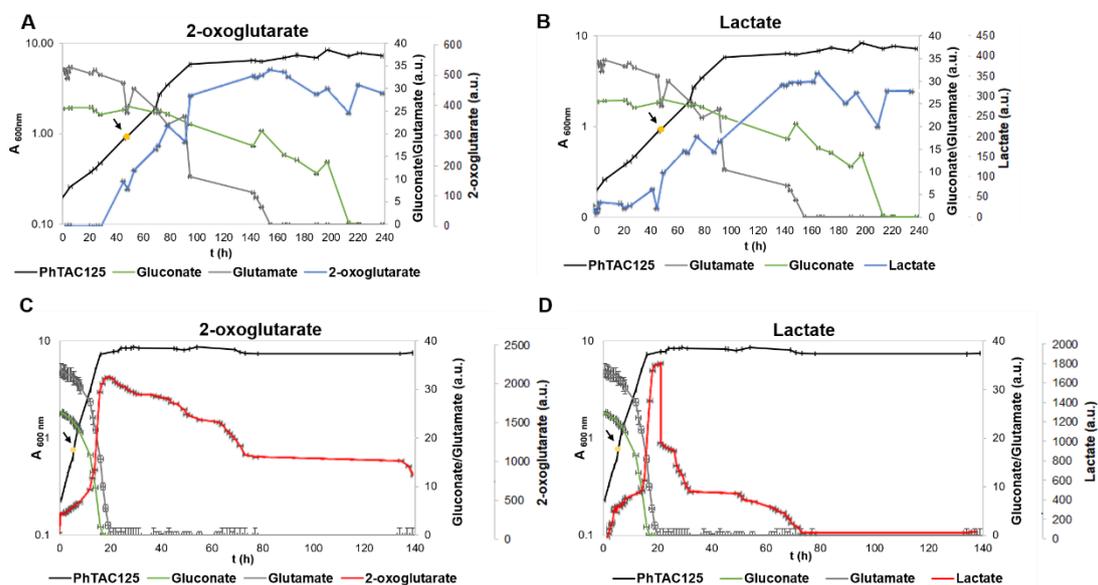


Figure 3.1.3. The growth curve of *PhTAC125* at 0 °C and 15 °C and metabolic profiles are compared to bring out the correlation between the depletion of carbon sources (gluconate and glutamate) and the internalization of 2-oxoglutarate (A - C) and lactate (B -D) at the end of the exponential growth phase. *PhTAC125* growth curve, and the profiles of gluconate, and glutamate are represented in black, green, grey, and red lines, respectively. The profiles of 2-oxoglutarate and lactate at 0 °C and 15 °C are in blue and red lines, respectively. The transcriptome of *PhTAC125* growing cells was sampled at 50h (0°C) and 7h (15 °C), as indicated with a black arrow. Error bars represent the standard deviation between duplicate samples.

The free ferric ion (Fe^{3+}) concentration was determined at 5 different time points taken during the growth of *PhTAC125* in the GG medium (with an initial Fe^{3+} concentration: 1000 $\mu\text{g/L}$) at 0 and 15 °C by ICP-AES, performed in collaboration with the University of Florence. At 0 °C, the samples were collected the following time points: 69, 78, 95, 142, and 245 hours. On the other hand, the 15 °C curve was sampled at 6, 12, 16, 39, and 68 hours. The qualitative data on ferric ions determination are listed in Table 3.2.1. An abnormal result can be observed in the first sample at 15 °C since the value is higher than that initial. This phenomenon may occur when iron traces in the environment affect the determination of the measure.

In 1999, Granger *et al.*²² reported the exact concentration of Fe-limiting in *PhTAC125* corresponding to 500 $\mu\text{g/L}$ Fe^{3+} when the Antarctic bacterium is grown in batch culture in the artificial seawater medium at 20 °C. Considering this study, although the growth conditions are not similar, it is possible to suggest that the psychrophilic bacterium is under Fe depletion after 16 h and 142 h at 15 °C and 0 °C, respectively. This result is surprising because the Fe-limitation condition at 15 °C corresponds to the exact point of the carbon sources starvation and the lactate and 2-oxoglutarate secretion, as indicated with a black arrow in Figure 3.2.1 A - B. That might not be a simple matter of coincidence, but it could suggest that *PhTAC125* rearranged its metabolism in order to cope with Fe stress.

Detection time (h)	Temperature (°C)	Fe^{3+} Concentration ($\mu\text{g/L}$)
6h	15°C	1273
12h	15°C	841
16h	15°C	617
39h	15°C	330
68h	15°C	243
69h	0°C	885
78h	0°C	850
95h	0°C	739
142h	0°C	504
245h	0°C	376

Table 3.2.1. Determination of extracellular Fe^{3+} concentration in 5 different time points during the *PhTAC125* growth in GG medium at 15°C and 0 °C.

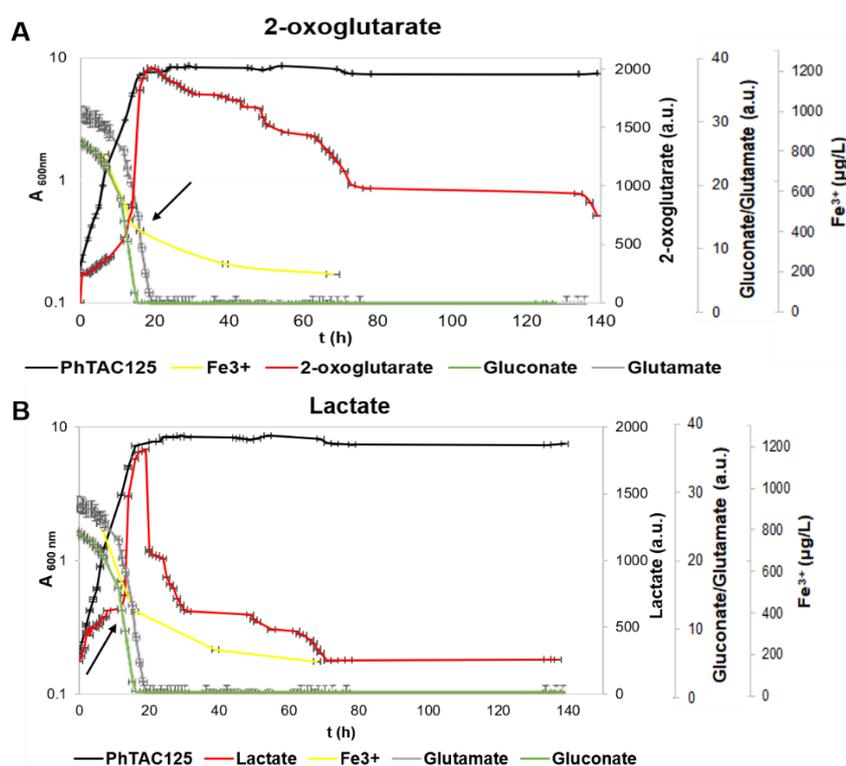


Figure 3.2.1 The graph shows the determination of the extracellular iron concentration about the growth curve of *PhTAC125* at 15 °C and metabolic profiles of gluconate, glutamate, lactate, and 2-oxoglutarate. *PhTAC125* growth curve, gluconate, glutamate, Fe³⁺, 2-oxoglutarate (A), and lactate (B) are represented in black, green, grey, yellow, and red lines, respectively. The exact point of the carbon sources starvation and the lactate and 2-oxoglutarate-scavenging is indicated with a black arrow. Error bars represent the standard deviation between duplicate samples.

To understand if changes in the lactate and 2-oxoglutarate profiles were associated with iron deficiency, *PhTAC125* was grown in the GG medium with a concentration of iron sulfate (FeSO₄·7H₂O) four times greater than that initially at 15 °C. This experiment was carried out only at 15 °C since the lactate and 2-oxoglutarate profiles did not bring out changes at 0 °C. In Figure 3.2.2, the growth curve of *PhTAC125* at 15 °C in the GG medium with a higher iron concentration (20 mg/L iron sulfate) (in orange line) is compared to that obtained from growing the bacterium with the classical iron sulfate concentration (5 mg/L iron sulfate) (in black line). The supplementation of iron did not show any sign of toxicity or hampering the growth of the psychrophilic bacterium. Specifically, the average growth rate of the bacterial cultures in the two iron concentrations did not vary, being 0.27 (s.d. 0.0009) h⁻¹ and 0.25 (s.d. 0.0003) h⁻¹ at 5 mg/L iron sulfate and 20 mg/L iron sulfate, respectively. Furthermore, the free ferric ion concentration determination showed that utilizing 20 mg/L of iron sulfate could already be sufficient to overcome the iron deficiency condition detected previously at 16h (Table 3.2.2). The extracellular metabolic analysis detected a significant reduction of the extracellular levels of lactate and 2-oxoglutarate, remaining the profiles of the other metabolites unperturbed (Figure 3.2.3). These findings would confirm the hypothesis that the secretion of lactate and 2-oxoglutarate is linked to iron limitation. The psychrophilic bacterium was grown with a higher iron sulfate concentration (70 mg/L FeSO₄·7H₂O) to assess the effect of a higher concentration of iron on growth

parameters. Also in this condition, the iron incrementation is not toxic to bacterial growth, as indicated by the growth curve in the green line in Figure 3.2.2, and the average growth rate (0.29, s.d. 0.0005 h⁻¹) is similar to those encountered in the other two conditions. Therefore, the increment of iron sulfate concentration in the GG medium involves the reduced extracellular levels of the lactate and 2-oxoglutarate but doesn't enhance the growth kinetic parameters. These achievements suggest that the growth of *PhTAC125* can be limited, beyond iron, by other nutrients.

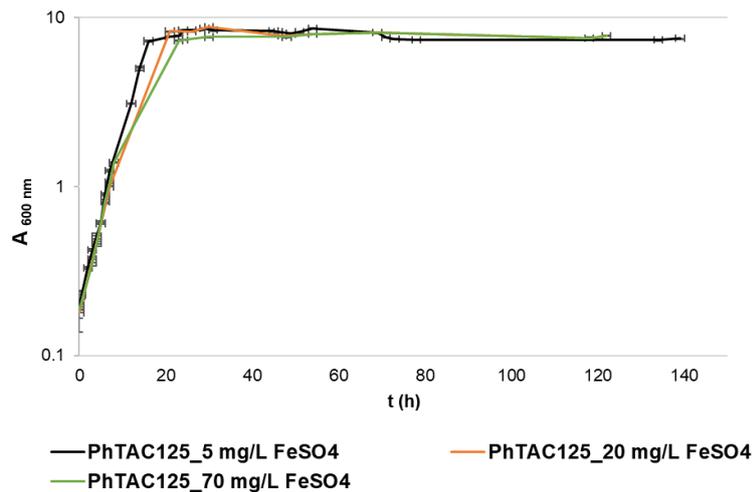
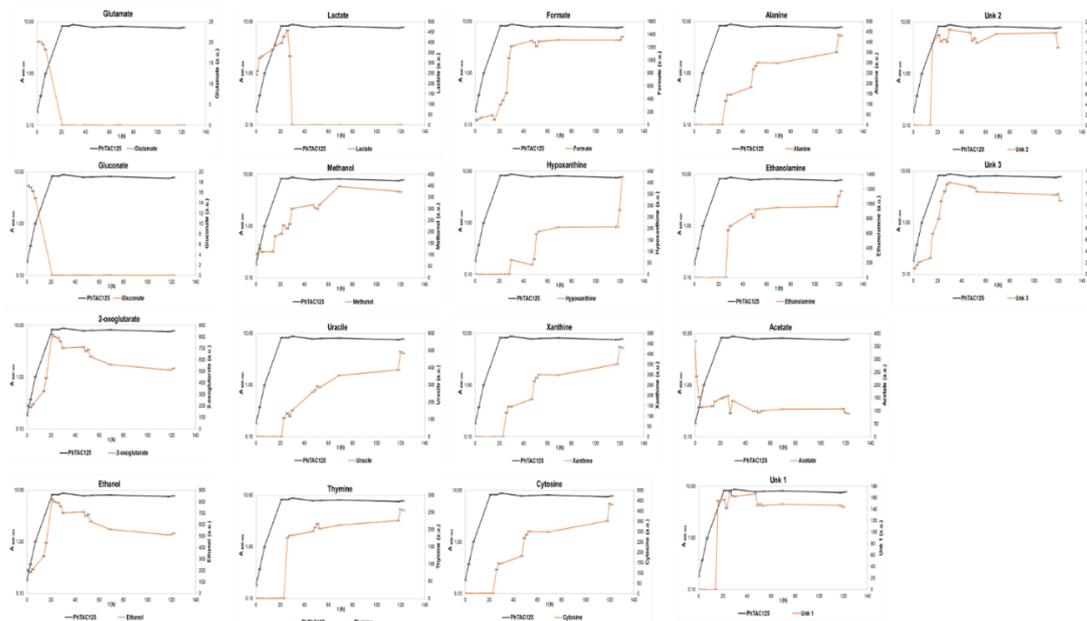


Figure 3.2.2 Growth curves of *PhTAC125* in the GG medium supplemented with different $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ concentrations at 15 °C. The bacterium was grown in the GG medium with three different iron sulfate concentrations, 5 mg/L iron sulfate, 20 mg/L iron sulfate, and 70 mg/L iron sulfate in the black, orange, and green line, respectively. Error bars represent the standard deviation between duplicate samples.



3.2.3 Growth curve and extracellular metabolite profiles of *PhTAC125* grown in the GG medium supplemented with 20 mg/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ at 15 °C. The growth curve of *PhTAC125* in the GG medium at 15 °C is shown as a black line. The optical density (600 nm) was measured until 122 hours. The time-resolved extracellular metabolite concentrations (a.u.) were measured using ¹H NMR spectroscopy

during the *PhTAC125* growth in the same growth conditions, indicated with orange lines. Error bars represent the standard deviation between duplicate samples.

Samples	Fe ³⁺ Concentration (µg/L)
PhTAC125 3h	3304
PhTAC125 5h	3243
PhTAC125 21h	2455
PhTAC125 30h	1987
PhTAC125 68h	793

Table 3.2.2. Determination of extracellular iron concentration during the *PhTAC125* growth in GG medium supplemented with 20 mg/L FeSO₄·7H₂O at 15 °C.

1.2.3.2 Transcriptomic analysis reveals differentially expressed genes associated to lactate and 2-oxoglutarate

In this section, the differential transcriptional analysis in *PhTAC125* aims to investigate changes in the expression of genes associated with lactate and 2-oxoglutarate and the molecular mechanisms that regulate them. This experiment was performed on *PhTAC125* at 7h (the beginning of the exponential growth phase) in the GG medium at 15 °C and 0 °C, as indicated with a black arrow in Figure 3.1.3. Differential gene expression data detected the down-regulation of the main enzymes involved in the TCA cycle and the up-regulation of the lactate permease and a candidate gene for the L-lactate dehydrogenase enzyme at 15 °C (Table 3.3). Given the key role of 2-oxoglutarate in signaling the nitrogen/carbon status of the cells²³, the transcriptomic analysis focused on the expression of genes whose products could be sensitive to 2-oxoglutarate levels. In many *Proteobacteria*, 2-oxoglutarate activates ammonium uptake via binding to GlnD protein, a uridylyl transferase/uridylyl-removing enzyme that senses and transmits the nitrogen status of the cell to the PII protein (GlnK)²⁴. The PII protein, in its uridylylated state, stimulates phosphorylation of the regulator NtrC by the NtrB kinase, which activates a large number of promoters for transcription of nitrogen assimilation genes that are recognized by the alternative σ factor σ^{N25} . Furthermore, further studies reported that citrate synthase, glutamate dehydrogenase, and 3-phosphoglycerate dehydrogenase enzymes are subjected to 2-oxoglutarate inhibition²⁶. From the differential transcriptomic analysis in *PhTAC125*, it is interesting to note the upregulation at 15 °C of the same genes regulated by 2-oxoglutarate in other microorganisms and principally involved in nitrogen assimilation (Table 3.3).

Metabolite	Gene ^a	Name ^b	Log ₂ Fold change (Log ₂ FC) ^c	P-value ^d	Annotation ^e
Lactate	<i>PSHAa0976</i>		2.9418	5.71E-14	putative dehydrogenase with NAD(P)-binding domain
	<i>PSHAa0978</i>		1.8689	2.52E-08	putative L-lactate permease
2-oxoglutarate	<i>PSHAa0159</i>	<i>acnA</i>	-1.4819	4.41E-07	aconitate hydratase
	<i>PSHAa1166</i>	<i>fumB</i>	-1.2879	2.17E-05	fumarate hydratase
	<i>PSHAa1644</i>	<i>sucD</i>	-0.9859	0.0031042	succinyl-CoA synthetase
	<i>PSHAa2002</i>	<i>ygfH</i>	-1.2793	1.56E-05	propionyl-CoA:succinate-CoA transferase
	<i>PSHAb0465</i>		-1.6465	5.13E-09	isopropylmalate dehydrogenase
2-oxoglutarate	<i>PSHAa0163</i>	<i>glnG</i>	0.7806	0.001294056	response regulator in two-component regulatory system with GlnL
	<i>PSHAa0666</i>	<i>serA</i>	-1.52318	5.94E-05	D-3-phosphoglycerate dehydrogenase
	<i>PSHAa1392</i>	<i>gdhA</i>	-0.9288	0.003669025	glutamate dehydrogenase
	<i>PSHAa1653</i>	<i>gltA</i>	-0.8985	0.007239206	citrate synthase
	<i>PSHAa2038</i>	<i>glnD</i>	0.7919	0.006501994	uridylyltransferase
	<i>PSHAa2258</i>	<i>glnK</i>	1.4150	0.007097977	Nitrogen regulatory protein P-II
	<i>PSHAa2551</i>	<i>rpnO</i>	0.7175	0.004784182	sigma N (sigma 54) factor of RNA polymerase
	<i>PSHAa2553</i>	<i>ptsN</i>	0.6786	0.014831172	PTS family enzyme IIA

Table 3.3. Genes involved in nitrogen metabolism that are differentially expressed during the growth of *PhTAC125* in GG medium at 15°C and 0°C.

^aLocus number of gene in *PhTAC125* genome (NC_007481.1/ NC_007482.1).

^bGenes in the nitrogen and carbon regulated transcriptome that were analyzed in this study.

^cGene expression profiles in log₂ fold change (log₂FC) obtained using REMI method. Genes with Log₂ Fold change greater than 0 are up-regulated genes at 15 °C, otherwise, up-regulated genes at 0 °C.

^d P-values of gene expression ratio from three biological replicates were corrected for multiple testing using Spearman correlation.

^eGene annotation from *PhTAC125* genome database (NC_007481.1/ NC_007482.1).

1.2.4 Discussion

The ability to survive in extreme environments requires that the Antarctic bacterium efficiently responds to changes in environmental conditions, such as temperature and nutrient fluctuation. As with other bacteria, the primary way that *PhTAC125* adapts to external changes is by altering gene expression to maintain unperturbed central metabolism (Riccardi *et. al* 2022 paragraph 1.1). However, one of the questions that remain to be answered is how the psychrophilic bacterium may regulate cellular adaptations and processes of acclimation in response to nutrient limitation. Integration of multiple omics technologies, such as transcriptomics and metabolomics, could be a valid approach to provide a more comprehensive view of what is going on in the bacterium under particular growth conditions. In addition, these technologies could shed light on possible limiting nutrients in the GG medium

affecting the growth performances of the *PhTAC125*. To this end, the extracellular metabolic analysis was performed to investigate the metabolite profiles of *PhTAC125* in the GG medium at 15 °C and 0 °C. The findings revealed no significant changes in metabolite profiles at the two temperatures, except for lactate and 2-oxoglutarate (Figure 3.1.1, Figure 3.1.2). To search for a possible explanation for the secretion of two metabolites during the exponential phase, the influence of other essential nutrients, which could be limiting in the GG medium, was investigated. In marine heterotrophic bacteria, it is demonstrated that the TCA cycle activity could be reduced under the Fe-limitation condition being iron an essential co-factor of TCA enzymes²⁷. To determine whether the psychrophilic bacterium senses an iron-limiting condition contributing to the secretion of lactate and 2-oxoglutarate, the iron concentration was detected during the *PhTAC125* growth in the GG medium at 15 °C and 0 °C. Considering the concentration of Fe-limiting in *PhTAC125* reported in a study by Granger²², the determination of the extracellular ferric ion concentration confirmed the hypothesis that the Antarctic bacterium is in an iron-restricted condition after 16 h at 15 °C (Table 3.2.1). Generally, the microorganisms have evolved a variety of mechanisms to rearrange their metabolism under Fe stress²⁸, such as an increment in the production of the siderophores²⁹ or the expression of outer-membrane receptors that target the iron-containing molecules³⁰. In *PhTAC125*, a complete siderophore biosynthesis gene cluster is yet to be found in its genome, even though it presents siderophore receptor genes. As in other marine bacteria, *PhTAC125* could use siderophore receptors on its surface to uptake Fe-containing molecules released by other microorganisms in the environment³¹. In addition, the negative results of the chrome azurol S (CAS) assay³² (a well-established colorimetric method used to detect siderophores independent of their structure) provided further experimental evidence that *PhTAC125* is not able to express siderophores (data not shown). After these considerations, *PhTAC125* was grown in the GG medium with higher iron concentrations (20 mg/L and 70 mg/L iron sulfate) at 15 °C. From extracellular metabolic analysis, it is possible to detect a significant reduction of levels of lactate and 2-oxoglutarate, maintaining unperturbed profiles of the other metabolites (Figure 3.2.3). Although these results suggest that the trends of lactate and 2-oxoglutarate could be an adaptive response of the psychrophilic bacterium in dealing with Fe-limitation, further studies are necessary to confirm that the profiles of these metabolites are a direct effect of Fe-limitation. After an extracellular accumulation during an initial exponential phase at 15 °C, the two metabolites are internalized following the exhaustion of the carbon sources (gluconate and glutamate). Since bacterial metabolism is intricately linked to the availability of carbon sources and cellular responses to this availability, this phenomenon might represent a physiological plasticity response of *PhTAC125* to reprogram its metabolic pathways when subjected to nutrient limitations. Therefore, lactate and 2-oxoglutarate could be viewed as signaling metabolites showing a limitation of carbon sources in the medium. This consideration is not surprising as several studies in *E. coli* reported that 2-oxoglutarate is a master regulator metabolite that signals the balance between carbon and nitrogen metabolism³³. Specifically, the 2-oxoglutarate levels are reduced when the key enzymes of the TCA cycle are affected by carbon starvation stress^{34,35}. So, the 2-oxoglutarate scavenging could indicate a reprogramming of *PhTAC125* metabolism to use this metabolite as a

carbon source. Lactate levels are also associated with carbon source availability, so its internalization might present an additional carbon substrate for *PhTAC125*. As reported in several studies, the initial response of marine bacteria to carbon starvation is to up-regulate the high-affinity lactate transport systems³⁶. Furthermore, other catabolic enzymes, such as lactate dehydrogenase, have also been shown to increase significantly under carbon starvation conditions³⁶. The understanding of metabolomic achievements was enhanced by the integration of transcriptomic data, thus linking the fluctuations of the two metabolites (lactate and 2-oxoglutarate) to the expression of genes associated with them. The transcriptomic analysis was carried out on *PhTAC125* at the beginning of the exponential growth phase in the GG medium at 15 °C and 0°C. The data revealed the overexpression of genes responsible for the transport and utilization of lactate at 15 °C (Table 3.3). The study by Toyoda³⁷ showed that the up-regulation of genes involved in lactate catabolism could represent a strategy for cells to use lactate as a carbon and energy source during energetically limited growth and low oxygen availability. Furthermore, the transcriptional analysis in *PhTAC125* revealed surprising overexpression of a variety of genes involved in the assimilation of nitrogen sources at 15 °C (Table 3.3). Extensively studies in *E. coli*, *Bacillus subtilis*, and *Salmonella typhimurium* reported that the first transcriptional response to nitrogen starvation is the upregulation of nitrogen assimilation-related genes through sensing the 2-oxoglutarate accumulation in the cell^{33,38}. This evidence provides a dominant role to 2-oxoglutarate concentration in triggering a transcriptional response to nitrogen starvation. Therefore, the overexpression of genes involved in nitrogen uptake and assimilation would indicate a status of nitrogen deficiency in *PhTAC125* at 15 °C, but further research is required to demonstrate the correlation between the 2-oxoglutarate and the down- and/or up-regulation of these nitrogen-regulated genes.

In conclusion, this study provides insights into the adaptation strategies in response to multiple nutrient deficiencies when *PhTAC125* is grown in the GG medium at two different temperatures. Specifically, the presented results may hypothesize that a combination of limiting nutrients, including carbon, nitrogen, and iron, leads to several acclimation responses that may be activated either simultaneously or sequentially in the psychrophilic bacterium. Indeed, the different nutritional deficiencies of *PhTAC125* at 15 °C and 0 °C could explain the unlike extracellular profiles of lactate and 2-oxoglutarate. In addition, these physiological responses can be mediated by fluctuations in the concentrations of principal metabolites, such as the 2-oxoglutarate, which enable the transfer of this information to regulatory proteins, thereby generating an appropriate metabolic response. Furthermore, the obtained information on iron concentration and nutrient catabolism could allow us to set up a new growth medium starting from the GG formulation, even though further studies on the metabolic regulation are necessary to answer many of the remaining questions that arose in this work.

Authors' Contribution

M. Calvanese and E. Parrilli designed all the experiments. M. Calvanese conducted growth experiments, RNA extraction, and wrote the work. C. Riccardi and M. Fondi at University of Florence carried out the transcriptomic analysis. P. Tufano and V.

Ghini at CERM (University of Florence) performed the metabolomic experiments. M. Calvanese and E. Parrilli contributed to collect and interpret the data. E. Parrilli contributed to revise the work.

References

1. Mashego, M. R. *et al.* Microbial metabolomics: Past, present and future methodologies. *Biotechnol. Lett.* 29, 1–16 (2007).
2. Pinu, F. R., Villas-Boas, S. G. & Aggio, R. Analysis of intracellular metabolites from microorganisms: Quenching and extraction protocols. *Metabolites* 7, (2017).
3. Sannino, F. *et al.* A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl. Microbiol. Biotechnol.* 101, 725–734 (2017).
4. Moriah Sandy and Alison Butler. Microbial Iron Acquisition: Marine and Terrestrial zzzzzczzz Moriah. *Chem Rev.* 23, 1–7 (2013).
5. Martin, J. H., Gordon, R. M. & Fitzwater, S. E. Iron in Antarctic waters. *Nature* 345, 156–158 (1990).
6. Abdul-Tehrani, H. *et al.* Ferritin mutants of *Escherichia coli* are iron deficient and growth impaired, and fur mutants are iron deficient. *J. Bacteriol.* 181, 1415–1428 (1999).
7. Andrews, S. C., Robinson, A. K. & Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* 27, 215–237 (2003).
8. Braun, V. & Braun, M. Active transport of iron and siderophore antibiotics. *Curr. Opin. Microbiol.* 5, 194–201 (2002).
9. Fourquez, M. *et al.* Effects of iron limitation on growth and carbon metabolism in oceanic and coastal heterotrophic bacteria. *Limnol. Oceanogr.* 59, 349–360 (2014).
10. Erdner, D. L. & Anderson, D. M. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol. Oceanogr.* 44, 1609–1615 (1999).
11. Kornberg, H. L. The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem. J.* 99, 1–11 (1966).
12. Medigue, C. *et al.* Coping with cold: The genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res.* 15, 1325–1335 (2005).
13. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890 (2018).
14. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419 (2017).
15. Bernacchioni, C. *et al.* NMR metabolomics highlights sphingosine kinase-1 as a new molecular switch in the orchestration of aberrant metabolic phenotype in cancer cells. *Mol. Oncol.* 11, 517–533 (2017).
16. Cortinagil, E. *et al.* Search for heavy neutral lepton production in k+decays. *Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys.* 778, 137–145 (2018).

17. Takis, P. G., Ghini, V., Tenori, L., Turano, P. & Luchinat, C. Uniqueness of the NMR approach to metabolomics. *TrAC - Trends Anal. Chem.* 120, 115300 (2019).
18. Vignoli, A. *et al.* High-Throughput Metabolomics by 1D NMR. *Angew. Chemie - Int. Ed.* 58, 968–994 (2019).
19. Perrin, E. *et al.* Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nat. Commun.* 11, 1–16 (2020).
20. Daly, G., Perrin, E., Viti, C., Fondi, M. & Adessi, A. Scaling down the microbial loop: data-driven modelling of growth interactions in a diatom–bacterium co-culture. *Environ. Microbiol. Rep.* 13, 945–954 (2021).
21. Tortell, P. D., Maldonado, M. T. & Price, N. M. bacteria in iron-limited ocean ecosystems. 383, 330–332 (1996).
22. Granger, J. & Price, N. M. The importance of siderophores in iron nutrition of heterotrophic marine bacteria. *Limnol. Oceanogr.* 44, 541–555 (1999).
23. Huergo, L. F. The Emergence of 2-Oxoglutarate as a Master Regulator Metabolite. 79, 419–435 (2015).
24. Jiang, P., Peliska, J. A. & Ninfa, A. J. Enzymological Characterization of the Signal-Transducing Uridylyltransferase / Uridylyl-Removing Enzyme (EC 2.7.7.59) of *Escherichia coli* and Its Interaction with the PII Protein †. 2960, 12782–12794 (1998).
25. Kamberov, E. S., Atkinson, M. R. & Ninfa, A. J. The *Escherichia coli* PII Signal Transduction Protein Is Activated upon Binding 2-Ketoglutarate and ATP *. *J. Biol. Chem.* 270, 17797–17807 (1995).
26. Zhao, M. *et al.* Structural basis for the allosteric control of the global transcription factor NtcA by the nitrogen starvation signal 2-oxoglutarate. 107, (2010).
27. Flint, D. H. & Allen, R. M. Iron – Sulfur Proteins with Nonredox Functions Contain a Fe-S Cluster That May Perform. *Chem. Rev.* (1996).
28. Caza, M. & Kronstad, J. W. Shared and distinct mechanisms of iron acquisition by bacterial and fungal pathogens of humans. 3, 1–23 (2013).
29. Sasnow, S. S., Wei, H. & Aristilde, L. Bypasses in intracellular glucose metabolism in iron- - limited *Pseudomonas putida*. 3–20 (2015) doi:10.1002/mbo3.287.
30. Hantke, K. & Braun, V. Control of Bacterial Iron Transport by Regulatory Proteins. 11–44 (1991).
31. Chatterjee, A. & Brian, M. R. O. Rapid evolution of a bacterial iron acquisition system. 108, 90–100 (2018).
32. Neilands, B. Universal Chemical Assay for the Detection Determination of Siderophores'. 56, 47–56 (1987).
33. Opin, C., Commichau, F. M. & Forchhammer, K. Commichau FM, Forchhammer K, St ?? lke J .. Regulatory links between carbon Regulatory links between carbon and nitrogen metabolism " rg Stu. 167–172 (2006) doi:10.1016/j.mib.2006.01.001.
34. Brauer, M. J. *et al.* Conservation of the metabolomic response to starvation across two divergent microbes. 103, (2006).
35. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nat. Publ. Gr.* 12, (2014).

36. Marine, E. O. N. T. H. E. THE TRANSIENT PHASE BETWEEN GROWTH AND NONGROWTH OF HETEROTROPHIC BACTERIA, WITH EMPHASIS ON THE MARINE. 25–49 (1987).
37. Toyoda, K., Teramoto, H., Inui, M. & Yukawa, H. The *ldhA* Gene, Encoding Fermentative L-Lactate Dehydrogenase of *Corynebacterium glutamicum*, Is under the Control of Positive Feedback Regulation Mediated by LldR. *J. Bacteriol.* 191, 4251–4258 (2009).
38. Wacker, I., Ludwig, H., Reif, I., Blencke, H. & Detsch, C. The regulatory link between carbon and nitrogen metabolism in *Bacillus subtilis*: regulation of the *gltAB* operon by the catabolite control protein CcpA. *Microb. Drug Resist.* 3001–3009 (2016) doi:10.1099/mic.0.26479-0.

Chapter 2: Improvement of *Pseudoalteromonas haloplanktis* TAC125 as host for recombinant protein production: development of high-copy number plasmids

Cold-adapted bacteria have received considerable attention in the last few decades in light of their biotechnological potential as cell factories for high-value proteins, and *PhTAC125* represents a model for such an application. In the last years, numerous studies have focused on the development of an efficient gene expression technology in *PhTAC125*, and a few plasmids have been successfully developed, allowing the production of difficult-to-express proteins in soluble and active form. Although major successes have been achieved, other bottlenecks are limiting the overall efficiency of the recombinant production system, as an example is the low copy number of the psychrophilic vectors. For this purpose, the replication origin (OriR) was mutated randomly, and the collection of OriR-sequences was cloned into a psychrophilic expression vector containing the gene coding for a fluorescent protein (GFP). The library was transferred into the *PhTAC125* by an electroporation protocol optimized for this bacterium, and successively, the induced cells were subjected to several rounds of selection by FACS. This random approach allowed the isolation and characterization of seven clones at higher PCN than the wild-type one, among which one of the selected clones allowed to achieve a two-order of magnitude enhancement in plasmid copy number. These achievements succeeded in the *PhTAC125* optimization as an unconventional platform for recombinant protein production.

Improvement of *Pseudoalteromonas haloplanktis* TAC125 as host for recombinant protein production: development of high-copy number plasmids

Marzia Calvanese¹, Cecilia Balestra², Andrea Colarusso^{1,3}, Concetta Lauro^{1,3}, Christopher Riccardi⁴, Marco Fondi⁴, Ermenegilda Parrilli¹, Maria Luisa Tutino^{1, *}

¹ Department of Chemical Sciences, Federico II University of Naples, Complesso Universitario Monte S.- Angelo, via Cintia, 80126 Napoli Italia

² Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Oceanography Division – OGS, Trieste, Italy

³ Istituto Nazionale Biostrutture e Biosistemi I.N.B.B, Viale Medaglie d'Oro, 305-00136, Roma, Italy

⁴ Department of Biology, Via Madonna del Piano 6, 50018 Sesto Fiorentino, Florence, Italy.

Abstract

The Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*) is considered an interesting alternative host for the recombinant protein production, that can be explored when the conventional bacterial expression systems fail. Indeed, the manufacture of all the difficult-to-express proteins produced so far in this bacterial platform gave back soluble and active products. Despite these promising results, the low yield of recombinant protein production achieved is hampering the wider and industrial exploitation of this psychrophilic cell factory. All the expression plasmids developed so far in *PhTAC125* are based on the origin of replication of the endogenous pMtBL plasmid and are maintained at a very low copy number. In this work, we set up an experimental strategy to select mutated OriR sequences endowed with the ability to establish recombinant plasmids at higher multiplicity per cell. The solution to this major production bottleneck was achieved by the construction of a library of psychrophilic vectors, each containing a randomly mutated version of pMtBL OriR, and its screening by fluorescence-activated cell sorting (FACS). The selected clones allowed the identification of mutated OriR sequences effective in enhancing the plasmid copy number of about two orders of magnitude, and the production of the recombinant green fluorescent protein was increased up to twenty times approximately. Moreover, the molecular characterization of the different mutant OriR sequences allowed us to suggest some preliminary clues on the pMtBL replication mechanism that deserve to be further investigated in the future.

3 Key Points

- Setup of an electroporation procedure for *Pseudoalteromonas haloplanktis* TAC125.
- Two order of magnitude improvement of OriR-derived psychrophilic expression systems.
- Almost twenty times enhancement in Green fluorescent protein production.

Keywords

High-copy plasmid, *Pseudoalteromonas haloplanktis* TAC125, recombinant protein production, cold-adapted bacteria

2.1 Introduction

Escherichia coli and members of the genus *Bacillus* are the most commonly used bacterial cell factories for the production of recombinant proteins for different purposes, ranging from basic research to industrial or therapeutic applications^{1,2}. These organisms are the prokaryotic golden standards for several good reasons, including their well-known genetic set-up, high growth rates in inexpensive media, and high-biomass yields³. However, these cell factories recurrently highlight significant product-related pitfalls, such as the formation of inclusion bodies, recombinant protein toxicity and the lack of biological activity⁴. Different strategies have been pursued to overcome the observed limitations, for instance, by adjusting the growth conditions or by genetically engineering either the host strain or the design of the expression vectors⁵. Unfortunately, despite the intensive application of the abovementioned interventions, a significant proportion of recombinant proteins (especially of human origin) remain unsuccessfully produced. In this respect, implementing new bacterial hosts as platforms for the production of difficult-to-express proteins is an attractive strategy. Although highly laborious, time-consuming, and expensive, this approach seems to be the only path to be pursued for the manufacture of high-value proteins, whose production in other bacterial hosts failed. In this regard, cold-adapted bacteria have received considerable attention in the last two decades for their biotechnological potential as "cell factories". *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*, recently renamed *P. translucida* TAC125)⁶ is the most intensively exploited psychrophilic bacterium for recombinant purposes. This Gram-negative strain, isolated from Antarctic seawater in 1992, was the first polar bacterium for which an efficient gene expression technology was developed⁷, exploiting the autonomous replication sequence (OriR) derived from one of its endogenous plasmids, pMtBL⁸. The implementation of either constitutive or inducible promoters⁸⁻¹², and the formulation of synthetic media based on the bacterial metabolism^{11,13}, allowed the production of several recombinant proteins at temperatures as low as - 2.5 °C¹¹. These successful examples posed *PhTAC125* on the stage as one of the most promising alternatives to conventional cell factories for the successful manufacture of difficult-to express proteins¹⁴⁻¹⁷. Recently, an efficient isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter was implemented in a pMtBL-derived psychrophilic vector, generating the pP79 vector¹². As the psychrophilic bacterium is devoid of any lactose import system, a genetically engineered bacterial strain was produced by genomically inserting of a functional copy of the *lacY* gene from *E. coli* into the *lon*-encoding gene of *PhTAC125*. The resulting mutant (KrPL *LacY*⁺) has two desired features as an improved host for recombinant protein production, i.e. a faster inducer internalization rate, thanks to the mesophilic lactose permease and the absence of the Lon protease activity generally considered the major protease for the degradation of recombinant products¹². Although the abovementioned approaches improved the efficiency of this unconventional cell factory, its industrial exploitation is still limited by the achievable average yields of recombinant products, which are often rather low¹². An experimental evaluation of the pMtBL copy number (PCN) revealed that at 15°C the average PCN is 1.1 ± 0.09 in the mid-exponential phase, and 1.34 ± 0.06

in the late exponential phase¹⁸. Even when the PCN is this low, pMtBL is extremely stable from generation to generation, thanks to the presence of a recently described partitioning system¹⁸. Given that all the *PhTAC125* expression plasmids developed so far are based on such replication origin, the recombinant production is likely limited by the low gene dosage achievable, regardless of the employed promoters and engineered strains. This observation raised our attention to pMtBL OriR and its replication efficiency making the development of psychrophilic replication origins endowed with higher copy number highly desirable and urgent. A rational approach toward this goal turned out unfeasible since the exact replication mechanism of pMtBL is still unknown, and the formulation of any hypothesis is made difficult by the lack of close relatives in the repository databases. Therefore, we decided to adopt a random approach, consisting in the construction of a library of psychrophilic plasmids, each containing a randomly mutated OriR. To obtain a fully representative number of independent recombinant psychrophilic clones an effective protocol for *PhTAC125* transformation was required. To fulfill this aim, we developed for the first time an electroporation procedure for the Antarctic bacterium transformation. The effect of each mutation on the average PCN was measured by the production of a fluorescent protein, upon IPTG induction of the culture. The psychrophilic cells were then sorted by fluorescence-activated cell sorting (FACS), and some cells with increased fluorescence were successfully isolated. Their molecular characterization allowed us to identify different mutant OriR sequences, all of which were able to establish the psychrophilic vector with a higher PCN than the wild-type one, reaching the enhancement of two orders of magnitude in the case of the mutant sequences we named Ori B40 and Ori Y18.

2.2 Materials and methods

2.2.1 Bacterial strains, media, and plasmids

The strains, plasmids, and oligonucleotides used in this study are listed in Table S1. *E. coli DH5* a strain was used for the cloning procedures. *E. coli* S17-1(λ pir) was used as a donor in intergeneric conjugation experiments⁸. *E. coli* cells were routinely grown in the LB (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L NaCl) broth containing 100 μ g/mL ampicillin (Sigma) if transformed at 37 °C. XL10-Gold ultracompetent cells were used for the transformation of the plasmid library after multi-site-directed mutagenesis. Transformation of XL10-Gold ultracompetent cells was optimized using NZY⁺ broth (10 g/L NZ amine (casein hydrolysate), 5 g/L yeast extract, and 5 g/L NaCl), complemented with 12.5 mL 1 M MgCl₂, 12.5 mL 1 M MgSO₄, 10 mL 2 M glucose, pH 7.5. KrPL, a cured *PhTAC125* strain, was used for protein expression and error-prone PCR (epPCR) library screening. The segregational stability assays and plasmid copy number (PCN) evaluation were also performed with the psychrophilic bacterium. KrPL was grown in the TYP broth (16 g/L bacto-tryptone, 16 g/L yeast extract, 10 g/L NaCl) during interspecific conjugations and precultures development. Expression experiments, plasmid stability, and PCN assays were carried out at 15 °C in the GG medium¹¹ and 100 μ g/mL ampicillin if transformed. Mutations of selected clones were confirmed via Sanger sequencing (Eurofins Genomics, Ebersberg, Germany).

2.2.2 Plasmid vector preparation

The pMAI79-*R9-gfp* plasmid used for the library construction was developed from the pP79-*R9-gfp* vector¹². Two unique restriction sites (NotI and Ascl) were introduced upstream and downstream of OriR into the original shuttle vector using the QuikChange Lightning Multi-Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) following the manufacturer's instructions. Two mutagenic primers, Left OriR NotI Fw and Right OriR Ascl Rv (Table S1), were designed to introduce the two restriction sites. A small volume (1.5 μ L) of the resulting PCR products was transformed into the XL10-Gold ultracompetent cells by chemical transformation. Then, the resulting plasmid, pMAI79-*R9-gfp*, was mobilized into KrPL by conjugation⁸. Selection of the recombinant transconjugants was performed at 15 °C in the presence of 50 μ g/mL kanamycin and 100 μ g/mL ampicillin.

2.2.3 Construction of the plasmid library

A library of plasmids was constructed by introducing random mutations in OriR (842 bp) of the pMAI79-*R9-gfp* expression system using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). To introduce mutations with low frequency (0 – 4.5 mutations/kb), 800 ng of target DNA was mutagenized by epPCR according to the manufacturer's protocol using both primers mut_pP79-*R9-gfp*_Fw and mut_pP79-*R9-gfp*_Rv (Table S1). The epPCR reaction was performed in 50 μ L containing 1 μ L 2.5 U/ μ L Mutazyme II DNA polymerase, 2 μ L 10 μ M each primer, 1 μ L 40 mM dNTP mix, 5 μ L 10X Mutazyme II reaction buffer, and 4 μ L distilled water. The obtained amplicons were subjected to NotI/Ascl double digestion and cloned into pMAI79-*R9-gfp*, which was previously digested with the same restriction enzymes. The plasmid library was transformed into the *E. coli* DH5 α competent cells by chemical transformation. Immediately after transformation, the cells were diluted 5-fold with LB medium, incubated for 3 hours at 37 °C under 220 rpm shaking, and an aliquot of the cells (100 μ L) was plated directly onto LB-agar plates to assess the complexity of the library. The remaining bacterial suspension was supplemented with glycerol at a final concentration of 15% (*vol/vol*), and 200 μ L aliquots were frozen at - 80 °C. An aliquot of the cells was then thawed and spread with two dilutions (1/10 and 1/100) onto LB-agar supplemented with ampicillin. After 24 h at 37 °C, the colonies were counted. The library was large enough to generate at least 250,000 independent plasmid clones in *E. coli*.

2.2.4 Electroporation of KrPL

The epPCR library was transformed into KrPL cells by electroporation. In this work, the electroporation protocol proposed by Kurusu and coworkers¹⁹ was optimized for KrPL as follows. The strain was streaked from a glycerol stock (stored at - 80 °C) over a TYP agar plate that was incubated for 2 days at 15 °C. Then, one isolated colony was dispersed in 2 mL of the TYP and incubated

overnight at 15 °C with 200 rpm agitation. Afterwards, the culture was diluted twice (1/100) in the same medium within the following 24 h. Finally, the last preculture was diluted in 50 mL of the TYP to start the growth at 15 °C until a value of 0.5 OD₆₀₀ was reached. Then, the cells were recovered by centrifugation (15 min, 4,000 rpm at 4 °C), and resuspended in 50 mL electroporation buffer (252 mM sucrose)¹⁹. Cells were washed twice with smaller volumes of electroporation buffer and finally resuspended in 1 mL of the same buffer. For electroporation, 100 µL of electrocompetent KrPL cells were mixed with 1 µg of DNA and electroporated in 2 mm gap cuvettes (Gene Pulser/MicroPulser Electroporation Cuvettes, Bio-Rad Laboratories, Hercules, CA, USA) at 2.5 kV using a MicroPulser Electroporator (Bio-Rad Laboratories, Hercules, CA, USA). Cells were then immediately transferred to 900 µL TYP, incubated overnight at 15 °C with 200 rpm agitation, and then serial dilutions of the electroporated cell culture were plated out on the TYP-agar selective plates that were kept at 15 °C for 96 h to assess the complexity of the library. The remaining bacterial suspension was supplemented with glycerol at a final concentration of 15% (*vol/vol*), and frozen at – 80 °C. The number of transformants was estimated by counting the colonies on the plates and the total library size represented in the transformation was calculated (a library size of 10⁵ mutants approximately). A stock culture was used to inoculate 200 mL of the TYP supplemented with 100 µg/mL ampicillin and incubated overnight at 15 °C to complete library enrichment. When the culture medium became turbid (~2.5 OD₆₀₀), it was supplemented with glycerol at a final concentration of 15% (volume/volume), and 5 mL aliquots were frozen at – 80 °C. Plasmid DNA was isolated from a small volume (4 mL) of the amplified library using a maxiprep DNA column (Qiagen, Hilden, Germany).

2.2.5 FACS sorting

A stock culture (~10⁵ cells) of KrPL was slowly thawed on ice and inoculated into 50 mL of the TYP broth plus ampicillin (100 µg/mL) at 15 °C. After 24 h, the bacterial culture was diluted (100-fold dilution) in the GG medium-plus ampicillin (100 µg/mL) and kept overnight at 15 °C. After a series of serial dilutions in a crescent volume of the GG medium, the inoculum was grown in the GG medium-plus ampicillin (100 µg/mL). The expression was induced with 5 mM IPTG when the optical density of the sample ranged between 1 and 1.5. After 24 h from the induction, bacteria were analyzed and sorted using a Becton Dickinson Influx cell sorter (BD Biosciences, Franklin Lake, USA), at the cytometry laboratory of the Anton Dohrn Zoology Station, equipped with a 488 nm argon laser for excitation and 100 µm nozzle orifice. For maximum sorting purity, the "1 drop pure" sorting mode was used, ensuring the absence of non-target particles within the target cell droplet and the droplets immediately surrounding the cell. Accuracy was verified microscopically by examining the presence of fluorescent bacteria on a slide. The combination of Side Scatter (SSC) and Green fluorescence (530/40 nm) was used to detect and discriminate fluorescent induced cells from not induced ones. Within the induced cells, three main populations were detected and only the one with the highest green fluorescence signal was sorted. A total of 200,000 fluorescent bacterial cells were selected, collected in a tube containing 500 µL of GG medium, and cultivated in 2

mL of the TYP-selective medium at 4 °C with agitation. After five days, the bacterial culture was transferred into the GG liquid medium at 15 °C and prepared for subsequent sorting under the same conditions described above. An aliquot of each sorted population was harvested during the exponential growth phase ($OD_{600} = 1.0 - 1.5$) in the GG medium, supplemented with 15% glycerol, and then stored at – 80 °C. Data acquisition and recording were achieved with the BD FACS Software software; graphs were drawn with FCS Express 6 Flow v 6.06.0025 (DeNovo Software, USA).

2.2.6 Selection of mutants with higher plasmid copy number and fluorescence assays

Sorted cells were spread onto the TYP agar selective plates and incubated at 15 °C. After 96 h, individual clones were randomly selected and in a sterile 12-well flatbottomed polystyrene plate filled with 1.5 mL TYP medium under selective growth conditions. Then, the cells were transferred to the GG medium with a starting OD_{600} of 0.1 and induced into the exponential phase ($OD_{600} = 1.0 - 1.5$) with 5 mM IPTG. To determine the fluorescence intensity of R9-GFP, 2 OD_{600} of liquid cultures were centrifuged at 13,000 rpm for 10 min at 4 °C and the pellets were resuspended in 0.5 mL PBS. The fluorescence was measured (488-nm excitation - slit 3, 509-nm emission - slit 6) using a JASCO FP-750 spectrofluorometer at 25 °C with a 1 cm path length. Fluorescence was normalized by dividing the measured fluorescence intensity for appropriate dilutions of the cell suspensions. The data were processed using the Origin 81 software.

2.2.7 SDS-PAGE

Recombinant protein production was also monitored by SDS-PAGE. Cell pellets (1 OD_{600}) were harvested after 24 h from the induction and collected by centrifugation (15 min, 12,000 rpm, 4 °C). The pellets were solubilized in 60 μ L of Laemmli buffer 4X and boiled at 95 °C for 20 min and 1.5 μ L of samples were analyzed with 12.5% polyacrylamide gel (SDS-PAGE). After the electrophoresis, the gels were stained with Coomassie brilliant blue dye for 2 h and destained with distilled water.

2.2.8 Plasmid Copy Number determination via qPCR

Relative plasmid copy number (PCN) was calculated by the quantitative PCR (qPCR) method. Amplification and analysis were performed with a StepOne Real-time PCR System (Applied Biosystems, Foster City, CA, USA) and the SYBR® Green PCR Kit (Applied Biosystems, Foster City, CA, USA). Particularly, the *PSHA_RS10135* gene was always used to detect the chromosome in the samples. The *R9-gfp* gene was used for the detection of the plasmid. Each couple of primers was selected using the free Primer 3 web (Table S1). For the PCN

estimation, total DNA was extracted from 1 OD₆₀₀ pellet collected at 8 h from induction using the Bacterial DNA kit (D3350-02, E.Z.N.A™, OMEGA bio-tek, Norcross, GA, USA) following the manufacturer's instructions. DNA concentrations were measured with a NanoDrop™ 1000 Sp spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) at the absorption of 260 nm. The ratios of A260/280 and A260/230 were calculated to estimate the purity of the extracted DNA. The integrity of the extracted DNA was assessed by gel electrophoresis and visualized using the ChemiDoc MP Imaging System (Bio-Rad Laboratories, Hercules, CA, USA). PCR reactions were prepared in 10 µL mixtures containing 1X PowerUp SYBR Green Master Mix (Applied Biosystems, Foster City, CA, USA) with ROX as passive reference dye and Uracil-DNA glycosidase (UDG) to eliminate contaminations, 400 nM of each primer and 1 µL of sample and the reaction master mixes were aliquoted in three wells of a reaction plate. Finally, the plate was sealed with an adhesive cover (Applied Biosystems, Foster City, CA, USA). The thermal cycling protocol was as follows: UDG activation for 2 min at 50 °C; initial denaturation for 10 min at 95 °C; 40 cycles of denaturation for 15 sec at 95 °C alternated with annealing/extension steps for 1 min at 60 °C. Each reaction was performed in triplicate. Cycle threshold (Ct) values were determined after automatic adjustment of the baseline and manual adjustment of the fluorescence threshold using the LightCycler® 96 Software. Standard curves were generated using 10-fold serial dilutions of either *R9-gfp* or *PSHAa2051* genes. The amplification efficiency (E) of each gene was calculated from the slope of the relative standard curve ($E = 10^{(-1/\text{slope})}$). The PCN was determined considering Ct values for the two amplicons (chromosome-c and plasmid-p) and the amplification efficiency of the plasmid (E_p) and chromosomal gene (E_c) with the following equation: $PCN = (E_c)^{C_{tc}} / (E_p)^{C_{tp} 20}$.

2.2.9 Plasmid loss frequency assays (Plasmid stability assay)

The plasmid segregational stability of the progenitor and some mutants was investigated under non-selective growth conditions. A single colony of each strain was collected from TYP agar selective plates and inoculated in the TYP plus antibiotic at 15 °C under agitation. Then, the cells were diluted to 0.1 OD₆₀₀ in the GG medium containing the selective agent. After 24 h of growth at the same temperature, the cultures were 1/20 diluted daily in antibiotic-free GG to keep them constantly in the exponential phase (0.2 – 1.5 OD₆₀₀). At precise intervals of time, the cultured cells were diluted by a factor of 10⁴ and spread onto antibiotic-free-TYP agar plates. After two days of incubation at 15 °C at least 30 colonies were selected and plated onto the TYP agar with and without ampicillin and incubated at 15 °C for two days. The plasmid stability was determined until 80 generations using duplicates. The maintenance of each plasmid in the KrPL bacterium was calculated by the number of colonies grown on the TYP-agar plates with ampicillin divided by the number of colonies grown on the TYP-agar plates without ampicillin.

2.2.10 Sequence analysis

Detection of potential open reading frames (ORFs) was performed with the ORFfinder software at NCBI²¹. The standard genetic code and “atg, gtg, ttg, ctg” as alternative start codons were used. The minimum ORF size was restricted to 30 amino acids (aa). Multiple sequence alignments were conducted in Clustal Omega – EMBL. The predicted secondary structure of DNA and RNA molecules was performed by the Mfold web server²².

2.2.11 Deletions of DNA sequences in the replication origin

The PCR-mediated plasmid DNA deletion method was used to delete DNA sequences in OriR of pMtBL-derived plasmids. Primers were designed to amplify the entire circular sequence of the plasmid except for the specific region that was to be deleted. The 5' ends of the primers include the cutting sequence of the chosen restriction enzyme. The restriction sites, NsiI and BsiWI, were chosen to remove the putative secondary structures, named STEM-LOOP I and STEM-LOOP II. The deletion of the fragment of 342 bp (from 1918 bp to 2259 bp) was performed using the NotI restriction site. The primers used for deletions are listed in Table S1. The PCR reactions were carried out using the Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) and the pMAI79-*R9-gfp* plasmid as the template. The PCR products were then digested with the chosen restriction enzyme (described above) and individually cloned into pMAI79-*R9-gfp* previously digested with the same restriction enzymes. The three mutated vectors were mobilized into KrPL by intergeneric conjugation⁸.

2.2.12 RNA isolation and RNA-Seq data analysis of pMtBL

For the RNA-Seq experiment, the *PhTAC125* bacterium was grown in the GG medium at 15 °C, and 1 OD₆₀₀ pellets were collected during the exponential growth phase (~1/1.5 OD/mL) by centrifugation (10 min, 13,000 rpm, 4 °C), washed in RNase-free PBS three times and stored at – 80 °C. Biological triplicates were performed. Total RNA was extracted using the Direct-zol RNA Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions. Contaminating genomic DNA was then removed through treatment with RNase-free DNase I (Roche, Mannheim, Germany). The sequencing service was provided by the Genome Research Center for Health (Campus of Medicine of the University of Salerno, Baronissi, Italy). Afterwards, all samples were analyzed and assessed for base call quality and adapter content using fastp²³ version 0.22.0, allowing down to a mean quality threshold of 20 (i.e., probability of incorrect base call of 1 in 100) and minimum read length of 40 nucleotides. A median of 98.9% of the reads across all samples passed the quality check, indicating that sequencing was carried out pristinely, and our data were biologically reliable. An index file for reads alignment was produced through Bowtie2²⁴ version 2.2.9 using the entire genomic sequence of *PhTAC125* (NCBI Reference Sequence ASM2608v1) with the additional concatenation of its recently discovered

plasmids, pMEGA (NZ_MN400773.1) and pMtBL (NZ_AJ224742.1). Reads mapping (also with Bowtie2, in very sensitive mode) produced a 99.98% alignment rate. Samtools version 1.11²⁵ with flags $-f$ 99/147 and $-f$ 83/163, respectively, were used to partition sequencing reads that mapped to the forward and reverse strand. After reads partitioning, the same suite was utilized to calculate per-nucleotide depth. This stage of the analysis confirmed a forward stranded protocol (first strand library, data not shown), and allowed to draw a strand-sensitive profile of sequencing depth (Fig. S4). All command line software was run in multithreading on at least one AMD Opteron Processor 6380, 2.5 GHz.

2.3 Results

2.3.1 Set up of a library of psychrophilic plasmids containing randomly mutated pMtBL replication origins

The preparation of the plasmid library containing the randomly mutated OriR started with the construction of pMAI79-*R9-gfp*, a direct derivative of the IPTG-inducible plasmid pP79-*R9-gfp*¹², which harbors the *R9-gfp* gene under the control of the *PhTAE79 lacR-lacZ* regulatory elements¹² (Fig. S1). The pMAI79-*R9-gfp* plasmid differs from its ancestor as two restriction sites (NotI and Ascl, respectively) were added by site-directed mutagenesis upstream and downstream of the pMtBL OriR sequence (corresponding to the sequence from nucleotide 1918 to 2759 of the GenBank database entry NZ_AJ224742.1)⁸. As for the OriR sequence random mutagenesis, the psychrophilic origin of replication was amplified by *error-prone* PCR with low mutation frequency, suitably digested by NotI and Ascl, and then ligated into pMAI79-*R9-gfp* corresponding sites. The library was first transformed into *E. coli* cells, and then a fully representative number of independent plasmids were extracted and used to transform KrPL cells, a *PhTAC125* strain devoid of the endogenous pMtBL plasmid¹², with an electroporation procedure suitably developed for the Antarctic bacterium. The strategy to achieve an effective electroporation was set up by modifying Kurusu's protocol¹⁹. In particular buffer composition and proper voltage settings were found to be key elements for the successful electroporation of Antarctic cells.

2.3.2 FACS sorting of psychrophilic cells harboring the randomly mutated plasmid library

The KrPL (pMAI79-*R9-gfp*) recombinant strain (from now on called “the progenitor”) was preliminarily analyzed by FACS. It was grown in the GG medium at 15 °C, and the expression was induced with 5 mM IPTG when the optical density of the sample ranged between 1 and 1.5. After 24 h from the induction, the cell culture was analyzed by FACS, and the Side Scatter (SSC) and Green fluorescence (530/40 nm) of each cell were correlated (Fig. 2.3.2.1 a). The progenitor population could be divided into three populations (M1, M2, and M3) based on differences both in size signal and fluorescence intensity. The distribution of R9-GFP fluorescence among the cells (Fig. 2.3.2.1 b) also confirmed that they could be divided into the main population (M1) and two subpopulations with higher fluorescence (M2 – M3). Subsequently, KrPL cells harboring the randomly mutated psychrophilic plasmid

library were subjected to FACS analysis, where the recombinant cells were sorted based on a combination of cell size (SSC) and their fluorescence as a proxy for their respective PCN (Fig. 2.3.2.2). A stock culture of recombinant KrPL cells (containing a number of clones representative of the whole library) was first propagated in the TYP medium at 15 °C, and then diluted in the GG medium. Finally, the R9-GFP recombinant production was induced by adding 5 mM IPTG for 24 h. To isolate the most fluorescent clones, the whole library was sorted by FACS. The collected subpopulation (called X population) was propagated in a small volume of the TYP medium at 4 °C; after five days, it was diluted in the GG medium at 15 °C and prepared for the round of sorting under the same conditions as described above. After the second sorting run, the subpopulation of the most fluorescent cells was collected, obtaining the Y population. Repeating this procedure, we collected the Z population after the third sorting run, the A population after the fourth one, and the B population after the fifth sorting run. As shown in Figure 2, the fluorescent cells began to be enriched after the second run compared with the progenitor. The maximum increase in fluorescence was seen after the third run of sorting.

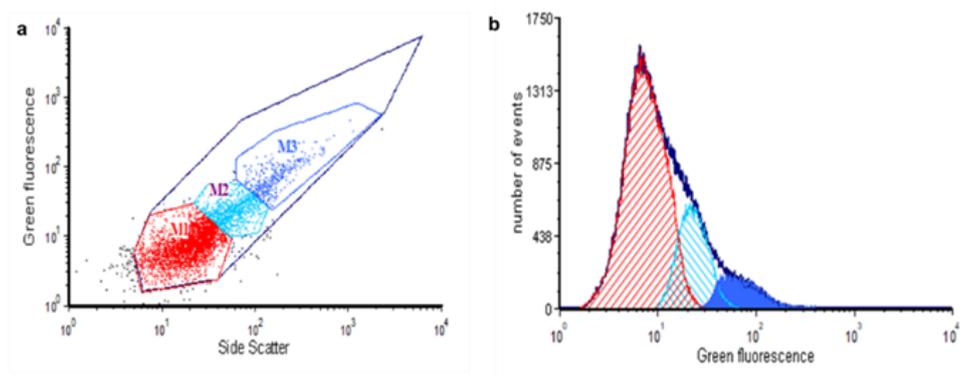


Fig. 2.3.2.1 FACS analysis of progenitor cells. (a) Cytofluorogram of Side Scatter (SSC, x-axis in the graph) and Green fluorescence (530/40[488], y-axis in the graph) shows fluorescent (dark blue line) of the progenitor, in which three main populations were discriminated based on fluorescence signal M1 red, M2 light blue, and M3 blue. (b) Histogram representing the fluorescence intensity of the three populations; The solid region indicates the population M3 which is the one with the highest green fluorescence signal. The graphs were drawn with FCS Express 6 flow v6.06.0033

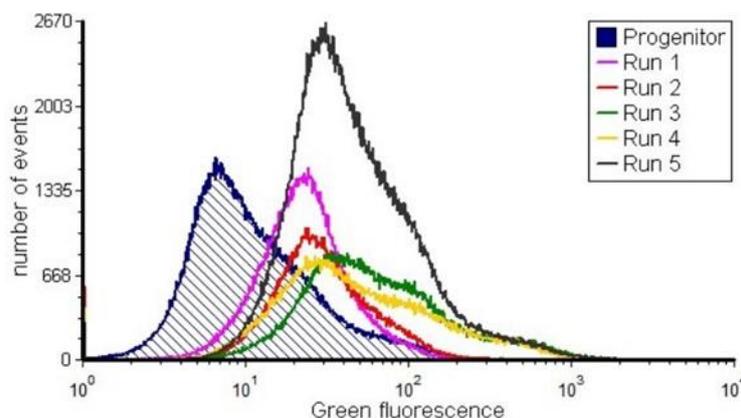


Fig. 2.3.2.2 Histogram of FACS sorting. The random library was subjected to five runs of FACS. The dark blue line shows the progenitor. The first run, second run, third run, fourth run, and fifth run are

represented by pink, red, green, yellow, and black curves, respectively. The maximum increase in Green fluorescence (530/40[488] on the x-axis) was seen after the third sorting (green line). The graphs were drawn with FCS Express 6 flow v6.06.0033

2.3.3 Analysis of FACS-sorted psychrophilic populations and molecular characterization of some selected clones.

Approximately 60 - 100 clones coming from each sorted population (X, Y, Z, A, B,) were randomly selected, cultured in the selective GG medium at 15 °C in 12-well plates, IPTG-induced following the abovementioned production conditions, and subjected to spectrofluorimetric analysis. Figure S2 shows, as an example, the distribution of fluorescence intensities observed in the 60 clones from the screening of the Y population. Once ruled out the occurrence of any deleterious effects on recombinant cell growth possibly due to either replication or segregation instability of the high-copy number plasmids (Fig. S3), 7 clones (X1, Y1, Y18, A31, A44, B15, and B40) were further selected after the population's analyses. Figure 3.3.1a (open bars in the graph) shows that all the selected mutants have a GFP-related fluorescence value higher than that of the progenitor, and the highest fluorescence intensity was recorded for the Y18 clone (8.55×10^6 Arbitrary Units), which is approximately 15-fold higher than the one of the progenitor. Quantitative PCR (qPCR) was performed to determine the direct correlation between the higher fluorescence intensity and each plasmid copy numbers. The obtained results are reported in Figure 2.3.3.1 a (solid bars in the graph, and Table 1). Interestingly, all the selected clones contain recombinant plasmids maintained at higher PCN than the progenitor, with the highest PCN detected in the case of the B40 clone and evaluated as 116 ± 21.69 plasmid copies per cell, which is approximately 30-fold higher than that of the progenitor. Finally, the production of R9-GFP was assessed in the total cellular extracts of IPTG-induced recombinant KrPL cells. Figure 3.3.1b shows a Coomassie-stained SDS-PAGE gel, in which the protein bands assigned to the overproduction of R9-GFP (29.25 kDa) are indicated by a black arrow. Interestingly, a distinct extra-band is visible only in the induced cells containing the higher PCN mutant plasmids, i.e., plasmids pY18_79-R9-gfp, pB15_79-R9-gfp, and B40_79-R9-gfp (lanes 6, 9, and 10, respectively, in Fig. 2.3.3.1 b). The lower production yield of the recombinant protein driven by the original plasmid and by the other mutants is consistent with the lower gene dosage of the recombinant cassette due to the lower PCN. Since mutations introduced in the replication origin may affect the plasmid segregational stability, we compared the loss rate over 80 generations of three vectors (pX1_79-R9-gfp, pA31_79-R9-gfp, and pY18_79-R9-gfp) with that of the progenitor under non-selective conditions (Fig. 2.3.3.1 c). These mutant vectors were selected because their PCN ranged from 20 (pX1_79-R9-gfp) to approximately 116 (pY18_79-R9-gfp). Results shown in Figure 3.3.1c indicate that the higher copy number plasmids (pY18_79-R9-gfp) have higher stability compared to the progenitor. The other two mutant plasmids, pX1_79-R9-gfp and pA31_79-R9-gfp, showed an average disappearance rate of resistant cells of 30% every 30 generations. Notably, vectors pX1_79-R9-gfp and pA31_79-R9-gfp exhibited stability of over 40% up to 45 generations, but this decreased to 10% after 80 generations.

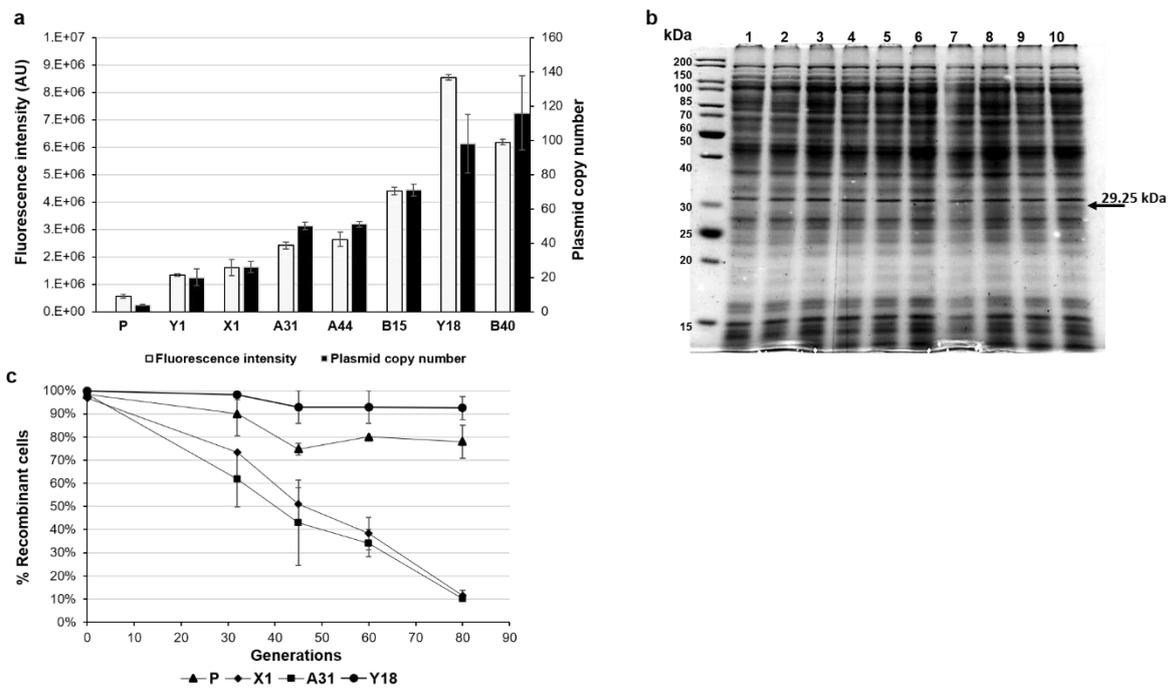


Fig. 2.3.3.1 Characterization of the seven selected psychrophilic clones carrying mutant OriR sequences. (a) Analysis of the R9-GFP expression by spectrofluorimetry (open bars) and determination of plasmid copy number by qPCR (solid bars) of the progenitor (P) and selected clones (X1, Y1, A31, A44, B15, B40, Y18) are plotted on bar graph. The fluorescence of the progenitor and the clones indicated on the x-axis was monitored on induced cells (5 mM IPTG) after 24 h from induction. Fluorescence intensities are reported in arbitrary units (AU) as mean \pm SD. (b) SDS-PAGE analysis of cell extracts of KrPL strains producing R9-GFP after 24 h from induction. lane 1, non-induced progenitor; lane 2, non-induced Y18; lane 3, induced progenitor; lane 4, induced X1; lane 5, induced Y1; lane 6, induced Y18; lane 7, induced A31; lane 8, induced A44; lane 9, induced B15; lane 10, induced B40. Black arrows on the right of the gel represent the expected molecular weights of the recombinant proteins. Black arrows inside the gel highlight the bands of the R9-gfp. (c) Plasmid stability assay. Plasmids pMA179-R9-gfp (P, \blacktriangle), pX1-R9-gfp (\blacklozenge), pA31-R9-gfp (\blacksquare), pY18-R9-gfp (\bullet) were separately propagated in KrPL for 80 generations without antibiotic selection. Plasmid stability was determined by replica plating onto selective media and presented as a percentage of cells that retain antibiotic resistance. Each experiment was carried out as biological duplicates and the error bars represent standard deviations.

2.3.4 Sequence analysis of the isolated mutant plasmids and deletion mutations affecting the replication sequence

The nucleotide sequences of the OriR mutants in the seven selected clones (X1, Y1, Y18, A31, A44, B15, and B40) were determined and compared with the wild type from pMtBL. The summary of the detected substitutions is listed in Table 2.3.4.1. Interestingly, we found two clones (X1 and Y1) in which a single substitution was responsible for the increase in the copy number of about 20 (T2399G and G2492T, respectively). Analysis of the sequencing data highlighted a nonhomogeneous distribution of mutations along the entire region (Fig. 2.3.4.1). In fact, all the mutations are located in the last four hundred nucleotides of the OriR sequence, except for one of the eight found in the A31 clone. DNA/RNA secondary structures can play a crucial role in regulating the plasmid copy number²⁶ and point mutations in these structures may affect the affinity of the interaction and thus alter regulatory mechanisms²⁶. Therefore, the Mfold WebServer tool was used to predict the DNA/RNA secondary structures that occur

in the second half of the pMtBL OriR sequence. The software identified several putative secondary structures throughout the entire sequence, but our attention was focused on those in which the single nucleotide substitution found in the X1 and Y1 clones occurs. In both cases, the substitution affected the complementarity of the putative stem forming sequences of STEM-LOOP I (position T2399G in the X1 clone) and STEM-LOOP II (position G2492T in the Y1 clone). To collect evidence on the potential destabilizing effect of the nucleotide substitutions on the suggested secondary structures, their minimum free energy value was computed and turned out to be higher than the value computed for the wild type pMtBL OriR (Table S2). Furthermore, many other nucleotide substitutions (found in the other selected clones) mapped in the same complementary sequences forming the stem of the two putative stem-loop structures. In all the cases, the mutations induced an increase in the calculated ΔG values of the corresponding secondary structures. To test whether the secondary structures contained herein are a necessary feature for the replication mechanism two OriR deletion mutants, i.e., devoid of the two putative stem-loop, was generated. The deleted origin turned out to be unable to support the plasmid replication in the Antarctic cells, suggesting that the deleted DNA region is essential for the mechanism of replication. As for the almost “invariant” OriR sequence (from nucleotide 1918 to 2259), an *in silico* prediction of open reading frames (ORFs) distribution was performed. At least 13 putative ORFs were found, that could potentially encode polypeptides between 10 and 66 amino acids. Unfortunately, a BLAST search revealed that none of them displays a worth mentioning homology with any other entry in publicly available databases. To understand whether this region is required for pMtBL replication, a pMAI79-*R9-gfp* derivative carrying an OriR mutant (devoid of the sequence from nucleotide 1918 to 2259) was constructed. Also, in this case, this specific deletion in pMtBL OriR resulted in a sequence unable to sustain the recombinant plasmid replication.

Clone	PCN	Mutations							
X1	20±4.99	T2399G							
Y1	26±3.29	G2492T							
A44	50±2.32	G2405A	A2472G						
A31	51±1.54	T2051A	A2348G	G2452A	C2496T	T2563C	A2607C	A2698G	T2712C
B15	71±3.45	T2446G	C2505A	A2608T	T2653A	T2660C	G2753T		
B40	98±17.13	G2428A	G2506T	A2556T	T2569C	A2625G	G2684A		
Y18	116±21.69	T2260C	C2491A	T2517A	T2544G	C2602A			

Table 2.3.4.1 Summary of nucleotide substitutions found in the selected OriR mutants. Nucleotide numbers are in agreement with the pMtBL sequence numbering (NZ_AJ224742.1).

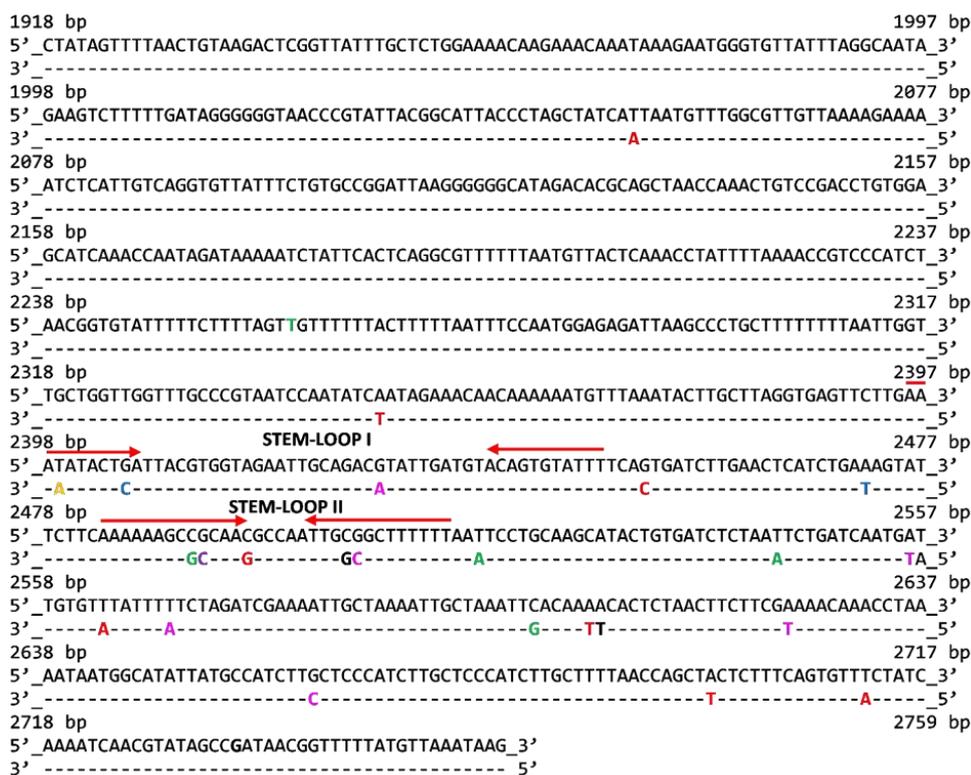


Fig. 2.3.4.1 Localization of nucleotide substitutions in the selected mutants OriR sequences. The nucleotide substitution/s found in each mutant OriR is reported below the corresponding wild type position using the following color code: yellow indicates substitution in clone X1; purple for clone Y1; green for Y18; red for A31; blue for A44; black for B15; pink for B40. The complementary sequences representing the stems of two putative stem-loop structures (STEM-LOOP I and STEM-LOOP II) are indicated with red arrows.

4 Discussion

The production of recombinant proteins is essential in industrial biotechnology, and the efficacy of this process is influenced by several parameters such as host features, process temperature, promoter strength, and the number of copies of the expression vector used²⁷. This latter aspect has a strong impact on the productivity of the system as a high plasmid copy number (PCN) enables the achievement of the gene dosage required for efficient recombinant protein accumulation. In bacteria, large differences in the PCN are associated with different replication origins²⁸. For example, pMB1 is a medium-copy number episome (15–20 copies/cell)²⁹, but some pMB1 derivatives, such as pUC, produce up to 700 copies/cell²⁹. In the case of the plasmid pMtBL⁸, isolated from the Antarctic marine bacterium *PhTAC125*, the copy number is very low¹⁸, and the PCN of pMtBL-derived vectors may limit the full exploitation of this bacterium as an unconventional host for recombinant protein production. Therefore, to improve the performance of the psychrophilic expression system, we decided to modify the origin of replication of pMtBL-based vectors to obtain a new generation of plasmids characterized by a higher copy number. Since the replication mechanism of pMtBL is unknown, a random mutagenesis strategy appeared to be the most suitable approach to select more efficient origins of replication. A library of plasmids containing random mutations in OriR was constructed and analyzed in the KrPL strain. The collection of recombinant psychrophilic cells was

then sorted by a FACS-based high-throughput screening method (HTS). Preliminary FACS analysis of the progenitor cells, i.e., KrPL cells carrying the wild-type pMAI79-*R9-gfp* plasmid, revealed a remarkable heterogeneity within this genetically homogeneous population. Similar heterogeneity has been previously observed in cells harboring low-copy number plasmids³⁰. As plasmids follow a discrete distribution, a low mean PCN may also lead to higher cell-to-cell heterogeneity in PCN and gene expression. Kittleson and colleagues demonstrated a clear relationship between the PCN and the degree of heterogeneity in plasmid encoded gene expression³¹. Therefore, the observed progenitor heterogeneity was expected and prompted us to use a screening strategy that combined the FACS sorting with a subsequent screening step based on determining the fluorescence intensity of some isolated clones. By applying this approach several times, we isolated seven recombinant clones from different sorting steps characterized by psychrophilic plasmids maintained at different copy numbers, ranging from 20 to more than 110 (Fig. 2.3.3.1 a). The ability to use genetic systems with different PCNs is advantageous under certain experimental conditions when tighter control of gene expression is required, and low-to-medium copy number plasmids may offer many more advantages compared with high copy number plasmids³². Furthermore, KrPL cells carrying the selected plasmids drive R9-GFP production to varying degrees, up to 15-fold higher than that observed in the progenitor (Fig. 2.3.3.1 a), without showing any signal of metabolic stress (Fig. S3). Metabolic load generally correlates with plasmid copy number and is associated with cellular resource utilization for plasmid replication and recombinant protein production³³⁻³⁵. It's interesting to note that the recombinant production of human cyclin-dependent kinase-like 5 (*hCDKL5*) in *PhTAC125*³⁶, obtained thanks to a vector containing the B40 OriR, resulted in a metabolic burden confirming that this effect is not related to the plasmid features but it is due to the recombinant protein characteristics. Another critical aspect, associated to high copy number plasmids, is their intrinsic genetic instability. We examined the maintenance of the plasmid carrying the mutated OriR sequences over several generations. The plasmid stability assay (Fig. 2.3.3.1 c) showed that the plasmid with a higher copy number (pY18_79-*R9-gfp*) had higher stability than the progenitor. In contrast, plasmids (pA31_79-*R9-gfp* and pX1_79-*R9-gfp*) with a medium-copy number showed lower stability, although about 50% of plasmid-containing cells are still present after 40 generations. The segregation of low-copy number plasmids is usually due to partition systems that actively distribute plasmid copies into the daughter cells. In the case of pMtBL, the occurrence of a potential partitioning *parABS* system was recently demonstrated¹⁸. Encoded by a plasmid sequence outside the OriR region, this system cannot contribute to the segregation stability of OriR-derived plasmids in KrPL cells¹⁸. In the absence of a partitioning system, successful plasmid segregation depends on the multi-copy state and the physical distribution of the plasmids inside the cell^{37,38}. Therefore, higher stability of pY18_79-*R9-gfp* was expected while in the case of pA31_79-*R9-gfp* and pX1_79-*R9-gfp* the lower stability compared to the progenitor might be related to the specific mutation(s) present in these plasmids and/or in the formations of plasmid multimers³⁹ (data not shown). Indeed, the formation of plasmid multimers may lead to a reduction in the number of heritable units during cell division, hence reducing the chance of successful segregation and

consequently also the plasmid persistence over time⁴⁰. To obtain further information on the actual mechanism of pMtBL replication initiation and copy number regulation, the DNA sequence of the seven selected mutated OriRs was determined, and the identified base substitutions are listed in Table 2.3.4.1, where the experimental PCN values are also given. Interestingly, two of the seven analyzed clones are characterized by a single point mutation (X1 mutation in yellow, Y1 mutation in purple in Fig. 2.3.4.1), each occurring in the stem forming sequence of a predicted stem-loop region. These results pointed the attention to these regions (indicated as inverted arrows in Fig. 2.3.4.1), and the observation that other clones, such as Y18 and B40, had mutations in these regions strengthened the hypothesis that the predicted DNA secondary structures might be involved in PCN control. consistent with these considerations, our *in silico* analysis showed that the mutations that occurred destabilize the predicted secondary structures (Table S2). Moreover, almost all the recorded mutations (except for one in the A31 clone) map in the last 350 nucleotides of OriR (region 2398-2712). The observed uneven distribution of the mutations along the OriR sequence from one side highlights the occurrence of a region of hot spot mutations, likely critical for the control of plasmid copy number, but also that the rest of the sequence does not allow the occurrence of any point mutation. Determining the functional role of this region (from 1918 bp to 2259 bp) and of the pMtBL replication mechanism is beyond the scope of this work, but what can be speculated is that the conservation of the original nucleotide sequence may underscore either the presence of small open reading frames (coding for small essential proteins) or the occurrence of functional sequences essential for the start or the progression of the plasmid replication. In light of this, the deletion of the first 342 bp of OriR abolished the ability of the remaining fragment to drive the psychrophilic plasmid replication. However, a transcriptomic analysis of the pMtBL plasmid showed that almost the whole plasmid is transcribed, with a higher than the average transcription involving the OriR region (1918-2759 bp) (Fig. S4). Interestingly, both the DNA strands are transcribed, although with different intensities, suggesting the occurrence of a replication mechanism based on complementary antisense RNA sequences, which may resemble what happens in the ColE1 plasmid. More focused work will be needed to describe the exact mechanism of pMtBL replication in detail. The mutant OriR sequences studied in this work have several applications, fostering the development of the *PhTAC125* genetics and its industrial exploitation. Indeed, the availability of mutant OriR sequences, ensuring the establishment of higher copy number expression vectors, was essential to setting up the first conditional gene silencing system in a wild-type Antarctic bacterium⁴¹. The transcription efficiency obtained by combining the IPTG-inducible promoter¹² with the B40 PCN allowed establishing a sufficient paired termini antisense RNA concentration to switch off the expression of a medium transcribed gene. This tool paves the way for the functional study of *PhTAC125* essential genes. Furthermore, the implementation of Ori B40 in the pP79 vector proved to be quite effective also in achieving very high recombinant production yields of a difficult-to-express human protein, the *hCDKL5* kinase¹⁷, whose production process was efficiently modeled by an *ad hoc* evolved genome-scale metabolic model³⁶. All the achieved results reported

so far demonstrate the potential of this psychrophilic microorganism as a novel host for recombinant protein production.

Fundings

This research was partially funded by the Italian “Programma Nazionale di Ricerca in Antartide” (PNRA project number PNRA18_00335) and by the Italian Parents’ Association “La fabbrica dei sogni 2 - New developments for Rett syndrome APS”.

Availability of data and materials: Correspondence and material requests should be addressed to MLT (tutino@unina.it) and MC (marzia.calvanese@unina.it).

Competing interests. Authors declare no competing interests.

Acknowledgements

Authors’ Contribution

M.C., A.C., E.P., and M.L.T. designed the project. M.C. conducted the selection experiments and all the molecular characterization of the high copy number plasmids and generated all the OriR mutants. C.L. set up the bacterial growth and induction conditions. A.C. developed the electroporation protocol for *Pseudoalteromonas haloplanktis* TAC125. C.B. carried out the sorting experiments. C.R and M.F. carried out the bioinformatics analysis. M.C., E.P, C.R., M.F., and M.L.T. collected and interpreted the data, and wrote the manuscript. All the authors revised the manuscript and approved the content.

References

1. Tripathi, N. K. & Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front. Bioeng. Biotechnol.* 7, (2019).
2. Westers, L., Westers, H. & Quax, W. J. *Bacillus subtilis* as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta - Mol. Cell Res.* 1694, 299–310 (2004).
3. Mühlmann, M., Forsten, E., Noack, S. & Büchs, J. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microb. Cell Fact.* 16, 1–12 (2017).
4. Adrio, J. L. & Demain, A. L. Microbial enzymes: tools for biotechnological processes. *Biomolecules* 4, 117–139 (2014).
5. Francis, D. M. & Page, R. Strategies to optimize protein expression in *E. coli*. *Curr. Protoc. Protein Sci.* 1–29 (2010) doi:10.1002/0471140864.ps0524s61.
6. Medigue, C. *et al.* Coping with cold: The genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res.* 15, 1325–1335 (2005).

7. Parrilli, E., De Vizio, D., Cirulli, C. & Tutino, M. L. Development of an improved *Pseudoalteromonas haloplanktis* TAC125 strain for recombinant protein secretion at low temperature. *Microb. Cell Fact.* 7, 1–10 (2008).
8. Tutino, M. L. *et al.* A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles* 5, 257–264 (2001).
9. Duilio, A. *et al.* Promoters from a cold-adapted bacterium: definition of a consensus motif and molecular characterization of UP regulative elements. *Extremophiles* 8, 125–32 (2004).
10. Papa, R., Rippa, V., Sannia, G., Marino, G. & Duilio, A. An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125. *J. Biotechnol.* 127, 199–210 (2007).
11. Sannino, F. *et al.* A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl. Microbiol. Biotechnol.* 101, 725–734 (2017).
12. Colarusso, A., Lauro, C., Calvanese, M., Parrilli, E. & Tutino, M. L. Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a cell factory: lptg-inducible plasmid construction and strain engineering. *Microorganisms* 8, 1–24 (2020).
13. Fondi, M. *et al.* Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Environ. Microbiol.* 17, 751–766 (2015).
14. Vigentini, I., Merico, A., Tutino, M. L., Compagno, C. & Marino, G. Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J. Biotechnol.* 127, 141–50 (2006).
15. Dragosits, M. *et al.* Influence of growth temperature on the production of antibody Fab fragments in different microbes: a host comparative analysis. *Biotechnol. Prog.* 27, 38–46 (2011).
16. Unzueta, U. *et al.* Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A. *Appl. Microbiol. Biotechnol.* 99, 5863–5874 (2015).
17. Calvanese, M., Colarusso, A., Lauro, C., Parrilli, E. & Tutino, M. L. Soluble Recombinant Protein Production in. 2406, 219–232 (2022).
18. Qi, W. *et al.* New insights on *Pseudoalteromonas haloplanktis* TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies. *Sci. Rep.* 9, 1–14 (2019).
19. Kurusu, Y. *et al.* Genetic transformation system for a psychrotrophic deep-sea bacterium: Isolation and characterization of a psychrotrophic plasmid. *Mar. Biotechnol.* 3, 96–99 (2001).

20. Škulj, M. *et al.* Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. *Microb. Cell Fact.* 7, 1–12 (2008).
21. Wheeler, D. L. *et al.* Database resources of the national center for biotechnology. *Nucleic Acids Res.* 31, 28–33 (2003).
22. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415 (2003).
23. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890 (2018).
24. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).
25. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
26. Camps, M. Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Patents DNA Gene Seq.* 4, 58–73 (2010).
27. Lozano Terol, G., Gallego-Jara, J., Sola Martínez, R. A., Cánovas Díaz, M. & De Diego Puente, T. Engineering protein production by rationally choosing a carbon and nitrogen source using *E. Coli* BL21 acetate metabolism knockout strains. *Microb. Cell Fact.* 18, 1–19 (2019).
28. Kim, J. W., Bugata, V., Cortés-Cortés, G., Quevedo-Martínez, G. & Camps, M. Mechanisms of Theta Plasmid Replication in Enterobacteria and Implications for Adaptation to Its Host. *EcoSal Plus* 9, 1–35 (2020).
29. Rossi, M., Brigidi, P., Gonzalez Vara y Rodriguez, A. & Matteuzzi, D. Characterization of the plasmid pMB1 from *Bifidobacterium longum* and its use for shuttle vector construction. *Res. Microbiol.* 147, 133–143 (1996).
30. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. & Müller, S. Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. *Microb. Cell Fact.* 15, 1–12 (2016).
31. Kittleson, J. T., Cheung, S. & Anderson, J. C. Rapid optimization of gene dosage in *E. coli* using DIAL strains. *J. Biol. Eng.* 5, 1–7 (2011).
32. Carrier, T., Jones, K. L. & Keasling, J. D. mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. *Biotechnol. Bioeng.* 59, 666–672 (1998).
33. Tan, J. *et al.* Independent component analysis of *E. coli*'s transcriptome reveals the cellular processes that respond to heterologous gene expression. *Metab. Eng.* 61, 360–368 (2020).
34. Li, Z. & Rinas, U. Recombinant protein production-associated metabolic burden reflects anabolic constraints and reveals similarities to a carbon overfeeding response. *Biotechnol. Bioeng.* 118, 94–105 (2021).
35. Silva, F., Queiroz, J. A. & Domingues, F. C. Evaluating metabolic stress and

- plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnol. Adv.* 30, 691–708 (2012).
36. Fondi, M. *et al.* Modelling hCDKL5 heterologous expression in bacteria. *Metabolites* 11, (2021).
 37. Wang, Y., Penkul, P. & Milstein, J. N. Quantitative Localization Microscopy Reveals a Novel Organization of a High-Copy Number Plasmid. *Biophys. J.* 111, 467–479 (2016).
 38. Reyes-Lamothe, R. *et al.* High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. *Nucleic Acids Res.* 42, 1042–1051 (2014).
 39. Bedbrook, J. R. & Ausubel, F. M. Recombination between bacterial plasmids leading to the formation of plasmid multimers. *Cell* 9, 707–716 (1976).
 40. Summers, D. Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol. Microbiol.* 29, 1137–1145 (1998).
 41. Lauro, C., Colarusso, A., Calvanese, M., Parrilli, E. & Tutino, M. L. Conditional gene silencing in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Res. Microbiol.* 103939 (2022) doi:10.1016/j.resmic.2022.103939.

Supplementary materials

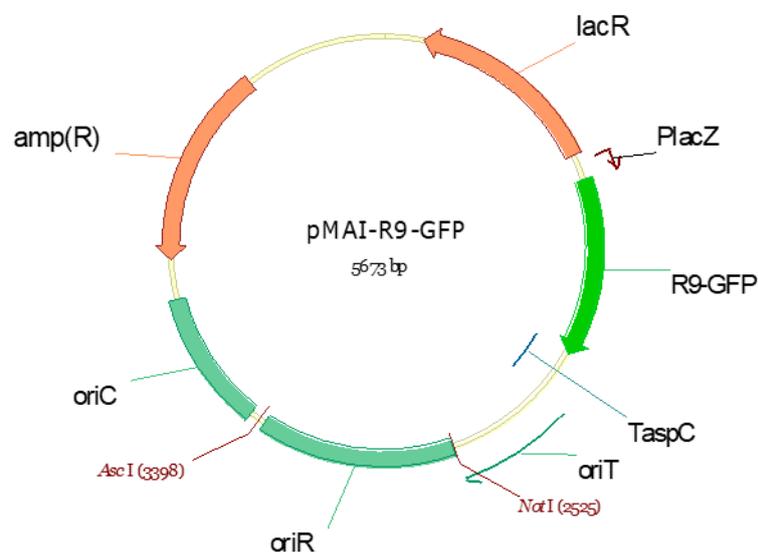


Fig. S1 Map of the pMAI79-*R9-gfp* plasmid. The two unique restriction sites (NotI and AscI) introduced upstream and downstream of the OriR sequence for the construction of the pMAI79-*R9-gfp* vector are highlighted in red. The map illustrates the genetic structure of the plasmid backbone including the pMtBL-derived replication origin for the maintenance in *PhTAC125* (OriR), the promoter (*PlacZ*), and the regulatory sequence (*PhlacR*) of lac operon recovered from the *PhTAE79*, the transcriptional terminator of the *PhTAC125 aspC* gene (*TaspC*), the conjugational DNA transfer origin (OriT), the β -lactamase encoding gene (*amp^R*) and the pUC18-derived replication origin for the propagation in *E. coli* (OriC).

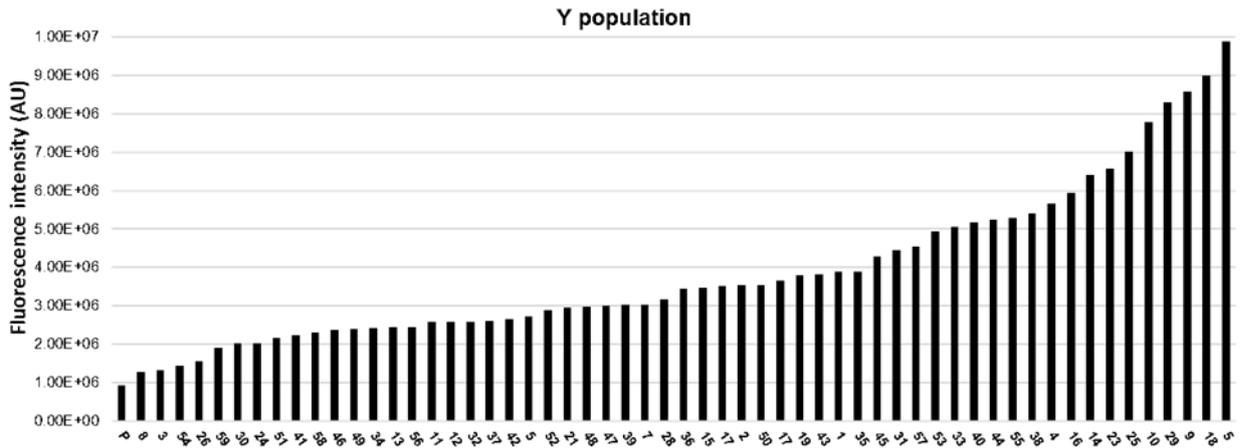


Fig. S2 Analysis of the R9-GFP expression by spectrofluorimetry. 60 clones of the Y population were randomly selected, cultivated in 96 deep well plates filled with the GG medium at 15°C, and harvested after 24 h from induction. The fluorescence of the progenitor (P) and the clones indicated on the x-axis are reported in arbitrary units (AU) as mean \pm SD, n = 3.

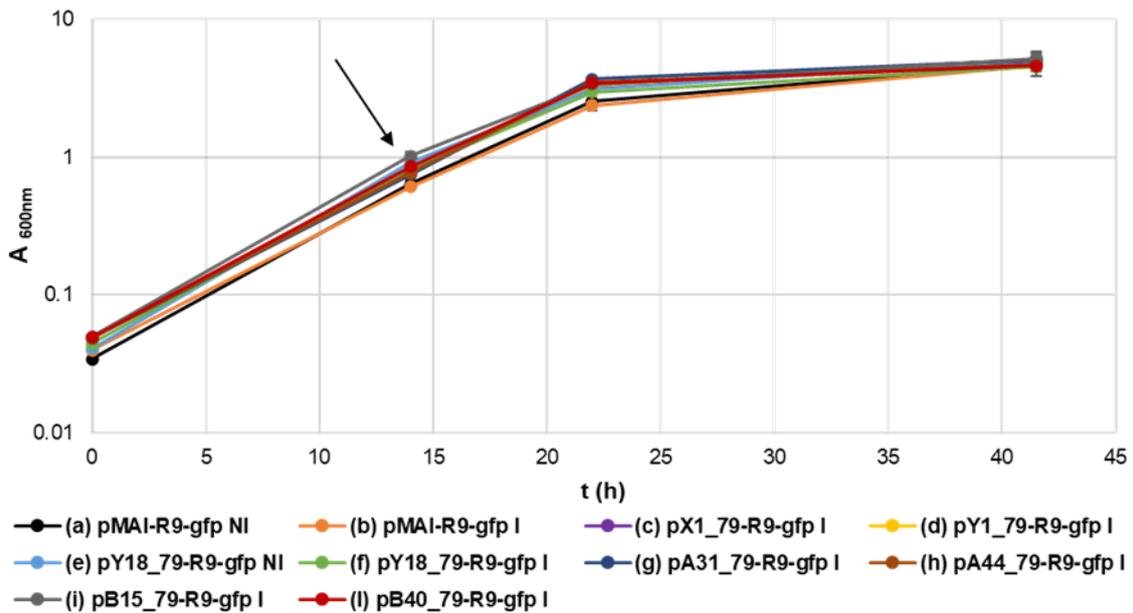


Fig. S3 Growth curves of KrPL harboring the original plasmid (*pMAI-R9-gfp*) and the high-copy plasmids. The recombinant cells were grown in the GG medium at 15 °C. Growth curves of non-induced cells harboring *pMAI-R9-gfp* (a) and *pY18_79-R9-gfp* (e). Growth curves of induced cells (5 mM IPTG) harboring *pMAI-R9-gfp* (b), *pX1_79-R9-gfp* (c), *pY1_79-R9-gfp* (d), *pY18_79-R9-gfp* (f), *pA31_79-R9-gfp* (g), *pA44_79-R9-gfp* (h), *pB15_79-R9-gfp* (i), and *pB40_79-R9-gfp* (l). The moment of the induction is represented with a black arrow.

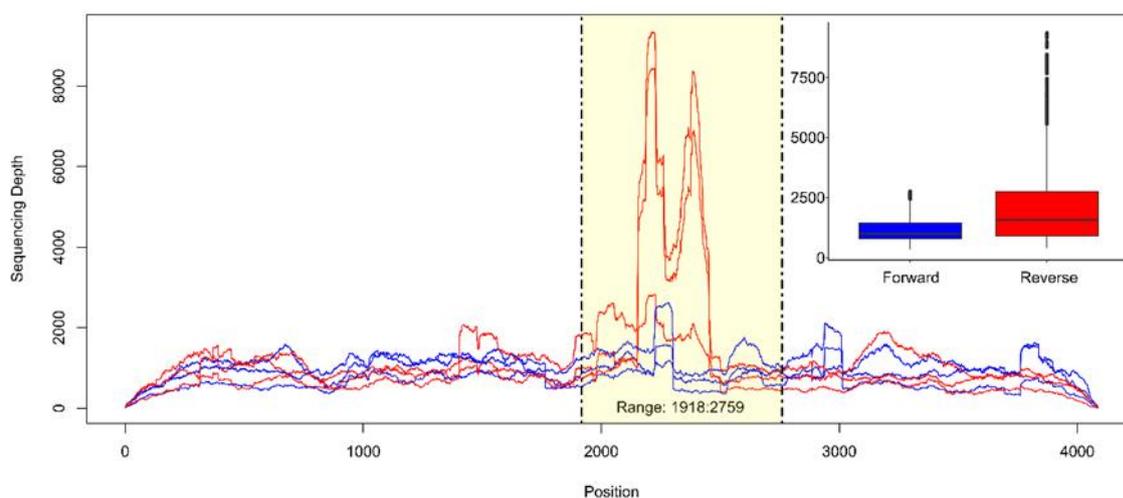


Fig. S4 Display of RNA-Seq data. Data from pMtBL transcriptome, the total coverage is displayed as a plot showing nucleotide-level sequence depth for the forward (blue) and reverse strand (red). The highlight focuses on OriR region, which stretches from nucleotides 1918 to 2759. Boxplot in the inset shows the quartiles of total forward (blue, mean: 1008.218, SD: 420.3342) and total reverse (red, mean: 2242.505, SD: 2123.815) strand sequencing depths for the highlighted region.

Strain	Relevant characteristics	References or Source
<i>E. coli</i> strains		
<i>DH5α</i>	[supE44, Δ lacU169 (ϕ 80 lacZ Δ M15) hsdR17, recA1, endA1, gyrA96, thi-1, relA1]	Lab stock
S17-1 (<i>λpir</i>)	thi, pro, hsd(r- m+) recA::RP4-2-TCr::Mu Kmr::Tn7 Tpr Smr λ pir	Tascon et al. 1993
XL10-Gold Ultracompetent Cells	TetR Δ (mcrA)183 Δ (mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacIqZ Δ M15 Tn10 (TetR) Amy CamR]	Lab stock
<i>PhTAC125</i> strains		
<i>PhTAC125 wt</i>	Possesses two endogenous plasmids	Médigue et al. 2005
KrPL	<i>PhTAC125</i> cured strain without pMtBL plasmid	Lab stock (Unpublished data)
Plasmid		
pP79- <i>R9-gfp</i>	Expression vector with <i>PhTAE79 lacZ</i> regulative sequences producing enhanced GFP	Colarusso et al. 2020
pMAI79- <i>R9-gfp</i>	pP79- <i>R9-gfp</i> vector, introducing two unique restriction sites, <i>NotI</i> and <i>AscI</i>	This work
Primer name	Sequence (5' - 3')	
Multiple-site mutagenesis		
Left OriR NotI Fw	CGGAAAAGATCCGTCGATTGCGGCCGCCTTATTTAACATAAAAAAC	

Right OriR Ascl Rv	TCAATTCATGTGAGCAAAAAGGCGCGCCAAAGGCCAGGAACCGTAA
Random mutagenesis	
mut pP79-R9-gfp Fw	GAAAAGATCCGTCGATTGCGGCCGC
mut pP79-R9-gfp Rv	GTTCTGGCCTTTGGCGCGCCTTT
Quantitative PCR	
<i>PSHAa2051</i> Fw	AACCGCACACAGACCCGAA
<i>PSHAa2051</i> Rv	AACGCACATTGGCATGACTGG
<i>R9-gfp</i> Fw	GGAGAGGGTGAAGGTGATGCT
<i>R9-gfp</i> Rv	GGTCAGAGTAGTGACAAGTGTGG
Deletions	
OriR_BsiWI Fw	GGCCGTACGGAAGAATACTTTTCAGATGAGTT
OriR_BsiWI Rv	CGGCGTACGAATTCCTGCAAGCATACT
OriR_NsiI Fw	ggcCGTACGgaagaatactttcagatgagtt
OriR_NsiI Rv	cggCGTACGaattcctgcaagcatact
OriR_NotI Fw	CTCGCAGAGCAGGATTCCCGTTGAG
OriR_AscI Rv	CGCGGCGCGCCTGTTTTTACTTTTTTAA

Table S1 List of bacterial strains, plasmids, and primers used in this study.

Sample	RNA structure		DNA structure	
	Stem-loop I Free energy minimum (kcal/mol)	Stem-loop II Free energy minimum (kcal/mol)	Stem-loop I Free energy minimum (kcal/mol)	Stem-loop II Free energy minimum (kcal/mol)
wild-type	-4.12	-17.45	-3.50	-14.67
X1	-3.81	/	-1.60	/
Y1	/	-11.66	/	-9.24
Y18	/	-11.47	/	-10.52
A31	/	-19.57	/	-16.05
A44	-2.57	/	-1.75	/
B15	-3.08	-11.77	-2.92	-10.40
B40	-2.69	-11.26	-2.07	-9.64

Table S2 Free energy values (kcal/mol) of RNA and DNA secondary structures in the selected mutants compared to wild-type OriR were estimated with the Mfold tool

Chapter 3: Study of condensation in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 producing a recombinant intrinsically disordered protein, the case of *hCDKL5*

Recently, it has been demonstrated that *PhTAC125* is the only prokaryotic host able to produce the full-length of human cyclin-dependent kinase-like 5 (*hCDKL5*). *hCDKL5* is a serine/threonine-protein kinase, which is mainly a disordered protein, except for its N-terminal catalytic domain. Mutations in its coding sequence are involved in CDKL5 Deficiency Disorder (CDD), a severe neurodevelopmental pathology. An integrated approach of plasmid design and strain engineering in *PhTAC125* was set up to overcome critical factors that might alter the expression of the full length of the *hCDKL5*. Despite the successful recombinant production of the *hCDKL5* protein in *PhTAC125*, its expression still highlights some bottlenecks related to its solubility. One possible hypothesis is that the apparent insolubility of *hCDKL5* could be due to the formation of condensates via liquid-liquid phase separation (LLPS). The condensation properties of this protein were investigated *in vivo* during the recombinant expression of a fluorescent protein fused to the N-terminal of *hCDKL5* into *PhTAC125*. Furthermore, several rationally designed *hCDKL5* mutants were constructed to uncover the contribution of each domain to its condensation. The findings demonstrated that the *hCDKL5* forms intracellular clusters in the psychrophilic bacterium, and a region of the catalytic domain drives the protein assembly into condensates. Further studies will be necessary to understand the relationship between the structure and dynamics of *hCDKL5* clusters.

3.1 Introduction

3.1.1 The intrinsically disordered proteins (IDPs) and their recruitment into protein condensates

Until recently, the well-established structure-function paradigm in structural biology studies (a protein requires a native folded structure to perform its biological function) was subjected to an intensive reassessment. These observations resulted in the recognition of biologically active proteins without a stable three-dimensional structure under physiological conditions, named intrinsically disordered proteins (IDPs)^{1–3}. In some cases, only some regions in the protein are disordered, and they are known as intrinsically disordered regions (IDRs)^{3,4}. Such IDPs constitute about 40% of the human proteome and are involved in many crucial cellular processes, such as signal transduction, transcription, translation, and macromolecular transport^{5–7}. As a consequence, mutations and dysregulation of IDPs are implicated in various devastating diseases^{1,6}. These proteins or relatively long regions (>30 residues) lack well-defined folding and typically do not exhibit stable secondary structures^{8,9}. These structural features make them dynamic and flexible, allowing them to adopt distinct conformations upon binding to different partners or based on environmental conditions^{10,11}. The IDPs usually interact with partner molecules (e.g., DNA, RNA, and proteins) through short linear motifs (SLiMs) or motif recognition features (MoRFs)^{6,12}. Many IDPs present multiple SLiMs or MoRFs that function synergistically to enable multivalent interactions with their molecular partners^{6,7}. This multispecific and multivalent property of IDPs would be linked to their disorderly and flexible structure, as well as the ability to involve partners that are already occupied by other ligands. However, poor conservation of these short recognition sequences makes their exact identification challenging^{13,14}. This is an aspect of considerable interest because there is growing evidence that interactions between IDPs/IDRs and macromolecular binding partners are attractive targets for drug development^{7,15,16}.

Recent investigations suggest that multivalent interactions between IDPs and their protein or nucleic acid partners drive protein clustering and subsequent condensation, forming various membrane-less organelles in cells (MLOs)¹⁷. For many years, the field of Liquid–liquid phase separation (LLPS) was limited to the study of functionally and structurally distinct MLOs in eukaryotic cells. However, with recent advances in super-resolution fluorescence imaging techniques, it has been proven the existence of MLOs in bacteria¹⁸. These organelles lack a membrane barrier allowing them to tune their composition, size, and structure, exhibit dynamic diffusivity inside, and undergo continuous exchange with the outside in response to environmental alterations¹⁹. Furthermore, condensates form in a concentration-dependent manner on the molecular components, when these macromolecules reach a saturation threshold¹⁹. A growing understanding of the molecular mechanisms and the physicochemical principles that drive the formation of organelles have provided evidence of their diverse functions in cellular

processes, including the formation of classical membranellar organelles, signaling complexes, the cytoskeleton, stress response, the regulation of gene expression, and the control of signal transduction^{20–22}. However, the biochemical functions and assembly mechanisms of these compartments have not yet been completely characterized. In the last few years, there has been increasing evidence that mutations of key residues in SLiMs/MoRFs, protein aggregation, and the dysregulation of MLOs are associated with human diseases, such as age-related disorders, and amyotrophic lateral sclerosis (ALS)^{19,23–26}. Understanding the physical principles and molecular interactions behind protein phase separation, how the organization of the different components into membrane-less organelles contributes to function, and how their dysregulation leads to disease, could inspire novel therapeutic strategies^{6,7,15,16}.

3.1.2 The study of the human cyclin-dependent kinase-like 5 (*hCDKL5*) containing an intrinsically disordered region (IDR)

Human Cyclin-dependent kinase-like 5 (*hCDKL5*) is a serine/threonine kinase, widely expressed in the human body with the highest abundance in the brain^{27,28}. Several mutations in its coding sequence often end up in the enzyme absence or the production of loss-of-function variants, and both conditions are causative of CDKL5 Deficiency Disorder (CDD; OMIM 300203; 300672), a rare and severe neurodevelopmental disorder for which no cure is available^{29,30}. The typical symptoms associated with CDD are early-onset epilepsy, intellectual disability, and cortical vision disorders³¹. Furthermore, the patients show a significant phenotypic variation, related to the type and location of the mutation³¹. CDD is more frequently developed in women since the gene encoding *hCDKL5* is located on the X chromosome (Xp22.3). Furthermore, the transcript of this gene is subjected to alternative splicing leading to the production of five isoforms²⁹. In particular, isoforms 1, 2, 3, and 4 are widely expressed in different organs of the human body with a different relative abundance (brain, heart, liver, lung, muscle, spleen, and kidney), rather isoform 5 is exclusively expressed in testis²⁹. All known *hCDKL5* isoforms contain a conserved N-terminal kinase domain and variable C-terminal³². The kinase domains of CDKL1, CDKL2, CDKL3, and CDKL5 were crystallized, except for the first 12 residues of the protein³². Furthermore, the structural studies carried out on the catalytic domain showed an unusual amphipathic helix, α J, at the C-terminal. The deletion of this region implied the reduction of the CDKL2 and CDKL3 activities, whereas CDKL1 and CDKL5 were unchanged³². The catalytic domain (aa 13–297) is characterized by an ATP-binding region (aa 19–43), a serine/threonine protein kinase active site (aa 131–143), and a Thr–Xaa–Tyr motif (TEY) (aa 169–171), involved in the autophosphorylation and consequent activation of the protein (Figure 3.1.2.1). The C-terminal is characterized by an extended unstructured region and contains two nuclear localization signals (NLS) (NLS1, aa 312–315; NLS2, aa 784–789) and a nuclear export signal (NES) (aa 836–845) (Figure 3.1.2.1)^{32,33}. Furthermore, the long disorder region contains multiple SLiMs determining the protein localization, stability, and degradation; these motifs are also

able to promote the binding with different partners and facilitate post-translational modifications³⁴. Previous studies *in vitro* have reported distinct subcellular localization of *hCDKL5* to the cytoplasm or nucleus of the neuronal cells during different stages of neuronal development³³. The molecular role of *hCDKL5* in the cytoplasm of neurons is related to the development of dendrites. In contrast, the protein in the nucleus is involved in RNA storage and modification, and splicing regulation³¹.

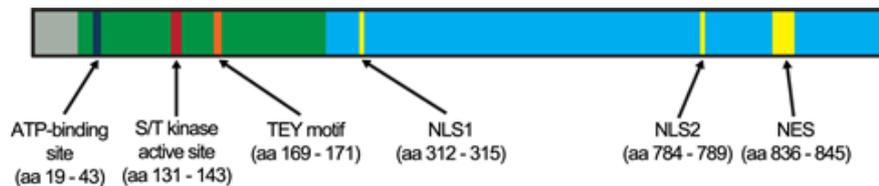


Figure 3.1.2.1 The N-terminus harbors the sequences relevant to the kinase catalytic function within the kinase domain (green), including the ATP-binding site, the S/T kinase active site, and a Thr–Glu–Tyr (TEY) motif. The C-terminus (blue) contains sequences involved in nuclear localization, including NLS (nuclear localization signal) and NES (nuclear export signal) sequences. Figure was adapted from Van Bergen *et. al*,³⁵.

3.1.3 An integrated approach of plasmid design and strain engineering finalized to improve the *hCDKL5* production in *Pseudoalteromonas haloplanktis TAC125*

Protein replacement therapy could be an alternative therapeutic approach to gene therapy. Therefore, this approach needs the large-scale recombinant production of the interest protein. The heterologous expression and purification of the full-length *hCDKL5* are considered challenging for several reasons: (i) two-thirds of its sequence is intrinsically disordered, except for its N-terminal catalytic domain; (ii) the lack of a precise 3D structure makes this region more susceptible to proteolytic attack by host-encoded proteases. Indeed, *hCDKL5* is accumulated in inclusion bodies in an insoluble and inactive form, when it is expressed in the most common prokaryotic host, *E. coli*³⁶. On the other hand, the production of this kinase enzyme is limited in mammalian cells by its toxicity in the cytoplasm and undesired post-translational modifications. Recently, it is demonstrated that *PhTAC125* is the only prokaryotic host able to produce the full length of *hCDKL5*^{37,38}. An integrated approach of plasmid design and strain engineering in *PhTAC125* was set up to overcome critical factors that might alter the expression of the full length of the *hCDKL5* 1 isoform. Preliminary experiments demonstrated a low translation efficiency and stability of the *hCDKL5* transcript. The transcript levels were triplicated after the expression of the protein in an IPTG-inducible bicistronic plasmid³⁸. The expression in this new system led to the identification of an internal initiation translation site at the Methionine 43. The integral form of the protein was restored through the mutagenic replacement of this residue by Valine, obtaining a new construct, named CDKL5-M43V. The stability of the transcript was ulteriorly improved using the SUMO tag at the N-terminal of the construct. Surprisingly, the introduction of the SUMO tag protected the N-terminal domain from proteolytic

degradation³⁸. Furthermore, the *hCDKL5* yield was significantly increased by the replacement of the replication origin of the plasmid with a new origin with a high copy number, named B40 (described in Chapter 2). Further improvements were obtained by engineering the strain through the genomic insertion of a functional copy of the *E. coli lacY* gene into the *PhTAC125 Lon* encoding gene to overcome the lack of an efficient and active lactose import system in the psychrophilic bacterium³⁸. The resulting mutant (KrPL *LacY*⁺) has two desired features as an improved host for recombinant protein production, i.e., a faster inducer internalization rate, thanks to the mesophilic lactose permease, and the absence of the *Lon* protease activity, which is generally recognized as the main protease involved in recombinant products degradation. Finally, the preliminary purification trials of the full-length of *hCDKL5* allowed obtaining a fraction of the soluble protein to set up an *in vitro* activity assay which demonstrated the capability of the recombinant *hCDKL5* to phosphorylate its substrate EB2³⁸. The overall efforts succeeded in the construction of an unconventional cell factory exploited for the expression of a hybrid unstructured/structured protein, *hCDKL5*, with a high production yield in an active form.

Despite the successful recombinant production of *hCDKL5* in *PhTAC125*, the *hCDKL5* expression still highlights some limiting factors related to its solubility. These issues arose with the increment of production levels but at the expense of the solubility of the recombinant kinase³⁸. Furthermore, some peculiar features were observed during several attempts at *hCDKL5* recovery in the soluble form: (i) the addition of the fusion tags doesn't enhance the *hCDKL5* solubility; (ii) the insolubility emerges when the intracellular protein concentration raises; (iii) the use of trehalose in the lysis buffer improves the solubility of *hCDKL5* promoting dissolution of existing "aggregates". These phenomena make us think that the apparent insolubility of *hCDKL5* was not associated with the formation of inclusion bodies (insoluble protein aggregates stabilized via interactions between exposed hydrophobic surfaces and never observed in *PhTAC125*). Instead, the observed features of the insoluble fraction behave as a liquid-like condensate. In support of this hypothesis, several studies established that proteins that are enriched in these droplets tend to have extensive intrinsically disordered regions as in the case of *hCDKL5*, which is mainly a disordered protein, except for its N-terminal catalytic domain^{18,39,40}. Furthermore, it is believed that weak multivalent interactions between undefined motifs in IDRs and macromolecules (e.g., DNA, RNA, and proteins) generate condensates via liquid-liquid phase separation (LLPS)⁴¹. In this regard, it has been recently demonstrated that *hCDKL5* is recruited to sites of DNA damage in actively transcribed regions of the nucleus, interacting with poly(ADP-ribose) with its C-terminal extremity⁴².

To study whether the human intrinsically disordered protein, *hCDKL5*, can self-assemble into condensates when produced in KrPL *LacY*⁺, I spent a period during my Ph.D. in the research group of Dr. Lasker from the Scripps Research Institute (La Jolla, CA, US). Dr. Lasker's lab possesses the required expertise and uses high-resolution imaging, computer modeling, and cellular biology to study the structure and dynamics of bacterial condensates. The condensation properties of

the protein were investigated *in vivo* during the recombinant expression of a fluorescent protein fused to the N-terminal of *hCDKL5* into KrPL *LacY*⁺. Furthermore, several rationally designed *hCDKL5* mutants were constructed to uncover the contribution of each domain to its condensation. The findings demonstrated that the *hCDKL5* forms intracellular clusters in the psychrophilic bacterium and a region of the catalytic domain drives the protein assembly into condensates.

3.2 Materials and Method

3.2.1 Bacterial strains, media, and plasmids

The strains, plasmids, and oligonucleotides used in this study are listed in Table 1. *E. coli DH5* α strain was used for the cloning procedures. *E. coli* S17-1(λ *pir*) was used as a donor in intergeneric conjugation experiments⁴⁵. *E. coli* cells were routinely grown in the LB (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L NaCl) broth containing 25 μ g/mL chloramphenicol (Sigma) if transformed at 37 °C. KrPL *LacY*⁺, a cured *PhTAC125* strain, was used for protein expression of the recombinant proteins and grown in the TYP broth (16 g/L bacto-tryptone, 16 g/L yeast extract, 10 g/L NaCl).

3.2.3 Construction of the *flCDKL5* mutants

The vector, pB40-79C-BCD2-*flCDKL5*, was used as a template for the generation of all constructs. Primer pairs used were designed with gene-specific sequences along with a portion of the vector sequences, each with 18–30 bp overlap (Table 1). PCR was usually performed in a 25 μ L total volume reaction containing 12.5 μ L of Q5® High-Fidelity 2X Master Mix (New England, Biolabs, US), 1.25 μ L 10 μ M each primer, and 20 ng of DNA template. The PCR cycling program was as follows: 1st step: 98 °C for 30 sec; 2nd step: 98 °C for 10 sec; 3rd step: 62 °C for 20 sec; 4th step: 72 °C for 2–3 min (40 sec /kb); 5th step: go back to step 2 for 24 times; 6th step: 72 °C for 5 min. All amplified fragments were purified using the PCR clean-up Gel extraction kit (New England, Biolabs, US) according to the manufacturer's instructions. The PCR fragments and the vectors were assembled according to the one-step isothermal DNA assembly method described by Gibson⁴⁴: 160 ng of each DNA fragment, 80 ng of the linearized vector, and 10 μ L Gibson Assembly® Master Mix (New England, Biolabs, US). The mixture was incubated at 50 °C for 1 h in a thermocycler. All volumes (20 μ L) of the Gibson assembly reaction were used for the transformation of the *E. coli DH5* α competent cells by chemical transformation. The recombinant vectors were mobilized into KrPL *LacY*⁺ by intergeneric conjugation⁴³.

3.2.3 Growth conditions

Glycerol stocks (- 80 °C) of KrPL *LacY*⁺ strains were streaked on TYP agar plates and incubated at 15 °C. A single colony was inoculated in 1-2 mL TYP at 15 °C for 24-48 hours at 220 rpm. Then, the cultures were diluted 1/100 in 10 mL TYP medium and incubated overnight at 15 °C. The inoculum was finally performed at 0.4 OD/mL in the liquid medium filling an Erlenmeyer flask with 20% of its capacity. The induction of the recombinant expression was carried out at the start of the growth with 5 mM IPTG.

3.2.4 Imaging KrPL *LacY*⁺ cells

For the cell imaging experiment, strains were imaged at the mid-exponential growth phase ($OD_{600} = 0.8-1.6$). 1 μ L of cells was added to 1.5% agarose/GG medium pads and imaged on 22x22 mm² no. 1.5 glass coverslips (in single-molecule experiments, Fisher 12-541-B or VWR 48366-227 cleaned by Ar plasma etching). Images were collected using a Leica DMI8 S microscope equipped with a Hamamatsu C9100 EM-CCD camera, a 100x oil-immersion objective (1.63 NA), and a SPECTRA X light engine (Lumencor). The Fiji tool was used to analyze the fluorescence images.

3.2.5 Analysis of the recombinant proteins production

To verify the expression levels, 1 OD_{600} cell pellets were collected during the production experiments by centrifugation and solubilized in 60 μ L of Laemmli buffer 4X. Then, the samples were boiled at 95 °C for 20 min, quickly cooled on ice for 1 minute, and finally centrifuged at 10,000 x *g* for 5 min at RT. The equivalent of 1/20 OD and 1/12 OD (3-5 μ L of total protein extract samples, respectively) were analyzed by SDS-PAGE using 4 - 15% Mini-Protean TGX (Biorad) gels in TGS buffer, setting the power supply to constant 120 V. For electroblotting, the Biorad Transblot Turbo system with Biorad PVDF mini membranes was used. After the transfer, the membrane was blocked with PBS, 0.05% v/v Triton X-100, and 5% w/v milk for one hour and incubated with a specific primary antibody.

The 3xFlag peptide was detected using Monoclonal ANTI-FLAG M2, Clone M2 (F1804, Sigma) antibody diluted 1:1,000 in PBS, 0.2% v/v Tween 20, 5% w/v milk. After one hour of the incubation at RT, the membrane was washed with PBS, 0.2% v/v Tween 20 three times (5 min each) and incubated with an anti-mouse antibody diluted 1:5,000 in PBS, 0.2% v/v Tween 20, 5% w/v milk for one hour at RT. Then, the membrane was washed again with PBS, 0.2% v/v Tween 20 three times (5 min each), and the secondary antibody was detected using the ECL method.

The 6xHis tag was highlighted with Monoclonal Anti-polyHistidine-Peroxidase clone HIS-1 antibody (A7058, Sigma) diluted 1:2,000 in PBS, 0.05% v/v Tween 20, 5% w/v milk. After one hour of incubation at RT with the antibody, the membrane was washed with PBS, 0.05% v/v Tween 20 three times (5 min each), and it was developed using the ECL method.

Strain	Relevant characteristics	References or Source
--------	--------------------------	----------------------

E. coli strains

<i>DH5α</i>	[supE44, ΔlacU169 (φ80 lacZΔM15) hsdR17, recA1, endA1, gyrA96, thi-1, relA1]	Lab stock
S17-1(<i>λpir</i>)	thi, pro, hsd(r- m+) recA::RP4-2-TCr::Mu Kmr::Tn7 Tpr Smr λpir	Tascon et al. 1993

PhTAC125 strains

KrPL <i>LacY</i> ⁺	<i>PhTAC125</i> cured strain without pMtBL plasmid	Colarusso et al. 2020
-------------------------------	--	-----------------------

Plasmid

pB40-79C-BCD2- <i>flCDKL5</i>	Expression vector possesses a replication origin with a high copy number (B40), an IPTG-inducible promoter (P79), a bicistronic expression cassette (BCD2) and encodes the flCDKL5 protein.	Colarusso et al. 2022
pB40-79C-BCD2- <i>mCherry</i>	Expression vector expressing mCherry	In this study
pB40-79C-BCD2- <i>mCherry-flCDKL5</i>	Expression vector expressing the fusion product, mCherry-flCDKL5	In this study
pB40-79C-BCD2- <i>mCherry-CD</i>	Expression vector expressing the fusion product, mCherry-CD	In this study
pB40-79C-BCD2- <i>mCherry-K2</i>	Expression vector expressing the fusion product, mCherry-K2	In this study
pB40-79C-BCD2- <i>mCherry-K3</i>	Expression vector expressing the fusion product, mCherry-K3	In this study
pB40-79C-BCD2- <i>mCherry-D2</i>	Expression vector expressing the fusion product, mCherry-D2	In this study
pB40-79C-BCD2- <i>mCherry-D3</i>	Expression vector expressing the fusion product, mCherry-D3	In this study
pB40-79C-BCD2- <i>mCherry-CD-D1</i>	Expression vector expressing the fusion product, mCherry-CD-D1	In this study
pB40-79C-BCD2- <i>mCherry-CD-D2</i>	Expression vector expressing the fusion product, mCherry-CD-D2	In this study
pB40-79C-BCD2- <i>mCherry-CD-D1-D2</i>	Expression vector expressing the fusion product, mCherry-CD-D1-D2	In this study
pB40-79C-BCD2- <i>mCherry-CD-D1-D3</i>	Expression vector expressing the fusion product, mCherry-CD-D1-D3	In this study
pB40-79C-BCD2- <i>mCherry-CD-D2-D3</i>	Expression vector expressing the fusion product, mCherry-CD-D2-D3	In this study
pB40-79C-BCD2- <i>mCherry-K1-K2</i>	Expression vector expressing the fusion product, mCherry-K1-K2	In this study
pB40-79C-BCD2- <i>mCherry-K2-K3</i>	Expression vector expressing the fusion product, mCherry-K2-K3	In this study

Primer name**Sequence (5' - 3')**

pB40-79C-BCD2- flCDKL5.Vect.Fw	TGCATGGAACAGAGAATTTATATTTCCAAGGTGATTACAAG
pB40-79C-BCD2- flCDKL5.Vect.Rv	TTTACTCACCATACCTTTATCATCATCATCAATATGACCG
mCherry.Fragm.Fw/ CD.Fragm.Fw/CD- D1.Fragm.Fw/ CD- D2.Fragm.Fw/ CD- D1-D2.Fragm.Fw/ CD-D1- D3.Fragm.Fw/ CD- D2-D.Fragm.Fw	ATGATAAAGGTATGGTGAGTAAAGGTGAAGAAGACAATATGG
mCherry.Fragm.Rv	TATAAATTCTCTGTTCCATGCATGTTAATCAGAGCCC
CD.Fragm.Rv/ CD.Fragm.Rv/CD- D1.Fragm.Rv/ CD- D2.Fragm.Rv/ CD- D1-D2.Fragm.Rv/ CD-D1- D3.Fragm.Rv/ CD- D2-D.Fragm.Rv	TGAGTCCGATGTTCCATGCATGTTAATCAGAGCC
K3.Fragm.Rv/K2- K3.Fragm.Rv/D2.Fra gm.Fw/D3.Fragm.Fw	CCGCCGCAAGACCAACAGAAAGATTCTGGATGT
K2.Fragm.Rv/K1- K2.Fragm.Rv/ D2.Fragm.Rv/D3.Fra gm.Rv	CCGCCGCCACGTTGTGTCTGAAAAGTCGGAT

Table 1 List of bacterial strains, plasmids, and primers used in this study.

3.3 Results and Discussion

3.3.1 Design and construction of *hCDKL5* mutants

To probe the behavior of *hCDKL5 in vivo*, a fusion product containing *hCDKL5* fused to the fluorescent protein mCherry was generated. The present study was focused on isoform 1 of the human protein produced in KrPL *LacY*⁺ harboring the pB40-79C-BCD2-*hCDKL5* vector³⁷. This vector optimized for the production of this variant (from now on called flCDKL5) in the psychrophilic strain possesses a replication origin with a high copy number (B40), an IPTG-inducible promoter (P79), a bicistronic expression cassette (BCD2) (Figure 3.3.1 A). In particular, the flCDKL5 sequence was fused, at the 5'-region, to a sequence coding for a TATk (*HIV-1 Transactivator of Transcription*) peptide. This aminoacidic sequence is used as a drug delivery system through the blood-brain barrier for therapeutic purposes⁴² and the Sumo sequence (Figure 3.3.1B). Colarusso *et. al.*³⁷ demonstrated that the presence of Sumo allowed the protection of the N-terminal extremity of the protein from the proteolysis, resulting in a higher accumulation of integral protein. Furthermore, the 6xHis and 3xFlag tag coding sequences were fused at the 5' and

3' regions of the construct, respectively, using Enterokinase and TEV protease recognition sequences as spacers.

The fusion product mCherry-flCDKL5 was generated by adding the sequence encoding the mCherry protein between the 6xHis tag and the Sumo sequence (Figure 3.3.1 C), obtaining the new vector pB40-79C-BCD2-mCherry-flCDKL5. In addition, a psychrophilic vector expressing only mCherry was constructed to investigate the behavior of the fluorescent protein when produced in KrPL *LacY*⁺ (Figure 3.3.1 D). The resulting plasmid was named pB40-79C-BCD2-mCherry.

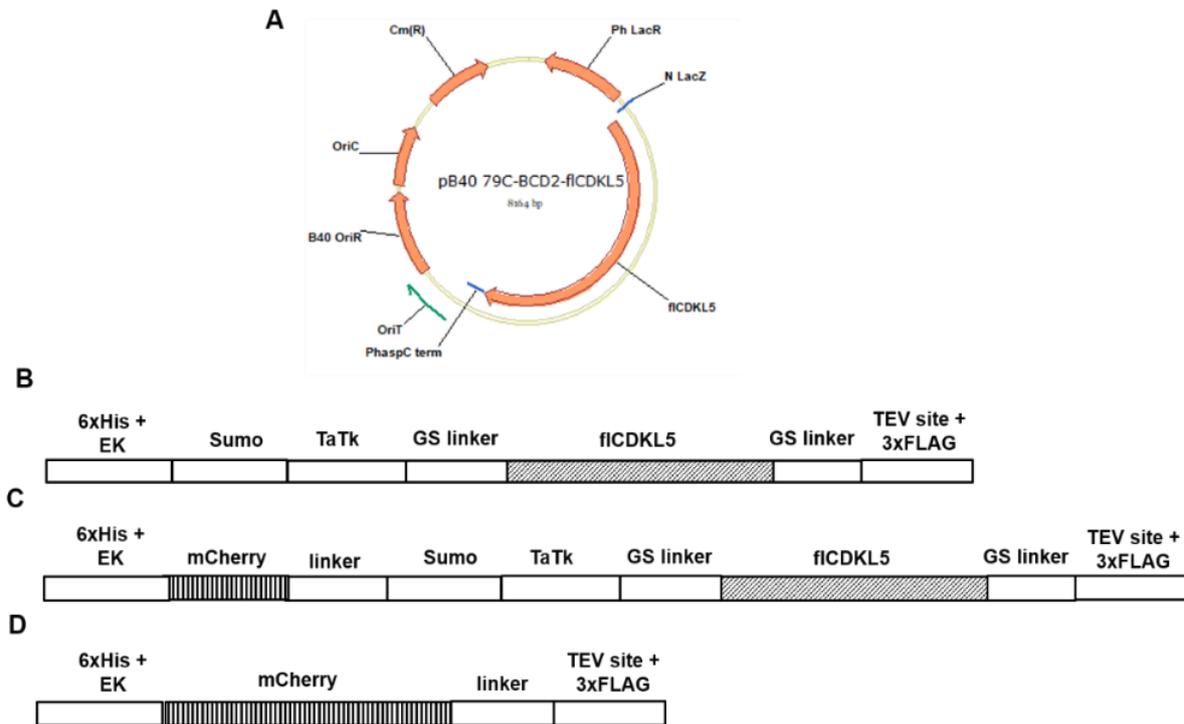


Figure 3.3.1. (A) Schematic map of the pB40-79C-BCD2-flCDKL5 expression vector. OriC, pUC18-derived origin of replication; B40_OriR, the mutated pMtBL-derived autonomous replication sequence; OriT, conjugational DNA transfer origin; P.h.aspC term, transcriptional terminator; P79 and Ph LacR, psychrophilic promoter inducible by IPTG and its regulator, respectively; BCD2, bicistronic expression cassette derived from *lacZ* and *trpA* genes; flCDKL5, gene coding the human CDKL5. (B) Schematic representation of flCDKL5 construct. The 6xHis tag, Enterokinase site (EK), sumo, TaTk, and GS linker are upstream of the flCDKL5 sequence, followed by another GS linker, Tobacco Each Virus protease (TEV) site, and 3xFlag tag. (C) The cDNA encoding flCDKL5 was modified by adding the mCherry sequence and a spacer between the enterokinase site and the sumo sequence. (D) The construct encoding mCherry is characterized by 6xHis and 3xFlag tags at 5' and 3', respectively.

To uncover the contribution of each domain to flCDKL5 condensation in the Antarctic bacterium, a series of flCDKL5 variants were generated. As shown in Figure 3.3.2A, the catalytic domain (CD), constituted by 352 aa, was divided into three regions, named K1 (1-12 aa), K2 (13-300), and K3 (301-352 aa). Also, the unstructured region was subdivided into D1 (380-585 aa), D2 (640-833 aa), and D3 (846-959 aa) (Figure 3.3.2A). Every region was amplified by Gibson assembly using as a template the vector pB40-79C-BCD2-mCherry. PCR products were assembled in the same vector used as a template. New plasmids contained the

sequence of the just catalytic domain (CD) or the single regions (K2, K3, D2, and D3) fused to mCherry at the N-terminal. Furthermore, it was constructed a series of combinations between (i) different regions in the catalytic domain (K1-K2; K2-K3) and (ii) the catalytic domain fused to different regions of the unstructured domain (CD-D1, CD-D2, CD-D1-D2, CD-D1-D3, and CD-D2-D3) leading to the generation of further plasmids. The schematic representation of all constructs is shown in Figure 3.3.2B. All vectors were transformed in the KrPL *LacY*⁺ by conjugation.

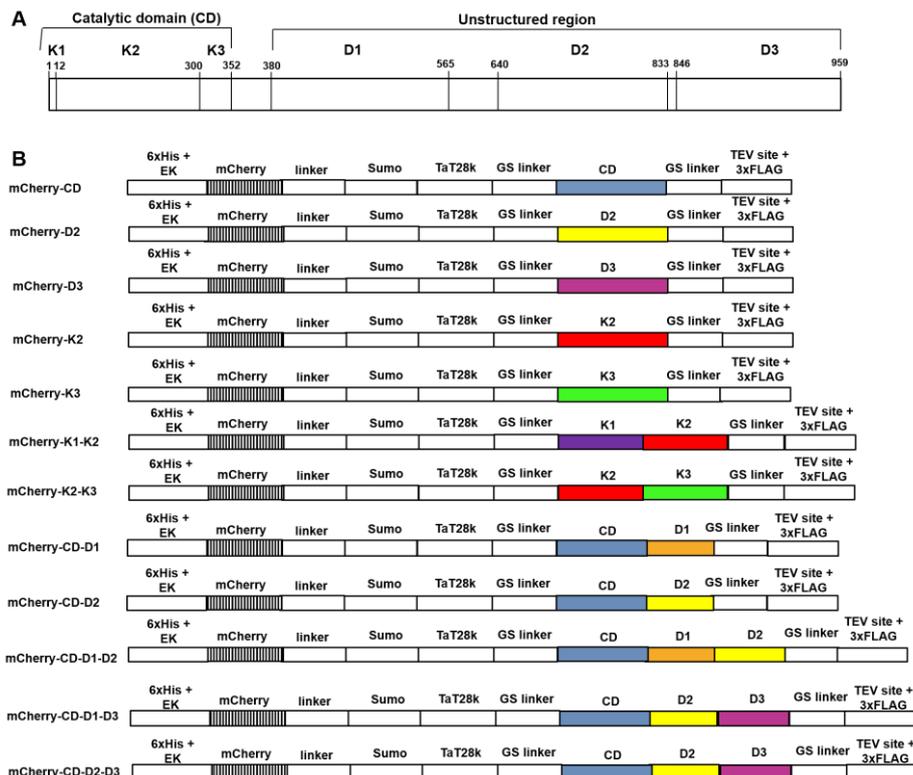


Figure 3.3.2. (A) Schematic representation of the flCDKL5 amino acid sequence to describe the strategy used for the development of the mutants. The flCDKL5 protein is characterized by an N-terminal catalytic domain (1 – 352 aa) and a long disorder structure at C-terminal (353 – 959 aa). The catalytic domain (CD) is divided into three regions, named K1 (1 – 12 aa), K2 (13 – 300), and K3 (300 – 352 aa). The unstructured region is subdivided into D1 (380-585 aa), D2 (640-833 aa), and D3 (846-959 aa). (B) The new constructs are constituted by the just CD (blue box) or K2 (red box), K3 (green box), D2 (yellow), and D3 (purple box) fused to mCherry at the N-terminal. A combination of different regions in the catalytic domain and the catalytic domain fused to different regions of the unstructured domain generated the following constructs: K1-K2, K2-K3, CD-D1, CD-D2, CD-D1-D2, CD-D1-D3, and CD-D2-D3.

3.3.2 Cell imaging experiments *in vivo* of mCherry-flCDKL5 and mutants expressed in KrPL *LacY*⁺ cells

The protein production was evaluated through SDS-PAGE and Western blot analyses (anti-His and anti-Flag) on total cellular extracts recovered after 4, 8, and 24 hours from the induction (Figure 3.3.2.3). Western blot analysis using the anti-His antibody was performed for the detection of the full-length form of the protein,

while the Western blotting with the anti-Flag antibody verified the integrity at the C-terminal of the protein produced. The analysis performed with the anti-Flag for the detection of mCherry pointed out an extra band corresponding to a lower molecular weight of the fluorescent protein attributable to a truncated form of the protein (Figure 3.3.2.3C). Not surprisingly, the production of truncated forms is also visible in other mutants (CD, K2, K3, K1-K2, K2-K3, D3, CD-D1, CD-D2) by the Western blot anti-Flag, as represented in Figure 3.3.2.3 C - F - I. Recently, a study by Maxime Fages-Lartaud⁴⁵ revealed that the red-fluorescent protein possesses an alternative translation initiation site that produces a short functional protein isoform. In future experiments, the mCherry sequence will be optimized through a mutagenic replacement of Methionine in position 10. Moreover, the analysis of the anti-His Western blot revealed low chemiluminescent signals of the proteins principally in fCDKL5, CD-D1-D3, and CD-D2-D3 mutants (Figure 3.3.2.3 B – M). This suggested that the His-tag in these mutants could be highly proteolyzed by endogenous proteases of the psychrophilic bacterium, from the proteolysis, which was most predominant in the CDKL5-M43V variant (lane 1), resulting in a higher accumulation of integral protein.

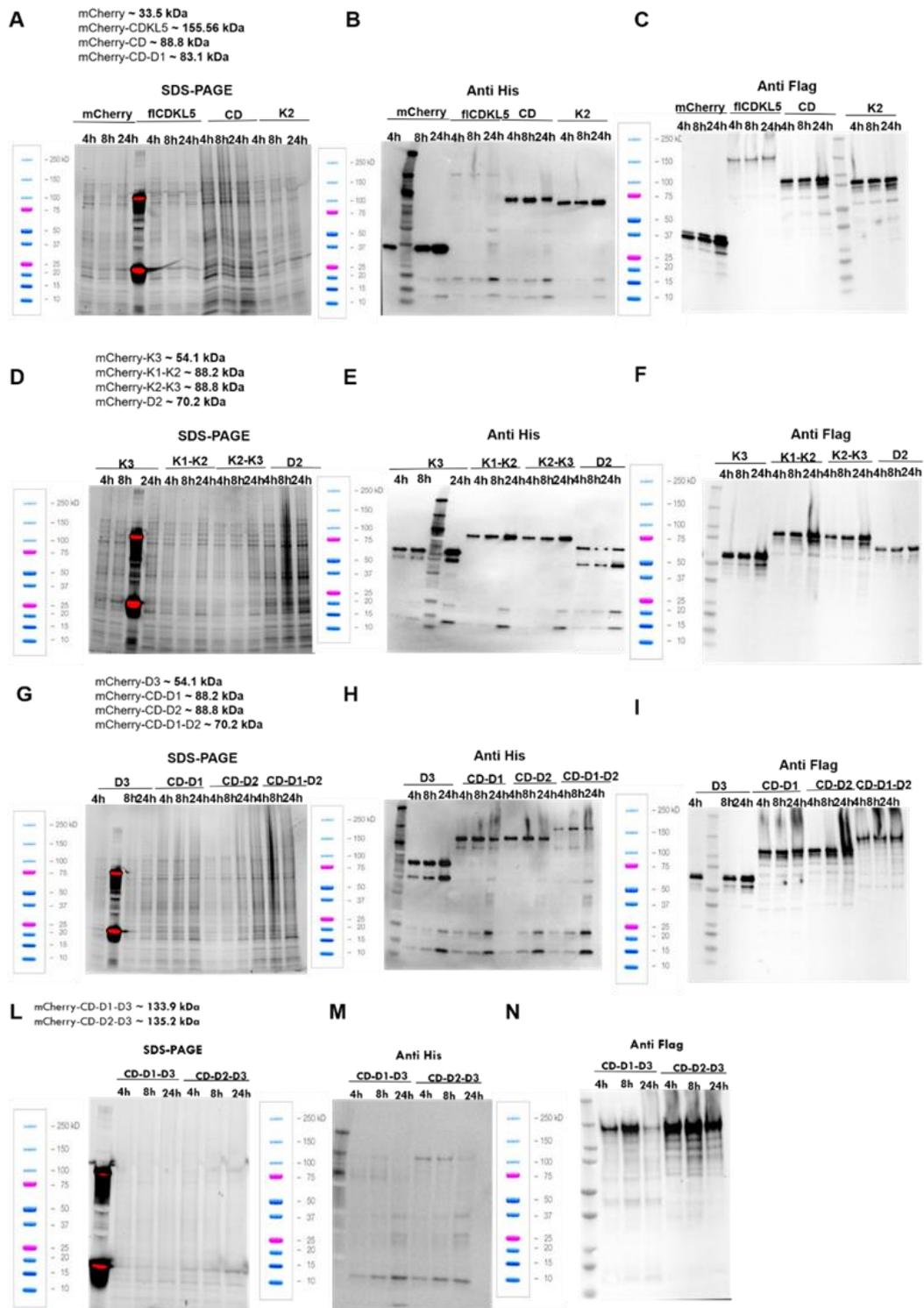


Figure 3.3.2.3 Production profiles of mCherry and CDKL5 fusion products in KrPL *LacY*⁺. 2 μ L of total cellular extracts of recombinant cells were analyzed after 4, 8, and 24 hours from the induction by SDS-PAGE, Western blot analysis anti-His, and Western blot analysis anti-Flag. (A) SDS-PAGE detects the production profiles of mCherry, mCherry-fICDKL5, mCherry-CD, and mCherry-K2. (B) Western blotting with the anti-His antibody shows the detection of mCherry, mCherry-fICDKL5, mCherry-CD, and mCherry-K2. (C) Western blotting with the anti-Flag antibody shows the detection of mCherry, mCherry-fICDKL5, mCherry-CD, and mCherry-K2. (D) SDS-PAGE detects the production profiles of mCherry-K3, mCherry-K1-K2, mCherry-K2-K3, and mCherry-D2. (E) Western blotting with the anti-His antibody shows the detection of mCherry-K3, mCherry-K1-K2, mCherry-K2-K3, and mCherry-D2. (F) Western blotting with the anti-Flag antibody shows the detection of mCherry-K3, mCherry-K1-K2, mCherry-K2-

K3, and mCherry-D2. (G) SDS-PAGE detects the production profiles of mCherry-D3, mCherry-CD-D1, mCherry-CD-D2, and mCherry-CD-D1-D2. (H) Western blotting with the anti-His antibody shows the detection of mCherry-D3, mCherry-CD-D1, mCherry-CD-D2, and mCherry-CD-D1-D2. (I) Western blotting with the anti-Flag antibody shows the detection of mCherry-D3, mCherry-CD-D1, mCherry-CD-D2, and mCherry-CD-D1-D2. (L) SDS-PAGE detects the production profiles of mCherry-CD-D1-D2, and mCherry-CD-D2-D3. (M) Western blotting with the anti-His antibody shows the detection of mCherry-CD-D1-D2, and mCherry-CD-D2-D3. (N) Western blotting with the anti-Flag antibody shows the detection of mCherry-CD-D1-D2, and mCherry-CD-D2-D3.

Cell imaging experiments *in vivo* with fluorescence microscopy were conducted first in the psychrophilic cells expressing mCherry. The recombinant strain was grown in the TYP medium at 15 °C and the expression was induced with 5 mM IPTG at the beginning of the growth. In the current study, TYP was chosen as a medium since previous experiments demonstrated that fICDKL5 yields and solubility are higher in TYP than those in the GG medium. The cells were observed during the exponential growth phase after 4h and 8h from the start of the growth with a high-resolution fluorescence microscope. In Figure 3.3.2.2A, it is possible to detect that the fluorescence distribution is diffused in the bacterial cell expressing mCherry. Next, the psychrophilic strain expressing the fusion product, mCherry-fICDKL5, was grown and explored in the same conditions. Surprisingly, the fluorescence imaging revealed the formation of mCherry-fICDKL5 clusters in the cells, as reported in Figure 3.3.2.2B. To identify one or more regions driving the protein assembly, the fICDKL5 mutants were analyzed. The image analysis shows the formation of clusters in the cells expressing the mCherry-K2 mutant and in all constructs containing this region (mCherry-CD, mCherry-CD-D1, mCherry-CD-D2, mCherry-CD-D1-D2, mCherry-CD-D1-D3, and mCherry-CD-D2-D3) (Figure 3.3.2.2C). On the other hand, the fluorescence is diffuse in mCherry-K3, mCherry-D2, and mCherry-D3 mutants.

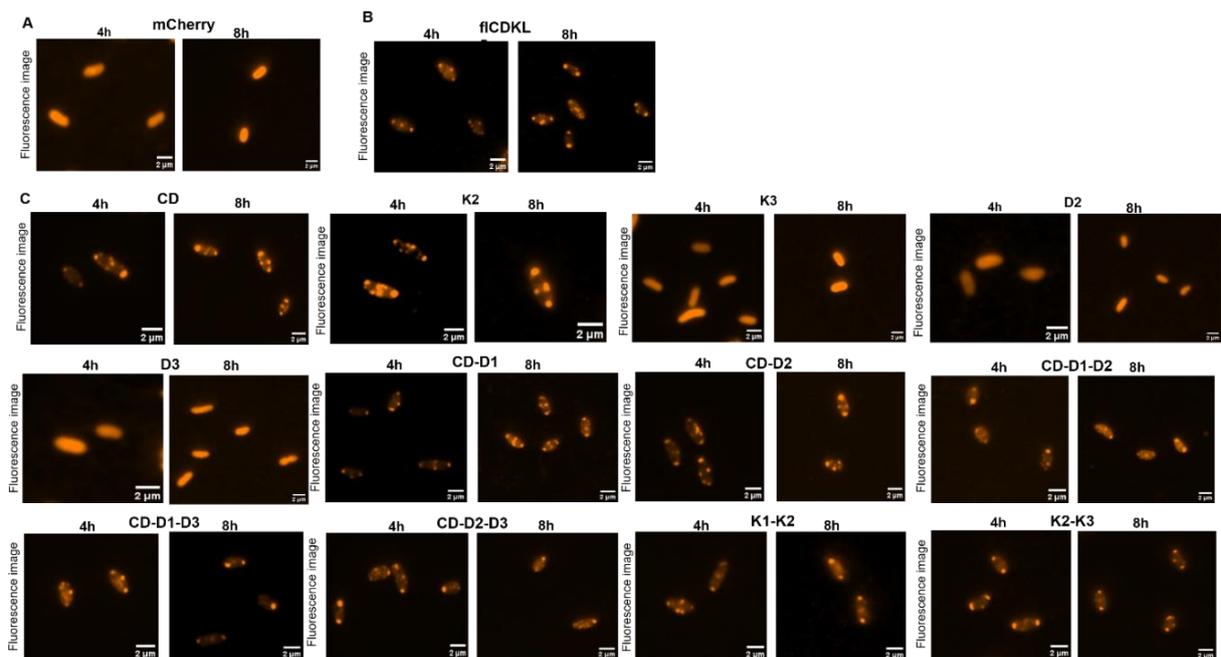


Figure 3.3.2.2 Fluorescence microscopy images. The recombinant cells were observed after 4 and 8 hours of the induction using a super-resolution fluorescence microscope. (A) images of cells expressing

the mCherry protein. The fluorescence distribution is diffuse in the cells. (B) Fluorescent images of cells expressing the full-length CDKL5 fused to mCherry (fICDKL5). The expression of the fusion product (mCherry-fICDKL5) forms clusters. (C) Fluorescent images of the mutants generated: catalytic domain (CD), K2, K3, D2, D3, CD-D1, CD-D2, CD-D1-D2, CD-D1-D3, CD-D2-D3, K1-K2, and K2-K3. The K2 mutant drives the formation of clusters. Indeed, the clusters in the cells are observed in the mutants constituted by the K2 region.

The findings of cell imaging *in vivo* demonstrated that a region of the catalytic domain, named K2, may contribute to the formation of fICDKL5 clusters in KrPL *LacY*⁺. This dynamic behavior was also observed in other kinase proteins. Sampson and coworkers⁴³ reported that the kinase domain activation of anaplastic lymphoma kinase (ALK) protein fused to a microtubule-associated protein drives cytoplasmic cluster formation. Furthermore, the use of ALK inhibitors blocks kinase activation and dissolves these clusters. These results suggested that kinase activation and active conformation promoted protein dimerization and phase separation⁴³.

These observations could be valid also for fICDKL5 since the K2 region (13 – 300 aa) of the catalytic domain possesses the elements essential for the protein activation (serine/threonine protein kinase active site (131 – 141 aa) and TEY motif (169 – 171 aa))³². Furthermore, the structure of isoform 1 of the kinase domain was crystallized in the presence of inhibitors, which formed an extra hydrogen bond to the catalytic loop residue N131. In addition, structural studies on the catalytic domain highlighted the presence of a disordered activation segment (152 -163 aa) and that the protein activation induces conformational changes in the activation loop³². However, further research on the structure of fICDKL5 in an active conformation is required to provide insights into the activation mechanism and cluster formation.

3.4 Conclusion

This study demonstrated the ability of fICDKL5 to form clusters within the KrPL *LacY*⁺ cells. Furthermore, the imaging experiments revealed that a region of the catalytic domain (named K2) drives the protein condensation. This dynamic behavior could be correlated to the active-like conformation in fICDKL5 protein since the K2 region is characterized by elements essential for the autophosphorylation and activation of the human protein. This new knowledge on fICDKL5 may be leveraged to advance more specific molecular studies will enable us to understand the relationship between the structure and dynamics of fICDKL5 clusters. Furthermore, these findings further create the need to improve the purification process of the human protein since *in vitro* assays are necessary to confirm condensate formation and investigate collective interactions between fICDKL5 and its partners.

Authors' Contribution

M. Calvanese, E. Parrilli, M. L. Tutino, and K. Lasker designed all the experiments. M. Calvanese conducted all experiments, construction of fusion product and the mutants, bacterial growth experiments, imaging experiments *in vivo* using high-

super resolution microscopy, and data analysis. M. Calvanese E. Parrilli, M. L. Tutino, and K. Lasker contributed to collecting and interpreting the data.

References

1. Uversky, V. N. Intrinsically disordered proteins and their 'Mysterious' (meta)physics. *Front. Phys.* 7, 8–23 (2019).
2. Ghag, G. *et al.* Disulfide bonds and disorder in granulin-3: An unusual handshake between structural stability and plasticity. *Protein Sci.* 26, 1759–1772 (2017).
3. Bondos, S. E., Dunker, A. K. & Uversky, V. N. On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Commun. Signal.* 19, 1–9 (2021).
4. Correspondence, G. Intrinsically disordered protein regions at membrane contact sites (1) Université Côte d ' Azur et CNRS , Institut de Pharmacologie Moléculaire et Cellulaire , 660 route des Lucioles 06560 Valbonne , France. 1–34.
5. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* 114, 6561–6588 (2014).
6. Van Der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631 (2014).
7. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29 (2015).
8. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* 337, 635–645 (2004).
9. Csizmok, V., Follis, A. V., Kriwacki, R. W. & Forman-Kay, J. D. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chem. Rev.* 116, 6424–6462 (2016).
10. Metskas, L. A. & Rhoades, E. Conformation and Dynamics of the Troponin I C-Terminal Domain. *Biophys. J.* 104, 482a (2013).
11. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8868–8873 (2000).
12. Tompa, P. & Han, K. H. Intrinsically disordered proteins. *Phys. Today* 65, 64–65 (2012).
13. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32, 102–112 (2015).
14. Bah, A. & Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* 291, 6696–6705 (2016).
15. Neira, J. L. *et al.* Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma. *Nat. Publ. Gr.* 1–15 (2017) doi:10.1038/srep39732.
16. Yu, C. *et al.* Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc. *Nat. Publ. Gr.* 1–11 (2016) doi:10.1038/srep22298.
17. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular

- phase transitions. 11, 899–904 (2015).
18. Azaldegui, C. A., Vecchiarelli, A. G. & Biteen, J. S. The emergence of phase separation as an organizing principle in bacteria. *Biophys. J.* 120, 1123–1138 (2021).
 19. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* 28, 420–435 (2018).
 20. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. 1–25 (2016) doi:10.7554/eLife.18413.
 21. Li, P. *et al.* signalling proteins. *Nature* 1–6 (2012) doi:10.1038/nature10879.
 22. Boulay, G. *et al.* HHS Public Access. 171, 163–178 (2019).
 23. Burke, K. A. *et al.* HHS Public Access. 60, 231–241 (2016).
 24. Conicella, A. E. *et al.* HHS Public Access. 24, 1537–1549 (2017).
 25. Murakami, T. *et al.* Article ALS / FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function Article ALS / FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogel. 678–690 (2015) doi:10.1016/j.neuron.2015.10.030.
 26. Patel, A. *et al.* Article A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation Article A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162, 1066–1077 (2015).
 27. Montini, E. *et al.* Identification and Characterization of a Novel Serine – Threonine Kinase Gene from the Xp22 Region. 433, 427–433 (1998).
 28. Yooseph, S. *et al.* Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* 468, 60–66 (2010).
 29. Hector, R. D. *et al.* Characterisation of CDKL5 Transcript Isoforms in Human and Mouse. 5, 1–22 (2016).
 30. Krishnaraj, R., Ho, G. & Christodoulou, J. RettBASE: Rett syndrome database update. 23263, 922–931 (2017).
 31. Encephalopathy, E., Jakimiec, M. & Paprocka, J. brain sciences CDKL5 Deficiency Disorder — A Complex. 1–9.
 32. Canning, P., Park, K., Pelletier, L., Bullock, A. N. & Leroux, M. R. Report CDKL Family Kinases Have Evolved Distinct Structural Features and Ciliary Function Report CDKL Family Kinases Have Evolved Distinct Structural Features and Ciliary Function. 885–894 (2018) doi:10.1016/j.celrep.2017.12.083.
 33. Rusconi, L. *et al.* CDKL5 Expression Is Modulated during Neuronal Development and Its Subcellular Distribution Is Tightly Regulated by the C-terminal Tail *. 283, 30101–30111 (2008).
 34. Fahmi, M. *et al.* In silico study of rett syndrome treatment-related genes, MECP2, CDKL5, and FOXG1, by evolutionary classification and disordered region assessment. *Int. J. Mol. Sci.* 20, 1–19 (2019).
 35. Bergen, N. J. Van *et al.* CDKL5 deficiency disorder: molecular insights and mechanisms of pathogenicity to fast-track therapeutic development. 0, (2022).
 36. Kameshita, I. *et al.* Biochemical and Biophysical Research Communications Cyclin-dependent kinase-like 5 binds and phosphorylates DNA

- methyltransferase 1. *Biochem. Biophys. Res. Commun.* 377, 1162–1167 (2008).
37. Calvanese, M., Colarusso, A., Lauro, C., Parrilli, E. & Tutino, M. L. Soluble Recombinant Protein Production in. 2406, 219–232 (2022).
 38. Colarusso, A., Lauro, C., Calvanese, M., Parrilli, E. & Tutino, M. L. Active human full-length CDKL5 produced in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Microb. Cell Fact.* 21, 1–18 (2022).
 39. Yeong, V., Werth, E. G., Brown, L. M. & Obermeyer, A. C. Formation of Biomolecular Condensates in Bacteria by Tuning Protein Electrostatics. *ACS Cent. Sci.* 6, 2301–2310 (2020).
 40. Gao, Z. *et al.* Liquid-Liquid Phase Separation: Unraveling the Enigma of Biomolecular Condensates in Microbial Cells. *Front. Microbiol.* 12, 1–17 (2021).
 41. André, A. A. M. & Spruijt, E. Liquid–liquid phase separation in crowded environments. *Int. J. Mol. Sci.* 21, 1–20 (2020).
 42. Dendooven, T. *et al.* A cooperative PNPase-Hfq-RNA carrier complex facilitates bacterial riboregulation. *Mol. Cell* 81, 2901-2913.e5 (2021).
 43. Tutino, M. L. *et al.* A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles* 5, 257–264 (2001).
 44. Sampson, J., Richards, M. W., Choi, J., Fry, A. M. & Bayliss, R. Phase-separated foci of EML 4 -ALK facilitate signalling and depend upon an active kinase conformation. 1–18 (2021) doi:10.15252/embr.202153693.
 45. Fages-lartaud, M. *et al.* mCherry contains a fluorescent protein isoform that interferes with its reporter function. 1–11 (2022) doi:10.3389/fbioe.2022.892138.

Communications and Publications

Communications:

1. **M. Calvanese**, A. Colarusso, C. Lauro, G.A. Apuzzo, P. D'Alessio, G. Del Vecchio, E. Parrilli, M.L. Tutino. Exploiting the use of sulforaphane to relieve both BDNF production and NRF2-dependent antioxidant potential in CDKL5 deficiency disease: a case study. Oral presentation at the 5th International Research and Family Conference, Edinburgh (UK), 22-23/06/2019.
2. A. Colarusso, **M. Calvanese**, C. Lauro, E. Parrilli, S. Clark and M.L. Tutino. Why the recombinant production of human TAT-CDKL5 in bacteria is so complex? Oral presentation at the 5th International Research and Family Conference, Edinburgh (UK), 22-23/06/2019
3. A. Colarusso, **M. Calvanese**, C. Lauro, E. Parrilli, S. Clark, M.L. Tutino. The recombinant production of full-length human TAT-CDKL5 in an Antarctic marine bacterium. Poster at the 5th CDKL5 Forum of LouLou Foundation, Boston (USA), 4-5/11/2019.
4. A. Colarusso, **M. Calvanese**, C. Lauro, E. Parrilli, M.L. Tutino. The recombinant production of full-length human CDKL5_1 in an Antarctic marine bacterium. Poster at the 9th meeting of the Neapolitan Brain group, Napoli (Italy), 12-12-2019.
5. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. Structural/functional characterization of full-length hCDKL5 isoform 1 produced in recombinant marine bacteria and use of pharmacological chaperones to stabilize hCDKL5 missense mutants. Oral presentation to the 2020 online CDKL5 - FORUM PRE-MEETING October 9th, 2020.
6. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. Pathological significance of hCDKL5 missense mutants: set up of an "ad hoc" toolbox. 2021 online CDKL5 Forum Grantee Pre-Meeting October 29th, 2021
7. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. The 5th ed. of Biotech France 2022 International Conference and Exhibition. Paris (France), 15-17/06/2022.
8. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. Study of pathological mutations in human CDKL5 protein through a co-expression kinase assay in a marine polar bacterium. FEMS Conference on Microbiology. Belgrade (Serbia), 06/06-02/07/2022. <https://www.femsbelgrade2022.org/abstract-book>
9. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. CDD mutations have variable impacts on CDKL5 activity: a bacterial model.

Poster at the CDKL5 Forum 2022 of LouLou Foundation, Boston (USA), 7-8/11/2022.

10. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. Active human full-length CDKL5 produced in *Pseudoalteromonas haloplanktis* TAC125. Oral presentation at the CDKL5 Forum 2022 of LouLou Foundation, Boston (USA), 7-8/11/2022.
11. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. Production and purification of human full-length CDKL5₁ and its pathological missense mutations: where are we now? Oral presentation to the 2022 CDKL5 Forum November 6th, 2022 Boston (USA)

Publications:

1. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli & M.L. Tutino, Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a cell factory: IPTG-inducible plasmid construction and strain engineering. *Microorganisms* 8, 1–24 (2020).
2. M. Fondi, S. Gonzi, M. Dziurzynski, P. Turano, V. Ghini, **M. Calvanese**, A. Colarusso, C. Lauro, E. Parrilli, M. L. Tutino. Modelling hCDKL5 heterologous expression in bacteria. *Metabolites* 11, (2021).
3. C. Riccardi, C. D'Angelo, **M. Calvanese**, A. Ricciardelli, A. Sellitto, G. Giurato, M.L. Tutino, A. Weisz, E. Parrilli, M. Fondi. Whole-genome sequencing of *Pseudomonas* sp. TAE6080, a strain capable of inhibiting *Staphylococcus epidermidis* biofilm. *Mar. Genomics* 60, 100887 (2021).
4. C. Riccardi, C. D'Angelo, **M. Calvanese**, A. Ricciardelli, M.L. Tutino, E. Parrilli, M. Fondi. Genome analysis of a new biosurfactants source: The Antarctic bacterium *Psychrobacter* sp. TAE2020. *Mar. Genomics* 61, 100922 (2022).
5. **M. Calvanese**, A. Colarusso, C. Lauro, E. Parrilli, M.L. Tutino. Soluble recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125: the case study of the full-length human CDKL5 protein *Methods Mol Biol.* 2022; 2406:219-232.
6. C. Lauro, A. Colarusso, **M. Calvanese**, E. Parrilli, M.L. Tutino. 2022. Conditional gene silencing in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Res Microbiol* 103939.

7. A. Colarusso, C. Lauro, **M. Calvanese**, E. Parrilli, M.L. Tutino. (2022) Active human full-length CDKL5 produced in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Microb Cell Fact* 21:1–18.
8. C. Riccardi, **M. Calvanese**, V. Ghini, T. Alonso-Vásquez, E. Perrin, P. Turano, G. Giurato, A. Weisz, E. Parrilli, M.L. Tutino, M. Fondi. (2022) Metabolic robustness to growth temperature of a cold adapted bacterium. *bioRxiv* (202)
9. **M. Calvanese**, C. Balesta, A. Colarusso, C. Lauro, C. Riccardi, M. Fondi, E. Parrilli, M.L. Tutino (2022) Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a host for recombinant protein production: development of high-copy number plasmids. *Applied Microbiology and Biotechnology* (Amab) (accepted).

Research activity in foreign laboratories

My research activities were carried out at the laboratories of Prof. Keren Lasker from the Scripps Research Institute (La Jolla, CA, US). Dr. Lasker's lab possesses the required expertise and uses high-resolution imaging, computer modeling, and cellular biology to study the structure and dynamics of bacterial condensates.

My work aimed to study the dynamics of bacterial condensates in *PhTAC125*. The condensation properties of the protein were investigated *in vivo* during the recombinant expression of a fluorescent protein fused to the N-terminal of *hCDKL5* into *KrPL LacY⁺*. Furthermore, several rationally designed *hCDLK5* mutants were constructed to uncover the contribution of each domain to its condensation.

Appendix

During my Ph.D, I performed side studies on different fields involving:

- Genomic analysis on psychrophilic bacteria, *Pseudomonas sp.* TAE6080 and *Psychrobacter sp.* TAE2020 (“Whole-genome sequencing of *Pseudomonas sp.* TAE6080, a strain capable of inhibiting *Staphylococcus epidermidis* biofilm” and “Genome analysis of a new biosurfactants source: The Antarctic bacterium *Psychrobacter sp.* TAE2020”).
- Optimization and genetic engineering of the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC125 as host for the production of the recombinant proteins (“Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a cell factory: IPTG-inducible plasmid construction and strain engineering”, “Conditional gene silencing in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125”).
- Production of the human full-length CDKL5 produced in *Pseudoalteromonas haloplanktis* TAC125 (“Modelling hCDKL5 heterologous expression in bacteria”, “Soluble recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125: the case study of the full-length human CDKL5 protein “, and “Active human full-length CDKL5 produced in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125”).



Article

Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a Cell Factory: IPTG-Inducible Plasmid Construction and Strain Engineering

Andrea Colarusso [†], Concetta Lauro [†], Marzia Calvanese, Ermenegilda Parrilli and Maria Luisa Tutino ^{*}

Dipartimento di Scienze Chimiche, Complesso Universitario Monte Sant'Angelo, Via Cintia, 80126 Napoli, Italy; andrea.colarusso@unina.it (A.C.); concetta.lauro@unina.it (C.L.); marzia.calvanese@unina.it (M.C.); erparrilli@unina.it (E.P.)

^{*} Correspondence: tutino@unina.it; Tel.: +39-081-674317

[†] These authors equally contributed to this work.

Received: 22 August 2020; Accepted: 22 September 2020; Published: 24 September 2020



Abstract: Our group has used the marine bacterium *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*) as a platform for the successful recombinant production of “difficult” proteins, including eukaryotic proteins, at low temperatures. However, there is still room for improvement both in the refinement of *PhTAC125* expression plasmids and in the bacterium’s intrinsic ability to accumulate and handle heterologous products. Here, we present an integrated approach of plasmid design and strain engineering finalized to increment the recombinant expression and optimize the inducer uptake in *PhTAC125*. To this aim, we developed the IPTG-inducible plasmid pP79 and an engineered *PhTAC125* strain called KrPL *LacY*⁺. This mutant was designed to express the *E. coli* lactose permease and to produce only a truncated version of the endogenous Lon protease through an integration-deletion strategy. In the wild-type strain, pP79 assured a significantly better production of two reporters in comparison to the most recent expression vector employed in *PhTAC125*. Nevertheless, the use of KrPL *LacY*⁺ was crucial to achieving satisfying production levels using reasonable IPTG concentrations, even at 0 °C. Both the wild-type and the mutant recombinant strains are characterized by an average graded response upon IPTG induction and they will find different future applications depending on the desired levels of expression.

Keywords: *Pseudoalteromonas haloplanktis*; strain engineering; Lon protease; *EcLacY*; recombinant protein production; IPTG; pP79 vector

1. Introduction

Over recent years, both constitutive promoters [1] and inducible cassettes [2,3] have been established for the recombinant expression in *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*) in a wide range of temperatures. Regulatable systems are particularly desirable in industrial processes where the decoupling of the biomass accumulation from the recombinant expression could be crucial to guarantee satisfactory yields. Although they proved to be useful for a series of studies [2–6], the two inducible expression vectors used in *PhTAC125* so far showed some major drawbacks. The L-malate inducible pUCRP plasmid guaranteed a remarkable protein accumulation [2], but its efficacy resulted in be strongly influenced by the medium composition. In particular, the use of L-glutamate as carbon source negatively affected pUCRP induction, making it necessary to formulate bacterial media devoid of this amino acid [5]. Given the pivotal contribution of such a carbon source to *PhTAC125* specific growth rate and metabolic regulation [7], its depletion might limit the versatility of this recombinant

system for industrial purposes. On the other hand, the D-galactose regulatable pMAV expression vector showed a good versatility in terms of the temperature range of use when *P. haloplanktis* TAE79 β -galactosidase was employed as a reporter [3]. Nevertheless, the amount of enzyme that could be accumulated in recombinant *PhTAC125* pMAV-*lacZ* was lower than the yield achievable in the nonrecombinant parental *PhTAE79* strain [8], suggesting a low strength of the used inducible promoter.

The β -galactosidase production in *PhTAE79* wt was indeed sufficiently high to guarantee its purification from *PhTAE79* extracts and its industrial exploitation for lactose treatment without the use of any recombinant technology [8,9]. This data induced us to evaluate the potential translatability of the regulatory sequences of *PhTAE79 lacZ* in *PhTAC125* for recombinant purposes. The preferable choice of *PhTAC125* rather than *PhTAE79* as a host descends from the wider available information in terms of genomic organization and annotation [10], genetic modification strategies [11], and metabolic networks [5,7,12,13] for the first bacterium.

Based on an in silico analysis of the *PhTAE79 lacZ* expression cassette and on previously published data from other authors, we were persuaded of the feasibility of the use of this regulatory system to develop a new expression vector in *PhTAC125*. This plasmid, called pP79, proved to be IPTG-inducible, to outperform our previous regulated expression vector pMAV, and to allow the detection of the production of a fluorescent protein in *PhTAC125* for the first time. To better the performance of this system, we genetically engineered the host for the expression of a mesophilic lactose permease. Such a mutant strain guaranteed a higher recombinant production using a lower IPTG concentration range in comparison with the parental strain. Collectively, our results emphasize the remarkable flexibility of *PhTAC125*, a polar host capable of combining a heterologous psychrophilic expression system with a mesophilic inducer transporter for the recombinant production of proteins.

2. Materials and Methods

2.1. Bacterial Strains and Growth Media Formulations

The strains used in this study are listed in Table S1. *E. coli* DH5 α was used for cloning procedures, while *E. coli* S17-1(λ pir) was employed in intergeneric conjugations as a donor strain for KrPL transformations [14]. KrPL—a cured *PhTAC125* strain—was used in all the recombinant expression and mutagenesis experiments. *E. coli* was cultured in LB broth (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37 °C and the recombinant strains were treated with either 34 μ g/mL chloramphenicol or 100 μ g/mL ampicillin, depending on the selection marker of the vector. KrPL was grown in TYP (16 g/L bacto-tryptone, 16 g/L yeast extract, 10 g/L NaCl) during conjugations and precultures development, and in GG [3] in expression growths. For the propagation and culture of KrPL recombinant strains, either chloramphenicol or ampicillin was used. In detail, chloramphenicol was added to solid and liquid media at 12.5 μ g/mL and 25 μ g/mL concentrations, respectively. Ampicillin was always used with a concentration of 100 μ g/mL, instead.

2.2. Construction of pP79 and p79C Expression Plasmids

The AUTL01000130.1 contig containing *PhTAE79 lacZ* (Figure 1) was automatically annotated with RAST [15] and the annotations were refined using BlastP [16]. For all digestion/ligation reactions, NEB enzymes were used (New England Biolabs, Hitchin, UK). The restriction sites in pP79 and p79C that were hydrolyzed for cloning purposes are visible in the maps in Figure 2. The pP79 inducible expression vector was designed by cloning the DNA fragment from *PhTAE79* encompassing the *lacR* gene and the divergent *lacZ* promoter + 5' UTR into pUCLT/Rterm vector [2,17]. To this aim the pSP73- β -gal vector [8] was used as a template in a PCR involving the use of p79_fw and p79_rv primers (Table S2). The resulting ~1.2 kb amplified sequence covered the 18'782–19'909 region of AUTL01000130.1 and was characterized by the addition of SphI restriction site to its 5' extremity and NdeI, SalI and XbaI to its 3' terminus. Both pUCLT/Rterm and the p79 amplicon were double digested with SphI/XbaI and ligated. The resulting plasmid was pP79.

p79C is a variant of pP79 harboring a chloramphenicol resistance marker rather than an ampicillin selection gene. For its construction, pP79 regulatory sequences, its MCS and the *aspC* transcriptional terminator were extracted with *SphI* and *SacI* from pP79 and ligated with pUCC [4] digested in the same sites.

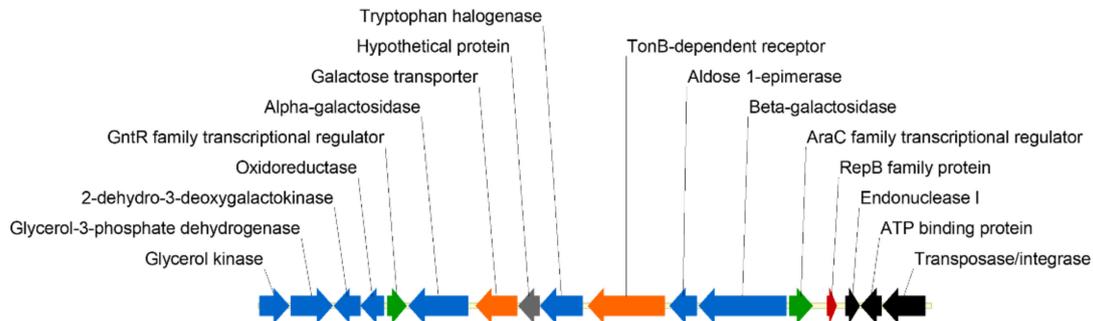


Figure 1. Disposition of genes surrounding the β -galactosidase encoding gene in *Pseudoalteromonas haloplanktis* TAE79 (*PhTAE79*) AUTL01000130.1 contig. Blue arrows indicate genes involved in metabolism, in orange receptors and transporters, while in green transcriptional regulators. Black arrows depict genes involved in DNA rearrangements, the red arrow highlights the presence of a gene involved in plasmid replication and the gray arrow indicates a coding sequence with unknown functions.

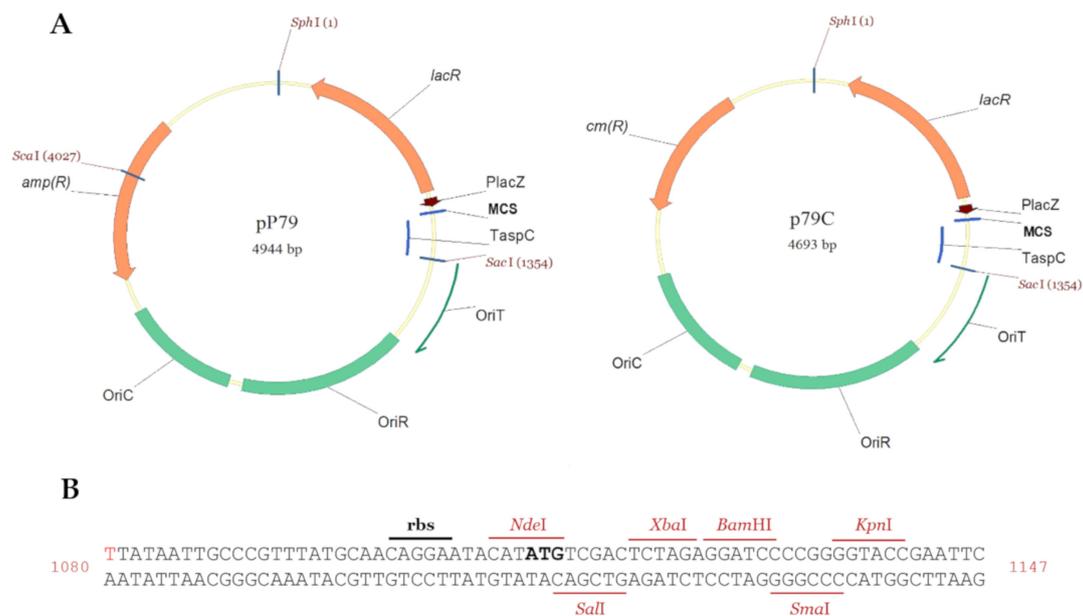


Figure 2. Maps of pP79 and p79C shuttle vectors. **(A)** The two plasmids differ for the selection markers, which are a β -lactamase encoding gene (*amp*(R)) in pP79 and a chloramphenicol acetyltransferase encoding sequence (*cm*(R)) in p79C. The common elements to the plasmids are *PhTAE79* regulatory gene *lacR*, the promoter of *PhTAE79 lacZ* gene (*PlacZ*), the transcriptional terminator on *PhTAC125 aspC* gene (*TaspC*), an origin of conjugative transfer (*OriT*), the pMtBL-derived replication origin for the maintenance in *PhTAC125* (*OriR*) and the pUC18-derived replication origin for the propagation in *E. coli* (*OriC*). Restriction sites outside the MCS that have been used for cloning purposes are indicated. **(B)** The sequence encompassing the 5' UTR and the MCS of the two plasmids included between *PlacZ* and *TaspC* is illustrated using the coordinates of pP79. The +1 of the mRNA and the start ATG are indicated in red and bold black, respectively.

2.3. Sub-Cloning of Heterologous Genes into the Expression Plasmids

The expression plasmids used in this work are reported in Table S3. pMAV-*lacZ* was prepared in a previous study [3] and was used to isolate the psychrophilic β -galactosidase encoding gene for the construction of pP79-*lacZ*. In particular, *lacZ* was split into two different fragments: one of 1.2 kb with NdeI/NcoI extremities and the second of 2.3 kb with NcoI/XbaI extremities. The two gene fragments were ligated with pP79 opened with NdeI and XbaI restriction sites. For the conversion of pP79-*lacZ* into p79C-*lacZ*, the same approach described in Section 2.2 was employed: the regulatory *lacR* gene and *lacZ* were isolated using SphI/SacI double digestion and inserted into pUCC hydrolyzed with the same enzymes.

The *R9-gfp* gene was taken from pET-21b-*R9-gfp* [18] using NdeI and HindIII restriction sites. In the detail, the HindIII digestion was performed first and then the extremities of the hydrolyzed vector were filled by Klenow reaction. After NdeI digestion, the *R9-gfp* gene was cloned into pMAV with NdeI/filled-EcoRI extremities. pMAV-*R9-gfp* was then converted into pP79-*R9-gfp* by replacing pMAV typical expression sequences with the ones of pP79. To do so, ScaI/NdeI double digestion was used to isolate the pP79 fragment encompassing its promoter and its regulatory gene (2.1 kb). This was cloned into the pMAV-*R9-gfp* backbone devoid of the *galT* expression sequences isolated with the same restriction sites (3.6 kb). pP79-*pGFP* was designed to drive the expression of a codon optimized version of the eGFP [19]. Composition optimization of the *pGFP* for the codon usage of *PhTAC125* was automatically performed with the Optimizer web tool using the “guided random” method [20]. The synthesized gene (Thermo Fisher Scientific, Waltham, MA, USA) was cloned into pP79 using NdeI and KpnI restriction sites.

p13C-*lacY* and pFC-*lacY* were the two constructs used for the constitutive expression of *E. coli lacY* gene. Briefly, the gene encoding the mesophilic lactose permease was synthesized by Thermo Fisher Scientific (Waltham, MA, USA) following a sequence optimization for the codon usage of *PhTAC125* [20] and adding a *c-myc* encoding sequence at its 3' extremity. The insert harbored NdeI and KpnI restriction sites at its 5' and 3' ends, respectively, and was cloned into p13C and pFC vectors using the same sites. p13C is a plasmid containing the P13 promoter and a chloramphenicol resistance gene. It was built by fusing P13 sequence taken from pPM13 [1] using HindIII/XbaI double digestion with pUCC [4] hydrolyzed with the same enzymes. pFC contains the constitutive promoter of the *PhTAC125 aspC* gene and was already available [21].

The complete sequences of genes introduced in this study are reported in the Appendix A and Supplementary Material.

2.4. Preparation of pVS-*lon* and pVS-*lacY* Suicide Vectors

For the construction of *lon* mutant, two DNA fragments of *PhTAC125 lon* gene (A and B) were amplified by PCR using bacterial genomic DNA as the template. Two primer pairs were designed to amplify a 305 bp region at the 5' end (*lonA_SphI fw*, *lonA_SacI rv*) and a 233 bp region at the 3' end (*lonB_SacI fw*, *lonB_EcoRI rv*) of *lon* gene. The obtained amplicons were subjected to SphI/SacI and SacI/EcoRI double digestions respectively and cloned into the pVS [22] previously digested with SphI and EcoRI, resulting in pVS-*lon* vector.

The construction of pVS-*lacY* was performed starting from the recovery of the fragment P13-*lacY* from p13C-*lacY* vector through hydrolysis with HindIII and KpnI. Then two fragments (B and B') at the 3' end of *PhTAC125 lon* gene were amplified by PCR. The reactions were carried out using the genomic DNA as the template and allowed the amplification of a fragment of 233 bp (*lonB_SphI fw*, *lonB_HindIII rv*) and one of a 170 bp region (*lonB'_HindIII fw*, *lonB'_EcoRI rv*). SphI/HindIII and KpnI/EcoRI double digestions were performed on the obtained amplicons, respectively. The fragment B carrying SphI/HindIII extremities and P13-*lacY* hydrolyzed with HindIII/KpnI were cloned into pUCC vector, previously digested with SphI and KpnI. Afterward, the obtained intermediate vector pUCC-*lonB*-P13-*lacY* was digested with SphI and KpnI to extract the fragment *lonB*-P13-*lacY*. This fragment was finally cloned together with the second amplicon B' hydrolyzed KpnI/EcoRI into pVS adequately digested with SphI and EcoRI, resulting in pVS-*lacY*.

2.5. Transformation of KrPL and Selection of the *lon* and *lacY*⁺ Mutant Strains

The recombinant vectors were mobilized into KrPL by intergeneric conjugation [14]. The selection of recombinant transconjugants was performed at 15 °C in the presence of 50 µg/mL kanamycin and either 100 µg/mL ampicillin or 12.5 µg/mL chloramphenicol, depending on the specifically employed vector.

As for the mutant strains, the transconjugants were selected at 15 °C in the presence of 50 µg/mL kanamycin and 30 µg/mL carbenicillin.

2.6. gDNA Extraction from the Mutant Strains and Sequence Analysis

Genomic DNA extraction from KrPL mutants and *PhTAE79* was performed using the Bacterial DNA kit (D3350-02, E.Z.N.A™, OMEGA bio-tek, Norcross, GA, USA) following the manufacturer's instructions. The insertion of the suicide vectors into the cells was verified by PCR analysis with a NEB Taq DNA polymerase (New England Biolabs, Hitchin, UK). The genomic DNA was used as the template of the reactions and two couples of primers were used for the amplification of *amp*(R) (*bla_fw*, *bla_rv*) and *pheS* (*pheS_fw*, *pheS_rv*) genes. Then, further PCR analysis was performed to identify the insertion site into *lon* gene. The couples of primers used for this purpose are:

lonA_SphI fw, *lon_rv* and *lon_fw*, *lonB_EcoRI rv*, for the analysis of *lon* mutants;
lonY_fw, *lacY_rv* and *lacY_fw*, *lonY_rv*, for the analysis of *lacY*⁺ mutants.

2.7. Recombinant Production of the Reporter Proteins

Glycerol stocks (−80 °C) of KrPL recombinant strains were streaked over TYP agar selective plates. After three-five days of incubation at 15 °C, a single colony was inoculated in 2–3 mL of TYP at 15 °C for one day. To grow the bacteria in GG, they were routinely trained in the same medium with two subsequent 1/100 dilutions within a time frame of 24 h. The actual inoculum was generally performed in the liquid medium filling an Erlenmeyer flask by 20% of its volume and with a starting OD₆₀₀ of 0.1. For recombinant expression at 15 °C, the cells were generally induced in late exponential phase (OD₆₀₀ = 1) about 13 h after the initial dilution. Strains harboring pMAV derived vectors were induced with 10 mM D-galactose, while pP79 and p79C carrying strains were treated with different concentrations of either IPTG or lactose. Expression trials were attempted also at 0 °C in a similar way as described above. In this case, Erlenmeyer flasks were filled by 35% of their volume to stem oxidative stress and the growths lasted several days considering that KrPL generation time was about 24 h at 0 °C. In most of the experiments, a Biosan PSU-20i orbital shaker was used setting the agitation at 180–220 rpm.

2.8. Analysis of the Production of the Recombinant Proteins

For the analysis of the β-galactosidase production, 10 OD₆₀₀ pellets were harvested during the cultures by centrifugation (5000× g for 5 min at 4 °C) and resuspended in 0.4 mL of Lysis buffer (100 mM sodium phosphate buffer pH 7.5, 2% (v/v) Triton X-100, 1 mM DTT, 5 mg/mL lysozyme). After 20 min of incubation at 15 °C, the samples were centrifuged (10,000× g, 15 min, 4 °C) and the supernatants were used in the following enzymatic measurements with ONPG as a substrate. The spectrophotometric assays were performed in triplicate as reported by Hoyoux et al. [8] and the data analysis was carried out using the ONPG extinction coefficient at 410 nm (3.5 mM^{−1} cm^{−1}) and the total protein concentration measured with the Bradford assay (Bio-Rad Laboratories, Hercules, CA, USA).

To monitor the production of fluorescent proteins, 1 OD₆₀₀ of liquid cultures was centrifuged at 5000× g for 5 min at 4 °C and the pellets were resuspended in 0.5 mL PBS. Then, the samples were serially diluted to achieve the best signal to noise ratio in fluorescence measurements and the dilution factor was used for normalization. Fluorescence measurements were conducted with a JASCO FP-750 spectrofluorometer at 25 °C with an excitation wavelength of 488 nm (slit 3 nm), an emission wavelength of 509 nm (slit 6 nm) and an integration time of 0.10 s.

The production of the recombinant proteins was also monitored by SDS-PAGE, by loading 20 µg of soluble cellular extracts onto the wells of 10% denaturant gels. The cellular homogenization was carried out through the chemical-enzymatic method indicated at the beginning of this section and the total protein concentration in the soluble fractions was estimated with the Bradford method. In the case of GFP producing strains, we also checked for the synthesis of the recombinant proteins in total lysates to control their solubility. Nevertheless, the presence of R9-GFP and pGFP was never visible both in the soluble and total extracts when run onto SDS-PAGE.

To verify the presence of the truncated form of Lon protease, 1 OD₆₀₀ cell pellets were collected by centrifugation and solubilized in 60 µL of Laemmli buffer 4X. Then, the samples were boiled at 95 °C for 20 min, quickly cooled on ice for 5 min and finally centrifuged at 10,000× g for 5 min at RT. 5 µL of samples were analyzed by SDS-PAGE. 4–15% Mini-Protean TGX (Biorad) gels were used in TGS buffer setting the power supply to constant 120 V. For electroblotting, the Biorad Transblot Turbo system with Biorad PVDF mini membranes was used employing the mixed molecular weight setting. After the transfer, the membrane was blocked with PBS, 0.05% Triton X-100, 5% (*w/v*) milk for one hour. Then, an anti-Lon antiserum (ab103809) was diluted 1:1,000 in the same buffer. After one hour of incubation at RT with the primary antibody, the membrane was washed with PBS, 0.05% Triton X-100 three times (5 min each) and incubated with an anti-rabbit antibody diluted 1:30,000 in PBS, 0.05% Triton X-100, 5% (*w/v*) milk for one hour at RT. Then, the membrane was washed again with PBS, 0.05% Triton X-100 three times (5 min each) and the secondary antibody was detected using the ECL method.

2.9. mRNA Extraction and qPCR

Total RNA was isolated from the cells using the Direct-zol RNA Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions. Contaminating genomic DNA was then removed through treatment with RNase-free DNase I (Roche, Mannheim, Germany). Total RNA was reverse transcribed using SuperScript IV (Invitrogen, Carlsbad, CA, USA) according to the recommended protocol. The primers used for this reaction are listed in Table S2. Quantitative real-time PCR was performed on cDNA from each sample by using PowerUp SYBR Green Master Mix (Applied Biosystems, Foster City, CA, USA) implemented with the specific primers (listed in Table S2) in StepOne Real-time PCR System (Applied Biosystems, Foster City, CA, USA). The housekeeping gene *PSHA_RS01090* was chosen as the normalizer. The expression level of the gene of interest was assayed for up-regulation in experimental samples in comparison to a calibrator sample (NI). The relative quantification of mRNA was expressed as fold-change and was calculated through the standard curve method [23]. Three independent sets of experiments were performed.

3. Results

3.1. Analysis and Cloning of the *PhTAE79 lacZ* Expression Sequences

The *PhTAE79* genome has been sequenced in the framework of a WGS project involving several Antarctic *Pseudoalteromonadales* [24], but it was neither assembled nor annotated. The *lacZ* gene is in the AUTL01000130.1 contig according to the GenBank notation, whose predicted genes distribution has been schematized in Figure 1. In this ~24 kb region, *lacZ* clusters with other genes involved in carbohydrates and amino acids metabolism (Figure 1, blue arrows) and, intriguingly, with three predicted CDSs involved in cut and paste mechanisms (black arrows) and a putative RepB protein involved in plasmid replication (red arrow) [25]. In particular, the RepB is highly conserved (65–95%

nucleotide identity) in plasmids harbored by three marine *Pseudoalteromonadales*, *P. haloplanktis* TAC125 (MN400773.1), *P. nigrifaciens* KMM 661 (CP011038.1), *P. arctica* A 37-1-2 (CP011027.1), whose reciprocal similarities have been recently examined [26]. Considering that the whole analyzed contig is almost totally conserved in *P. nigrifaciens* plasmid (88% coverage with 99% identity), it is very likely that the DNA containing the *lacZ* gene is the result of horizontal gene transfer also in *PhTAE79*.

Upstream and divergent to *PhTAE79 lacZ* is a gene predicted to encode an AraC family transcriptional regulator (one of the two green arrows in Figure 1), which probably regulates the β -galactosidase mRNA synthesis and, for this reason, it will be named LacR from now on. Hoyoux et al. used a combination of lactose and IPTG to induce the production of the β -galactosidase in *PhTAE79* and reported that IPTG addition led to an increased protein yield [8]. We confirmed this outcome by inducing *lacZ* expression in *PhTAE79* using IPTG as the only inducer molecule (data not shown). This suggests that LacR is probably regulated by this small allolactose analog.

Persuaded by this preliminary data, we developed a shuttle vector, named pP79, containing *PhTAE79 lacR-lacZ* regulatory elements. In detail, a PCR was designed to amplify the LacR CDS with its putative transcriptional terminator and promoter together with the predicted *lacZ* divergent promoter, its 5' UTR and initial ATG. Then, the amplicon was cloned into the shuttle vector pUCLT/Rterm [2,17], so to have the pP79 plasmid. To make this expression system compatible with other constructs, we also developed its chloramphenicol resistant version, p79C, by ligating pP79 expression cassette with the pUCC vector (Figure 2A) [4]. The transcription start of the *lacZ* gene indicated in red in Figure 2B was identified with a primer extension assay (data not shown).

3.2. Quantification of pP79 Activity Using β -Galactosidase and R9-GFP Reporters

3.2.1. Comparison between pP79 and pMAV Efficiencies

To test the usefulness of pP79, we compared its performance with our most recent inducible expression system, pMAV [3]. To this aim, we used two different reporter genes, *PhTAE79 lacZ* that has been employed for the characterization of all the expression plasmids in *PhTAC125* so far [1–3], and *R9-gfp* which encodes a GFP variant tagged with an N-terminal R9 peptide [18]. In particular, the GFP protein encoded by this construct harbors the eGFP mutations for enhanced fluorescence [19] and the Cycle 3 mutations for improved folding [27]. To ensure plasmids stability, the recombinant constructs were mobilized into KrPL, a *PhTAC125* strain cured of its endogenous plasmid pMtBL (unpublished results from this laboratory). The β -galactosidase production in KrPL pMAV-*lacZ* was carried out by D-galactose induction in the defined medium GG at 15 °C [3]. In the case of KrPL pP79-*lacZ* strain, a 1–10 mM IPTG range was tested for induction in the same growth conditions and the levels of accumulated recombinant protein were measured after overnight expression. pP79 proved to guarantee a higher enzymatic specific activity at all the tested inducer concentrations than pMAV, reaching a 20-fold higher production when 10 mM IPTG was used (Figure 3A). This result was confirmed by assessing the fluorescence emitted by R9-GFP producing strains at 15 °C (Figure 3B). When induced with 10 mM D-galactose, pMAV-R9-*gfp* bearing cells had a fluorescence that was at the same level as the autofluorescence of non-recombinant *PhTAC125*. Conversely, KrPL pP79-R9-*gfp* showed a detectable protein accumulation over time when 10 mM IPTG was added to the culture.

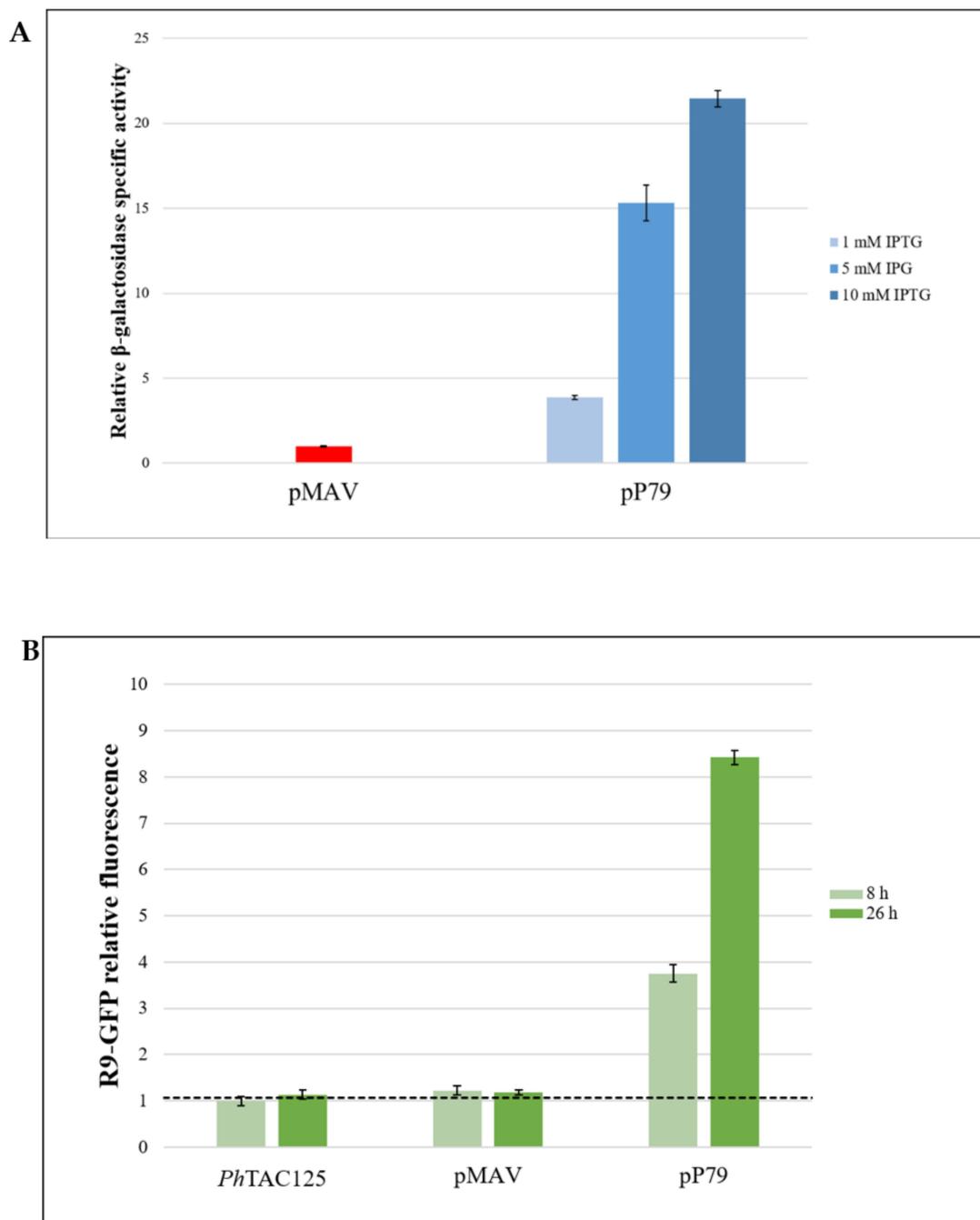


Figure 3. Quantification of the relative strengths of pMAV and pP79-driven expression of *lacZ* and *R9-gfp* genes at 15 °C. **(A)** β -galactosidase production measured with an enzymatic assay. Logarithmic cultures of KrPL pMAV-*lacZ* and KrPL pP79-*lacZ* strains were exposed to 10 mM D-galactose and 1–10 mM IPTG, respectively. After 26 h expression, the β -galactosidase activities were assayed. The enzymatic specific activities are reported as measures normalized by pMAV-*lacZ*. **(B)** R9-GFP synthesis was triggered with 10 mM D-galactose in the case of pMAV-*R9-gfp* bearing strain and with 10 mM IPTG in the case of KrPL pP79-*R9-gfp*. The recorded fluorescence intensities were scaled to the autofluorescence of wild-type cells (*PhTAC125* bars and horizontal dashed line). Levels of β -galactosidase activity and fluorescence are expressed as mean \pm SD, $n = 3$.

3.2.2. Evaluation of the Reliability of *lacZ* and *R9-gfp* as Reporter Systems

Although the two reporter systems consistently demonstrated that pP79 guarantees a higher recombinant production than pMAV, the fluorescence-based approach suffered from lower sensitivity because pMAV-driven R9-GFP production could not be distinguished from the background noise (Figure 2B). To define if this drawback was due to the intrinsic different sensitivities of the two assays or to a discrepancy in the absolute production of the two reporters, further analyses were carried out. First, we monitored the β -galactosidase productions via SDS-PAGE, which allowed us to detect the presence of the recombinant enzyme both in KrPL pMAV-*lacZ* and KrPL pP79-*lacZ* induced strains (Figure 4A). As expected, even when the lower IPTG concentration of 1 mM was used, the intensity of the estimated 118 kDa band was higher in the cellular extract of KrPL pP79-*lacZ* than the one observable in KrPL pMAV-*lacZ* induced lysate (lanes 2 and 4 in Figure 4A, respectively). However, no protein band was visible in induced KrPL strains producing R9-GFP at the expected molecular weight of 28 kDa, regardless of the employed plasmid and inducer concentration (Lanes 5 and 6 in Figure 4A). To define if this different accumulation of the two recombinant proteins was related to transcriptional issues, *lacZ* and *R9-gfp* mRNAs produced by pP79 were quantified through quantitative real-time PCR. The results were expressed as fold-changes to compare the relative amount of *lacZ* and *R9-gfp* mRNA produced both in the presence and in the absence of IPTG. As reported in Figure 4B, a rapid increase of mRNA and a significant accumulation during the time were observed for both the reporters.

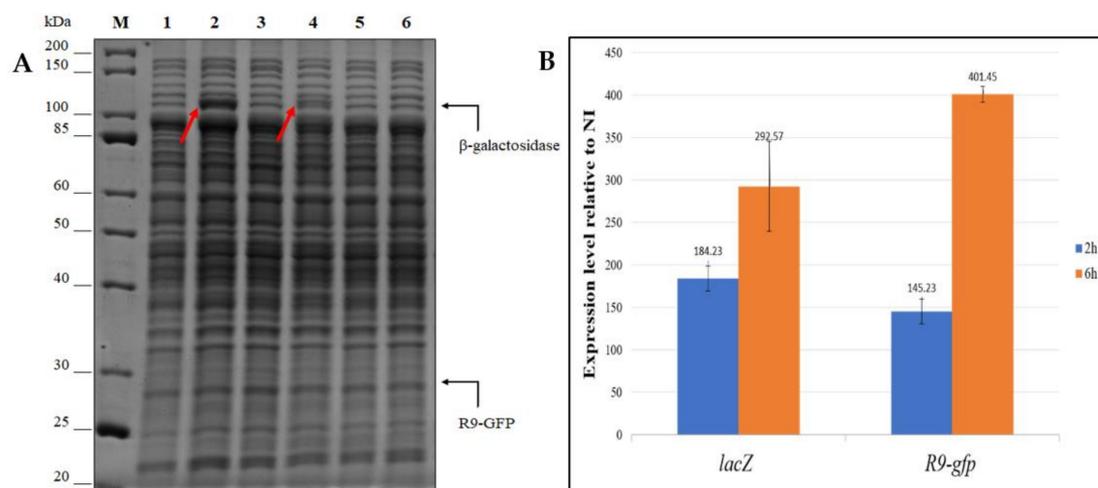


Figure 4. β -galactosidase and R9-GFP proteins productions (panel A) and their respective mRNA transcriptions (panel B). (A) SDS-PAGE analysis of cell extracts of KrPL strains producing the β -galactosidase (lanes 1–4) and R9-GFP (lanes 5–6). The induction was performed with 10 mM D-galactose for pMAV carrying strains and with 1 mM IPTG for pP79 bearing cells and protracted for 26 h. M, molecular weight marker; 1, non-induced pP79-*lacZ*; 2, induced pP79-*lacZ*; 3, non-induced pMAV-*lacZ*; 4, induced pMAV-*lacZ*; 5, induced pP79-*R9-gfp*; 6, induced pMAV-*R9-gfp*. Black arrows on the right of the gel represent the expected molecular weights of the recombinant proteins. Red arrows inside the gel highlight the bands of the β -galactosidase. (B) Relative quantification by RT-qPCR of mRNA expression levels of *lacZ* and *R9-gfp*. Two genes under the control of the PlacZ promoter were analyzed for their mRNA expression levels after 2 and 6 h from the induction in comparison to the non-induced condition. The reported results are the mean of three independent experiments.

Taken together, these results prove that R9-GFP production is considerably less efficient than the one of the psychrophilic β -galactosidase in *PhTAC125* and that this phenomenon is unrelated to transcriptional issues. To determine if the N-terminal polyarginine moiety of R9-GFP and its Cycle 3 mutations were the cause of such a low protein yield, we designed an automatically codon optimized gene encoding the eGFP devoid of any N-terminal tag and sequence mutations other than the ones needed for increased fluorescence. This protein was named pGFP and its fluorescence levels were compared with the ones of R9-GFP using pP79 plasmid. The fluorescence of KrPL cultures was significantly higher when R9-GFP was produced (Figure S1), suggesting that this construct is characterized by improved properties in comparison to the canonical eGFP also at low temperatures.

3.2.3. Influence of Medium Composition on pP79 Efficiency

We performed all our expression trials in GG, a defined medium whose only carbon sources are D-gluconate and L-glutamate [3]. The uptake of substrates is hierarchical in *PhTAC125* [7,28] and specific amino acid combinations had to be formulated in the past to guarantee the optimal induction of another psychrophilic expression plasmid in this bacterium [5]. To understand if this kind of interference could be experienced also using pP79, we measured the achievable levels of β -galactosidase production when KrPL pP79-*lacZ* was grown in TYP, a complex medium containing yeast extract and bacto-tryptone. The growth curves of the recombinant bacteria cultivated in GG and TYP at 15 °C are reported in Figure S2A,B, respectively. As observable in Figure S2C, the performance of the expression plasmid was similar in the two media when the lower IPTG concentration (1 mM) was reached in the cultures. On the other hand, at the two higher tested IPTG concentrations the β -galactosidase production was less efficient in the complex broth, indicating that in those conditions some negative effects took place.

3.3. Optimization of IPTG Transport Mechanism

3.3.1. Attempts in the Plasmidic Expression of a Lactose Permease

For an in-depth study of the pP79 system, the influence of the IPTG transport mechanism was evaluated in relationship to protein expression. The requirement of a high concentration of IPTG (10 mM) to reach full induction and the absence of the lactose metabolic pathway in *PhTAC125* [29] suggest the potential absence of high-affinity IPTG transporters in this bacterium. If so, the mechanism by which the inducer penetrates within the cells is supposed to be either simple or facilitated diffusion.

This observation led us to examine whether the heterologous expression of a suitable lactose permease could deliver a significant contribution to the optimization of the pP79 system. Since the *E. coli* LacY transporter has already been successfully used in other Gram negative bacteria [30], its encoding gene was optimized for the codon usage of *PhTAC125* and cloned into pFC [21] and p13C, a pPM13 derivative [1]. These constitutive psychrophilic expression plasmids possess a medium and a strong promoter, respectively [1]. However, no transconjugant clones resulted from the mobilization of both plasmids into KrPL, probably due to toxic effects on the cell membrane deriving from the excessive production of the permease (data not shown). An alternative strategy was then applied through the integration of *E. coli lacY* gene into the genome of KrPL.

3.3.2. Construction of KrPL *lon* and *lacY*⁺ Mutant Strains

To ensure a subtoxic level of LacY in KrPL a mutant strain was developed so that the production of the permease derived from a single copy of *lacY*, integrated within the host genome.

Firstly, we focused on the selection of the target gene for the integration of *lacY*. To obtain a mutant strain displaying improved features as a host for recombinant protein production, the centerpiece of our analysis was the set of genes coding for proteases that are constitutively expressed in *PhTAC125* and are involved in the proteolytic process of recombinant proteins. The Lon protease encoding gene (*PSHA_RS10175*) was chosen as the target of mutagenesis because it represents the major protein quality control protease and, as such, is responsible for most of the ATP-dependent degradation

of misfolded proteins in bacteria [31]. Despite this protease is involved in a wide range of cellular functions (from proteins degradation to DNA replication and recombination, stress response, motility and biofilm formation), it is not an essential enzyme in many bacterial species such as *E. coli* [32]. To evaluate the consequence of *lon* disruption in *PhTAC125*, a first mutant strain was constructed through a two-step integration–segregation approach using pVS, a suicide vector suitably constructed for *PhTAC125* [22]. Two internal gene fragments of *lon* were chosen as homologous sequences for the recombination event, amplified by PCR and cloned into pVS, resulting in pVS-*lon*. The first fragment (named A) is located into the sequence encoding the N-terminal domain of Lon, while the second fragment (named B) includes the sequence encoding the region straddling the ATPase domain and the proteolytic domain (Figure 5A). Since two different crossing-over events could occur, depending on which fragment underwent recombination, the two homologous regions were selected in order to provide the disruption of the whole Lon protease in the case of recombination of fragment A or the deletion of its proteolytic domain in the case of recombination of fragment B (Figure 5B). Once obtained, pVS-*lon* was mobilized into KrPL by intergeneric conjugation and a single recombination event allowed the vector insertion on the genome. PCR analyses demonstrated that the insertion occurred in fragment B of the gene (data not shown) resulting in a mutant that contains two non-functional copies of the *lon* gene. The first one encodes a truncated form of the Lon protease because it is devoid of the fragment B downstream sequence, coding for the active site domain of the protease. The second copy is not transcribed because it lacks its promoter and the 5'-encoding region (Figure 5C).

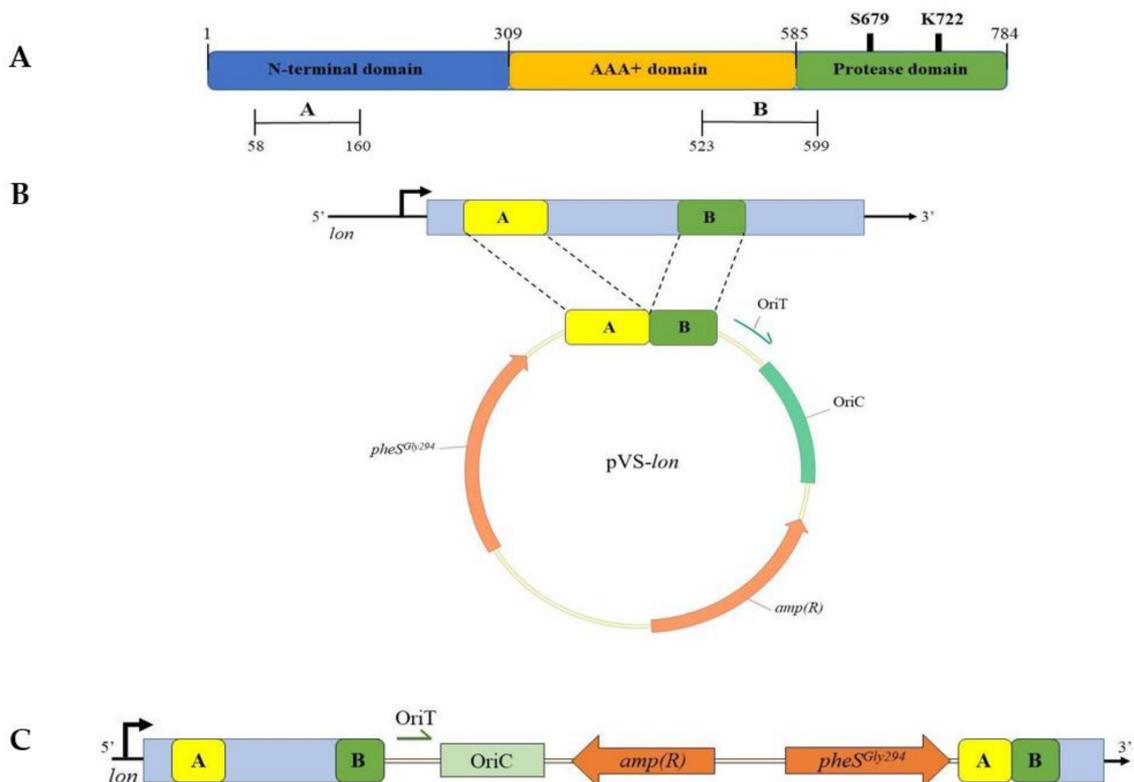


Figure 5. Construction of KrPL *lon* mutant strain. (A) Domain organization of Lon protease. The N-domain is involved in substrate recognition and binding; the AAA+ domain contains the ATPase module; the Protease domain is responsible for proteins degradation. S679 and K722 represent the catalytic dyad of the proteolytic domain. Fragments A and B encoding sequences were chosen as recombination regions. (B) Schematic representation of pVS-*lon* vector. (C) Genetic organization of KrPL *lon* selected mutant.

The presence of the truncated form of Lon was evaluated through Western blot analysis carried out on total KrPL cellular extracts with a polyclonal anti-Lon antiserum. As shown in Figure 6, the Lon band signal was detected at different heights in the wild-type and the mutant strains: the first one exhibited a band compatible with the expected molecular weight of the full-length protein (87.4 kDa, lane 1 in Figure 6), whereas the lower band of the mutant strain was clearly the truncated form (lane 2 in Figure 6). Indeed, the partial deletion of the first copy of the *lon* gene caused the loss of 564 bp at the 3'-region, generating a truncated protein with a theoretical size of 66 kDa.

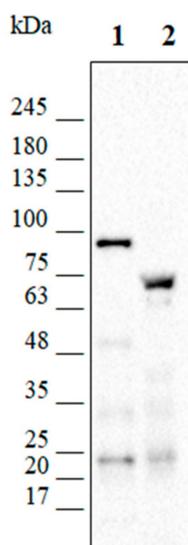


Figure 6. Western blot analysis carried out with anti-Lon antiserum. Total cellular extracts of KrPL wt and *lon* mutant were analyzed through Western blot analysis. Lane 1 shows a signal corresponding to the full-length form of Lon protease (expected size 87.4 kDa) in the wt strain. A lower band (theoretical size of 66 kDa) is detected in the selected *lon* mutant strain corresponding to the truncated form of the protein.

The growth behavior and the fitness of *lon* mutant were then compared to the wt strain and no deleterious effects took place (Figure S3). Thus, *lon* was confirmed as the target of insertion of *EclacY* gene and the strain mutated in *lon* was used as its isogenic control. With this purpose, an expression cassette—consisting of the strong constitutive psychrophilic promoter P13 [1] and the *E. coli lacY* gene—was designed and included between two intragenic fragments of the *lon* gene. The entire construct was then cloned into pVS resulting in pVS-*lacY*. With a similar strategy used for the obtaining of the *lon* mutant strain, pVS-*lacY* was designed to obtain a truncated Lon protease devoid of its proteolytic domain. To do this, the fragment B, already used for the construction of KrPL *lon*, and the fragment B', including the upstream region of the active site of Lon, were chosen as target sequences (Figure 7A).

The obtained pVS-*lacY* was then mobilized into KrPL through interspecies conjugation and the genomes of the mutant clones were analyzed by PCR to define the presence and the orientation of the insertion (data not shown). Here too, the recombination event occurred in fragment B, succeeding in the disruption of *lon* and insertion of *lacY*. The obtained KrPL *lacY*⁺ mutant strain is potentially capable of producing a lactose permease and a truncated form of Lon protease (Figure 7C).

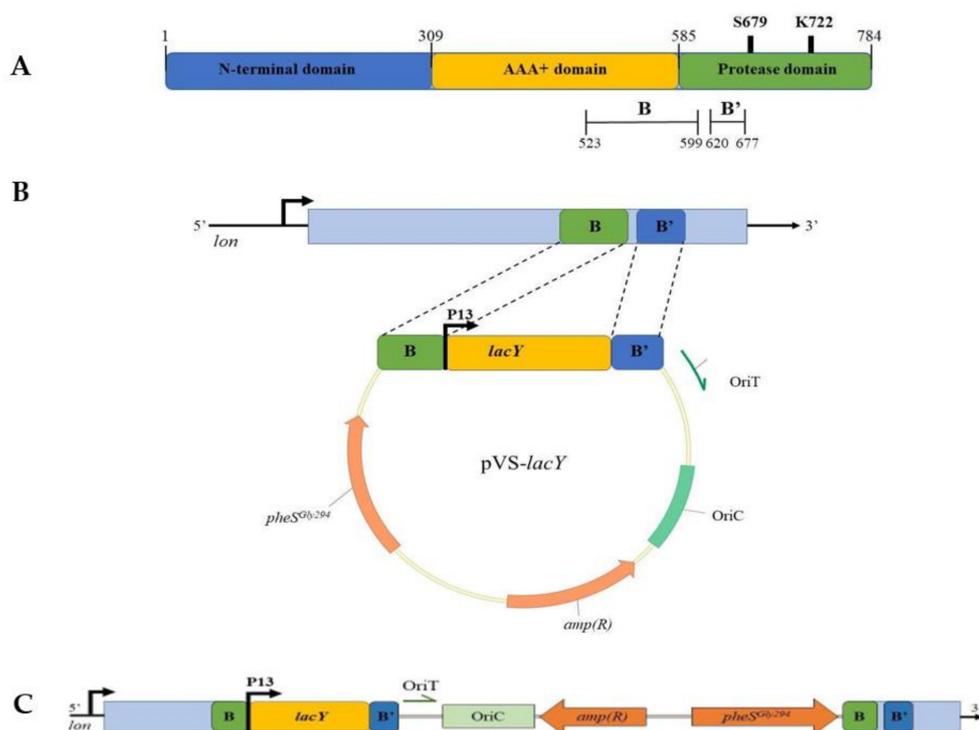


Figure 7. Construction of KrPL *lacY*⁺ mutant strain. (A) Domain organization of Lon protease. The N-domain is involved in substrate recognition and binding; the AAA+ domain contains the ATPase module; the Protease domain is responsible for proteins degradation. S679 and K722 represent the catalytic dyad of the proteolytic domain. Fragments B and B' encoding sequences were chosen as recombination regions. (B) Schematic representation of pVS-*lacY* vector. (C) Genetic organization of KrPL *lacY*⁺ selected mutant.

3.4. Comparison between the Performance of KrPL *lon* and KrPL *lacY*⁺ Strains

3.4.1. Evaluation of the Production Improvement at Different Temperatures

To study whether the lactose permease provides an improvement to the transport of IPTG within the cell, the recombinant production of β -galactosidase was performed in *lacY*⁺ in comparison to its isogenic control, *lon* mutant. Both strains were transformed with the expression vector p79C-*lacZ* and grown in GG medium at 15 °C. During the middle exponential growth phase, the induction was performed with different concentrations of IPTG. In particular, 0.05 mM, 0.1 mM, 0.5 mM, 1 mM IPTG was added to the culture to examine the difference in β -galactosidase production. The β -galactosidase activity was then assayed in the soluble cellular extracts recovered 8, 24, 32 and 48 h after the induction. As shown in Figure 8A, the highest production was achieved with the *lacY*⁺ mutant and proved to be about 5-fold superior to *lon* strain. In the tested range of IPTG concentration, a direct proportionality between the inducer amount and production level was observed in both strains, but with a higher slope in *lacY*⁺ mutant. As an example, Figure 8B highlights the linear correlation between β -galactosidase activity and IPTG concentration for the last data points. Bacterial cells containing the lactose permease yielded high levels of production both with 0.5 mM and 1 mM IPTG. However, the increase of inducer concentration in this range only drove a slight improvement in recombinant production. This is probably due to the decrease in the contribution of the lactose permease in the IPTG uptake when its concentration is relatively high, compatibly with its saturation. Furthermore, these results highlight that the minimum concentration of IPTG needed for the induction of expression in the strain containing the LacY transporter is 10-fold lower in comparison with the strain lacking the permease. When the induction of *lacZ* expression was performed with 0.05 mM and 0.1 mM IPTG in *lon* mutant, no difference

in β -galactosidase activity is observed in comparison to the non-induced cells (NI). Hence, the lactose permease is a very important contributor in transporting IPTG across the *PhTAC125* membrane.

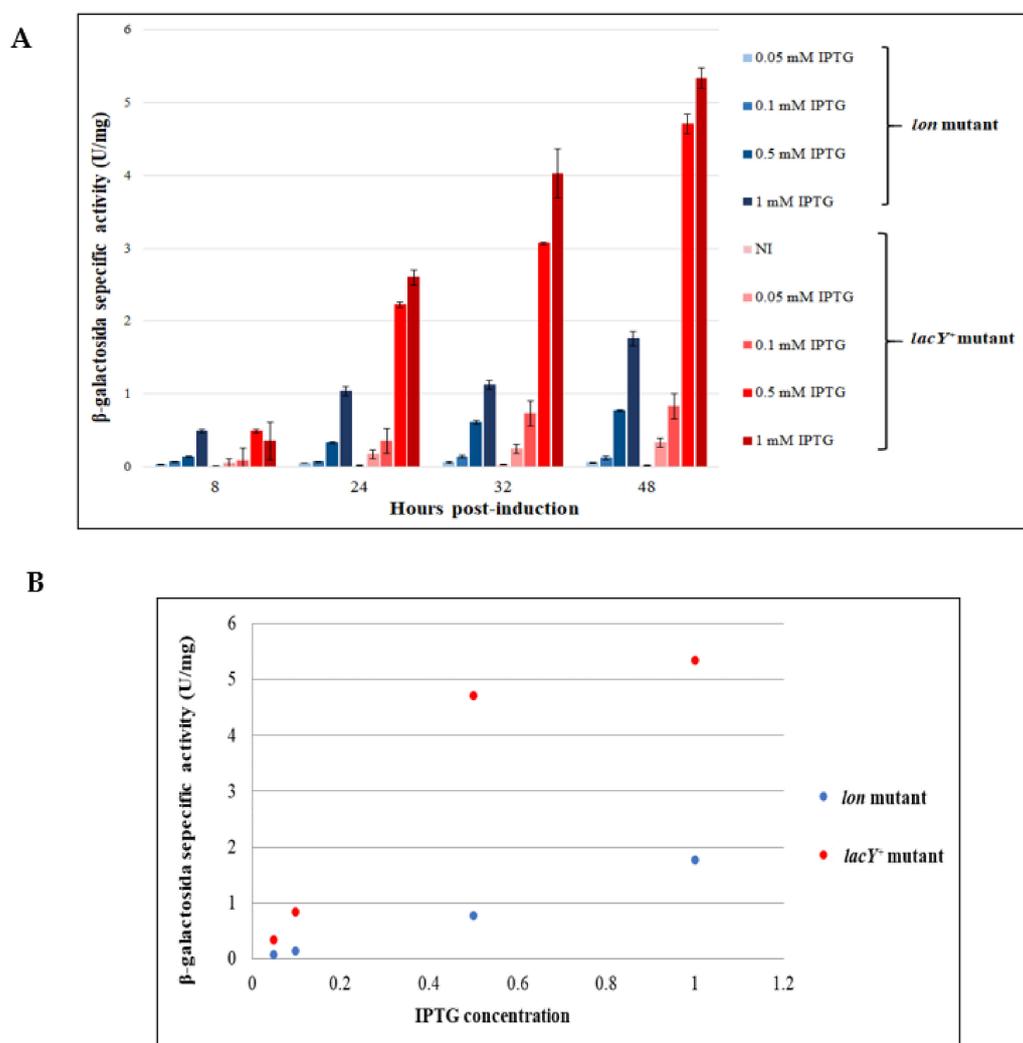


Figure 8. Evaluation of production performance of KrPL *lon* and *lacY*⁺ mutant strains. **(A)** β -galactosidase specific activity (U/mg) in *lacY*⁺ and *lon* mutant cells harboring p79C-*lacZ*, collected after progressive times of induction, in GG medium at 15 °C using different concentrations of inducer. Levels of β -galactosidase activity are expressed as mean \pm SD, $n = 3$. **(B)** Analysis of the relationship between IPTG concentrations and β -galactosidase specific activity measured after 48 h of expression in *lacY*⁺ and *lon* strains.

To verify that the mesophilic membrane protein is produced and functioning in the psychrophilic bacterium even at ultra-low temperatures, the recombinant production of β -galactosidase was performed at 0 °C using 0.5–1 mM IPTG. The levels of the reporter protein were then assayed after 24, 48 and 72 h from the induction. The specific activity of β -galactosidase measured in *lon* mutant highlights a poor accumulation of the protein, suggesting that the response of the system is owing to the basal expression of the protein (Figure 9). On the contrary, the effect of LacY in the transport of IPTG is noticeable already after 24 h of expression, with an enhancement of the production in *lacY*⁺ strain of about 1.5-fold in comparison to *lon* cells treated with the same inducer concentration. As with the expression trials at 15 °C, 0.5 and 1 mM IPTG triggered the same expression levels in *lacY*⁺ mutant also at 0 °C, except for the first time point where a higher recombinant production was guaranteed by increasing amounts of inducer.

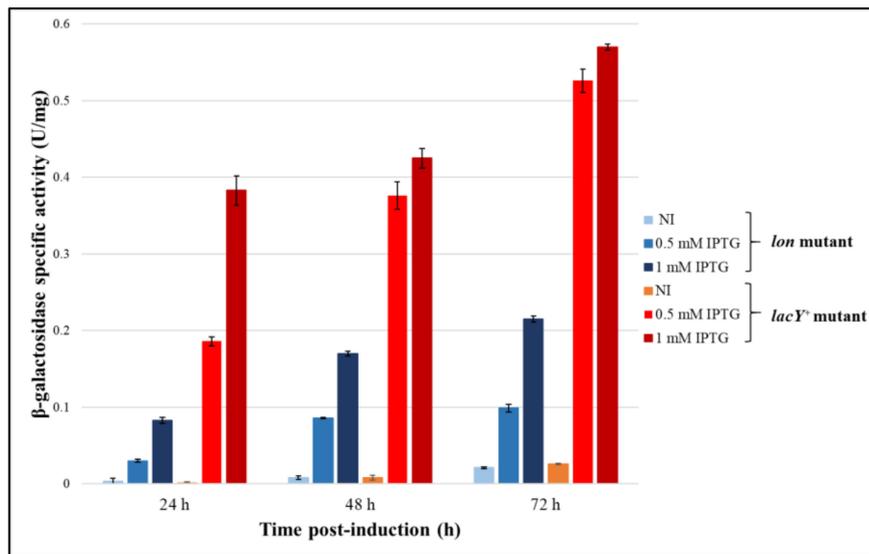


Figure 9. Evaluation of production performance of KrPL *lon* and *lacY*⁺ mutant strains at 0 °C. β -galactosidase specific activity (U/mg) in *lacY*⁺ and *lon* mutant cells harboring p79C-*lacZ*, collected after progressive times of induction, in GG medium at 0 °C using different concentrations of inducer. Levels of β -galactosidase activity are expressed as mean \pm SD, *n* = 3.

3.4.2. Evaluation of β -Galactosidase Production Using Lactose as an Inducer

A further demonstration of the functioning of *EcLacY* transporter in the mutant strain was performed through the recombinant expression of *lacZ* by using 2% (*w/v*) lactose as an inducer. As shown in Figure 10A, a rapid increase in β -galactosidase activity was observed in *lacY*⁺ strain after the first 4 h of expression, while no production was detected in *lon* mutant. Nonetheless, after 8 h of induction, the amount of recombinant protein took a decreasing trend. This can be traced back to a toxic effect observed when the cells can transport the lactose within the cells (Figure 10B). To better understand the reason for this occurrence, the growth behavior of the cells bearing pP79-*lacZ* and pP79 in the presence of lactose was compared. As reported in Figure 10B, only the cells capable of producing β -galactosidase showed cell death, suggesting that this effect was potentially caused by the metabolism and degradation of the disaccharide.

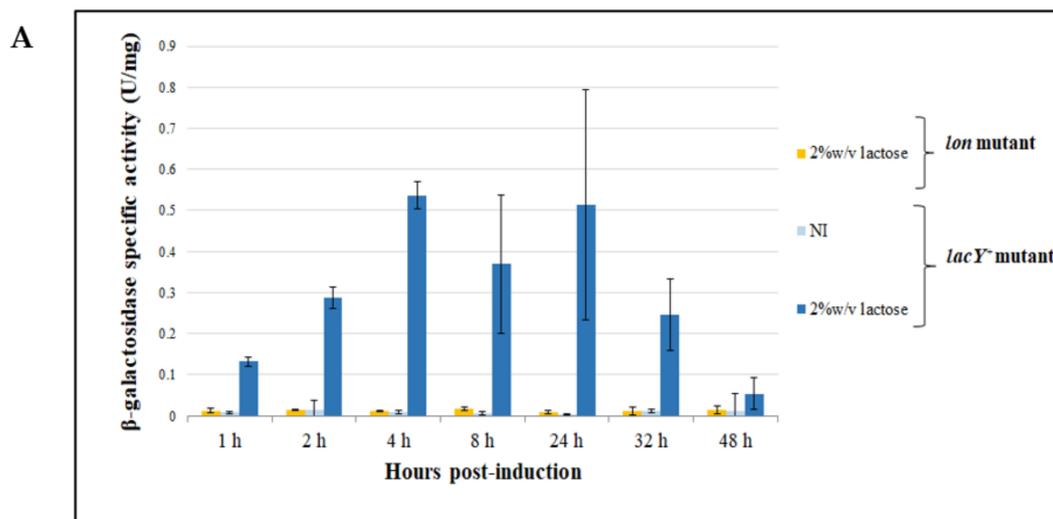


Figure 10. Cont.

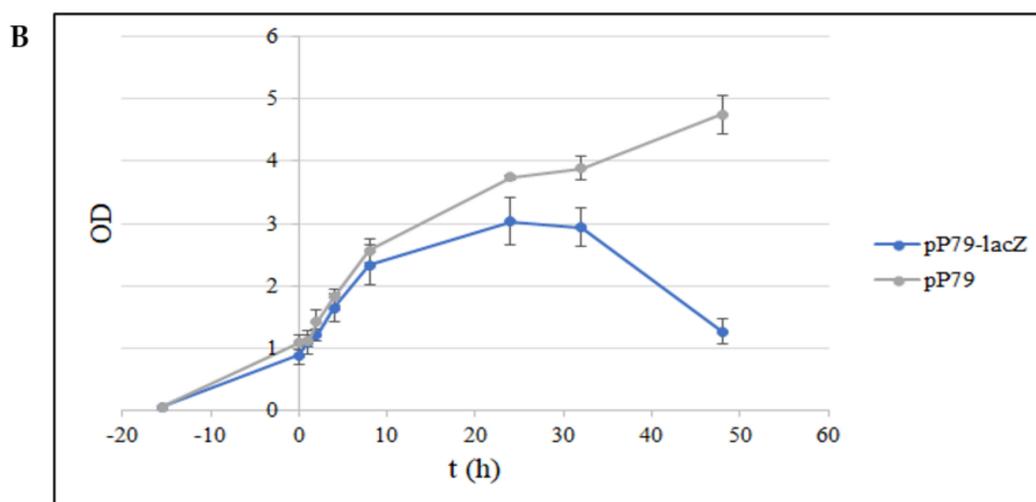


Figure 10. (A) Evaluation of β -galactosidase production in KrPL $lacY^+$ strain using lactose as the inducer. β -galactosidase specific activity (U/mg) in $lacY^+$ and *lon* mutant cells harboring p79C-*lacZ* collected after progressive times of induction in GG medium at 15 °C using 2% (*w/v*) lactose as inducer. Levels of β -galactosidase activity are expressed as mean \pm SD, $n = 3$. (B) Growth curves of KrPL $lacY^+$ harboring p79C-*lacZ* and pP79 in the presence of 2% (*w/v*) lactose. The growth was performed at 15 °C in GG medium. The moment of the induction is represented by the intersection of the axes. The measures of optical density are expressed as mean \pm SD, $n = 2$.

4. Discussion

The ability to produce heterologous proteins with high yields is a prerequisite for the exploitation of a microorganism as a cell factory [33]. The psychrophilic bacterium *PhTAC125* represents a model as a non-conventional host for the production of difficult to express proteins in a soluble and active form [2,5,34,35]. In the present paper, the set of plasmids for controlled gene expression in *PhTAC125* and KrPL—a pMtBL deficient strain—has been expanded with the IPTG-inducible plasmid pP79. Moreover, the usability of this plasmid for different purposes has been widened by the development of an engineered KrPL strain. All our experiments were carried out in the cured *PhTAC125* strain to avoid instability issues possibly arising from the coexistence of pMtBL and pP79 or pMAV.

In our selection of a new expression system, we looked for characterized psychrophilic genes involved in carbohydrate catabolism with a clear regulator-catabolic gene asset. In this sense, the choice of the *lacR-lacZ* gene couple of *PhTAE79* was immediate, considering the high levels of β -galactosidase produced by this bacterium [8] and the lack of *lacZ*-based inducible plasmids in other prokaryotes [33]. Rather than a disadvantage, the lack of lactose metabolism in the chosen host *PhTAC125* [29] can be seen as a possible prerequisite for a more predictable and tunable expression. As a matter of fact, the integration of a heterologous regulatory network in a new context can provide the basis to avoid undesired autocatalytic phenomena as the ones due to uneven and uncontrollable inducer transport [36] causing either bistable or “all-or-none” responses [37–39]. The main prerequisite needed for the functioning of *lacZ* induction in KrPL was the possibility of the internalization of its inducer. Both Hoynoux et al. [8] and our group demonstrated that IPTG could be used as a molecule regulating LacR activity. This reinforced our idea of implementing this recombinant system in KrPL, given the capability of this inducer to penetrate biological membranes in a diffusive manner also in transporter-deficient strains [30,36,40].

The pP79 vector proved to be more efficient than pMAV—a D-galactose inducible plasmid previously used in *PhTAC125* [3]. In particular, with the new system, we could accumulate a 20-fold higher quantity of β -galactosidase than pMAV-*lacZ* and we could detect the production of a fluorescent reporter for the first time in this bacterium. Moreover, we demonstrated that the growth broth composition had an impact on the levels of expression, i.e., a rich medium caused partial repression of the IPTG mediated induction of pP79. Understanding the underlying mechanisms of this negative

regulation (e.g., inducer exclusion or repression of the regulator) will be important in the future to increase the extent of fine regulation that can be applied to pP79 [41,42]. However, it is worth noting that the deeply different formulations of the two used media gave rise to drastic different behaviors of the bacterial growths in general (Figure S2A,B). In particular, the use of TYP guaranteed the doubling of the specific growth rate and the final biomass concentration, suggesting that diversified metabolic networks are activated in this “feast” condition. Hence, to dissect the processes that interfere with pP79 activity, slighter progressive modifications of the medium composition must be applied in the future.

Despite their promising features, KrPL pP79 recombinant strains needed high IPTG concentrations to reach maximal expression, a demand that could be prohibitive in large scale applications requiring induction levels as high as possible.

To approach this problem, we applied a strain engineering strategy based on the combination of the psychrophilic regulatory elements derived from *PhTAE79* with the mesophilic lactose permease LacY from *E. coli*. Despite the use of plasmids with a low copy number [26] and two different constitutive promoters with medium and high strength, the accumulation of the permease and the alteration of membrane properties probably caused serious toxic phenomena in *PhTAC125* [43]. This led to the failure of the first attempt of heterologous expression of LacY.

A further effort was made to modify the host cell to accommodate the production of the membrane protein through the integration of the *lacY* gene into the KrPL genome. Emphasis was concomitantly given to the control of the deleterious proteolysis of recombinant products with the final aim to design a more robust cell factory with improved features for various biotechnological applications. The novel mutant strain *lacY*⁺, constructed through a genome-scale manipulation, was characterized by a deletion in the proteolytic domain of Lon protease and the capability to produce the lactose transporter LacY in a functional form (Figure 7).

Previous analyses performed to characterize the truncated Lon protease suggest that it could act as a molecular chaperone [44]. Indeed, mutations in the active site abolish proteolysis but not ATPase activity, resulting in a protease that is still able to bind its substrates without degrading them [44]. The occurrence of the same phenomenon in KrPL mutant strains has to be proved by analyzing the production of more complex and unstable proteins than the ones studied in this work.

The functional characterization of the *lacY*⁺ mutant highlighted great differences in the levels of the reporter protein produced in comparison to its isogenic control, *lon* strain (Figure 8A). Despite no changes in the growth behavior and kinetics were observed between the two strains (Figure S3), a 5-fold increased protein accumulation was observed for *lacY*⁺, showing a higher slope of the direct proportionality between the production level and the inducer concentration. Owing to the cost and the possible cytotoxic effect of high concentrations of IPTG, this feature remarks the potentiality of *lacY*⁺ strain as a cell factory given the advantage to use less inducer to reach the same level of recombinant production.

Furthermore, our novel mutant succeeded in the production of the mesophilic membrane protein in a functional form also at 0 °C, allowing the enhancement of the production levels of a reporter in comparison to its counterpart *lon* strain (Figure 9). This result is quite impressive, considering that the expressed lactose permease is naturally used to work at 37 °C and the known effects of the low temperatures on the membrane structure and composition [45].

Additional evidence to support the functionality of the transporter in *lacY*⁺ strain was then gained by inducing the recombinant production of the β-galactosidase with lactose. This experiment confirmed that little or no lactose molecule penetrates inside *lon* cells by either other facilitated transport systems or diffusion and LacY is required for the disaccharide internalization (Figure 10A). However, a deleterious effect on cell viability was observed in *lacY*⁺ cells, resembling a “lactose killing” phenomenon [46]. Surprisingly, the comparison of the growth behaviors of recombinant KrPL *lacY*⁺ bearing pP79-*lacZ* and pP79 highlights that the cause of the observed stress is not related to the de-energization caused by elevated transmembrane lactose transport. Instead, a toxic accumulation of its catabolic products likely takes place (Figure 10B). As a matter of fact, the only difference between the two used strains

consists of the production of the β -galactosidase which is causative of the conversion of lactose to catabolic intermediates. For this reason, these experiments carried out with lactose must be taken as proof of LacY functioning in KrPL, rather than as an example of its actual use for induction purposes. *PhTAE79* LacR is predicted to be an AraC-type protein and we experimentally demonstrated that its activity is regulated by IPTG, an allolactose analog. Considering that we used a β -galactosidase as a reporter and that allolactose is a product of the reaction catalyzed by this enzyme [47], it is very likely that part of the observed lactose-mediated induction is due to the peculiar activity of this protein (see Appendix A and Figure A1). Accordingly, when we used the same approach for the expression of other genes than *lacZ* in KrPL *lacY*⁺ pP79, we could observe a lactose-mediated induction, but it was considerably lower than the one achievable with IPTG (data not shown), suggesting again that allolactose is probably the main inducer of this system. Our analysis of KrPL mutants demonstrated that LacY contributes to improving the recombinant production yield in *PhTAC125* when the *lacZ* promoter is employed. This finding is in agreement with most studies performed in *E. coli* and other Gram negative bacteria [30] and is noteworthy given that the IPTG uptake mechanism can be mediated by lactose permease, passive diffusion, or other types of permeases [40].

Altogether, our data certify that in both KrPL *lon* and KrPL *lacY*⁺ strains an average graded production is possible upon IPTG induction (Figure 8B). Depending on the particular application, the two strains can be differently employed. As with the *E. coli* Tuner(DE3) strain (Novagen, [33]), either KrPL or KrPL *lon* could be successfully used for those studies requiring low production levels with a clear linear response over a wide IPTG concentration range, as in the case of metabolic engineering and of the production of toxic proteins. On the other hand, KrPL *lacY*⁺ can be useful for those processes where an average graded response is still guaranteed, though not always linear (Figure 8B), and high production is triggered by a low concentration of the inducer [39]. This might be the case of the synthesis of non-problematic proteins and of low added value products requiring the containment of the production costs. However, one has to keep in mind that all our studies recorded the average expression levels of the cultures, and certain conclusions about the homogeneity of the induction cannot be deduced. Nevertheless, the introduction of the GFP as a reporter in *PhTAC125* for the first time opens to the possibility of single-cell studies as FACS screenings that can address this question.

Finally, although it is beyond the scope of this study, it is worth drawing some considerations about the selection of reporter genes to study promoter strengths at 15 °C. The *PhTAE79* β -galactosidase and R9-GFP were the main tools used to study pP79. Despite they both demonstrated a higher protein accumulation in pP79 recombinant strains than pMAV bearing cells, the production of the fluorescent reporter was way more inefficient than the β -galactosidase and this phenomenon was not related to transcriptional efficiency. More reasons can justify this difference. First, in pP79 the β -galactosidase encoding sequence is directly fused to its natural 5' UTR, while the R9-GFP CDS is artificially joined with the psychrophilic 5' UTR. Different groups have widely reported how the 5' UTR composition and the fusion with heterologous translated sequences can cause translational issues [48–50]. In this sense, even the comparison of pMAV and pP79 relative strengths might be partially biased by the fact that they harbor the *galT* and *lacZ* 5' UTR sequences, respectively.

Nevertheless, in the past a similar disparity in terms of protein accumulation was also observed when the production of the psychrophilic β -galactosidase and a mesophilic α -glucosidase were compared in recombinant *PhTAC125* pUCRP strains [2]. Also, in that case the synthesis of the cold-adapted enzyme was higher than the mesophilic one, indicating that even when fused to a heterologous 5' UTR, the *lacZ* CDS is efficiently translated. Hence, the remarkable accumulation of *PhTAE79* β -galactosidase may be strictly related to either a high translation efficiency or protein stability at low temperatures. This hypothesis is corroborated by a study effectively demonstrating that the *PhTAE79* β -galactosidase showed a particularly high activity when produced in a heterologous bacterium at low temperatures, despite the low levels of transcription [51]. Collectively, these observations may suggest that even if *PhTAE79 lacZ* can be a useful tool for comparative studies, as done in this work, it might not reliably represent the absolute

achievable amounts of other proteins at low temperatures using the same system, especially in the case of mesophilic proteins.

On the other hand, the production of fluorescent reporters in KrPL was so low not to allow the detection of the expression starting from weak promoters, as the one of pMAV vector. Originally, we used an R9-GFP variant, possessing an N-terminal oligoarginine peptide and the mutations characteristic of the eGFP [19] and of Cycle 3 GFP [27]. We intended to combine the enhanced fluorescence of eGFP and improved folding features of Cycle 3 mutations to the aggregation-prone polyarginine peptide [18,52]. In this way, we wanted to confer an increased in vivo stability to the protein, typical of full-protein nanoparticles [53]. However, all these modifications were predicted to improve the GFP properties at 37 °C with no further information about their consequences at lower temperatures. That is why we also produced a plain codon optimized eGFP—called pGFP—with neither R9 nor Cycle 3 mutations. The R9-gfp gene has a Codon Adaptation Index (CAI) of 0.66, while the pGFP gene has a CAI equal to 0.73 [20], but R9-GFP derived fluorescence was considerably higher. This indicates that at least some of the modifications introduced in R9-GFP had a positive effect. Nevertheless, the actual necessity to use the R9 peptide has still to be tested, considering the translational and degradation issues possibly deriving from N-terminal oligoarginines [50,54] and that GFP nanoparticles can sometimes show worsened fluorescent properties [55]. To make this fluorescent system more sensitive, other approaches have to be pursued, such as the shift of the emission spectrum to a wavelength for which *PhTAC125* experiences a lower autofluorescence and the application of protein mutations that improve the fluorophore maturation at low temperatures. However, both pGFP and R9-GFP are already detectable in combination with pP79, allowing for new kinds of study.

5. Conclusions

There is still a significant number of predicted protein products whose recombinant production in conventional gene expression systems is unsuccessful, making their structural/functional characterization and their biotechnological application impossible. Almost 20 years ago, our research group suggested the use of *PhTAC125* and its derived genetic tools for the setup of a novel cell factory working at low temperatures [14]. Till then, much evidence highlighted the notable skills of the Antarctic bacterium in the high quality production of human and/or eukaryotic complex proteins, reinforcing our original idea [2,3,5,6,11,34,35].

In this paper, we achieved a further considerable improvement toward the actual application of *PhTAC125* as an industrial cell factory. Based on the study of the regulated genetic elements in the psychrophilic bacterium *PhTAE79*, we developed pP79, a novel IPTG-inducible plasmid. By using this expression system, we obtained about 20-fold higher production of the recombinant β -galactosidase in comparison to pMAV, the previous best inducible genetic system exploited in *PhTAC125* [3]. For the first time, the detection of a fluorescent protein was achieved in *PhTAC125* pP79 recombinant cells, paving the way for a variety of sensitive and innovative approaches of study.

Another essential aim of this work was to demonstrate the feasibility of a rational approach toward the host improvement. The inducer internalization and the control of proteolytic events were addressed, constructing the engineered strain *lacY*⁺ capable of producing a mesophilic lactose permease and a truncated form of Lon protease. This mutant strain allowed a 5-fold higher production than its isogenic *lon* mutant using a lower IPTG concentration. Furthermore, the heterologous permease showed its positive contribution to induction at 0 °C, widening the applicability of KrPL *lacY*⁺ also as a host for the recombinant protein production at ultra-low temperatures.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-2607/8/10/1466/s1>, Figure S1: Production levels of R9-GFP and pGFP using pP79 expression plasmid, Figure S2: Growths of KrPL pP79-*lacZ* and β -galactosidase production in GG and TYP, Figure S3: Growths curves of KrPL wt, *lon* and *lacY*⁺ strains, Table S1: Strains used in this study, Table S2: Oligonucleotides used in this study, Table S3: Plasmids used in this study.

Author Contributions: A.C. had the idea of using the IPTG-inducible gene expression system in *PhTAC125*, set up the gene expression systems; C.L. constructed the genome inserted mutants, and tested their growth and production performance; M.C. constructed the psychrophilic gene expression vectors for the production of the reporter proteins, and tested the production performance; E.P. and A.C. were responsible for data curation; A.C. and C.L. wrote the original draft; all the authors contributed to reviewing and editing of the final version; M.L.T. had the supervision and was responsible for funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by two grants from Progetto Nazionale di Ricerca in Antartide, number PNRA 2013/B1.04 and PNRA18_00335.

Acknowledgments: The generous contribution of the Italian parent's association "La fabbrica dei sogni 2-New developments for Rett Syndrome" is warmly acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Allolactose-Dependent AraC-Like Transcriptional Regulators Predicted on the Basis of β -galactosidases Features

The only AraC-type transcriptional regulator of a β -galactosidase gene that has been characterized at the molecular level is BgaR of the gram positive *Clostridium perfringens* [56–58], which has been demonstrated to be regulated by lactose and partially lactulose, but not by IPTG. On the other hand, we have clearly demonstrated that pP79 induction is mainly mediated by IPTG. The different regulatory mechanisms influencing the β -galactosidase production in *Clostridium* and *Pseudoalteromonas* genera are probably ascribable to the different capacities of the two enzymes of catalyzing intramolecular transglycosylation.

β -galactosidases are known to catalyze two main reactions: the hydrolysis of lactose into D-galactose and D-glucose, and the allolactose synthesis using the two monosaccharides as substrates [47]. While intermolecular allolactose synthesis is a common feature to all β -galactosidases, only a restricted class of these enzymes is capable of intramolecular transglycosylation, a process requiring that the same glucose moiety produced upon lactose hydrolysis remains entrapped in the catalytic site of the enzyme and is used to attack a galactose molecule for the generation of allolactose [59]. Only when intramolecular transglycosylation takes place, ~50% of the initial product of β -galactosidase-mediated reactions is made of allolactose. In the other cases the allolactose production is so inefficient that D-galactose and D-glucose release becomes instantaneously predominant. Wheatley and co-workers suggested for the first time that intramolecular allolactose synthesis could be a property typical of those β -galactosidases whose expression is regulated by this disaccharide [59]. They, indeed, demonstrated that β -galactosidases harboring at least 7 of the 14 residues required for glucose binding in *E. coli* β -galactosidase are in proximity of LacI-GalR family repressor sequences, statistically in 60% of the cases.

The β -galactosidase from *PhTAE79* possesses 9/14 of the residues included in this motif (asterisks in Supplementary Figure S1). It is worth noting that all the amino acids directly involved in the interaction with glucose are preserved, while only some residues predicted to form the catalytic loop are divergent (798-802 in the *E. coli* enzyme). On the other hand, β -Gal from *C. perfringens*—the other well-characterized β -galactosidase whose expression is regulated by an AraC-type transcriptional regulator—is evolutionarily distant from *E. coli* and *PhTAE79* enzymes (~30% identity with the *E. coli* enzyme) and harbors only 3/14 of the residues of the motif defined for β -galactosidases classification.

In their analysis, Wheatley et al. already recognized that various β -galactosidases from marine Gammaproteobacteria are potentially capable of synthesizing allolactose, but they found that in most cases their genes are not coupled with LacI canonical regulators. They hypothesized that this feature could be the outcome of horizontal gene transfer and that the allolactose synthesis might be a vestigial feature of enzymes whose expression might no more be induced by allolactose. Nevertheless, considering that *PhTAE79 lacZ* is IPTG-inducible, as mentioned by Hoyoux et al. [8] and by us, and *C. perfringens* β gal gene is regulated by lactose but not by IPTG, we might deduce that Wheatley's considerations can be further extended to other regulatory systems. This hypothesis is limited by the little information available about AraC-like transcriptional regulators of *lacZ* genes. However, it is tantalizing to consider that part of those genes encoding β -galactosidases with an allolactose synthetic activity originally considered to be orphan of an allolactose-dependent regulation (40% of the genes belonging to this category at

the time of Wheatley’s analysis), might still be induced by this disaccharide through the action of less explored regulators.

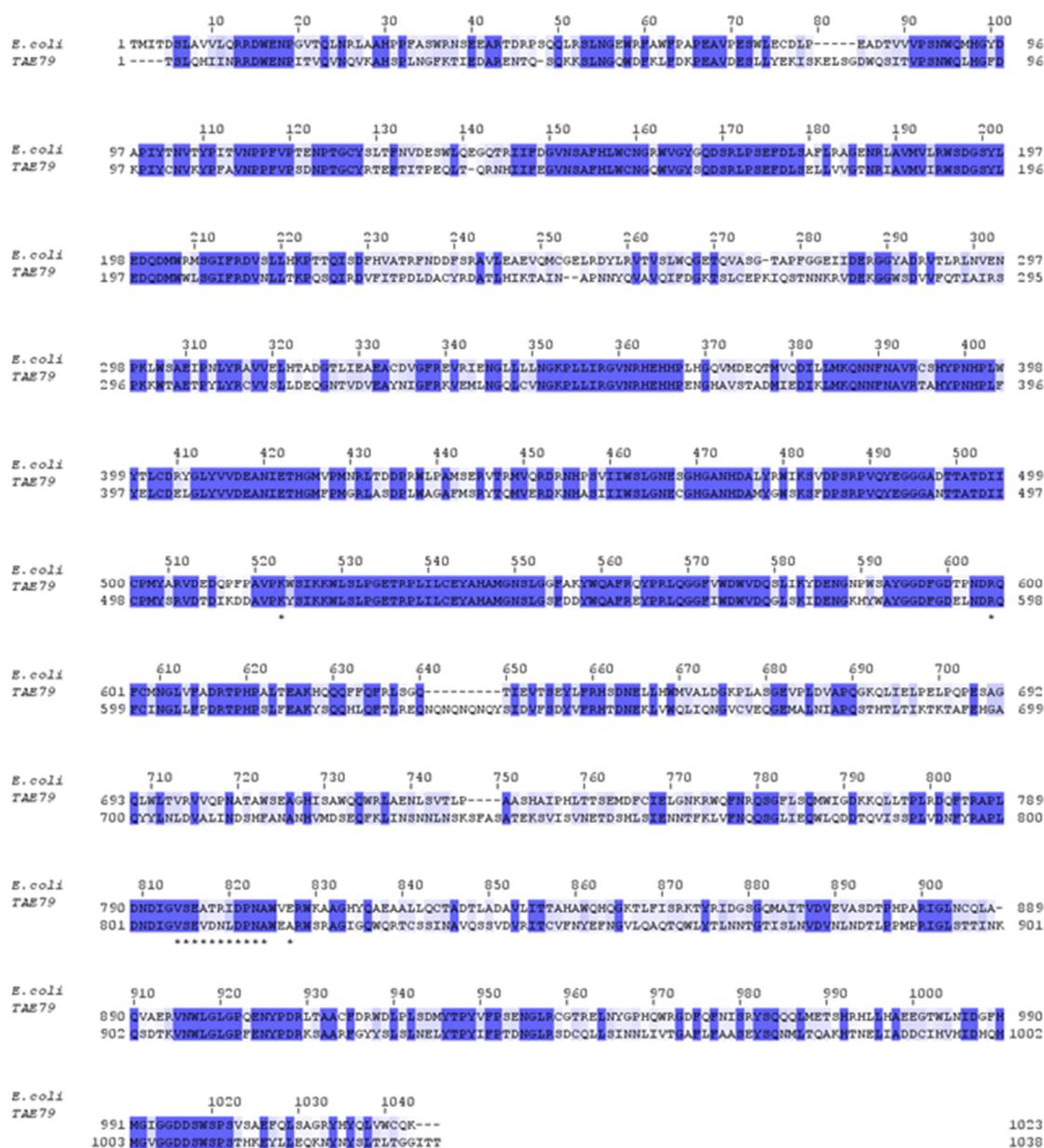


Figure A1. Protein sequence alignment of the β -galactosidases of *E. coli* and *PhTAE79*. Identical and similar amino acids are highlighted in dark and light blue, respectively. The residues involved in glucose binding for intramolecular allolactose synthesis are marked with asterisks.

References

1. Duilio, A.; Madonna, S.; Tutino, M.L.; Pirozzi, M.; Sannia, G.; Marino, G. Promoters from a cold-adapted bacterium: Definition of a consensus motif and molecular characterization of UP regulative elements. *Extremophiles* **2004**, *8*, 125–132. [CrossRef]
2. Papa, R.; Rippa, V.; Sannia, G.; Marino, G.; Duilio, A. An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125. *J. Biotechnol.* **2007**, *127*, 199–210. [CrossRef] [PubMed]
3. Sannino, F.; Giuliani, M.; Salvatore, U.; Apuzzo, G.A.; De Pascale, D.; Fani, R.; Fondi, M.; Marino, G.; Tutino, M.L.; Parrilli, E. A novel synthetic medium and expression system for subzero growth and

- recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl. Microbiol. Biotechnol.* **2016**, *101*, 725–734. [[CrossRef](#)] [[PubMed](#)]
4. Parrilli, E.; Giuliani, M.; Tutino, M.L. General Secretory Pathway from marine Antarctic *Pseudoalteromonas haloplanktis* TAC125. *Mar. Genom.* **2008**, *1*, 123–128. [[CrossRef](#)]
 5. Giuliani, M.; Parrilli, E.; Ferrer, P.; Baumann, K.; Marino, G.; Tutino, M.L. Process optimization for recombinant protein production in the psychrophilic bacterium *Pseudoalteromonas haloplanktis*. *Process. Biochem.* **2011**, *46*, 953–959. [[CrossRef](#)]
 6. Giuliani, M.; Parrilli, E.; Sannino, F.; Apuzzo, G.A.; Marino, G.; Tutino, M.L. Recombinant production of a single-chain antibody fragment in *Pseudoalteromonas haloplanktis* TAC125. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 4887–4895. [[CrossRef](#)] [[PubMed](#)]
 7. Wilmes, B.; Hartung, A.; Lalk, M.; Liebeke, M.; Schweder, T.; Neubauer, P. Fed-batch process for the psychrotolerant marine bacterium *Pseudoalteromonas haloplanktis*. *Microb. Cell Factories* **2010**, *9*, 72. [[CrossRef](#)]
 8. Hoyoux, A.; Jennes, I.; Dubois, P.; Genicot, S.; Dubail, F.; François, J.M.; Baise, E.; Feller, G.; Gerday, C. Cold-Adapted β -Galactosidase from the Antarctic Psychrophile *Pseudoalteromonas haloplanktis*. *Appl. Environ. Microbiol.* **2001**, *67*, 1529–1535. [[CrossRef](#)]
 9. Van De Voorde, I.; Goiris, K.; Syryn, E.; Bussche, C.V.D.; Aerts, G. Evaluation of the cold-active *Pseudoalteromonas haloplanktis* β -galactosidase enzyme for lactose hydrolysis in whey permeate as primary step of d-tagatose production. *Process. Biochem.* **2014**, *49*, 2134–2140. [[CrossRef](#)]
 10. Médigue, C.; Krin, E.; Pascal, G.; Barbe, V.; Bernsel, A.; Bertin, P.N.; Cheung, F.; Cruveiller, S.; D’Amico, S.; Duilio, A.; et al. Coping with cold: The genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res.* **2005**, *15*, 1325–1335. [[CrossRef](#)]
 11. Parrilli, E.; De Vizio, D.; Cirulli, C.; Tutino, M.L. Development of an improved *Pseudoalteromonas haloplanktis* TAC125 strain for recombinant protein secretion at low temperature. *Microb. Cell Factories* **2008**, *7*, 2. [[CrossRef](#)] [[PubMed](#)]
 12. Wilmes, B.; Kock, H.; Glagla, S.; Albrecht, D.; Voigt, B.; Markert, S.; Gardebrecht, A.; Bode, R.; Danchin, A.; Feller, G.; et al. Cytoplasmic and Periplasmic Proteomic Signatures of Exponentially Growing Cells of the Psychrophilic Bacterium *Pseudoalteromonas haloplanktis* TAC125. *Appl. Environ. Microbiol.* **2010**, *77*, 1276–1283. [[CrossRef](#)] [[PubMed](#)]
 13. Fondi, M.; Maida, I.; Perrin, E.; Meller, A.; Mocali, S.; Parrilli, E.; Tutino, M.L.; Lió, P.; Fani, R. Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Environ. Microbiol.* **2014**, *17*, 751–766. [[CrossRef](#)] [[PubMed](#)]
 14. Tutino, M.L.; Duilio, A.; Parrilli, E.; Remaut, E.; Sannia, G.; Marino, G. A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles* **2001**, *5*, 257–264. [[CrossRef](#)] [[PubMed](#)]
 15. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **2015**, *5*, 8365. [[CrossRef](#)]
 16. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
 17. Tutino, M.L.; Parrilli, E.; Giaquinto, L.; Duilio, A.; Sannia, G.; Feller, G.; Marino, G. Secretion of α -Amylase from *Pseudoalteromonas haloplanktis* TAB23: Two Different Pathways in Different Hosts. *J. Bacteriol.* **2002**, *184*, 5814–5817. [[CrossRef](#)]
 18. Vázquez, E.; Roldán, M.; Diez-Gil, C.; Unzueta, U.; Domingo-Espín, J.; Cedano, J.; Conchillo, O.; Ratera, I.; Veciana, J.; Daura, X.; et al. Protein nanodisk assembling and intracellular trafficking powered by an arginine-rich (R9) peptide. *Nanomedicine* **2010**, *5*, 259–268. [[CrossRef](#)]
 19. Zhang, G.; Gurtu, V.; Kain, S.R. An Enhanced Green Fluorescent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells. *Biochem. Biophys. Res. Commun.* **1996**, *227*, 707–711. [[CrossRef](#)]
 20. Puigbo, P.; Guzmán, E.; Romeu, A.; Garcia-Vallve, S. OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* **2007**, *35*, W126–W131. [[CrossRef](#)]
 21. Cusano, A.M.; Parrilli, E.; Duilio, A.; Sannia, G.; Marino, G.; Tutino, M.L. Secretion of psychrophilic $\hat{I}\pm$ -amylase deletion mutants in *Pseudoalteromonas haloplanktis* TAC125. *FEMS Microbiol. Lett.* **2006**, *258*, 67–71. [[CrossRef](#)]

22. Giuliani, M.; Parrilli, E.; Pezzella, C.; Rippa, V.; Duilio, A.; Marino, G.; Tutino, M.L. A Novel Strategy for the Construction of Genomic Mutants of the Antarctic Bacterium *Pseudoalteromonas haloplanktis* TAC125. *Adv. Struct. Saf. Stud.* **2011**, *824*, 219–233. [[CrossRef](#)]
23. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**, *29*, e45. [[CrossRef](#)]
24. Bosi, E.; Fondi, M.; Orlandini, V.; Perrin, E.; Maida, I.; De Pascale, D.; Tutino, M.L.; Parrilli, E.; Giudice, A.L.; Filloux, A.; et al. The pangenome of (Antarctic) *Pseudoalteromonas* bacteria: Evolutionary and functional insights. *BMC Genom.* **2017**, *18*, 93. [[CrossRef](#)]
25. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [[CrossRef](#)]
26. Qi, W.; Colarusso, A.; Olombrada, M.; Parrilli, E.; Patrignani, A.; Tutino, M.L.; Toll-Riera, M. New insights on *Pseudoalteromonas haloplanktis* TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies. *Sci. Rep.* **2019**, *9*, 16444. [[CrossRef](#)] [[PubMed](#)]
27. Cramer, A.; Whitehorn, E.A.; Tate, E.; Stemmer, W.P. Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nat. Biotechnol.* **1996**, *14*, 315–319. [[CrossRef](#)] [[PubMed](#)]
28. Perrin, E.; Ghini, V.; Giovannini, M.; Di Patti, F.; Cardazzo, B.; Carraro, L.; Fagorzi, C.; Turano, P.; Fani, R.; Fondi, M. Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nat. Commun.* **2020**, *11*, 3135. [[CrossRef](#)] [[PubMed](#)]
29. Mocali, S.; Chiellini, C.; Fabiani, A.; Decuzzi, S.; De Pascale, D.; Parrilli, E.; Tutino, M.L.; Perrin, E.; Bosi, E.; Fondi, M.; et al. Ecology of cold environments: New insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. *Sci. Rep.* **2017**, *7*, 839. [[CrossRef](#)]
30. Hansen, L.H.; Knudsen, S.; Sørensen, S.J. The effect of the lacY gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*. *Curr. Microbiol.* **1998**, *36*, 341–347. [[CrossRef](#)]
31. Tsilibaris, V.; Maenhaut-Michel, G.; Van Melderen, L. Biological roles of the Lon ATP-dependent protease. *Res. Microbiol.* **2006**, *157*, 701–713. [[CrossRef](#)]
32. Van Melderen, L.; Aertsen, A. Regulation and quality control by Lon-dependent proteolysis. *Res. Microbiol.* **2009**, *160*, 645–651. [[CrossRef](#)] [[PubMed](#)]
33. Rosano, G.L.; Ceccarelli, E.A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **2014**, *5*, 1–17. [[CrossRef](#)]
34. Vigentini, I.; Merico, A.; Tutino, M.L.; Compagno, C.; Marino, G. Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J. Biotechnol.* **2006**, *127*, 141–150. [[CrossRef](#)] [[PubMed](#)]
35. Unzueta, U.; Vázquez, F.; Accardi, G.; Mendoza, R.; Toledo-Rubio, V.; Giuliani, M.; Sannino, F.; Parrilli, E.; Abasolo, I.; Schwartz, S.; et al. Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: Production of human alpha-galactosidase A. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 5863–5874. [[CrossRef](#)] [[PubMed](#)]
36. Khlebnikov, A.; Keasling, J. Effect of lacY Expression on Homogeneity of Induction from the Ptac and Ptrc Promoters by Natural and Synthetic Inducers. *Biotechnol. Prog.* **2002**, *18*, 672–674. [[CrossRef](#)]
37. Özbudak, E.M.; Thattai, M.; Lim, H.N.; Shraiman, B.I.; Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **2004**, *427*, 737–740. [[CrossRef](#)]
38. Afroz, T.; Biliouris, K.; Kaznessis, Y.N.; Beisel, C.L. Bacterial sugar utilization gives rise to distinct single-cell behaviours. *Mol. Microbiol.* **2014**, *93*, 1093–1103. [[CrossRef](#)]
39. Afroz, T.; Biliouris, K.; Boykin, K.E.; Kaznessis, Y.N.; Beisel, C.L. Trade-offs in Engineering Sugar Utilization Pathways for Titratable Control. *ACS Synth. Biol.* **2014**, *4*, 141–149. [[CrossRef](#)]
40. Fernández-Castané, A.; Vine, C.E.; Caminal, G.; López, C. Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. *J. Biotechnol.* **2012**, *157*, 391–398. [[CrossRef](#)]
41. Guzman, L.M.; Belin, D.; Carson, M.J.; Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **1995**, *177*, 4121–4130. [[CrossRef](#)] [[PubMed](#)]
42. Stülke, J.; Hillen, W. Carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* **1999**, *2*, 195–201. [[CrossRef](#)]

43. Gubellini, F.; Verdon, G.; Karpowich, N.K.; Luff, J.D.; Boël, G.; Gauthier, N.; Handelsman, S.K.; Ades, S.E.; Hunt, J.F. Physiological Response to Membrane Protein Overexpression in *E. coli*. *Mol. Cell. Proteom.* **2011**, *10*, 10. [[CrossRef](#)] [[PubMed](#)]
44. Van Melderen, L.; Gottesman, S. Substrate sequestration by a proteolytically inactive Lon mutant. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6064–6071. [[CrossRef](#)] [[PubMed](#)]
45. D’Amico, S.; Collins, T.; Marx, J.-C.; Feller, G.; Gerday, C. Psychrophilic microorganisms: Challenges for life. *EMBO Rep.* **2006**, *7*, 385–389. [[CrossRef](#)]
46. Grube, M.; Dimanta, I.; Gavare, M.; Strazdina, I.; Liepins, J.; Juhna, T.; Kalnenieks, U. Hydrogen-producing *Escherichia coli* strains overexpressing lactose permease: FT-IR analysis of the lactose-induced stress. *Biotechnol. Appl. Biochem.* **2014**, *61*, 111–117. [[CrossRef](#)]
47. Juers, D.H.; Matthews, B.W.; Huber, R.E. LacZ β -galactosidase: Structure and function of an enzyme of historical and molecular biological importance. *Protein Sci.* **2012**, *21*, 1792–1807. [[CrossRef](#)]
48. Mutalik, V.K.; Guimaraes, J.C.; Cambay, G.; Lam, C.; Christoffersen, M.J.; Mai, Q.-A.; Tran, A.B.; Paull, M.; Keasling, J.D.; Arkin, A.P.; et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **2013**, *10*, 354–360. [[CrossRef](#)]
49. Mirzadeh, K.; Martinez, V.; Toddo, S.; Guntur, S.; Herrgård, M.J.; Elofsson, A.; Nørholm, M.H.H.; Daley, D.O. Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector–Coding Sequence Junction. *ACS Synth. Biol.* **2015**, *4*, 959–965. [[CrossRef](#)]
50. Boel, G.; Letso, R.; Neely, H.; Price, W.N.; Wong, K.-H.; Su, M.; Luff, J.D.; Valecha, M.; Everett, J.K.; Acton, T.B.; et al. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* **2016**, *529*, 358–363. [[CrossRef](#)]
51. Welsch, N.; Homuth, G.; Schweder, T. Suitability of different β -galactosidases as reporter enzymes in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **2011**, *93*, 381–392. [[CrossRef](#)] [[PubMed](#)]
52. Tesei, G.; Vazdar, M.; Jensen, M.R.; Cragnell, C.; Mason, P.E.; Heyda, J.; Skepö, M.; Jungwirth, P.; Lund, M. Self-association of a highly charged arginine-rich cell-penetrating peptide. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11428–11433. [[CrossRef](#)] [[PubMed](#)]
53. Verma, D.; Gulati, N.; Kaul, S.; Mukherjee, S.; Nagaich, U. Protein Based Nanostructures for Drug Delivery. *J. Pharm.* **2018**, *2018*, 1–18. [[CrossRef](#)] [[PubMed](#)]
54. Tasaki, T.; Sriram, S.M.; Park, K.S.; Kwon, Y.T. The N-end rule pathway. *Annu. Rev. Biochem.* **2012**, *81*, 261–289. [[CrossRef](#)] [[PubMed](#)]
55. Favaro, M.T.P.; Sánchez-García, L.; Sánchez-Chardi, A.; Roldan, M.; Unzueta, U.; Serna, N.; Cano-Garrido, O.; Azzoni, A.R.; Ferrer-Miralles, N.; Villaverde, A.; et al. Protein nanoparticles are nontoxic, tuneable cell stressors. *Nanomedicine* **2018**, *13*, 255–268. [[CrossRef](#)]
56. Hartman, A.H.; Liu, H.; Melville, S.B. Construction and Characterization of a Lactose-Inducible Promoter System for Controlled Gene Expression in *Clostridium perfringens*. *Appl. Environ. Microbiol.* **2011**, *77*, 471–478. [[CrossRef](#)]
57. Caron, K.; Trowell, S.C. Highly Sensitive and Selective Biosensor for a Disaccharide Based on an AraC-Like Transcriptional Regulator Transduced with Bioluminescence Resonance Energy Transfer. *Anal. Chem.* **2018**, *90*, 12986–12993. [[CrossRef](#)]
58. Newman, J.; Caron, K.; Nebl, T.; Peat, T.S. Structures of the Transcriptional Regulator BgaR, a Lactose Sensor. *Acta Crystallogr. Sect. D* **2019**, *75*, 639–646. [[CrossRef](#)]
59. Wheatley, R.W.; Lo, S.; Jancewicz, L.J.; Dugdale, M.L.; Huber, R.E. Structural Explanation for Allolactose (Lac Operon Inducer) Synthesis by LacZ β -Galactosidase and the Evolutionary Relationship between Allolactose Synthesis and the Lac Repressor. *J. Biol. Chem.* **2013**, *288*, 12993–13005. [[CrossRef](#)]



Article

Modelling hCDKL5 Heterologous Expression in Bacteria

Marco Fondi ^{1,2,*} , Stefano Gozi ¹, Mikolaj Dziurzynski ³ , Paola Turano ^{4,5} , Veronica Ghini ^{4,5},
Marzia Calvanese ⁶ , Andrea Colarusso ^{6,7} , Concetta Lauro ^{6,7} , Ermenegilda Parrilli ⁶ 
and Maria Luisa Tutino ⁶ 

- ¹ Department of Biology, University of Florence, Sesto F.no Florence, 50019 Florence, Italy; stefano.gozi@stud.unifi.it
- ² Centro Studi Dinamiche Complesse (CSDC), University of Florence, Sesto F.no Florence, 50019 Florence, Italy
- ³ Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Faculty of Biology, University of Warsaw, 02-096 Warsaw, Poland; mikolaj.dziurzynski@biol.uw.edu.pl
- ⁴ Magnetic Resonance Center (CERM) and Department of Chemistry “Ugo Schiff”, University of Florence, via Sacconi 6, Sesto Fiorentino, 50019 Fiorentino, Italy; turano@cerm.unifi.it (P.T.); veronica.ghini@unifi.it (V.G.)
- ⁵ Consorzio Interuniversitario Risonanze Magnetiche MetalloProteine (CIRMMP), via Sacconi 6, Sesto Fiorentino, 50019 Fiorentino, Italy
- ⁶ Dipartimento di Scienze Chimiche, Complesso Universitario Monte Sant’Angelo, 80126 Napoli, Italy; marzia.calvanese@gmail.com (M.C.); andrea.colarusso@unina.it (A.C.); concetta.lauro@unina.it (C.L.); erparril@unina.it (E.P.); tutino@unina.it (M.L.T.)
- ⁷ Istituto Nazionale Biostrutture e Biosistemi—I.N.B.B., Viale Medaglie d’Oro, 305-00136 Roma, Italy
- * Correspondence: marco.fondi@unifi.it



Citation: Fondi, M.; Gozi, S.; Dziurzynski, M.; Turano, P.; Ghini, V.; Calvanese, M.; Colarusso, A.; Lauro, C.; Parrilli, E.; Tutino, M.L. Modelling hCDKL5 Heterologous Expression in Bacteria. *Metabolites* **2021**, *11*, 491. <https://doi.org/10.3390/metabo11080491>

Academic Editor:
Hunter N. B. Moseley

Received: 18 June 2021
Accepted: 23 July 2021
Published: 28 July 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: hCDKL5 refers to the human cyclin-dependent kinase like 5 that is primarily expressed in the brain. Mutations in its coding sequence are often causative of hCDKL5 deficiency disorder, a devastating neurodevelopmental disorder currently lacking a cure. The large-scale recombinant production of hCDKL5 is desirable to boost the translation of preclinical therapeutic approaches into the clinic. However, this is hampered by the intrinsically disordered nature of almost two-thirds of the hCDKL5 sequence, making this region more susceptible to proteolytic attack, and the observed toxicity when the enzyme is accumulated in the cytoplasm of eukaryotic host cells. The bacterium *Pseudoalteromonas haloplanktis* TAC125 (PhTAC125) is the only prokaryotic host in which the full-length production of hCDKL5 has been demonstrated. To date, a system-level understanding of the metabolic burden imposed by hCDKL5 production is missing, although it would be crucial for upscaling of the production process. Here, we combined experimental data on protein production and nutrients assimilation with metabolic modelling to infer the global consequences of hCDKL5 production in PhTAC125 and to identify potential overproduction targets. Our analyses showed a remarkable accuracy of the model in simulating the recombinant strain phenotype and also identified priority targets for optimised protein production.

Keywords: CDKL5; genome-scale metabolic modelling; protein production

1. Introduction

The possibility to heterologously express and purify specific recombinant proteins in large amounts permits their biochemical characterisation, the development of commercial goods and their use in industrial processes. With the development of recombinant insulin and its production in *Escherichia coli* in the 1980s [1], a multi-billion dollar market was launched, leading to current large-scale applications that are nowadays capable of releasing products ranging from protein biologics to industrial enzymes [2]. Ideally, the practical steps that lead to recombinant protein production are pretty straightforward and include the identification of the gene of interest, its cloning into an expression vector, its transformation into the host of choice, the induction of protein synthesis and its final purification and characterisation [3]. The intrinsic complexity of biological systems, however,

usually poses problems down the pipeline of bacterial heterologous protein production. Indeed, as a consequence of the induction of the production of the foreign protein, the biochemistry and physiology of the host may be dramatically altered. The numerous physiological changes that may occur often lower the amount of the target foreign protein that is produced and eventually recovered from the recombinant organism [4]. In bacteria, high levels of recombinant protein production frequently lead to an impact on host cell metabolism; this is usually detectable through growth retardation and is generally known as ‘metabolic burden’ [5]. This additional metabolic load on the microbial chassis has been defined as the portion of a host cell’s resources—either in the form of energy, such as ATP or GTP, or raw materials, such as amino acids—that is required to maintain and express foreign DNA, as either RNA or protein, in the cell [4]. In *E. coli*, for example, the overexpression of an unnecessary protein results in a linear decrease in the growth rate, with the zero-growth limit occurring when the overexpressed protein occupies a mass fraction equal to $1-\phi_{fixed}$, with ϕ_{fixed} representing the growth-rate invariant fraction of the proteome [6]. There are many factors contributing to the emergence of this burden on growing cells that, at the same time, express a heterologous protein. These mainly include the transcription, translation and folding of the foreign protein [5,7,8] and the processes associated with plasmid maintenance, expression and amplification [7,8]. In addition, the expression of recombinant proteins may induce a system-level stress response that downregulates key metabolic pathway genes, leading to a decline in cellular health and feedback inhibition of both growth and protein expression [9]. Finally, from an energetic perspective, the expression of a foreign protein in a cell may use a significant fraction of its metabolic resources and precursors, removing them away from its central metabolism and placing a metabolic drain on the host [4]. Thus, upon protein production induction, an overall cellular reprogramming has to occur in order to ensure an adequate supply of energy and charged amino acids to the process of protein synthesis [9]. The identification of these system-level adjustments following heterologous protein production requires the use of computational representations of microbial metabolism that are able to consider the entire cellular metabolic network. In addition, these computational models may help identify the most suitable approaches to getting to target (protein) overproduction. Indeed, it has been recently acknowledged that the most innovative approach currently available to improving the yield of recombinant proteins, while minimising wet-lab costs, relies on the combination of in silico studies to reduce the experimental search space [10]. Among all the available in silico approaches, genome-scale metabolic models (GEMs) offer the possibility to predict a cellular phenotype from a genotype under certain environmental conditions and, importantly, to identify possible metabolic targets to improve the production of valuable compounds, while ensuring sufficiently high growth rates [11–13]. GEMs can also be used for descriptive purposes, including the identification of specific metabolic rewiring strategies following external perturbations and/or a nutrient switch [14,15]. Thus, not surprisingly, GEMs have been extensively exploited in the context of recombinant protein production, mostly with the aim of optimising either the cultivation conditions or the strain genetic background for improved recombinant protein production [16–18].

Although *E. coli* is arguably the bacterium of choice for the production of recombinant proteins, the emergence of a novel bacterial chassis is an important fact, especially considering the possible unique properties of their physiology and metabolism and the practical applications in which they are expected to outperform other microbial platforms [19]. Among them, *Pseudoalteromonas haloplanktis* TAC125 (PhTAC125), the first Antarctic bacterium in which an efficient gene expression technology was established [20], is particularly promising for a number of reasons. Firstly, several generations of cold-adapted gene expression vectors allow one to produce recombinant proteins either by constitutive or by inducible systems and to address the product towards any cell compartment or to the extracellular medium. Secondly, the development of synthetic media and efficient fermentation schemes allows upscaling the recombinant protein production in automatic bioreactors. Finally, the recently reported possibility to produce proteins within a range of tempera-

ture from 15 to -2.5 °C enhances the chances to improve the conformational quality and solubility of recombinant proteins. Up to now, PhTAC125 has been used for the production of several recombinant proteins, such as a psychrophilic β -galactosidase, *S. cerevisiae* α -glucosidase, human nerve growth factor and the lysosomal enzyme α -galactosidase A (hGLA) [21–23].

Recently, PhTAC125 was found to be a potential chassis for the production of human CDKL5 (hCDKL5). hCDKL5 is a cyclin-dependent-like protein kinase abundantly expressed in the brain, and it exerts its function in different neuron districts, such as the nucleus, the cytoplasm and the synaptosome. Mutations in the X-linked *cdkl5* gene often end up in the enzyme absence or in the production of loss-of-function variants, and both conditions are causative of hCDKL5 deficiency disorder (CDD), a rare and severe neurodevelopmental disorder for which no cure is available [24]. Recently, a protein replacement therapy was suggested, consisting of the administration of protein transduction domain (TAT)-fused hCDKL5 (TAT-CDKL5). When injected in *cdkl5*-knockout mice, TAT-CDKL5 was able to rescue many anatomical and behavioural deficits [25]. The translation of this promising therapeutic approach to clinics needs the large-scale recombinant production of TAT-CDKL5. However, full-length human CDKL5 is a difficult-to-produce enzyme for two main reasons:

- (i) Almost two-thirds of its sequence is predicted to be intrinsically disordered, and the lack of a precise 3D structure makes this region more susceptible to proteolytic attack by host-encoded proteases.
- (ii) The cytoplasmic accumulation of the enzyme in eukaryotic cells is associated to considerable toxicity, and the only permissive production strategy is its extracellular secretion, often accompanied with unwanted glycosylation [26]. PhTAC125 is the only prokaryotic cell factory in which full-length hCDKL5 production has been demonstrated, and the implementation of its efficient production process is the obligatory step towards any possible application (Calvanese et al., 2021, in press).

In this work, we modelled the heterologous production of the hCDKL5 protein in the bacterium PhTAC125. The genome-scale model of the recombinant strain was based on its original formulation [27] and further refined/updated and constrained with experimental data on hCDKL5 production and substrate consumption. This recombinant model was then used to study the global metabolic consequences of the induction of hCDKL5 production as well to identify potential targets for its overproduction.

2. Results and Discussion

2.1. An Updated Metabolic Reconstruction of PhTAC125

The latest version of the iMF721 metabolic model of *P. haloplanktis* TAC125 [27] was updated to be compatible with the current Systems Biology Markup Language Level 3 Version 2 Core specification [28] extended with the Flux Balance Constraints version 2 package specification [29]. The update was conducted using the *libsbml* Python library. It covered appropriate objective function declaration, compartment redefinition, model definition annotation with SBO terms, extension of species definitions with chemical formulas, update of gene names with the newest version of the *P. haloplanktis* genome and various minor syntax changes. The update increased the iMF721 Memote Total Score from 30% to 78% (Memote reports are available at <https://github.com/mdziurzynski/tac125-metabolic-model>, accessed on 16 July 2021). Additionally, we used BOFdat [30] to revise the original definition of the biomass composition in iMF721 using available experimental data. We also used the revised genome sequence of *P. haloplanktis* [31] and a compendium of transcriptomics data from previously published works [32] to improve the formulation of the biomass assembly reaction originally proposed [27]. After updating the model, we checked whether it could quantitatively reproduce growth phenotypes, as done by the original metabolic reconstruction. Growth simulations on defined media revealed an overall accuracy that matched the one of the original iMF721 reconstruction (Supplementary Material, Figure S1). This updated version of the model was referred to as

iMF721_v2 in subsequent sections and is available at https://github.com/mfondi/CDKL5_recombinant_production, accessed on 16 July 2021.

2.2. CDKL5 Production in Controlled Growth Conditions

Human CDKL5 was recombinantly expressed as an N-terminally His-tagged engineered construct to allow for easy Western blot detection and quantification. Its gene was expressed under the control of an IPTG-regulatable promoter [33] cloned in a high-copy-number plasmid, named pB40_79C-CDKL5 (average copy number equal to 100, manuscript in preparation), in a mutant version of PhTAC125-KrPI LacY+ capable of fast IPTG internalisation [33]. hCDKL5 synthesis was induced in the late exponential phase with 5 mM IPTG at 15 °C in bacteria grown in GG medium [34] for 8 h. Total production of the target protein was estimated to be 5.2 mg/L of the culture by Western blot using a commercial His-tagged calibrator with a similar MW as hCDKL5.

2.3. Estimation of Average hCDKL5 Production Flux and Nutrients Uptake Rates

Here, we computed the actual (average) production and growth rates from the experimental data. As for hCDKL5 (molecular weight 128,082.77 mg mmol⁻¹), after 8 h, a total amount of 5.2 mg (for 1 L of culture) was obtained. After the same amount of time, the OD of the culture was measured to be 2.55, which, when multiplied by 0.74 (i.e., the factor for converting PhTAC125 OD to grams of biomass [34]), corresponds to 1.887 g of cell dry weight (CDW). Putting everything together, we can compute the average production flux of hCDKL5 as follows:

$$[(5.2 \text{ mg}/8 \text{ h})/128,082.77 \text{ mg mmol}^{-1}]/1.887 \text{ g}_{\text{CDW}} = 2.7 \times 10^{-6} \text{ mmol/g}_{\text{CDW}} \text{ h}^{-1}$$

The average growth rate for the recombinant strain across the 8 h period was computed using initial and final OD values:

$$(\ln 2.55 - \ln 0.94) \div 8 = 0.125 \text{ h}^{-1}$$

The same calculation led to an average growth rate of 0.169 h⁻¹ for the WT strain. According to these data, the production of hCDKL5 imposes an overall burden on growing PhTAC125 cells, which leads to a 26% reduction in biomass production in the hCDKL5 strain (Figure 1A).

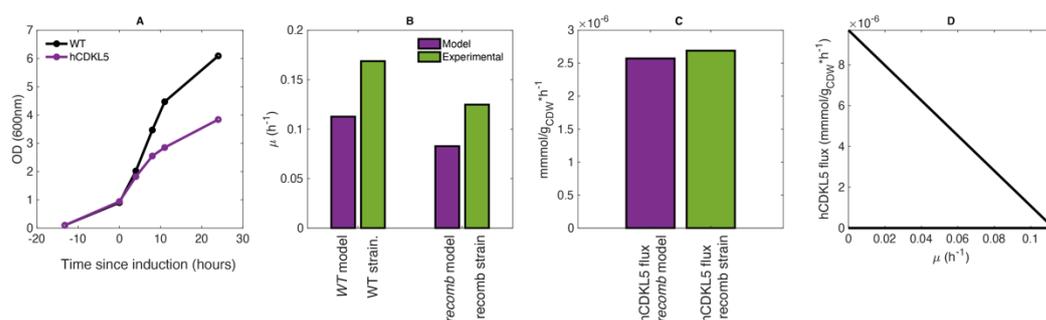


Figure 1. (A) Growth curves of WT and hCDKL5 strains, as experimentally determined. (B) Comparison between the model-predicted and measured growth rates in the wild-type strain. (C) Comparison between the measured hCDKL5 production rate in the recombinant strain and the one predicted by the model. (D) Production envelope for hCDKL5.

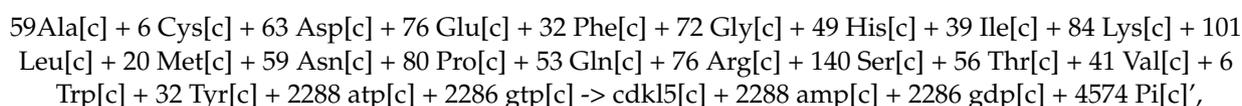
At this point, the only parameters that are missing to fully characterise the CDKL5 production dynamics are the uptake rates for glutamate and gluconate when they represent the only C sources on a minimal medium. To calculate these, we set up an ad hoc experiment (see Section 3 (Materials and Methods) and Supplementary Material) that revealed an uptake rate of 0.35 and 0.66 mmol/g_{CDW} h⁻¹ for glutamate and gluconate, respectively.

2.4. Recombinant Model Construction to Account for hCDKL5 Production

We then extended iMF721_v2 to include heterologous hCDKL5 production (leading to iMF721_v2_CDKL5 reconstruction (see Supplementary Material, Figure S3). The processes taken into account are (i) synthesis of the pB40 plasmid and (ii) synthesis of hCDKL5 mRNA and its translation into the corresponding protein sequence. As hCDKL5 is not secreted by PhTAC125, no energy-dependent hCDKL5 secretion reaction was added to the model. A plasmid copy number (Pcn) of 100 was used for pB40 because the latter is a high-copy-number plasmid. The reaction included in the metabolic network of PhTAC125 representing the synthesis of pB40 is the following:



The stoichiometric coefficients for dATP, dGTP, dCTP and dTTP were determined according to the GC composition of the pB40 plasmid. The ATP requirement for the synthesis of the pB40 plasmid was estimated based on the amount of ATP required for the synthesis of the chromosomal DNA, as previously described [16,18]. The obtained value (0.21) was multiplied by 100, the estimated copy number of pB40. Finally, pB40 was included in the biomass reaction of the model to account for the burden of the plasmid on the overall physiology of the cell. The stoichiometric coefficient of pB40 was again derived from the stoichiometric coefficient of chromosomal DNA in the biomass assembly reaction of iMF721_v2. This was done using the following proportion: $3850,272:0.001608 = 8166:100X$, where the first, second, third and fourth terms represent the size (in bp) of the PhTAC125 genome, the stoichiometric coefficient for DNA in the original formulation of the PhTAC125 biomass reaction, the length of the pB40 plasmid and the (unknown) actual stoichiometric coefficient for the 100 copies of the plasmid, respectively. This calculation led to a stoichiometric coefficient for pB40 of 0.000341. Concerning the reaction for hCDKL5 synthesis, this was formalised as follows:



where the stoichiometric coefficients for the amino acids were based on the composition of the protein sequence and the amount of ATP was computed considering the requirement of four ATP molecules for each amino acid added to the protein [35]. As said above, since hCDKL5 is not exported from the cell in vivo, no active transport reaction was included in the model.

At this point, we constrained this iMF721_v2_CDKL5 reconstruction with experimental data to build two further models, i.e., a *wt* model and a recombinant model (named *recomb* for brevity). More specifically, we constructed:

1. A *wt* model by constraining the iMF721_v2_CDKL5 reconstruction with glutamate/gluconate uptake rates to the values experimentally determined and setting the biomass assembly reaction as the BOF of the model
2. A *recomb* model by constraining the iMF721_v2_CDKL5 reconstruction with glutamate/gluconate uptake and growth rates to the values experimentally determined and setting the hCDKL5 production reaction as the BOF of the model

These two models were used for all the simulations described below. The schematic representation of the computational steps leading to the two models is reported in (Supplementary Material, Figure S3).

2.5. The PhTAC125 Recomb Model Accurately Simulates hCDKL5 Production

To account for the predictive capability of PhTAC125 reconstruction in the context of hCDKL5 production, we computed growth and hCDKL5 production rates in the *wt* and *recomb* models.

As said above, the *wt* model was obtained by setting the lower bound of glutamate and gluconate uptake reactions to 0.35 and 0.66 mmol/g_{CDW} h⁻¹, respectively, and performing an FBA simulation using biomass production as the objective function. This *wt* model predicted a growth rate of 0.119 h⁻¹, which closely resembles the one experimentally measured (Figure 1B, “WT”). Afterwards, to generate the *recomb* model, we maintained the same boundaries for the glutamate and gluconate reactions and constrained the growth rate to 74% of the optimal one predicted by the model (74% of 0.119 h⁻¹) and optimised for hCDKL5 production (Figure 1A). The simulations using this *recomb* model returned a hCDKL5 production flux of 2.67×10^{-6} mmol/g DCW h⁻¹, which accurately resembles the one measured experimentally (2.7×10^{-6} mmol/g DCW h⁻¹) (Figure 1C). A production envelope analysis correctly revealed that hCDKL5 production and biomass production compete for a common pool of nutrients and allowed us to sketch the current trade-off between these two cellular objectives (Figure 1D).

These data indicate that when constrained with experimental data, the *recomb* model is capable of providing a stoichiometrically reliable representation of hCDKL5 production in PhTAC125.

2.6. PhTAC125 Metabolic Rewiring Following hCDKL5 Induction

To explore the extent of PhTAC125 metabolic network rewiring upon the induction of hCDKL5 synthesis, we then analysed the differences in flux distributions between *wt* and *recomb* models. As expected, running an FBA simulation on the two models, we found a different number of flux-carrying reactions, with the *recomb* model showing a higher number of *core* reactions (491 vs. 484). However, since an FBA solution may not be unique (i.e., alternative flux distributions may still lead to an equally optimal solution), we used flux variability analysis (FVA) to assess the set of *core* reactions in each of the simulations (see Materials and Methods). A set of 84 *core* reactions was shared by the *wt* and *recomb* models. This set of reactions represented 74% and 97%, respectively, of the *core* reactions of the two models (i.e., the set of reactions remaining after removing the set of reactions showing a large variability range). Within this set, we identified 12 reactions (11 of them were gene encoded) shared by both models but that showed an increased flux in the *recomb* vs. the *wt* model (Table 1). The 11 gene-encoded reactions included the reactions involved in histidine biosynthesis and an ammonia transporter. The histidine biosynthetic reactions covered the entire pathway, i.e., from 5-phosphoribosyl 1-pyrophosphate (PRPP) to histidine. The higher flux predicted in the histidine biosynthetic pathway of the *recomb* model vs. the *wt* model can be explained by the different amino acid composition of recombinant protein with respect to the native PhTAC125 proteome (Figure 2). Indeed, as the abundance of this amino acid is double in hCDKL5 with respect to the PhTAC125 proteome, precursors used to produce histidine in the *recomb* model will be drained faster than in the *wt* model and fluxes around those precursors are expected to be significantly altered [36].

Table 1. Reactions showing an increased flux in the *recomb* vs. *wt* model simulations.

Reaction Model Code	Subsystem	Reaction Name
rxn00789	Histidine metabolism	1-(5-Phospho-D-ribosyl)-ATP:pyrophosphate phosphoribosyl-transferase
rxn00863	Histidine metabolism	L-histidinal:NAD + oxidoreductase
rxn02159	Histidine metabolism	L-histidinol:NAD + oxidoreductase
rxn02160	Histidine metabolism	L-histidinol-phosphate phosphohydrolase
rxn02320	Histidine metabolism	5-Amino-2-oxopentanoate:2-oxoglutarate aminotransferase
rxn02473	Histidine metabolism	D-erythro-1-(imidazol-4-yl)glycerol 3-phosphate hydrolyase
rxn02834	Histidine metabolism	Phosphoribosyl-ATP pyrophosphohydrolase
rxn02835	Histidine metabolism	1-(5-Phospho-D-ribosyl)-AMP 1,6-hydrolase
rxn03135	Histidine metabolism	Imidazole-glycerol-3-phosphate synthase
rxn03175	Histidine metabolism	N-(5'-phospho-D-ribosylformimino)-5-amino-1-
rxn05466	Ammonia transport	Ammonia transport via diffusion

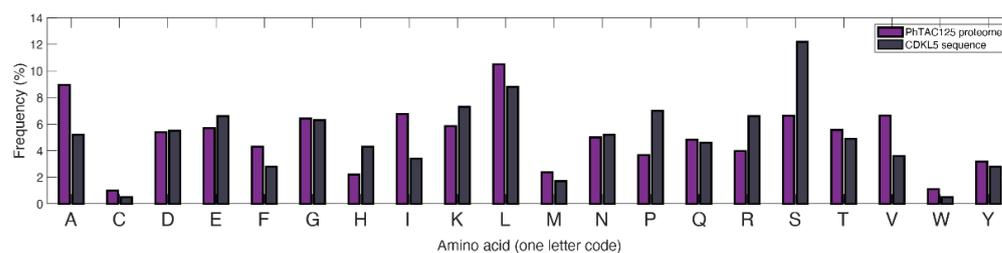


Figure 2. Difference in amino acid composition between the PhTAC125 proteome and CDKL5.

2.7. Finding the Optimal Growth Medium

We then sought to identify potential carbon sources whose inclusion in the original, optimised medium could boost the production of hCDKL5. To this purpose, we selected all the transport reactions present in the iMF721_v2 metabolic reconstruction and created a list including the transported compounds. We considered PhTAC125 as capable of taking up these compounds inside the cell because its genome encodes the corresponding transporters. We then performed one simulation for each of these compounds, adding it to the defined medium used during the previous simulations (Schatz salts plus glutamate and gluconate; see Materials and Methods), constraining the growth rate to the experimentally determined value and using hCDKL5 production as the objective function. In these simulations, the uptake rate of the extra carbon source was arbitrarily set to $0.5 \text{ mmol/g}_{\text{CDW}} \text{ h}^{-1}$. We then estimated the effect of the amended carbon source by computing the ratio between the hCDKL5 production flux in the new carbon source and the original one (i.e., with no amendments) and selected the first 30 compounds in the list (Figure 3).

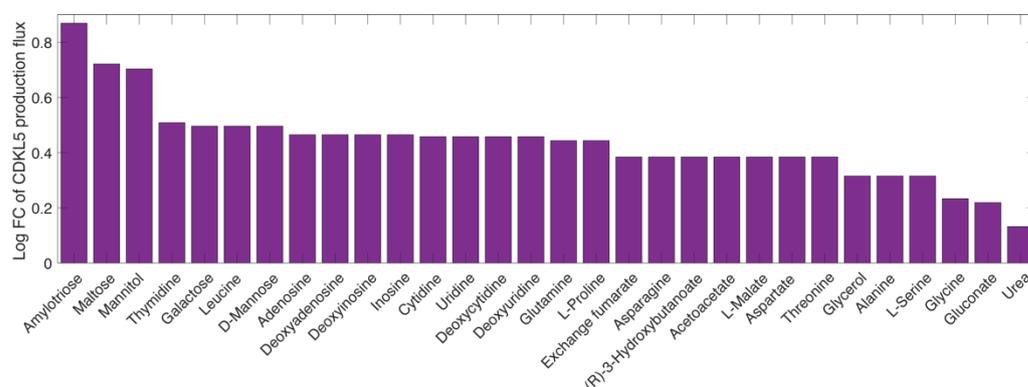


Figure 3. The effect of putative nutrients to be added to GG medium on the CDKL5 production flux. The y -axis indicates the log fold change of the CDKL5 production flux with respect to the *recomb* model grown on GG medium.

The three most promising compounds identified in our analysis were amylotriose, maltose and mannitol. The first catabolic steps of these three compounds led to the formation of D-glucose (amylotriose and maltose) or D-fructose in PhTAC125, thus suggesting that the strengthening of sugar metabolism might be the primary effect of adding these compounds to the growth medium of the recombinant strain and one of the possible ways to increase hCDKL5 production. To better address this point, we further investigated which part of the PhTAC125 metabolic network is specifically affected by the amendment of the best-performing nutrients to the growth medium. We thus checked which reactions increased their flux in the recombinant model growing in GG medium plus amylotriose compared to the same model grown in simple GG medium (Table 2). This list of reactions was filtered by removing those (*non-core*) reactions showing more than 30% variation between their minimum and maximum fluxes during an FVA, as described in Materials and Methods. Overall, we found 15 gene-encoded reactions displaying an increased flux in this condition: the majority of them (11) were involved in histidine biosynthesis;

3 in phenylalanine, tyrosine and tryptophan biosynthesis; and 1 in riboflavin metabolism. We found a similar scenario (i.e., the same involved pathways) for the other top four nutrients (maltose, mannitol, thymidine and galactose), with a majority of histidine and phenylalanine metabolism-related enzymes displaying an increased flux in the amended medium. Additional pathways that might be affected by these nutrients include nicotinate and nicotinamide metabolism (found when simulating the amendment of thymidine) and galactose metabolism (found when simulating the amendment of galactose).

Table 2. Reactions showing an increased flux in the *recomb* model growing in GG medium amended with amylotriase.

Reaction Model Code	Subsystem	Reaction Name
rxn05466	Ammonium transporter	Ammonia transport via diffusion
rxn02160	Histidine metabolism	L-histidinol-phosphate phosphohydrolase
rxn00863	Histidine metabolism	L-histidinal:NAD + oxidoreductase
rxn02159	Histidine metabolism	L-histidinol:NAD + oxidoreductase
rxn02834	Histidine metabolism	Phosphoribosyl-ATP pyrophosphohydrolase
rxn03175	Histidine metabolism	N-(5'-phospho-D-ribosylformimino)-5-amino-1-
rxn02473	Histidine metabolism	D-erythro-1-(imidazol-4-yl)glycerol 3-phosphate hydro-lyase
rxn02508	Phenylalanine, tyrosine and tryptophan biosynthesis	N-(5-phospho-beta-D-ribosyl)anthranilate ketol-isomerase
rxn02320	Histidine metabolism	5-Amino-2-oxopentanoate:2-oxoglutarate aminotransferase
rxn02507	Phenylalanine, tyrosine and tryptophan biosynthesis	1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose-5-phosphate
rxn00789	Histidine metabolism	1-(5-Phospho-D-ribosyl)-ATP:pyrophosphate phosphoribosyl-transferase
rxn03135	Histidine metabolism	Imidazole-glycerol-3-phosphate synthase
rxn00791	Phenylalanine, tyrosine and tryptophan biosynthesis	N-(5-phospho-D-ribosyl)anthranilate:pyrophosphate
rxn02835	Histidine metabolism	1-(5-Phospho-D-ribosyl)-AMP 1,6-hydrolase
rxn00392	Riboflavin metabolism	ATP:riboflavin 5'-phosphotransferase

Taken together, these results suggest that the main effect of adding extra nutrients to the medium would be an increased availability of histidine molecules inside the cell, which, in turn, would result in an improved production rate of hCDKL5. Again, this can be explained by the different histidine content of the overall PhTAC125 proteome and of the hCDKL5 sequence (Figure 2). The enzymes involved in phenylalanine, tyrosine and tryptophan biosynthesis that appear to increase their flux in the tested conditions include those responsible for the generation of 5-phospho-alpha-D-ribose 1-diphosphate (PRPP), which is a key pentose phosphate pathway (PPP) intermediate for purine, pyrimidine and histidine biosynthesis. Its connection to increased hCDKL5 production might thus be double: on the one side, it could fuel histidine biosynthesis for the reasons described above; on the other side, it could facilitate the synthesis of purines and pyrimidines required by plasmid replication and transcription during heterologous protein expression.

2.8. Finding Hypothetical Targets for hCDKL5 Overproduction

We then used the model to predict possible targets to improve the production of hCDKL5 in *P. haloplanktis* TAC125. We focused our attention on the use of the well-established FSEOF algorithm [37]. Briefly, FSEOF scans all the fluxes in the reconstruction and identifies the increasing ones when the flux towards product formation is set (enforced) as a further constraint during FBA. The reactions identified by FSEOF are primary overexpression targets that may lead to improved synthesis of the desired target (hCDKL5 in our case). By applying FSEOF, as described in Materials and Methods, we identified 70 target gene-encoded reactions whose overexpression may lead to improved target production. The complete list of these reactions is available in Supplementary Material. The top 10 target reactions identified by FSEOF are shown in Table 3. The first reaction in the list is represented by rxn05937, catalysing the formation of NADPH from NADP and reduced ferredoxin. Forcing the flux through this reaction would allow increasing the overall

NADPH pool of the cell, and this has been widely recognised as an important factor in the process of heterologous protein production in microorganisms [38]. Reduced ferredoxin necessary for the production of NADPH might be provided by L-glutamateferredoxin oxidoreductase (Table 3), catalysing the conversion of L-glutamate to L-glutamine with the reduction of ferredoxin. Reactions belonging to the Entner–Doudoroff (ED) branch of the PPP are also high-ranking overexpression targets according to FSEOF (Table 3, Figure 4A). These include the three enzymes catalysing the conversion to D-glucono-1,5-lactone 6-phosphate to pyruvate and D-glyceraldehyde 3-phosphate (encoded by *agal*, *edd* and *eda*). Overall, the degradation of one molecule of glucose through this pathway, as opposed to classical PPP leading to ribose-5P, leads to lower amounts of reducing equivalents (one NADPH produced instead of two) but ensures a greater and balanced production of precursors (namely pyruvate and glyceraldehyde-3P (G3P)) that can be used both to fuel the TCA cycle and for amino acid biosynthesis [39]. Indeed, it is known that the ED pathway, as a variant glycolysis pathway, produces equal amounts of G3P and pyruvate, and this superior stoichiometric feature makes the ED pathway a preferable route for precursor supply [40]. Importantly, targets within these metabolic pathways (i.e., ED and PPP in general) have been identified in other works aimed at identified optimisation production strategies [18,19,39,41]. Most of the other reactions identified by the FSEOF algorithm are involved in the metabolism of amino acids. In particular, our simulations suggest that the production of hCDKL5 might be improved by redirecting the catabolism of glutamate towards the production of aspartate (through the action of L-aspartate-2-oxoglutarate aminotransferase) and its subsequent conversion to 4-phospho-L-aspartate and L-aspartate-4-semialdehyde (Figure 4B), catalysed by ATPL-aspartate-4-phosphotransferase and L-aspartate-4-semialdehyde: NADP + oxidoreductase, respectively. L-aspartate-4-semialdehyde, in particular, serves as a substrate for the biosynthesis of many amino acids, including lysine, threonine and glycine (Figure 4B). Finally, our FSEOF simulation identified the enzyme serine O-acetyltransferase (catalysing the formation of serine from CoA and O-acetyl-L-serine) as a likely hCDKL5 overproduction target. Looking at the unbalanced distribution of S residues in the sequence of hCDKL5 with respect to the one of the PhTAC125 genome (Figure 2), it can be hypothesised that the meaning of this latter finding resides in the necessity to increase the production of serine to cope with the higher request of this amino acid following the induction of CDKL5 production.

Table 3. Top 10 reaction targets predicted by the FSEOF algorithm.

Reaction Model Code	Subsystem	Reaction Name	Formula
rxn05937	NA	Ferredoxin:NADP+ oxidoreductase	$\text{NADP} + \text{H}^+ + \text{reduced ferredoxin} \Rightarrow \text{NADPH} + \text{oxidised ferredoxin}$
rxn12822	Glyoxylate and dicarboxylate metabolism	L-glutamateferredoxin oxidoreductase (transaminating)	$2 \text{ L-glutamate} + 2 \text{ oxidised ferredoxin} \Rightarrow 2\text{-oxoglutarate} + \text{L-glutamine} + 2 \text{ H}^+ + 2 \text{ reduced ferredoxin}$
rxn01477	PPP	6-Phospho-D-gluconate hydro-lyase (edd)	$6\text{-Phospho-D-gluconate} \Rightarrow \text{H}_2\text{O} + 2\text{-keto-3-deoxy-6-phosphogluconate}$
rxn03884	PPP	2-Dehydro-3-deoxy-D-gluconate-6-phosphate D-glyceraldehyde-3-phosphate-lyase (eda)	$2\text{-Keto-3-deoxy-6-phosphogluconate} \Rightarrow \text{pyruvate} + \text{glyceraldehyde-3-phosphate}$
rxn01476	PPP	6-Phospho-D-glucono-1,5-lactone lactonohydrolase (Agal)	$\text{H}_2\text{O} + 6\text{-phospho-D-glucono-1-5-lactone} \Rightarrow \text{H}^+ + 6\text{-phospho-D-gluconate}$

Table 3. Cont.

Reaction Model Code	Subsystem	Reaction Name	Formula
rxn00260	Alanine, aspartate and glutamate metabolism	L-aspartate 2-oxoglutarate aminotransferase	2-Oxoglutarate + L-aspartate <= L-glutamate + oxaloacetate
rxn00337	Glycine, serine and threonine metabolism	ATPL-aspartate 4-phosphotransferase	ATP + L-aspartate => ADP + 4-phospho-L-aspartate
rxn01643	Glycine, serine and threonine-cysteine and methionine-lysine metabolism	L-aspartate 4-semialdehyde:NADP+ oxidoreductase (phosphorylating)	NADP + phosphate + L-aspartate 4-semialdehyde <= NADPH + 4-phospho-L-aspartate
rxn00285	Citrate cycle (TCA cycle)	Succinate-CoA ligase (ADP forming)	ATP + CoA + succinate => ADP + phosphate + succinyl-CoA
rxn00423	Cysteine and methionine metabolism	Serine O-acetyltransferase	Acetyl-CoA + L-serine <= CoA + O-acetyl-L-serine

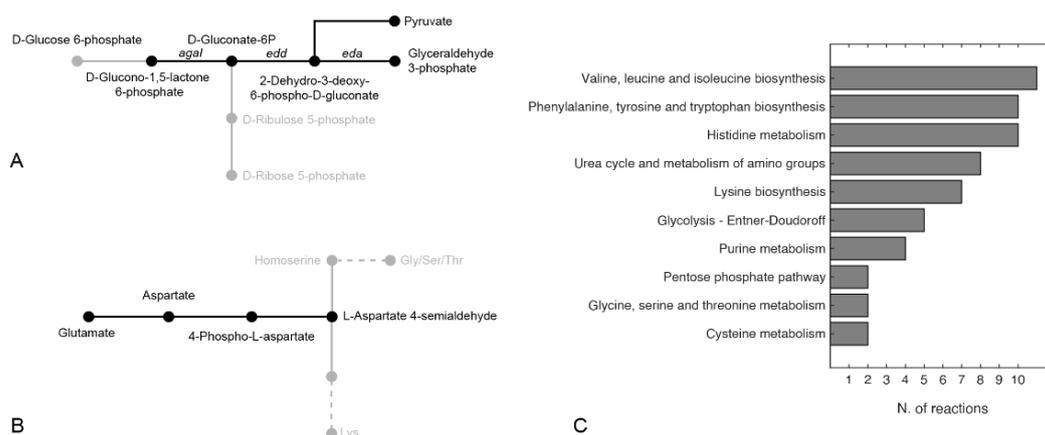


Figure 4. Pathways including reactions identified as potential overexpression targets by the FSEOF algorithm for (A) the Entner–Doudoroff pathway and (B) glutamate catabolism. Only pathways including 2 or more reactions are shown in (C).

Further, to provide a general view of the reactions identified as potential overexpression targets by FSEOF, we grouped them according to their corresponding metabolic pathway (Figure 4C). In line with the results illustrated above, 6 pathways (out of 10 with more than two reactions included) were representatives of amino acid metabolism, with 4 of them appearing in the top five pathways (i.e., Val/Leu/Ile, Phe/Tyr/Trp and Lys biosynthesis and His metabolism). In addition to amino acid metabolism, the other pathways represented were urea and amino group metabolism, glycolysis, the PPP and purine metabolism.

3. Materials and Methods

3.1. Bacterial Strains and Conjugation Experiments

The pB40_79C-CDKL5 plasmid was mobilised from *E. coli* S17-1(λ pir) to KrPL LacY+ [33] through standard conjugation techniques [41]. *E. coli* S17-1(λ pir)—a strain possessing *mob* and *tra* genes for plasmid mobilisation [42]—was routinely grown in LB (10 g/L of bacto-tryptone, 5 g/L of yeast extract, 10 g/L of NaCl) at 37 °C with the supplementation of 34 μ g/mL of chloramphenicol, if needed, for plasmid selection. KrPL LacY+, a *P. haloplanktis* TAC125 strain engineered for improved IPTG uptake [33], was grown at 15 °C in TYP (16 g/L of bacto-tryptone, 16 g/L of yeast extract, 10 g/L of NaCl) for conjugational experiments and initial pre-inocula. Recombinant KrPL LacY+ was selected with

25 and 12.5 µg/mL of chloramphenicol in liquid and solid media, respectively. Solid LB and TYP broths were prepared by the addition of 15 g/L of agar.

3.2. hCDKL5 Production

The pB40_79C-CDKL5 plasmid allows the IPTG-inducible expression of a PhTAC125 codon optimised gene coding for an engineered variant of human CDKL5 isoform 1. The translated protein possesses tandem His-Sumo [43] and TatK [25,44] N-terminal tags and a C-terminal 3xflag. The whole 1144 aa sequence was expressed as a cytosolic protein from the pB40 plasmid, which is characterised by an average copy number of 100 (manuscript in preparation). For recombinant gene expression, KrPL LacY+ was cultivated at 15 °C in a 100 mL Erlenmeyer flask containing 20 mL of GG medium [34]: 10 g/L of L-glutamic acid monosodium salt monohydrate, 10 g/L of gluconic acid sodium salt, 10 g/L of NaCl, 1 g/L of NH₄NO₃, 1 g/L of KH₂PO₄, 0.2 g/L of MgSO₄·7H₂O, 5 mg/L of FeSO₄·7H₂O and 5 mg/L of CaCl₂·2H₂O (pH 7.8). After inoculating at 0.10 OD₆₀₀, bacterial growth was followed for 13 h and the recombinant gene expression triggered at 1.00 OD₆₀₀ with 5 mM IPTG. Eight hours after induction, the bacterial cells were harvested by centrifugation (4 °C, 4000× g, 20 min) when they reached 2.55 OD₆₀₀. To check and estimate hCDKL5 intracellular production at the end of the culture, bacterial pellets equivalent to 1.00 OD₆₀₀ were resuspended in 60 µL of Laemmli Buffer 4× and denatured at 90 °C for 20 min. Denatured cellular extracts equivalent to 1/120 OD₆₀₀ were loaded onto a 7.5% precast Mini-Protean TGX (BioRad Laboratories, Hercules, CA, USA) and resolved by SDS-PAGE. Known amounts of His-Neuropilin (110 kDa; Immunological Sciences, Rome, Italy) were loaded onto adjacent lanes to develop a calibration curve. Then, separated proteins were transferred to a PVDF membrane using a semi-dry system, and His-tagged proteins (hCDKL5 and His-Neuropilin) were detected with an HRP-conjugated anti-His antibody (1:2000; Sigma-Aldrich) using the enhanced chemiluminescence (ECL) kit (BioRad, Hercules, CA, USA) and a ChemiDoc MP Imaging System (BioRad, Hercules, CA, USA). Quantitative analyses of blotted hCDKL5 and His-Neuropilin were carried out using Image Lab software (BioRad, Hercules, CA, USA), and the volumetric yield was derived considering the final biomass concentration (OD₆₀₀: 2.55).

3.3. Glutamate and Gluconate Consumption Experiment

PhTAC125 bacterial culture was grown in GG medium modified so to contain 5 g/L of L-glutamic acid monosodium salt monohydrate and 5 g/L of D-gluconic acid sodium salt in a stirred tank reactor with a 3 L fermenter (Applikon, Schiedam, The Netherlands) with a working volume of 1.5 L. The bioreactor was equipped with standard pH, pO₂, level and temperature sensors for bioprocess monitoring. Culture was carried out at 15 °C for 30 h under aerobic conditions (45% dissolved oxygen). Next, 1 mL samples for metabolomic analysis were collected during growth and centrifuged at 1300 rpm for 20 min at 4 °C. After centrifugation, supernatants were recovered, filtered through membranes with a pore diameter of 0.22 µm and stored at −80 °C.

3.4. Metabolomic Data

Metabolomic data on cell growth media were obtained by 1H nuclear magnetic resonance (NMR) spectroscopy. The supernatant samples were thawed at room temperature. Next, 540 µL of each sample was added with 60 µL of potassium phosphate buffer (1.5 M K₂HPO₄, 100% (v/v) 2H₂O, 10 mM sodium trimethylsilyl [2,2,3,3-²H₄] propionate (TMSP), pH 7.4). The mixture was transferred into 5 mm NMR tubes for subsequent analysis.

Spectral acquisition and processing were performed according to standard procedures [45,46]. One-dimensional (1D) 1H NMR spectra were recorded using a Bruker 600 MHz spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm PATXI 1H-13C-15N and 2H-decoupling probe, including a z-axis gradient coil, automatic tuning and matching and an automatic and refrigerate sample changer (SampleJet, Bruker BioSpin GmbH, Rhein-

stetten, Germany). A BTO 2000 thermocouple served for temperature stabilisation at the level of ~ 0.1 K at the samples. Before measurement, samples were kept for 5 min inside the NMR probe head for temperature equilibration at 300 K.

NMR spectra were acquired with water peak suppression using the 1D standard NOESY pulse sequence (128 scans, 65,536 data points, spectral width of 12,019 Hz, acquisition time of 2.7 s, relaxation delay of 4 s and mixing time of 0.01 s).

The raw data were multiplied by 0.3 Hz exponential line broadening before applying Fourier transformation. Transformed spectra were automatically corrected for phase and baseline distortions. All spectra were then calibrated to the reference signal of TMS at $\delta = 0.00$ ppm using TopSpin 3.5 (Bruker BioSpin GmbH, Rheinstetten, Germany).

The signals deriving from glutamate and gluconate were assigned using an internal NMR spectral library of pure organic compounds; matching between the present NMR spectra and the NMR spectral library was performed using AssureNMR software (Bruker BioSpin GmbH, Rheinstetten, Germany). Their concentrations were calculated by integrating the corresponding signals in the defined spectral range using a home-made R 3.0.2 script.

3.5. PhTAC125 Genome-Scale Metabolic Reconstruction and Constraint-Based Simulations

The original *P. haloplanktis* TAC125 genome-scale metabolic reconstruction [27] was used as the starting point of the modelling procedures. This metabolic reconstruction was then updated and quality-checked, as described above, using BOFdat [30] (for the biomass reaction) and Memote [47] (model consistency evaluation).

The recently published genome sequence of *P. haloplanktis* TAC125 [31] was fed into BOFdat *DNA.py* script in order to generate the updated stoichiometric coefficients for As, Ts, Cs, and Gs. Similarly, a compendium of expression (RNAseq) data from previously published [32] datasets was fed into the BOFdat *RNA.py* code in order to generate revised and experimentally based stoichiometric coefficients for RNA building blocks.

Constraint-based simulation (e.g., FBA) were performed using COBRA Toolbox v3.0 [48] in MATLAB 2020b and using Gurobi as a solver. Overexpression targets were identified using the latest FSEOF version implemented in Raven [49] and selecting 100 iterations and a ratio coefficient of the optimal target reaction flux of 0.9. The codes used to run all the simulations are available at https://github.com/mfondi/CDKL5_recombinant_production, accessed on 26 July 2021.

3.6. Identification of Core Reactions

Flux variability analysis (FVA) was used to assess the relevance of each reaction when simulating growth and hCDKL5 production. The *fluxVariability* function of the COBRA toolbox was used for this purpose. The following procedure was applied (separately) to both *wt* and *recomb* models. First, an FBA optimisation was run on the model to predict the flux across each reaction. Afterwards, an FVA simulation with exactly the same constraints as the previous FBA simulation was performed and the flux range for each reaction stored. Then, for each of the two models, only those reactions satisfying the following criterion were labelled as *core* reactions:

with $sol_{wt/recomb} > 0$:

$$f_{min, FVA} > 0.7 \times sol_{wt/recomb} \text{ AND } f_{max, FVA} < 1.3 \times sol_{wt/recomb}$$

with $sol_{wt/recomb} < 0$:

$$f_{max, FVA} < 0.7 \times sol_{wt/recomb} \text{ AND } f_{min, FVA} > 1.3 \times sol_{wt/recomb}$$

with sol_{wt} , $f_{min, FVA}$ and $f_{max, FVA}$ representing the FBA solution, the lower FVA solution value and the upper FVA solution value, respectively. According to this strategy, in each simulation, only those (*core*) reactions displaying a flux value different from zero and

with a narrow range of admissible flux (30%) during an FVA simulation were maintained, whereas those not satisfying this condition were considered unreliable and filtered away.

4. Conclusions

In this work, we combined experimental and computational approaches to characterising the production of recombinant hCDKL5 in the Antarctic marine bacterium *P. haloplanktis* TAC125. By constraining an updated genome-scale metabolic model of this bacterium with experimentally determined nutrient absorption rates, we were able to predict hCDKL5 production rates that matched those determined experimentally and to correctly estimate the burden (in terms of a reduction in biomass yield, about 25% compared to the *wt* strain) of protein production in this bacterium. Next, we used the model to describe the metabolic rewiring occurring in this bacterium upon the induction of hCDKL5 production and to identify possible overproduction strategies (both in terms of amendments to the original growth medium and in terms of overexpression targets). Despite the fact that each of these analyses highlighted specific pathways and/or targets that appear to be strongly connected to hCDKL5 production, common trends could be identified (e.g., the role played by reactions belonging to histidine metabolism and to the PPP). Taken together, our findings suggest that a possible future strategy for increasing the production of hCDKL5 in PhTAC125 may involve the overexpression of the target genes identified by the FSEOF algorithm and/or growth of the recombinant cells in media amended with one (or more) of the compounds that our simulations identified as the most promising in increasing the yield of the heterologous protein.

Work is currently in progress to experimentally verify both the hCDKL5 overproduction targets and the hypothetical amendments to the PhTAC125 growth medium capable of increasing the protein yield in silico.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/metabo11080491/s1>, Figure S1: Comparison among predicted growth rates of different versions of PhTAC125 metabolic reconstructions; Figure S2: Growth and glutamate and gluconate consumption of PhTAC125 cells over an 8 h period; Figure S3: Pipeline used to obtain *wt* and *recomb* metabolic reconstructions; Table S1. WT and CDKL5 strain growth parameters over an 8 h growth period; Table S2: Glutamate and gluconate consumption parameters during growth in GG medium with an initial volume of 1.6 l; Table S3: Complete list of target reactions identified through FSEOF.

Author Contributions: Conceptualisation, M.F.; methodology, M.F., E.P., M.L.T., P.T., V.G., S.G. and M.D.; formal analysis, M.F. and S.G.; data curation, M.F., M.D., S.G., M.C., A.C. and C.L.; writing—original draft preparation, M.F.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by a research grant from the University of Pennsylvania Orphan Disease Center on behalf of the Loulou Foundation to MLT (pilot award no. CDKL5-20-101-08) and by the Italian Parents' Association 'La fabbrica dei sogni 2—new developments for Rett syndrome' (to A.C., M.C., C.L., E.P. and M.L.T.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available as Supplementary Material and at https://github.com/mfondi/CDKL5_recombinant_production.

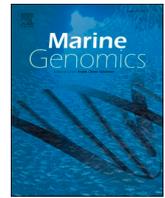
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goeddel, D.V.; Kleid, D.G.; Bolivar, F.; Heyneker, H.L.; Yansura, D.G.; Crea, R.; Hirose, T.; Kraszewski, A.; Itakura, K.; Riggs, A.D. Expression in *Escherichia Coli* of Chemically Synthesized Genes for Human Insulin. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 106–110. [[CrossRef](#)]
2. Burdette, L.A.; Leach, S.A.; Wong, H.T.; Tullman-Ercek, D. Developing Gram-Negative Bacteria for the Secretion of Heterologous Proteins. *Microb. Cell Factories* **2018**, *17*, 196. [[CrossRef](#)]

3. Rosano, G.L.; Ceccarelli, E.A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, *5*, 172. [[CrossRef](#)] [[PubMed](#)]
4. Glick, B.R. Metabolic Load and Heterologous Gene Expression. *Biotechnol. Adv.* **1995**, *13*, 247–261. [[CrossRef](#)]
5. Li, Z.; Rinas, U. Recombinant Protein Production Associated Growth Inhibition Results Mainly from Transcription and Not from Translation. *Microb. Cell Factories* **2020**, *19*, 83. [[CrossRef](#)] [[PubMed](#)]
6. Scott, M.; Hwa, T. Bacterial Growth Laws and Their Applications. *Curr. Opin. Biotechnol.* **2011**, *22*, 559–565. [[CrossRef](#)] [[PubMed](#)]
7. Hoffmann, F.; Rinas, U. Stress Induced by Recombinant Protein Production in Escherichia coli. In *Physiological Stress Responses in Bioprocesses*; Advances in Biochemical Engineering/Biotechnology; Springer: Berlin/Heidelberg, Germany, 2004; Volume 89, pp. 73–92. ISBN 978-3-540-20311-7.
8. Grabherr, R.; Nilsson, E.; Striedner, G.; Bayer, K. Stabilizing Plasmid Copy Number to Improve Recombinant Protein Production. *Biotechnol. Bioeng.* **2002**, *77*, 142–147. [[CrossRef](#)]
9. Sharma, A.K.; Shukla, E.; Janoti, D.S.; Mukherjee, K.J.; Shiloach, J. A Novel Knock out Strategy to Enhance Recombinant Protein Expression in Escherichia Coli. *Microb. Cell Factories* **2020**, *19*, 148. [[CrossRef](#)] [[PubMed](#)]
10. Santos, J.; Cardoso, M.; Moreira, I.S.; Gonçalves, J.; Correia, J.D.G.; Verde, S.C.; Melo, R. Integrated in Silico and Experimental Approach towards the Design of a Novel Recombinant Protein Containing an Anti-HER2 ScFv. *IJMS* **2021**, *22*, 3547. [[CrossRef](#)]
11. Fang, X.; Lloyd, C.J.; Palsson, B.O. Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications. *Nat. Rev. Microbiol.* **2020**, *18*, 731–743. [[CrossRef](#)]
12. Zielinski, D.C.; Patel, A.; Palsson, B.O. The Expanding Computational Toolbox for Engineering Microbial Phenotypes at the Genome Scale. *Microorganisms* **2020**, *8*, 2050. [[CrossRef](#)]
13. Gu, C.; Kim, G.B.; Kim, W.J.; Kim, H.U.; Lee, S.Y. Current Status and Applications of Genome-Scale Metabolic Models. *Genome Biol.* **2019**, *20*, 121. [[CrossRef](#)] [[PubMed](#)]
14. Presta, L.; Bosi, E.; Mansouri, L.; Dijkshoorn, L.; Fani, R.; Fondi, M. Constraint-Based Modeling Identifies New Putative Targets to Fight Colistin-Resistant *A. Baumannii* Infections. *Sci. Rep.* **2017**, *7*, 3706. [[CrossRef](#)] [[PubMed](#)]
15. Fondi, M.; Bosi, E.; Presta, L.; Natoli, D.; Fani, R. Modelling Microbial Metabolic Rewiring during Growth in a Complex Medium. *BMC Genom.* **2016**, *17*, 970. [[CrossRef](#)]
16. Sohn, S.B.; Graf, A.B.; Kim, T.Y.; Gasser, B.; Maurer, M.; Ferrer, P.; Mattanovich, D.; Lee, S.Y. Genome-Scale Metabolic Model of Methylophilic Yeast *Pichia Pastoris* and Its Use for in Silico Analysis of Heterologous Protein Production. *Biotechnol. J.* **2010**, *5*, 705–715. [[CrossRef](#)]
17. Nocon, J.; Steiger, M.G.; Pfeffer, M.; Sohn, S.B.; Kim, T.Y.; Maurer, M.; Rußmayer, H.; Pflügl, S.; Ask, M.; Haberhauer-Troyer, C.; et al. Model Based Engineering of *Pichia Pastoris* Central Metabolism Enhances Recombinant Protein Production. *Metab. Eng.* **2014**, *24*, 129–138. [[CrossRef](#)]
18. Lule, I.; D’Huys, P.-J.; Van Mellaert, L.; Anné, J.; Bernaerts, K.; Van Impe, J. Metabolic Impact Assessment for Heterologous Protein Production in *Streptomyces Lividans* Based on Genome-Scale Metabolic Network Modeling. *Math. Biosci.* **2013**, *246*, 113–121. [[CrossRef](#)]
19. Calero, P.; Nikel, P.I. Chasing Bacterial *Chassis* for Metabolic Engineering: A Perspective Review from Classical to Non-Traditional Microorganisms. *Microb. Biotechnol.* **2019**, *12*, 98–124. [[CrossRef](#)]
20. Cusano, A.M.; Parrilli, E.; Marino, G.; Tutino, M.L. A Novel Genetic System for Recombinant Protein Secretion in the Antarctic Pseudoalteromonas Haloplanktis TAC125. *Microb. Cell Factories* **2006**, *5*, 40. [[CrossRef](#)] [[PubMed](#)]
21. Unzueta, U.; Vázquez, F.; Accardi, G.; Mendoza, R.; Toledo-Rubio, V.; Giuliani, M.; Sannino, F.; Parrilli, E.; Abasolo, I.; Schwartz, S.; et al. Strategies for the Production of Difficult-to-Express Full-Length Eukaryotic Proteins Using Microbial Cell Factories: Production of Human Alpha-Galactosidase A. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 5863–5874. [[CrossRef](#)] [[PubMed](#)]
22. Vigentini, I.; Merico, A.; Tutino, M.L.; Compagno, C.; Marino, G. Optimization of Recombinant Human Nerve Growth Factor Production in the Psychrophilic Pseudoalteromonas Haloplanktis. *J. Biotechnol.* **2006**, *127*, 141–150. [[CrossRef](#)]
23. Papa, R.; Rippa, V.; Sannia, G.; Marino, G.; Duilio, A. An Effective Cold Inducible Expression System Developed in Pseudoalteromonas Haloplanktis TAC125. *J. Biotechnol.* **2007**, *127*, 199–210. [[CrossRef](#)] [[PubMed](#)]
24. Katayama, S.; Sueyoshi, N.; Inazu, T.; Kameshita, I. Cyclin-Dependent Kinase-Like 5 (CDKL5): Possible Cellular Signalling Targets and Involvement in CDKL5 Deficiency Disorder. *Neural Plast.* **2020**, *2020*, 6970190. [[CrossRef](#)] [[PubMed](#)]
25. Trazzi, S.; De Franceschi, M.; Fuchs, C.; Bastianini, S.; Viggiano, R.; Lupori, L.; Mazziotti, R.; Medici, G.; Lo Martire, V.; Ren, E.; et al. CDKL5 Protein Substitution Therapy Rescues Neurological Phenotypes of a Mouse Model of CDKL5 Disorder. *Hum. Mol. Genet.* **2018**, *27*, 1572–1592. [[CrossRef](#)] [[PubMed](#)]
26. Clark, S.; Sullivan, S.; Gray, H. Recombinant CDKL5 Proteins, Gene Therapy and Production Methods. WO/2021/087282, 6 May 2021.
27. Fondi, M.; Maida, I.; Perrin, E.; Meller, A.; Mocali, S.; Parrilli, E.; Tutino, M.L.; Liò, P.; Fani, R. Genome-Scale Metabolic Reconstruction and Constraint-Based Modelling of the Antarctic Bacterium *Pseudoalteromonas Haloplanktis* TAC125: Modelling of *P. Haloplanktis* TAC125 Metabolism. *Environ. Microbiol.* **2015**, *17*, 751–766. [[CrossRef](#)]
28. Hucka, M.; Bergmann, F.T.; Chaouiya, C.; Dräger, A.; Hoops, S.; Keating, S.M.; König, M.; Novère, N.L.; Myers, C.J.; Olivier, B.G.; et al. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. *J. Integr. Bioinform.* **2019**, *16*, 1–183. [[CrossRef](#)]
29. Olivier, B.G.; Bergmann, F.T. SBML Level 3 Package: Flux Balance Constraints Version 2. *J. Integr. Bioinform.* **2018**, *15*, 1–38. [[CrossRef](#)]

30. Lachance, J.-C.; Lloyd, C.J.; Monk, J.M.; Yang, L.; Sastry, A.V.; Seif, Y.; Palsson, B.O.; Rodrigue, S.; Feist, A.M.; King, Z.A.; et al. BOFdat: Generating Biomass Objective Functions for Genome-Scale Metabolic Models from Experimental Data. *PLoS Comput. Biol.* **2019**, *15*, e1006971. [[CrossRef](#)]
31. Qi, W.; Colarusso, A.; Olombrada, M.; Parrilli, E.; Patrignani, A.; Tutino, M.L.; Toll-Riera, M. New Insights on Pseudoalteromonas Haloplanktis TAC125 Genome Organization and Benchmarks of Genome Assembly Applications Using next and Third Generation Sequencing Technologies. *Sci. Rep.* **2019**, *9*, 16444. [[CrossRef](#)]
32. Perrin, E.; Ghini, V.; Giovannini, M.; Di Patti, F.; Cardazzo, B.; Carraro, L.; Fagorzi, C.; Turano, P.; Fani, R.; Fondi, M. Diauxie and Co-Utilization of Carbon Sources Can Coexist during Bacterial Growth in Nutritionally Complex Environments. *Nat. Commun.* **2020**, *11*, 3135. [[CrossRef](#)]
33. Colarusso, A.; Lauro, C.; Calvanese, M.; Parrilli, E.; Tutino, M.L. Improvement of Pseudoalteromonas Haloplanktis TAC125 as a Cell Factory: IPTG-Inducible Plasmid Construction and Strain Engineering. *Microorganisms* **2020**, *8*, 1466. [[CrossRef](#)]
34. Sannino, F.; Giuliani, M.; Salvatore, U.; Apuzzo, G.A.; de Pascale, D.; Fani, R.; Fondi, M.; Marino, G.; Tutino, M.L.; Parrilli, E. A Novel Synthetic Medium and Expression System for Subzero Growth and Recombinant Protein Production in Pseudoalteromonas Haloplanktis TAC125. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 725–734. [[CrossRef](#)] [[PubMed](#)]
35. Boyle, J. Lehninger Principles of Biochemistry (4th ed.): Nelson, D., and Cox, M. *Biochem. Mol. Biol. Educ.* **2005**, *33*, 74–75. [[CrossRef](#)]
36. Carneiro, S.; Ferreira, E.C.; Rocha, I. Metabolic Responses to Recombinant Bioprocesses in Escherichia Coli. *J. Biotechnol.* **2013**, *164*, 396–408. [[CrossRef](#)] [[PubMed](#)]
37. Choi, H.S.; Lee, S.Y.; Kim, T.Y.; Woo, H.M. In Silico Identification of Gene Amplification Targets for Improvement of Lycopene Production. *AEM* **2010**, *76*, 3097–3105. [[CrossRef](#)] [[PubMed](#)]
38. Driouch, H.; Melzer, G.; Wittmann, C. Integration of in Vivo and in Silico Metabolic Fluxes for Improvement of Recombinant Protein Production. *Metab. Eng.* **2012**, *14*, 47–58. [[CrossRef](#)] [[PubMed](#)]
39. Liu, H.; Sun, Y.; Ramos, K.R.M.; Nisola, G.M.; Valdehuesa, K.N.G.; Lee, W.; Park, S.J.; Chung, W.-J. Combination of Entner-Doudoroff Pathway with MEP Increases Isoprene Production in Engineered Escherichia Coli. *PLoS ONE* **2013**, *8*, e83290. [[CrossRef](#)] [[PubMed](#)]
40. Li, C.; Ying, L.-Q.; Zhang, S.-S.; Chen, N.; Liu, W.-F.; Tao, Y. Modification of Targets Related to the Entner–Doudoroff/Pentose Phosphate Pathway Route for Methyl-d-Erythritol 4-Phosphate-Dependent Carotenoid Biosynthesis in Escherichia Coli. *Microb. Cell Factories* **2015**, *14*, 117. [[CrossRef](#)]
41. Tutino, M.; Duilio, A.; Parrilli, E.; Remaut, E.; Sannia, G.; Marino, G. A Novel Replication Element from an Antarctic Plasmid as a Tool for the Expression of Proteins at Low Temperature. *Extremophiles* **2001**, *5*, 257–264. [[CrossRef](#)]
42. Tascon, R.I.; Rodriguez-Ferri, E.F.; Gutierrez-Martin, C.B.; Rodriguez-Barbosa, I.; Berche, P.; Vazquez-Boland, J.A. Transposon Mutagenesis in Actinobacillus Pleuropneumoniae with a Tn10 Derivative. *J. Bacteriol.* **1993**, *175*, 5717–5722. [[CrossRef](#)]
43. Marblestone, J.G. Comparison of SUMO Fusion Technology with Traditional Gene Fusion Systems: Enhanced Expression and Solubility with SUMO. *Protein Sci.* **2006**, *15*, 182–189. [[CrossRef](#)]
44. Flinterman, M.; Farzaneh, F.; Habib, N.; Malik, F.; Gäken, J.; Tavassoli, M. Delivery of Therapeutic Proteins as Secretable TAT Fusion Products. *Mol. Ther.* **2009**, *17*, 334–342. [[CrossRef](#)] [[PubMed](#)]
45. Takis, P.G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metabolomics. *TrAC Trends Anal. Chem.* **2019**, *120*, 115300. [[CrossRef](#)]
46. Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P.G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem. Int. Ed.* **2019**, *58*, 968–994. [[CrossRef](#)] [[PubMed](#)]
47. Lieven, C.; Beber, M.E.; Olivier, B.G.; Bergmann, F.T.; Ataman, M.; Babaei, P.; Bartell, J.A.; Blank, L.M.; Chauhan, S.; Correia, K.; et al. MEMOTE for Standardized Genome-Scale Metabolic Model Testing. *Nat. Biotechnol.* **2020**, *38*, 272–276. [[CrossRef](#)]
48. Heirendt, L.; Arreckx, S.; Pfau, T.; Mendoza, S.N.; Richelle, A.; Heinken, A.; Haraldsdóttir, H.S.; Wachowiak, J.; Keating, S.M.; Vlasov, V.; et al. Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0. *Nat. Protoc.* **2019**, *14*, 639–702. [[CrossRef](#)]
49. Wang, H.; Marčišauskas, S.; Sánchez, B.J.; Domenzain, I.; Hermansson, D.; Agren, R.; Nielsen, J.; Kerkhoven, E.J. RAVEN 2.0: A Versatile Toolbox for Metabolic Network Reconstruction and a Case Study on Streptomyces Coelicolor. *PLoS Comput. Biol.* **2018**, *14*, e1006541. [[CrossRef](#)]



Data



Whole-genome sequencing of *Pseudomonas* sp. TAE6080, a strain capable of inhibiting *Staphylococcus epidermidis* biofilm

Christopher Riccardi^{a,*}, Caterina D'Angelo^b, Marzia Calvanese^b, Annarita Ricciardelli^b, Assunta Sellitto^c, Giorgio Giurato^{c,d}, Maria Luisa Tutino^b, Alessandro Weisz^{c,d}, Ermenegilda Parrilli^b, Marco Fondi^a

^a Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, Italy

^b Department of Chemical Sciences, University of Naples Federico II, Complesso Universitario Monte S. Angelo, via Cinthia, 80125 Naples, Italy

^c Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Via S. Allende, 84081 Baronissi, Italy

^d Genome Research Center for Health - GRGS, Campus di Medicina, Via S. Allende, 84081 Baronissi, Italy

ARTICLE INFO

Keywords:

Anti-biofilm
Antarctic marine bacterium
Pseudomonas sp.
NGS

ABSTRACT

Antarctic bacteria are able to survive under extreme environmental conditions and have adapted to exploit some of the most ephemeral nutrient pockets. Importantly, such strains have been often shown to be capable of synthesizing compounds of valuable biotechnological importance. Here we show that *Pseudomonas* sp. TAE6080, a possibly new bacterium isolated in 1994 during water column samplings near the French Antarctic station Dumont d'Urville, is capable of inhibiting the formation of *Staphylococcus epidermidis* biofilm, known to be an important opportunistic pathogen in infections associated to medical devices. A better understanding of this bacterium can therefore provide useful insight on new bioactive molecules that could play a role against chronic infections. To this end, the anti-biofilm effect of cell-free supernatant of *Pseudomonas* sp. TAE6080 was evaluated on *S. epidermidis* RP62A biofilm formation, demonstrating that it significantly reduced its aggregation. Furthermore, genome sequencing, assembly and mining revealed a plethora of putative biosynthetic gene clusters that might be involved in biofilm disruption. The experimental and genomic data presented here open the venue to further investigations on the molecular basis underlying biofilm inhibition.

1. Introduction

Epithelial and mucosa microbiota is composed of a variety of bacterial colonizers that were first and foremost regarded as harmless. Among these, however, *Staphylococcus epidermidis* has recently emerged as an important opportunistic pathogen in infections associated with medical devices, localized mainly in joint prosthesis and fracture fixations (Heilmann et al., 2019). The capability of *S. epidermidis* to adhere on both eukaryotic cells and abiotic surfaces and to form biofilm is an essential virulence factor (Le and Otto, 2015) that contributes to the chronicization of infections that are particularly difficult to eradicate. Therefore, the prevention of *S. epidermidis* biofilm formation and the treatment of existing ones is currently a difficult challenge and the discovery of new anti-biofilm molecules is urgent (Hogan et al., 2015). In general, bacterial biofilms are complex, surface-attached communities of bacteria that are held together by a self-produced polymer

matrix. It is estimated that about 40–80% of bacterial cells on earth can form biofilms (Flemming et al., 2016), which are believed to significantly contribute to successful microbial survival in hostile environments and to equip bacteria with much greater resistance to environmental challenges, including antimicrobial agents, when compared to their free-living counterparts (Kumar et al., 2017). Microbial biofilms affect the world economy at the level of billions of dollars. In clinical settings, for example, biofilms have been shown to persist on medical device surfaces and on patient's tissues causing persistent infections (Roy et al., 2018). Marine microorganisms living in Polar regions are adapted to survive under extreme environmental conditions and these adaptations are often accompanied by peculiar metabolic pathways, making them a potential source of new metabolites (Margesin and Feller, 2010). One of the above-mentioned survival strategies may be represented by the production of molecules with anti-biofilm activity, which could be exploited to fight the biological competition of other

* Corresponding author.

E-mail address: christopher.riccardi@unifi.it (C. Riccardi).

bacteria. Recently, several works endorsed the potential of cold-adapted marine bacteria as a source of novel anti-biofilm agents (Papa et al., 2015; Artini et al., 2019) active against several pathogens, including *S. epidermidis* (Parrilli et al., 2015; Casillo et al., 2017). In this paper, we report the genome analysis of *Pseudomonas* sp. TAE6080, an Antarctic marine bacterium that is able to produce and secrete a bioactive molecule which inhibits and eradicates biofilm aggregation of the opportunistic pathogen *Staphylococcus epidermidis* (Le et al., 2014). Minimum information for the hereby presented sequence (MIxS) is reported in Table 1.

2. Data description

2.1. Bacterial strains and culture conditions

Pseudomonas sp. TAE6080 was grown in synthetic medium G (D-Gluconic acid sodium 10 g/L, NaCl 10 g/L; NH_4NO_3 1 g/L; $\text{KH}_2\text{PO}_4 \cdot 7\text{H}_2\text{O}$ 1 g/L; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 200 mg/L; $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 5 mg/L; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 5 mg/L) or in TYP medium at 15 °C under vigorous agitation (180 rpm). The bacterial strain *Staphylococcus epidermidis* RP62A is a reference strain isolated from an infected catheter (ATCC collection no. 35984). The bacterium was grown in Brain Heart Infusion broth (BHI, Oxoid, UK). Biofilm formation was assessed in static conditions at 37 °C, whereas planktonic cultures were performed under vigorous agitation (180 rpm) at 37 °C. All strains were maintained at -80 °C in cryovials with 15% glycerol.

2.2. Biofilm formation assay

Biofilm formation of *S. epidermidis* was evaluated in the presence and absence of *Pseudomonas* sp. TAE6080 cell-free supernatant. Quantification of in vitro biofilm production was based on a previously reported method (Casillo et al., 2017). Briefly, the wells of a sterile 96-well flat-bottomed polystyrene multiwell were filled with 100 μL of BHI broth. 1/100 dilution of *S. epidermidis* overnight bacterial culture were added into each well. The first row of the multiwell contained the untreated bacteria, while each of the remaining rows contained serial dilutions of *Pseudomonas* sp. TAE6080 cell-free supernatant, in this case, a concentrated BHI was used to ensure the same medium composition in each sample. After inoculation, the plates were incubated aerobically for 20 h

at 37 °C in a static condition. Biofilm formation was measured using crystal violet staining. After treatment, planktonic cells were gently removed; each well was washed three times with water. To quantify biofilm formation, each well was stained with 0.1% crystal violet and incubated for 15 min at room temperature, rinsed twice with double-distilled water, and thoroughly dried. The dye bound to adherent cells was solubilized with 20% (v/v) glacial acetic acid and 80% (v/v) ethanol. After 30 min of incubation at room temperature, OD590 was measured to quantify the total biomass of biofilm formed in each well. Each data point is composed of three independent experiments, each performed at least in 6-replicates.

2.3. Anti-biofilm activity of *Pseudomonas* sp. TAE6080

The anti-biofilm effect of cell-free supernatant of *Pseudomonas* sp. TAE6080, grown at 15 °C in G medium, was evaluated on *S. epidermidis* RP62A biofilm formation. As shown in Fig. 1A, *Pseudomonas* sp. TAE6080 cell-free was able to interfere with the biofilm formation of *S. epidermidis* leading to a reduction of 50%. The effect on *S. epidermidis* biofilm formation was dose-dependent.

2.4. DNA extraction and sequencing

Total DNA was extracted from the *Pseudomonas* sp. TAE6080 strain, grown in 2 mL TYP medium for 24 h at 15 °C, using the E.Z.N.A. Bacterial DNA kit (Omega Bio-Tek Inc) following the manufacturer's instructions. For *Pseudomonas* sp. TAE6080 DNA sequencing, 500 ng of genomic DNA were used for library preparation with Nextera DNA flex Prep Kit (Illumina, San Diego, CA, USA) and sequencing (paired-end, 2 \times 75 cycles) on the NextSeq 500 platform (Illumina, San Diego, CA, USA).

2.5. Statistics and reproducibility of results

Data reported were statistically validated using Student *t*-test comparing the mean absorbance of treated and untreated samples. The significance of differences between mean absorbance values was calculated using a two-tailed Student's *t*-test. A *p*-value <0.05 was considered significant.

2.6. Genome annotation and bioinformatics analysis

Following DNA sequencing, a quality check of 23273024 paired-end reads in FASTQ format was performed using FastQC (Andrews, 2010) and custom scripts. The sequencing of the genome produced reads with consistent score, suggesting that there were no major flaws in the extraction nor in the preparation protocol. Analyzing per-base sequence quality revealed a slight fraction of outliers towards the end of the reads (nucleotides 64–74) that were dragging down the overall quality score. Trimmomatic (Bolger et al., 2014) was used to filter out reads that were not suitable for assembly (average quality between 0 and 10). Other than dumping low quality reads, its sliding window algorithm ensured that a consecutive series of poor-quality bases triggered the trimming of the 3' end. A lower N count and a higher proportion of Phred33 quality scores in the range 32–42 (>80%, data not shown) were observed as a product of reads trimming and filtering with Trimmomatic. Given the high and stable quality of the reads, no additional filtering was applied. Forward and reverse sequencing reads were piped into SPAdes (Prjibelski et al., 2020) – for a de novo assembly, producing 417 scaffolds and 425 contigs with a coverage of 253 \times , an N50 of 122642 and L50 of 17. After filtering out contigs that counted less than 500 nucleotides, a total of 101 scaffolds were forwarded for genome mining analysis. Genomic features were retrieved using Prokka (Seemann, 2014), which makes use of Prodigal (Hyatt et al., 2010) gene recognition and translation initiation site identification software and Barrnap for ribosomal RNA prediction (github.com/tseemann/barrnap). The pipeline allowed for a

Table 1

MIxS Checklist for the newly sequenced strain.

Item	Description
Investigation type	Bacteria_archaea
Project name	Transcriptional Regulatory Networks in Antarctic bacteria as a proxy for global warming effects on microbial life in the Polar Oceans
Geographic location (latitude and longitude)	66° 40' S; 140° 01' E
Geographic location (sea, region)	French Antarctic station Dumont d'Urville, Terre Adélie
Collection date	1994-09
Broad-scale environmental context	Marine biome [ENVO:00000447]
Environmental medium	Terre Adélie [GAZ:00008877]
Local environment context	Coastal sea water [ENVO:00002150]
Observed biotic relationship	Free living
Encoded traits	Anti-biofilm activity
Relationship to Oxygen	Aerobe
Sequencing method	Illumina NextSeq 500
Assembly Software	SPAdes;3.12.0
Annotation	Prokka;1.12
Number of contigs	101
Feature prediction	Prodigal;2.6.3
16S recovered software	Barrnap;0.9
Number of standard tRNAs extracted	57
tRNA extraction software	Prokka;1.12

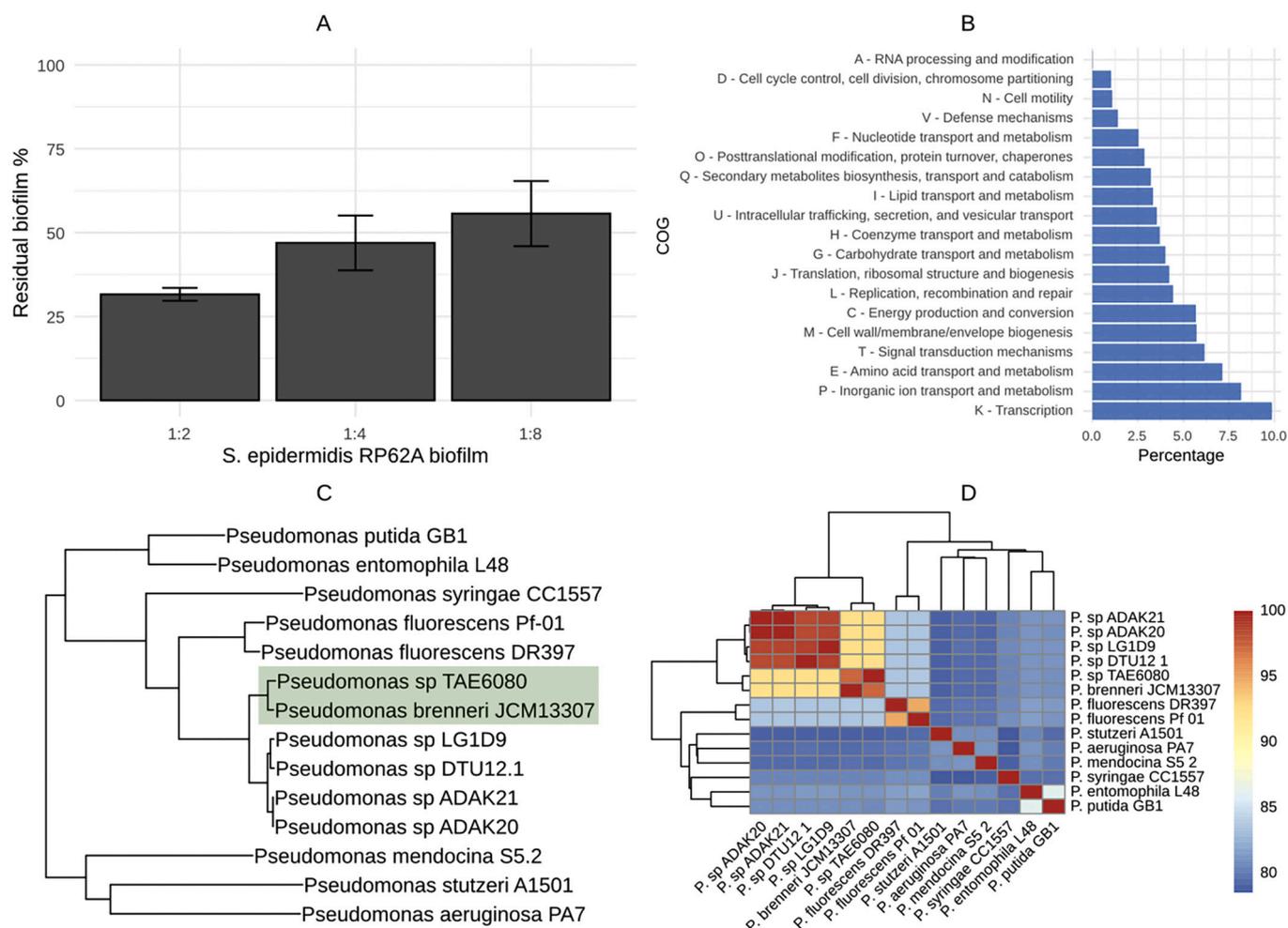


Fig. 1. Effect of *Pseudomonas* sp. TAE6080 supernatant on *S. epidermidis* RP62A biofilm formation (A). Biofilm formation was evaluated with a serial dilution of the Polar bacterium cell-free supernatant. Data are reported as the percentage of residual biofilm. Each data point represents the mean \pm SD of three independent samples. Biofilm formation was considered unaffected in the range of 90–100%. Data sets were statistically significant according to the t-Student test. (B) COG categories distribution in the newly sequenced *Pseudomonas* sp. TAE 80 strain. Phylogenetic analysis (C) and average nucleotide identity heatmap (D) showing the taxonomic relatedness among the newly sequenced *Pseudomonas* strain and the other *Pseudomonas* representatives examined in this work.

total of 5817 features to be identified as open reading frames, 2073 of which coding for polypeptides that have not yet been characterized. The annotation protocol also identified 59 RNA-coding genes, two of which were ribosomal RNAs: 5S and 16S ribosomal RNA respectively. When 5S ribosomal RNA (110 bp) and a partial 16S ribosomal RNA (1064 bp) were queried to the Nucleotide collection (nt) database (NCBI) they were consistently assigned to *Pseudomonas* (100% identity and query cover, *E*-values of $1e-40$ and 0.0 , respectively). To assess the presence of gene clusters involved in the production of bioactive compounds we submitted the newly sequenced genome and annotation files as generated by Prokka to antiSMASH 5 Web Server pipeline (Blin et al., 2019). Constrained strictness settings (strict) were selected to allow only well-defined clusters, i.e., those containing all required parts, to be inferred as putative biosynthetic gene clusters. These settings alone allowed for a plethora of operons involved in the biosynthesis of metabolites such as terpenes, bacteriocins, lipopeptides, and other non-ribosomal peptides to be detected. Specifically, seven cluster types were found distributed in 11 groups as displayed in Table 2, where each column respectively specifies: the region number, the product types as detected by antiSMASH, the nucleotide range within the region, the closest compound from the MiBIG database (Kautsar et al., 2020) and the fraction of genes within the closest known compound that have a significant BLAST hit to the genes within the current region (row). When antiSMASH queried its

Table 2
Summary of antiSMASH predicted BGCs.

Region	Type	From	To	Cluster	Similarity
Region 2.1	RiPP-like	244,708	255,553		
Region 6.1	NRPS	121,394	174,338	Pyoverdine	0.14
Region 7.1	terpene	111,962	134,187		
Region 8.1	NAGGN	8976	23,762		
Region 11.1	arylpolyene	57,043	100,617	APE Vf	0.35
Region 15.1	redox-cofactor	23,037	45,193	lankacidin C	0.13
Region 18.1	betalactone	60,126	83,391	Fengycin	0.13
Region 20.1	RiPP-like	21,680	32,561		
Region 25.1	terpene	37,359	58,387		
Region 36.1	NRPS	15,140	61,972	Pyoverdine	0.08
Region 47.1	arylpolyene	1	32,543		

underlying clusters repository for similar matches, the resulting similarities were strikingly skewed towards complete genomes of the *Pseudomonadaceae* family of microorganisms: regions 2.1 and 18.1 were

perfectly identical to that of multiple strains within the *Pseudomonas* genus, whereas region 8.1 displayed a perfect match with both *Pseudomonas fluorescens* strain PF08 and *Pseudomonas fluorescens* strain UK4. Region 6.1 showed up to 97% similarity to *Pseudomonas fluorescens* strain UK4 and region 11.1 had 100% identity to *Pseudomonas* sp. ADAK20 and *Pseudomonas lurida* strain MYb11. Region 20.1 is completely identical to that of *Pseudomonas fluorescens* strain PF08, *Pseudomonas* sp. LG1D9, *Pseudomonas* sp. DTU12.1, *Pseudomonas* sp. ADAK20, *Pseudomonas fluorescens* strain UK4 and *Pseudomonas trivialis* strain IHBB745. The remaining regions shared between 19% and 46% similarity to *Pseudomonas thivervalensis* strain DSM 13194 and *Pseudomonas brenneri* strain DSM 15294, respectively. Clusters of orthologous groups analysis was performed after submitting the proteome to eggNOG mapper 5 (Huerta-Cepas et al., 2019), and the results parsed through in-house scripts (github.com/chrisondakeys/eggNOG_cleanup). Relative abundance of COG categories for the newly assembled genome is depicted in Fig. 1 B. Not shown is the unknown function category S which represented 21.92% of the total; categories for which relative abundance was zero are also omitted. In order to phylogenetically characterize the newly sequenced strain, we operated in a stepwise fashion: we first queried 5S and 16S rDNA sequences as well as some key genomic portions, as the putative pyoverdine BGC, to the Nucleotide collection database (BLAST) to get a hint of its evolutionary uniqueness which allowed us to clarify to which genus the species belongs to (*Pseudomonas*). We then performed an average nucleotide identity (ANI) analysis to assess whether sufficient resolution at species level could be found. After downloading 654 genomes present under various assembly statuses, we made use of fastANI (github.com/ParBLISS/FastANI) to build a nucleotide identity matrix. One of the strains used for comparison, namely *P. brenneri* JCM13307, reached an ANI value greater than 97%. The second-best scoring was *P. sp.* ADAK21 which shared 92.5% identity with our newly sequenced strain. The four species that followed, ordered by decreasing nucleotide identity, were *P. sp.* ADAK20 (92.47%), *P. sp.* LG1D9 (92.43%), *P. sp.* DTU12.1 (92.41%) and *P. fluorescens* DR397 (91.46%). Considering that the next-in-rank scored 87.08% (not reported) we retained the genomic sequences of all the six aforementioned species for subsequent phylogenetic analysis. Not only did we retrieve the closest species in terms of nucleotide identity, but we also included seven strains belonging to genomes attributed to known species as seen in (Özen and Ussery, 2012) to have a more general understanding of the evolutionary history in the context of a wider set of *Pseudomonas* members. The phylogenetic tree (Fig. 1C) was inferred through the Neighbor-Joining method (Saitou and Nei, 1987) on a multiple sequence alignment of 9633 nucleotide positions from a core genome which was computed using Roary (Page et al., 2015). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test, consisting of 1000 replicates, resulted in a 100% accordance. Mega 7 (Kumar et al., 2016) was used for inferring the evolutionary history through NJ while the processing of raw data was performed through the awk scripting language and the images rendered using R (<https://www.r-project.org/>). As to confirm what was already observed during genome mining analysis, the newly sequenced species shares a cluster with *P. brenneri* JCM13307, *P. fluorescens* DR397, *P. sp.* ADAK21 and *P. sp.* ADAK20. The way the evolutionary distances were modelled allowed the maximum bootstrap support to be consistently spread throughout the nodes, fortifying the hypothesis that the newly sequenced strain is more closely related to *P. brenneri* JCM13307 and the subarctic-acclimated bacteria *Pseudomonas* sp. ADAK20 and *Pseudomonas* sp. ADAK21 (Haan et al., 2020), than it is to the other known species such as *P. aeruginosa* and *P. stutzeri*.

Data accessibility

The sequence described in this work was deposited on the GenBank, under the accession number JAHIDY000000000.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

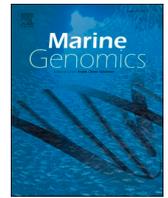
Acknowledgements

Research supported by MIUR (Programma Nazionale di Ricerche in Antartide 2018, grant PNRA18.00335) and by Regione Campania (POR Campania FESR 2014/2020, Azione 1.5, grant 'GENOMAeSALUTE').

References

- Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Artini, M., Papa, R., Vrenna, G., Lauro, C., Ricciardelli, A., Casillo, A., Corsaro, M.M., Tutino, M.L., Parrilli, E., Selan, L., 2019 Nov. Cold-adapted bacterial extracts as a source of anti-infective and antimicrobial compounds against *Staphylococcus aureus*. *Future Microbiol.* 14, 1369–1382. <https://doi.org/10.2217/fmb-2019-0147>. Epub 2019 Oct 9. 31596138.
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019 Jul 2. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47 (W1), W81–W87. <https://doi.org/10.1093/nar/gkz310>. PMID: 31032519; PMCID: PMC6602434.
- Bolger, A.M., Lohse, M., Usadel, B., 2014 Aug 1. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30 (15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.
- Casillo, A., Papa, R., Ricciardelli, A., Sannino, F., Ziaco, M., Tilotta, M., Selan, L., Marino, G., Corsaro, M.M., Tutino, M.L., Artini, M., Parrilli, E., 2017 Feb 23. Antibiofilm activity of a long-chain fatty aldehyde from antarctic pseudoalteromonas haloplanktis TAC125 against *Staphylococcus epidermidis* biofilm. *Front. Cell. Infect. Microbiol.* 7, 46. <https://doi.org/10.3389/fcimb.2017.00046>. PMID: 28280714; PMCID: PMC5322152.
- Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., Kjelleberg, S., 2016 Aug 11. Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* 14 (9), 563–575. <https://doi.org/10.1038/nrmicro.2016.94>. 27510863.
- Haan, T., Seitz, T.J., Francisco, A., Gliner, K., Gloger, A., Kardash, A., Matsui, N., Reast, E., Rosander, K., Sonnek, C., Wellman, R., Drown, D.M., 2020 Jun 18. Complete genome sequences of seven strains of *Pseudomonas* spp. Isolated from boreal forest soil in interior Alaska. *Microbiol. Resour. Annot.* 9 (25) <https://doi.org/10.1128/MRA.00511-20>. PMID: 32554792; PMCID: PMC7303412.
- Heilmann, C., Ziebuhr, W., Becker, K., 2019 Sep. Are coagulase-negative staphylococci virulent? *Clin. Microbiol. Infect.* 25 (9), 1071–1080. <https://doi.org/10.1016/j.cmi.2018.11.012>. Epub 2018 Nov 29. 30502487.
- Hogan, S., Stevens, N.T., Humphreys, H., O'Gara, J.P., O'Neill, E., 2015. Current and future approaches to the prevention and treatment of staphylococcal medical device-related infections. *Curr. Pharm. Des.* 21 (1), 100–113. <https://doi.org/10.2174/1381612820666140905123900>. 25189861.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019 Jan 8. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47 (D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>. PMID: 30418610; PMCID: PMC6324079.
- Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010 Mar 8. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11, 119. <https://doi.org/10.1186/1471-2105-11-119>. PMID: 20211023; PMCID: PMC2848648.
- Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hoof, J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2020 Jan 8. MIBIG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48 (D1), D454–D458. <https://doi.org/10.1093/nar/gkz882>. PMID: 31612915; PMCID: PMC7145714.
- Kumar, S., Stecher, G., Tamura, K., 2016 Jul. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>. Epub 2016 Mar 22. 27004904.
- Kumar, A., Alam, A., Rani, M., Ehtesham, N.Z., Hasnain, S.E., 2017 Dec. Biofilms: survival and defense strategy for pathogens. *Int. J. Med. Microbiol.* 307 (8), 481–489. <https://doi.org/10.1016/j.ijmm.2017.09.016>. Epub 2017 Sep 21. 28950999.
- Le, K.Y., Otto, M., 2015 Oct 27. Quorum-sensing regulation in staphylococci-an overview. *Front. Microbiol.* 6, 1174. <https://doi.org/10.3389/fmicb.2015.01174>. PMID: 26579084; PMCID: PMC4621875.
- Le, K.Y., Dastgheyb, S., Ho, T.V., Otto, M., 2014 Nov 26. Molecular determinants of staphylococcal biofilm dispersal and structuring. *Front. Cell. Infect. Microbiol.* 4, 167. <https://doi.org/10.3389/fcimb.2014.00167>. PMID: 25505739; PMCID: PMC4244807.

- Margesin, R., Feller, G., 2010 Jul-Aug. Biotechnological applications of psychrophiles. *Environ. Technol.* 31 (8–9), 835–844. <https://doi.org/10.1080/09593331003663328>. 20662375.
- Özen, A.I., Ussery, D.W., 2012 Feb. Defining the *Pseudomonas* genus: where do we draw the line with *Azotobacter*? *Microb. Ecol.* 63 (2), 239–248. <https://doi.org/10.1007/s00248-011-9914-8>. Epub 2011 Aug 3. PMID: 21811795; PMCID: PMC3275731.
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015 Nov 15. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 31 (22), 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>. Epub 2015 Jul 20. PMID: 26198102; PMCID: PMC4817141.
- Papa, R., Selan, L., Parrilli, E., Tilotta, M., Sannino, F., Feller, G., Tutino, M.L., Artini, M., 2015 Dec 14. Anti-biofilm activities from marine cold adapted bacteria against staphylococci and *Pseudomonas aeruginosa*. *Front. Microbiol.* 6, 1333. <https://doi.org/10.3389/fmicb.2015.01333>. PMID: 26696962; PMCID: PMC4677098.
- Parrilli, E., Papa, R., Carillo, S., Tilotta, M., Casillo, A., Sannino, F., Cellini, A., Artini, M., Selan, L., Corsaro, M.M., Tutino, M.L., 2015 Mar. Anti-biofilm activity of *pseudoalteromonas haloplanktis* tac125 against *staphylococcus epidermidis* biofilm: evidence of a signal molecule involvement? *Int. J. Immunopathol. Pharmacol.* 28 (1), 104–113. <https://doi.org/10.1177/0394632015572751>. 25816412.
- Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., Korobeynikov, A., 2020 Jun. Using SPAdes De novo assembler. *Curr. Protoc. Bioinformatics* 70 (1), e102. <https://doi.org/10.1002/cpbi.102>. 32559359.
- Roy, R., Tiwari, M., Donelli, G., Tiwari, V., 2018 Jan 1. Strategies for combating bacterial biofilms: A focus on anti-biofilm agents and their mechanisms of action. *Virulence.* 9 (1), 522–554. <https://doi.org/10.1080/21505594.2017.1313372>. PMID: 28362216; PMCID: PMC5955472.
- Saitou, N., Nei, M., 1987 Jul. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>. 3447015.
- Seemann, T., 2014 Jul 15. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 30 (14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>. Epub 2014 Mar 18. 24642063.



Data



Genome analysis of a new biosurfactants source: The Antarctic bacterium *Psychrobacter* sp. TAE2020

Christopher Riccardi^{a,*}, Caterina D'Angelo^b, Marzia Calvanese^b, Annarita Ricciardelli^c,
Maria Luisa Tutino^b, Ermenegilda Parrilli^b, Marco Fondi^a

^a Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, Italy

^b Department of Chemical Sciences, University of Naples Federico II, Complesso Universitario Monte S. Angelo, via Cinthia, 80125 Naples, Italy

^c Department of Marine Biotechnology, Stazione Zoologica "Anton Dohrn", 80121 Naples, Italy

ARTICLE INFO

Keywords:

Surfactants
Cold-adapted bacteria
Psychrobacter sp.
Next generation sequencing

ABSTRACT

Biosurfactants are considered a possible green alternative to chemical surfactants for countless commercial products including detergents and cleaners, personal care products, cosmetics, pharmaceuticals and therapeutics, food additives, emulsifiers, and dispersants for bioremediation. Organisms from extreme environments are well-adapted to the harsh conditions and represent an exciting avenue of discovery of naturally occurring biosurfactants. In this study, we report the genome analysis of *Psychrobacter* sp. TAE2020, an aerobic γ -proteobacterium isolated from an Antarctic coastal seawater sample collected in the vicinity of the French Antarctic station Dumont d'Urville, Terre Adelie (66° 40' S; 140° 01' E) which has been shown to produce biosurfactants. Biochemical assays indicate that *Psychrobacter* sp. TAE2020 can produce one or more excellent emulsifiers and a biosurfactant which is able to reduce the surface tension of a Gut medium. Next generation sequencing and genome mining allowed the identification of a plethora of biosynthetic gene clusters possibly involved in the production of emulsifying agents, just waiting to be isolated and characterized. This study paves the way for a more thorough investigation into the potential biotechnological applications of this new Antarctic strain.

1. Introduction

Chemical and microbial surfactants are amphiphilic compounds that are able to reduce both surface and interfacial tension, and to form emulsions from two immiscible liquids. These molecules have many industrial applications in pharmaceuticals, cosmetics, food additives, herbicides and pesticides fields (Tripathi et al., 2018). Compared to their chemically synthesized equivalents, biosurfactants have the advantage of being biodegradable and having low toxicity (Mukherjee et al., 2006), thus representing an environmentally friendly alternative. Moreover, their use has been extensively studied for the enhancement of oil recovery and remediation of hydrocarbon and metal-contaminated soil and water (Tripathi et al., 2018). Cold habitats are extraordinary reservoirs of molecules of biotechnological interest, such as cold-active enzymes, antibiotics, and biosurfactants (Trudgeon et al., 2020). Notably, *Pseudomonas*, *Pseudoalteromonas*, *Marinomonas*, *Halomonas*, *Rhodococcus* and *Cobetia* genera have been described as biosurfactant producers from polar oceans, deep sea, and marine sediments (Perfumo

et al., 2018). Biosurfactants play key ecological roles in many cold soils and marine environments (Trudgeon et al., 2020; Dang et al., 2016), for instance, they participate in carbon-cycling processes by enhancing the bioavailability of poorly soluble compounds. Furthermore, they have the ability to interact with multiple physical phases (water, ice, hydrophobic compounds, and gases) at low and freezing temperatures, and for this reason they can be used to develop new biotechnological products and processes that have low energy demand and operate under low-temperature regimes (Trudgeon et al., 2020; Malavenda et al., 2015). In this paper we report the genome sequence and analysis for *Psychrobacter* sp. TAE2020, a previously uncharacterized Antarctic marine bacterium that is able to produce surfactants, and whose biosynthetic gene clusters could make it an eligible candidate for biotechnological applications. Minimum information for the hereby presented sequence (MIxS) is reported in Table 1.

* Corresponding author.

E-mail address: christopher.riccardi@unifi.it (C. Riccardi).

<https://doi.org/10.1016/j.margen.2021.100922>

Received 31 August 2021; Received in revised form 28 December 2021; Accepted 28 December 2021

Available online 3 January 2022

1874-7787/© 2021 Published by Elsevier B.V.

Table 1
MIXS Checklist for the newly sequenced strain *Psychrobacter* sp. TAE2020.

Item	Description
Investigation type	Bacteria archaea
Project name	Transcriptional Regulatory Networks in Antarctic bacteria as a proxy for global warming effects on microbial life in the Polar Oceans
Geographic location (latitude and longitude)	66° 40' S; 140° 01' E
Geographic location (sea, region)	French Antarctic station Dumont d'Urville, Terre Adélie
Collection date	1994-09
Broad-scale environmental context	Marine biome [ENVO:00000447]
Environmental medium	Terre Adélie [GAZ:00008877]
Local environment context	Coastal sea water [ENVO:00002150]
Observed biotic relationship	Free living
Encoded traits	Biosurfactants producer
Relationship to Oxygen	Aerobe
Sequencing method	Illumina HiSeq 2500
Assembly Software	SPAdes;3.12.0
Annotation	Prokka;1.12
Number of contigs	33
Feature prediction	Prodigal;2.6.3
16S recovered software	Barrnap;0.9
Number of standard tRNAs extracted	42
tRNA extraction software	Prokka;1.12

2. Data description

2.1. Bacterial strains and culture conditions

Psychrobacter sp. TAE2020 was grown in the synthetic medium Gut (L-Glutamic acid 10 g/L, NaCl 10 g/L; NH₄NO₃ 1 g/L; KH₂PO₄ · 7 H₂O 1 g/L; MgSO₄ · 7 H₂O 200 mg/L; FeSO₄ · 7 H₂O 5 mg/L; CaCl₂ · 2 H₂O 5 mg/L) in planktonic conditions at 15 °C under vigorous agitation (180 rpm) up till the stationary phase of growth (72 h). The supernatant was recovered by centrifugation at 6000 rpm for 30 min at 4 °C, sterilized by filtration through membranes with a pore diameter of 0.22 µm and stored at 4 °C until use.

2.2. Drop-collapse assay

The drop-collapse assay (Jain et al., 1991) was performed as a qualitative test for the identification of a biosurfactants producer strain. In detail, drop collapse assay was performed by the deposition of 50 µL droplets of samples (supernatant or only medium) on a hydrophobic surface. Methylene blue was added to stain the samples for photographic purposes and did not influence the shape of the droplets. The spreading of the droplet on parafilm was observed after 1 h.

2.3. Emulsification index

The emulsification index (E24), (Blesic et al., 2018) of samples was determined by adding 2 mL of dextol (decane and toluene in 65:35 volume ratio) and 1 mL of the cell-free supernatant in a test tube, vortexed at maximum speed for 3 min and allowed to stand for 24 h. The emulsification index, E24, was determined by calculating the ratio between the height of emulsifying layer and the total height, multiplied by 100.

2.4. Surface tension measurements

The surface tension, γ , was measured through De Nouy ring method using a KSV Sigma 70 digital tensiometer (Dyne Testing Ltd., Newton House, Lichfield, UK) equipped with an automatic device to set the time between two consecutive measurements and to select the rising velocity of the platinum ring (D'Errico et al., 2005). The ring rising velocity was

set low enough to reach the equilibrium between the air–solution interface and the solution bulk. At least three surface tension measurements were performed on a 10 mL sample.

2.5. *Psychrobacter* sp. TAE2020 surfactants production

Cell-free supernatant of *Psychrobacter* sp. TAE2020 (*Psychrobacter* sp. TAE2020-SN) was analyzed by drop collapse assay (Fig. 1A), a sensitive and rapid method for the screening of surfactants producing bacteria. In general, if the liquid does not contain surfactants, polar water molecules are repelled from the hydrophobic surface and the drops remain stable, whereas if the liquid contains surfactants, the drops spread or even collapse because of the reduced interfacial tension between the liquid drop and the hydrophobic surface. As shown in Fig. 1A, the drop of *Psychrobacter* sp. TAE2020 cell-free supernatant collapsed, suggesting the presence of compound/s acting as surfactants. To evaluate if *Psychrobacter* sp. TAE2020 bioactive compound was also able to reduce the surface tension of an aqueous solution, the Du Nouy ring method was applied and the surface tension value obtained for *Psychrobacter* sp. TAE2020 cell-free supernatant resulted to be lower than that of Gut medium, 60.63 (\pm 0,5) and 77.49 (\pm 0,2) mN/m, respectively. Biosurfactants can also form emulsions that confer high long-term stability to the dispersed phase droplets in emulsions between two immiscible liquids. *Psychrobacter* sp. TAE2020-SN emulsification ability was tested in presence of 2 mL of Dectol, a “model oil”, compared to the negative control (Gut medium), by measuring the emulsification index E24. *Psychrobacter* sp. TAE2020-SN showed an E24 value of 70% (Fig. 1B). Taken together, these results demonstrate that *Psychrobacter* sp. TAE2020-SN contains an excellent emulsifier and a biosurfactant that is able to reduce the surface tension of Gut medium.

2.6. Genome annotation and bioinformatics analysis

A total of 1,225,606 paired ended sequencing reads were collected and submitted to Trimmomatic (Bolger et al., 2014) for quality filtering and trimming. Reads that were not suitable for analysis due to low quality were discarded, while those having a low score towards the 3' end were shortened, leaving a number of 1,218,486 idoneous reads, most of which measured 251 nucleotides in length (879,430, or 72.17% of the total). Genome de novo assembly was performed using SPAdes (Prjibelski et al., 2020), leading to a draft genome embedding 3,294,027 nucleotides (N50: 179582, L50: 7, contigs: 74, coverage: 89 \times , G + C content: 41.7%). The obtained scaffolds were then assayed for the presence of 16S rRNA gene using ContEst16S web server (Lee et al., 2017), which predicted the presence of one 1539 bp 16S rRNA gene that allowed to preliminarily assign the genome to the *Psychrobacter* genus, given the 99.22% identity at 100% query cover with *Psychrobacter urativorans* strain R10.10B. In order to further refine the ordering and orientation of the contigs obtained through the assembly procedure, all available *Psychrobacter* genomic sequences were downloaded from the NCBI and we made use of the multi draft-based scaffolder MeDuSa (Bosi et al., 2015), which remarkably improved the overall quality of the newly sequenced genome. When MeDuSa was fed with our newly assembled contigs and a set of 19 complete or draft *Psychrobacter* genomes, it was able to assemble the 74 contigs into 45 scaffolds, and increase the N50 statistic of an order of magnitude, from 179,582 to 3,206,207, thus allowing a more accurate representation of the genome. Contigs of size less than 200 bp were manually removed, leaving a final number of 3,293,789 base pairs distributed throughout 33 scaffolds. After assembly and refinement, we sought to locate the 16S rRNA gene in order to assess to which species the bacterium under investigation is more closely related to. Using Barrnap software for ribosomal RNA prediction (github.com/tseemann/barrnap) on the newly scaffolded genome we were able to retrieve six complete ribosomal RNA sequences (four 5S, one 16S and one 23S) which all displayed between 93 and 98% identity to *P. urativorans* strain R10.10B and *Psychrobacter cryohalolentis*

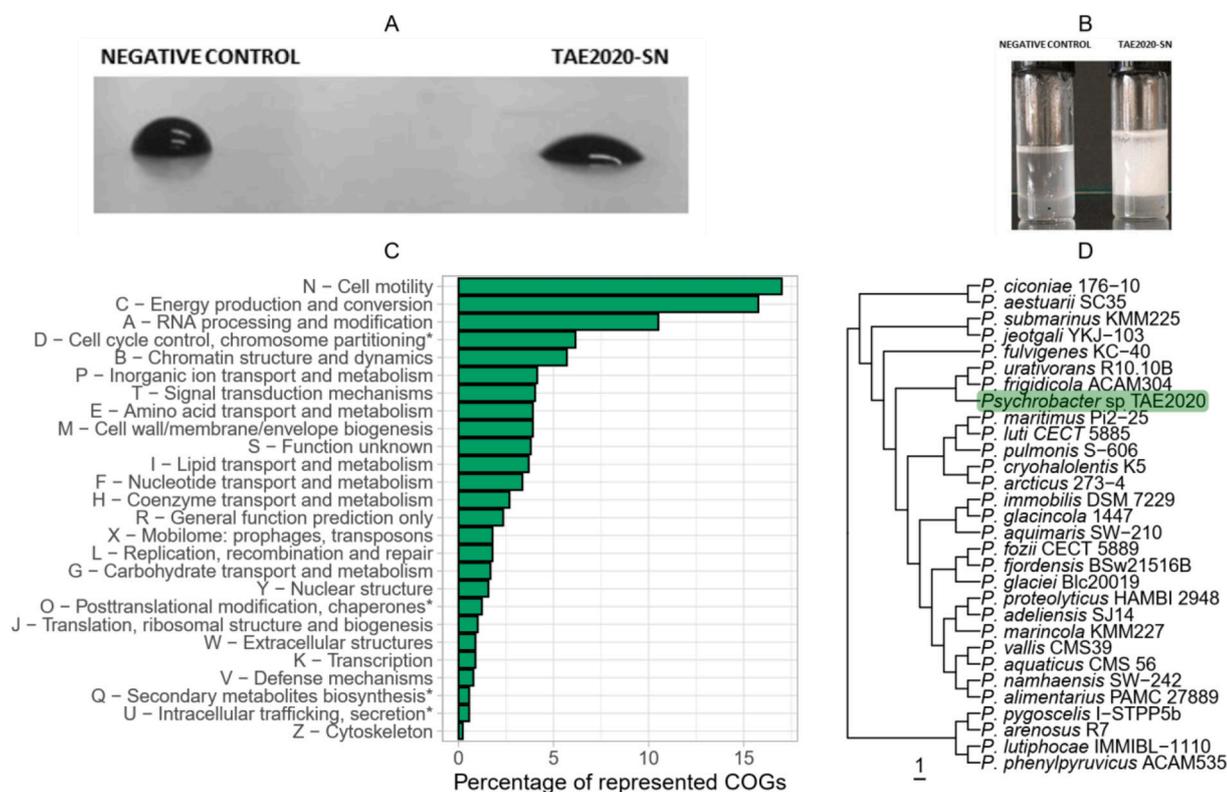


Fig. 1. (A) Drop collapse assay: drop of *Psychrobacter* sp. TAE2020 cell-free supernatant (TAE2020-SN) and negative control (L-Glutamate culture medium) on a hydrophobic surface; (B) Emulsification index: Emulsion of 1 mL of *Psychrobacter* sp. TAE2020 cell-free supernatant (TAE2020-SN) mixed to 2 mL of Dectol after 24 h, in comparison to the negative control (mixture of L-Glutamate culture medium and Dectol, left). (C) COG categories distribution in *Psychrobacter* sp. TAE2020. Descriptions marked with an asterisk were slightly simplified for better readability. Full descriptions available at <https://www.ncbi.nlm.nih.gov/research/cog#>. (D) Phylogenetic analysis showing the taxonomic relatedness among the newly sequenced strain and the other *Psychrobacter* representatives used within the context of this work. All nodes displayed a 100% bootstrap support therefore bootstrap values were omitted from the visualization. Isolate identifier was added to those species to which a strain has not been assigned yet.

strain FDAARGOS_308, respectively. When limiting the analysis to the 16S rRNA gene sequence, we obtained a match with the corresponding 16S sequence of *P. urativorans* strain DSM 14009 with a query coverage of 99% and sequence identity of 99.93%, and with that of *P. urativorans* strain R10.10B with a query coverage of 100% and sequence identity of 99.22%. To further investigate the genomic background that is behind *Psychrobacter* sp. TAE2020's biosurfactants production, a whole-genome annotation was first conducted using prokka (Seemann, 2014), which allowed to identify 2695 protein products, 931 of which were labelled as "hypothetical protein", and 42 tRNAs. The proteome file in FASTA format that was generated by prokka was then uploaded to eggNOG mapper genome functional annotation webserver (Huerta-Cepas et al.,

2019), while the genomic FASTA file was submitted to antiSMASH 6.0 webserver (Blin et al., 2021) for genome mining analysis. The most represented cluster of orthologous groups (COG) category is Cell Motility with 152 genes (17%), followed by Energy Production and conversion (15.7%) and RNA Processing and Modification (10.51%), (Fig. 1C). The predicted, biologically relevant gene cluster (BGC) that shares the highest similarity to a known cluster is involved in the production of a siderophore that is likely to resemble vibrioferrin (36% identity), whereas the remaining ones share a somewhat mild similarity to clusters producing emulsan, surfactin, TP-1161 (Thiopeptide), K53 capsular polysaccharide and, more in general, saccharides (Table 2). It did not escape our notice that the similarity values shared with other

Table 2

antiSMASH-predicted biosynthetic gene clusters. Each column respectively specifies: the region number, the product types as detected by antiSMASH, the nucleotide range within the region, the closest compound from the MiBIG database (Kautsar et al., 2020) and the fraction of genes within the closest known compound that have a significant BLAST hit to the genes within each region.

Region	Type	From	To	Most similar known cluster	Similarity
Region 1.1	fatty_acid	126,562	145,897		
Region 1.2	siderophore	325,289	340,438	vibrioferrin	Other 36%
Region 1.3	saccharide	350,792	372,207		
Region 1.4	saccharide	491,914	519,018		
Region 1.5	terpene	1,136,338	1,158,478		
Region 1.6	saccharide	1,218,766	1,242,740	K53 capsular polysaccharide	Saccharide 15%
Region 1.7	saccharide	1,913,419	1,930,991	emulsan	Saccharide 13%
Region 1.8	fatty_acid	2,367,607	2,386,641		
Region 1.9	fatty_acid	2,460,581	2,481,819	TP-1161	RiPP:Thiopeptide 16%
Region 1.10	saccharide	2,652,530	2,680,125		
Region 1.11	saccharide	2,857,505	2,882,929	surfactin	NRP:Lipopeptide 8%
Region 4.1	saccharide	1	8863	emulsan	Saccharide 9%

(known) gene clusters involved in biosurfactant/emulsan production is quite low (ranging from 8% to 15%). This raises questions on the actual structure of the biosurfactant gene cluster in *Psychrobacter* sp. TAE2020 and on the exact metabolic steps leading to the synthesis of this valuable compound in this strain. Further work will be necessary to elucidate this point in the future.

2.7. Phylogenetic analysis

To further investigate the evolutionary relatedness between *Psychrobacter* sp. TAE2020 and other members of the same genus, all *Psychrobacter* representative genomes were downloaded from the NCBI to build a phylogenetic tree on a core genome composed of 97 conserved markers (Supplementary Material). Despite being labelled as “Representative” on the RefSeq (O’Leary et al., 2016) database, some of those genomes were heavily fragmented. As Roary developers recommend against using highly fragmented scaffolds, the N50 statistic was used as a filter to select a final number of 30 genomes for which at least half of the nucleotides in the assembly were represented in contigs with length equal to or greater than the 1st Quartile (142680), (Supplementary Material). The phylogenetic tree was inferred through the Neighbor-Joining (Saitou and Nei, 1987) method on a multiple sequence alignment of 79,363 nucleotide positions from a core set of markers conserved across *Psychrobacter*, which was computed using Roary (Page et al., 2015). The percentage of (1000) replicate trees in which the associated taxa clustered together in the bootstrap test resulted in a 100% accordance throughout all the nodes. The obtained midpoint-rooted tree (Fig. 1D) shows that *Psychrobacter* sp. TAE2020 is a close relative of *Psychrobacter frigidicola* ACAM304 and *P. urativorans* R10.10B. Parallel to the phylogenetic analysis, we also sought to assess the Average Nucleotide Identity (ANI) between the newly sequenced strain and the 30 other selected genomes using FastANI (<https://github.com/ParBLiSS/FastANI>). The highest average genomic similarity 78.9% between *Psychrobacter* sp. TAE2020 and *P. frigidicola* ACAM304 (Supplementary Material). Mega 7 (Kumar et al., 2016) was used for inferring the evolutionary history through NJ while the processing of raw data was performed through the awk scripting language and the images rendered using the ggtree (Yu, 2020) package in R (<https://www.r-project.org/>).

Data accessibility

Biological source materials used in the herein presented genome announcement are available under the BioSample database accession JAHLMU000000000.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research supported by MIUR (Programma Nazionale di Ricerche in Antartide 2018, grant PNRA18_00335) and by Regione Campania (POR Campania FESR 2014/2020, Azione 1.5, grant ‘GENOMAeSALUTE’).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.margen.2021.100922>.

References

- Blesic, M., Dichiarante, V., Milani, R., Linder, M., Metrangolo, P., 2018. Conference paper evaluating the potential of natural surfactants in the petroleum industry: the case of hydrophobins. *Pure Appl. Chem.* 90, 305–314.
- Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M. H., Weber, T., 2021 May 12. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.*, gkab335 <https://doi.org/10.1093/nar/gkab335>. Epub ahead of print. PMID: 33978755.
- Bolger, A.M., Lohse, M., Usadel, B., 2014 Aug 1. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30 (15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.
- Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.F., Lió, P., Crescenzi, P., Fani, R., Fondi, M., 2015 Aug 1. MeDuSa: a multi-draft based scaffold. *Bioinformatics.* 31 (15), 2443–2451. <https://doi.org/10.1093/bioinformatics/btv171>. Epub 2015 Mar 25. PMID: 25810435.
- Dang, N.P., Landfald, B., Willassen, N.P., 2016. Biological surface-active compounds from marine bacteria. *Environ. Technol. (United Kingdom)* 37, 1151–1158. <https://doi.org/10.1080/09593330.2015.1103784>.
- D’Errico, G., Ciccarelli, D., Ortona, O., 2005 Jun 15. Effect of glycerol on micelle formation by ionic and nonionic surfactants at 25 degrees C. *J. Colloid Interface Sci.* 286 (2), 747–754. <https://doi.org/10.1016/j.jcis.2005.01.030> (PMID: 15897093).
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattai, T., Jensen, L.J., von Mering, C., Bork, P., 2019 Jan 8. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47 (D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>. PMID: 30418610; PMCID: PMC6324079.
- Jain, D.K., Collins-Thompson, D.L., Lee, H., Trevors, J.T., 1991. A drop-collapsing test for screening biosurfactant-producing microorganisms. *J. Microbiol. Methods* 13 (4), 271–279. [https://doi.org/10.1016/0167-7012\(91\)90064-W](https://doi.org/10.1016/0167-7012(91)90064-W).
- Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J. J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2020 Jan 8. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48 (D1), D454–D458. <https://doi.org/10.1093/nar/gkz882>. PMID: 31612915; PMCID: PMC7145714.
- Kumar, S., Stecher, G., Tamura, K., 2016 Jul. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. <https://doi.org/10.1093/molbev/msw054> (Epub 2016 Mar 22. PMID: 27004904).
- Lee, I., Chalitha, M., Ha, S.M., Na, S.I., Yoon, S.H., Chun, J., 2017 Jun. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 67 (6), 2053–2057. <https://doi.org/10.1099/ijsem.0.001872>. Epub 2017 Jun 22. PMID: 28639931.
- Malavenda, R., Rizzo, C., Michaud, L., Gerçe, B., Bruni, V., Sylødatk, C., Hausmann, R., Lo Giudice, A., 2015. Biosurfactant production by Arctic and Antarctic bacteria growing on hydrocarbons. *Polar Biol.* 38, 1565–1574. <https://doi.org/10.1007/s00300-015-1717-9>.
- Mukherjee, S., Das, P., Sen, R., 2006. Towards commercial production of microbial surfactants. *Trends Biotechnol.* 24, 509–515. <https://doi.org/10.1016/j.tibtech.2006.09.005>.
- O’Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badrettin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O’Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016 Jan 4. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44 (D1), D733–D745. <https://doi.org/10.1093/nar/gkv1189>. Epub 2015 Nov 8. PMID: 26553804; PMCID: PMC4702849.
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015 Nov 15. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 31 (22), 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>. Epub 2015 Jul 20. PMID: 26198102; PMCID: PMC4817141.
- Perfumo, A., Banat, I.M., Marchant, R., 2018. Going green and cold: biosurfactants from low-temperature environments to biotechnology applications. *Trends Biotechnol.* 36, 277–289. <https://doi.org/10.1016/j.tibtech.2017.10.016>.
- Prijbelski, A., Antipov, D., Meleshko, D., Lapidus, A., Korobeynikov, A., 2020 Jun. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinformatics* 70 (1), e102. <https://doi.org/10.1002/cpbi.102> (PMID: 32559359).
- Saitou, N., Nei, M., 1987 Jul. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454> (PMID: 3447015).
- Seemann, T., 2014 Jul 15. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 30 (14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>. Epub 2014 Mar 18. PMID: 24642063.

Tripathi, L., Irorere, V.U., Marchant, R., Banat, I.M., 2018. Marine derived biosurfactants: a vast potential future resource. *Biotechnol. Lett.* 40, 1441–1457. <https://doi.org/10.1007/s10529-018-2602-8>.

Trudgeon, B., Dieser, M., Balasubramanian, N., Messmer, M., Foreman, C.M., 2020. Low-temperature biosurfactants from polar microbes. *Microorganisms* 8, 1–14. <https://doi.org/10.3390/microorganisms8081183>.

Yu, G., 2020 Mar. Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinformatics* 69 (1), e96. <https://doi.org/10.1002/cpbi.96> (PMID: 32162851).



Soluble Recombinant Protein Production in *Pseudoalteromonas haloplanktis* TAC125: The Case Study of the Full-Length Human CDKL5 Protein

Marzia Calvanese, Andrea Colarusso, Concetta Lauro, Ermenegilda Parrilli, and Maria Luisa Tutino

Abstract

The Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 is an unconventional protein production host displaying a notable proficiency in the soluble production of difficult proteins, especially of human origin. Furthermore, the accumulation of recombinant products in insoluble aggregates has never been observed in this bacterium, indicating that its cellular physicochemical conditions and/or folding processes are rather different from those observed in mesophilic bacteria. The ability of this cell factory was challenged by producing a human protein, the cyclin-dependent kinase-like 5 (*h*CDKL5) in the bacterium cytoplasm at 0 °C. Human CDKL5 is a serine/threonine protein kinase characterized by the absence of a defined structure for the last two/third of its sequence, one of the largest intrinsically disordered regions so far observed in a human protein. This large unstructured domain makes difficult its production in most of the conventional hosts since the recombinant product accumulates as insoluble aggregates and/or is heavily proteolyzed. As the full-length *h*CDKL5 production is of great interest both for basic science and as protein drug for an enzyme replacement therapy, its production in the Antarctic bacterium was tested by combining the use of a regulated psychrophilic gene expression system with the use of a defined growth medium optimized for the host growth at subzero temperature. This is the first report of soluble and full-length recombinant production of *h*CDKL5 protein in a bacterium.

Key words *Pseudoalteromonas haloplanktis* TAC125, Human CDKL5_5, Psychrophilic gene expression system pMAV, GG medium, Intrinsically disordered Proteins

1 Introduction

Pseudoalteromonas haloplanktis TAC125, a psychrophilic Gram-negative bacterium isolated from Antarctic seawater [1], is considered one of the most interesting unconventional hosts for the recombinant production of difficult-to-express proteins, thanks to favorable physiological features, including fast growth at low

Marzia Calvanese, Andrea Colarusso, and Concetta Lauro contributed equally to this work.

temperatures and efficient protein synthesis [1, 2]. Furthermore, a series of proof-of-concept studies have already demonstrated its feasibility for the recombinant expression of human proteins [3–5]. During the last decade, the number of reliable genetic systems for the recombinant gene expression in *P. haloplanktis* TAC125 was significantly increased, and the recently reported possibility to produce proteins within a range of temperature from 15 to -2.5 °C enhances the chances to improve the conformational quality and solubility of recombinant proteins [6]. Moreover, the development of synthetic media and of a finely regulated gene expression system inducible by D-galactose [6] allowed for the production of a recombinant protein at subzero temperature for the first time, thus providing an innovative strategy for the recombinant production of “difficult” proteins.

In the present chapter, we describe the main issues encountered and the implemented strategies towards the successful production of a human protein by the psychrophilic cell factory using the regulated gene expression system inducible by D-galactose (pMAV) and the synthetic medium GG [6].

The object of the study is the human cyclin-dependent kinase-like 5 (CDKL5), a serine/threonine protein kinase almost ubiquitous in vertebrates and very conserved over evolution. In mammals, it is produced mainly in the brain and in some other tissues and has been reported to be essential for the normal development of the central nervous system. Mutations in its gene are known to cause a severe neurodevelopmental disorder (CDKL5 deficiency disorder) [7, 8] accompanied by intractable epilepsy.

This protein is involved in different molecular processes and many aspects of its basic biology are still not completely understood, but it is known that it exerts its action both in the cell nucleus, in the cytoplasm and at the level of the cytoplasmic membrane. Such heterogeneity in terms of cellular localization and function is a common feature of Intrinsically Disordered Proteins (IDPs) [9], which are characterized by uncommon flexibility allowing them to meet several different molecular partners. Human CDKL5 responds to these structural criteria, as only about one-third of the protein (the N-terminal one) is precisely structured, while the remaining part is predicted to be flexible [10]. This structural peculiarity makes human CDKL5 a difficult-to-express protein in the most conventional hosts. It tends to be either heavily proteolyzed or to accumulate in insoluble aggregates. As observed in previous reports, when CDKL5 was expressed in *E. coli* the recombinant protein was exclusively obtained in the insoluble fraction [11, 12].

Human CDKL5 is known to exist in five isoforms resulting from alternative splicing [13, 14]. In this chapter, the recombinant production of the isoform *h*CDKL5_5 (previously known as CDKL5 115) is reported [15].

In detail, we describe the procedure for the cloning of *bCDKL5* gene into the pMAV psychrophilic gene expression system and its mobilization into *P. haloplanktis* TAC125 cells. The recombinant Antarctic strain was then grown in optimized culture conditions in the synthetic medium (GG) [6] at 0 °C, and different induction conditions were explored to produce the human protein. The recombinant expression was followed by monitoring *bCDKL5_5* biosynthesis to define optimal process conditions to produce the human enzyme in soluble form.

2 Materials

2.1 Bacterial Strains

Bacteria can be stored indefinitely in cultures containing 20% (v/v) of sterile glycerol (sterilize by autoclaving for 20 min at 1 atm on liquid cycle) at low temperature (from –20 to –70 °C). Add 0.25 mL of 80% (v/v) glycerol to 0.75 mL of bacterial culture, mix gently and freeze rapidly placing the tube in a dry ice/acetone mix.

1. *P. haloplanktis* TAC125. This strain was kindly provided by C. Gerday, University of Liege, Belgium. The strain was isolated from the seawater in the surroundings of the Dumont d'Urville Antarctic station (66°40' S, 40°01' E) during the 1988 summer campaign of the “Expeditions Polaires Française” in Terre Adélie [1].
2. *E. coli* DH5 α (*supE44*, Δ *lacU169* (ϕ 80 *lacZ* Δ M15) *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*). This strain was used as the host for the gene cloning.
3. *E. coli* strain S17–1(λ *pir*) (*thi*, *pro*, *hsd* (*r*[–] *m*⁺) *recA*::RP4–2-TCr::Mu Kmr::Tn7 Tpr Smr λ *pir*). This strain was used as the donor strain in intergeneric conjugation experiments [16].

2.2 Solutions

1. 100 mg/mL ampicillin stock solution: Dissolve 1 g of ampicillin powder in 8 mL of deionized water. Adjust the volume of the solution to 10 mL with dH₂O and sterilize by filtration through a 0.22 μ m sterile filter. Split the obtained stock solution in 10 aliquots of 1 mL each in sterile polypropylene tubes and store them at –20 °C.
2. 1 \times TAE buffer for agarose gel electrophoresis: 40 mM Tris-acetate, 1 mM EDTA, pH 8.0. Make a 50 \times TAE stock solution by mixing 242 g of Tris base, 57.1 mL of glacial acetic acid, 100 mL of 0.5 M EDTA, pH 8.0 and adjust the volume of the solution to 1 L with dH₂O. Store at room temperature (RT) up to 1 year.
3. 0.5 M EDTA, pH 8.0: Dissolve 186.1 g of EDTA in 800 mL of dH₂O. Adjust the pH to 8.0 with NaOH (about 20 g of

NaOH pellets) and adjust the volume of the solution to 1 L with dH₂O.

4. 3 M NaCl stock solution: Dissolve 87.6 g of NaCl in 500 mL of dH₂O.
5. GelRed precast agarose gel. BiotiumGelRed 10,000× solution is diluted to 1× by adding 3.5 μL of the stock solution to 35 mL of molten agarose gel. Once casted, protect the gel from light with aluminum foil (*see Note 1*).
6. 1 M D-galactose stock solution: Dissolve 1.8 g of solid D-galactose in 8 mL of dH₂O. Adjust the volume of the solution to 10 mL with dH₂O and sterilize by filtration through a 0.22 μm sterile filter. Store at −20 °C up to 1 year.
7. 1 M DTT stock solution: Dissolve 3.09 g of DTT in 20 mL of dH₂O. Sterilize by filtration through a 0.22 μm sterile filter. Dispense into 1 mL aliquots and close them under nitrogen flow. Store the aliquots at −20 °C.
8. 4× SDS-PAGE loading buffer: 250 mM Tris–HCl pH 6.8, 40% (v/v) glycerol, 4% (w/v) SDS, 200 mM DTT, and 0.02% bromophenol blue. This buffer lacking DTT can be stored at RT. DTT should be added just before the use from a 1 M stock solution.
9. 0.5 M Tris–HCl, pH 6.8: Dissolve 60.55 g of Tris base in 800 mL of dH₂O. Adjust the pH to 6.8 with HCl and add dH₂O to make up a final volume of 1 L.
10. 0.5% (w/v) bromophenol blue: Dissolve 0.25 g bromophenol blue powder in 45 mL of dH₂O. Shake well to dissolve the dye and then adjust the volume of the solution to 50 mL with dH₂O. Store at RT.
11. 5× Running buffer: Dissolve 15.1 g of Tris base, 94 g of glycine, and 5 g of SDS in 900 mL of dH₂O. Adjust the volume of the solution to 1 L with dH₂O.
12. 0.5 M Sodium Phosphate buffer, pH 7.8: Dissolve 68.9 g of NaH₂PO₄ in 900 mL of dH₂O. Adjust the pH to 7.8 with NaOH and add dH₂O to make up a final volume of 1 L.
13. Lysis buffer: Mix 20 mL of 0.5 M Sodium Phosphate buffer, pH 7.8, with 16.67 mL of 3 M NaCl, 0.4 mL of 0.5 M EDTA, pH 8.0, 0.1 mL of 1 M DTT solution, and 1 mL of Triton X-100. Add dH₂O to 90 mL, and eventually adjust the pH to 7.8. Take the final solution volume to 100 mL with deionized water.
14. Western blot 1× Transfer buffer: Dissolve 3.03 g Tris base, 14.41 g glycine in 800 mL of dH₂O. Add 200 mL methanol. Adjust the volume of the solution to 1 L with dH₂O.

15. Western blot blocking buffer: Dissolve 50 g Skimmed Milk in 1 L of PBS buffer (5% w/v). Add 0.5 mL Triton X-100 (0.05% v/v) and mix.
16. Western blot washing buffer: Add 0.5 mL of Triton X-100 in 1 L of PBS buffer (0.05% v/v) and mix.
17. 1× PBS: Dissolve 8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, and 0.24 g KH₂PO₄ in 800 mL of dH₂O. Adjust the pH to 7.4 with HCl. Add dH₂O to 1 L. Sterilize by autoclave.
18. Schatz Salts stock solutions:
 - (a) 100× KH₂PO₄ (100 g/L): Dissolve 100 g of KH₂PO₄ in 800 mL of dH₂O. Adjust the pH to 7.0 by NaOH addition. Add dH₂O to 1 L. Sterilize by filtration through a 0.22 μm sterile filter. Store at RT up to 6 months.
 - (b) 300× MgSO₄*7H₂O (60 g/L): Dissolve 60 g of MgSO₄*7H₂O in 800 mL of dH₂O. Add dH₂O to 1 L. Sterilize by filtration through a 0.22 μm sterile filter. Store at RT up to 6 months.
 - (c) 500× FeSO₄*7H₂O (2.5 g/L): Dissolve 2.5 g of FeSO₄*7H₂O in 800 mL of dH₂O. Adjust the pH to 4.0–5.0 by NaOH addition. Add dH₂O to 1 L. Sterilize by filtration through a 0.22 μm sterile filter. Store at RT up to 6 months.
 - (d) 3000× CaCl₂*2H₂O (15 g/L): Dissolve 15 g of CaCl₂*2H₂O in 800 mL of dH₂O. Adjust the pH to 7 by NaOH addition. Add dH₂O to 1 L. Sterilize by filtration through a 0.22 μm sterile filter. Store at RT up to 6 months.

2.3 Media

1. LB medium (1 L) [17]: 10 g Bacto-Tryptone, 5 g Bacto-Yeast Extract, 10 g NaCl. Add 950 mL of dH₂O. Shake until the solutes have dissolved. Adjust the volume of the solution to 1 L with dH₂O. Sterilize by autoclaving for 20 min at 1 atm on liquid cycle. Let it cool down and store at RT. When required, add 1 mL of sterile ampicillin stock solution. To prepare solid medium, add 15 g/L Bacto-Agar just before autoclaving.
2. TYP Medium (1 L) [18]: 16 g Bacto-Tryptone, 16 g Bacto-Yeast Extract, 10 g NaCl, add 950 mL of dH₂O. Shake until the solutes have dissolved. Adjust the volume of the solution to 1 L with dH₂O. Sterilize by autoclaving for 20 min at 1 atm on liquid cycle. When required, add 1 mL of sterile ampicillin stock solution. To prepare solid medium, add 15 g/L Bacto-agar just before autoclaving.
3. GG medium (1 L) [6]: 10 g L-glutamic acid monosodium salt monohydrate, 10 g gluconic acid sodium salt, 10 g NaCl, 1 g NH₄NO₃ in 900 mL of dH₂O. Adjust the pH to 7.8 with

NaOH and reach 1 L before autoclaving. After sterilization, dilute Schatz Salts stock solutions to 1× concentration (*see Note 2*).

When required, add 1 mL of sterile ampicillin stock solution to 1 L medium to have a 100 µg/mL final concentration. For bacterial growths at 0 °C a lower amount of antibiotic was used, i.e., 0.25 mL per liter of medium (25 µg/mL final concentration).

2.4 Reagents

1. Restriction enzymes: *NdeI*, *XhoI*, *SaII*.
2. Calf Intestinal alkaline Phosphatase, CIP.
3. T4 DNA Ligase.
4. QIAquick® PCR Purification kit.
5. QIAprep® Spin Miniprep kit.
6. Anti-CDKL5 Antibody (D-12, Santa Cruz Biotechnology).
7. Peroxidase conjugated anti-mouse IgG.
8. Clarity Western ECL substrate.
9. cComplete mini Protease Inhibitor cocktail tablet.
10. DNA gel extraction kit.

2.5 Vectors

1. pMAV: a psychrophilic gene expression system in which the expression is induced by galactose supply (*see Note 3*).
2. pET28a-*hCDKL5*: an expression vector containing the *E. coli* codon-optimized sequence coding for the human CDKL5 isoform 5 (*see Note 4*).

2.6 Equipment

1. Nüve-cooled incubator (model ES120).
2. Spectrophotometer.
3. Bandelin Sonoplus HD 3200 sonicator with an MS72 probe or equivalent.
4. ChemiDoc MP Imager.
5. Tans-Blot SD Semi-Dry Transfer Cell.

3 Methods

3.1 *hCDKL5_5* Expression Vector Construction

HCDKL5_5 gene was extracted from pET28a-*hCDKL5* vector by double hydrolysis with *NdeI* and *XhoI* restriction enzymes and cloned into pMAV plasmid double digested with *NdeI/SaII*. *XhoI* and *SaII* produced complementary cohesive ends and this allowed for a direct ligation reaction without further modifications.

1. Digest the pET28a-*hCDKL5_5* vector with *NdeI* and *XhoI* enzymes. Restriction hydrolysis is performed by using 5 enzyme

Units/μg of DNA, in the reaction conditions defined by the manufacturer.

2. The DNA is purified after agarose gel electrophoresis using a commercial purification kit.
3. The pMAV vector is digested with *NdeI* and *SaII* enzymes.
4. The DNA is purified using a commercial purification kit.
5. The 5' phosphate groups of the cleaved vector are dephosphorylated by treatment with calf intestinal alkaline phosphatase (CIP) (1 U/pmol of 5' phosphate ends) for 30 min at 37 °C by using the appropriate buffer delivered with the enzyme.
6. The dephosphorylated DNA is then purified using a PCR purification kit following the manufacturer's instructions.
7. The cleaved dephosphorylated vector is then ligated to the digested fragment by using ligation reactions, by the means of T4 DNA ligase according to the supplier's instructions.
8. The ligation reaction mixture is used directly for the transformation of the chemically competent bacteria (*E. coli* DH5α strain).
9. Recombinant clones are selected on LB agar plates containing 100 μg/mL ampicillin as the selection agent.
10. Plasmids are isolated from ampicillin resistant clones by a suitable mini prep kit, and the presence of the appropriate insert is verified by restriction digestion analysis.

3.2 Construction of *P. haloplanktis* TAC125 (pMAV-hCDKL5_5) Recombinant Strain

1. The resulting vector pMAV-*hCDKL5_5* is mobilized into *P. haloplanktis* TAC125 by intergeneric conjugation [16]. After the mating step, cells are suitably diluted and plated on TYP solid medium containing 100 μg/mL ampicillin and incubated at 4 °C to select recombinant *P. haloplanktis* TAC125 (the low temperature hinders *E. coli* growth as colonies on solid media).
2. Three isolated colonies are picked and each one inoculated in 3 mL of TYP liquid medium containing 100 μg/mL ampicillin and incubated at 4 °C under shaking (250 rpm) for 48–72 h.
3. Plasmidic DNA is extracted from each clone by using a mini-prep kit and recombinant plasmid clones are screened by restriction digestion analysis.

3.3 hCDKL5-5 Production

All psychrophiles, including *P. haloplanktis* TAC125, are exposed to permanent oxidative stress at low temperatures, which originates from increased oxygen solubility. Therefore, *hCDKL5_5* production trials were performed at different liquid-to-gas ratios in 1 L shaken flasks. Recombinant *P. haloplanktis* (pMAV- *hCDKL5_5*) batch cultivation was performed in different conditions:

- 200 mL of GG supplemented with Schatz Salts (20% fill volume)
- 350 mL of GG supplemented with Schatz Salts (35% fill volume)
- 700 mL of GG supplemented with Schatz Salts (70% fill volume).

In all the above conditions, 25 µg/mL ampicillin was added to maintain a suitable selection for the growth of recombinant clones. All the production trials were performed in 1 L Erlenmeyer flasks at 0 °C with shaking at 160 rpm in Nüve-cooled incubator (model ES120). 40 mM of D-galactose was used as the inducer of recombinant expression and the production was evaluated after 8, 24, and 48 h post induction. Cell growth was monitored by measuring the optical density (OD) at 600 nm using a spectrophotometer.

3.3.1 Pre-culture

The viability of the pre-cultured cells is crucial for a satisfying process outcome. At the time of the inoculation of the production batch, the bacteria must be in middle exponential phase in the same medium that will be used for the expression growth. Furthermore, temperature control is crucial for the reproducibility of the experiments (*see Note 5*).

1. From a glycerol stock streak the *P. haloplanktis* TAC125 (pMAV-*hCDKL5_5*) strain onto a TYP agar plate containing 100 µg/mL ampicillin. Incubate it at 15 °C for about 36 h. The plate can be stored up to 3 days at 4 °C, carefully sealed with Parafilm to avoid oxygen availability to the cells (*see Note 6*).
2. Pick a single colony and inoculate it into 2 mL of liquid TYP medium supplemented with 100 µg/mL ampicillin in a 15 mL snap-cap inoculation tube, and incubate at 15 °C under vigorous shaking (180 rpm) for 24 h.
3. Perform a 1:100 dilution of the culture of **step 2** in 10 mL of GG medium supplemented with Schatz Salts and 100 µg/mL ampicillin in a 100 mL Erlenmeyer flask and incubate for 24 h at 15 °C under shaking (180 rpm).
4. Dilute the pre-culture of **step 3** to 0.2 OD_{600nm} in 50 mL of GG medium supplemented with Schatz Salts and 25 µg/mL ampicillin in a 250 mL Erlenmeyer flask and incubate for 48–72 h at 0 °C under shaking (160 rpm). The final biomass concentration should be about 2.5–3.0 OD_{600nm}.

3.3.2 Production Process

1. Inoculate the amount of pre-culture required to obtain a starting concentration of OD_{600nm} = 0.1. To calculate it, register the optical density of the pre-culture at 600 nm using a spectrophotometer. Calculate the volume of inoculum by using the following formula:

$$\text{mL inoculum} = (\text{culture OD}_{600\text{nm}} \times \text{mL culture volume}) / (\text{pre-culture OD}_{600\text{nm}}).$$

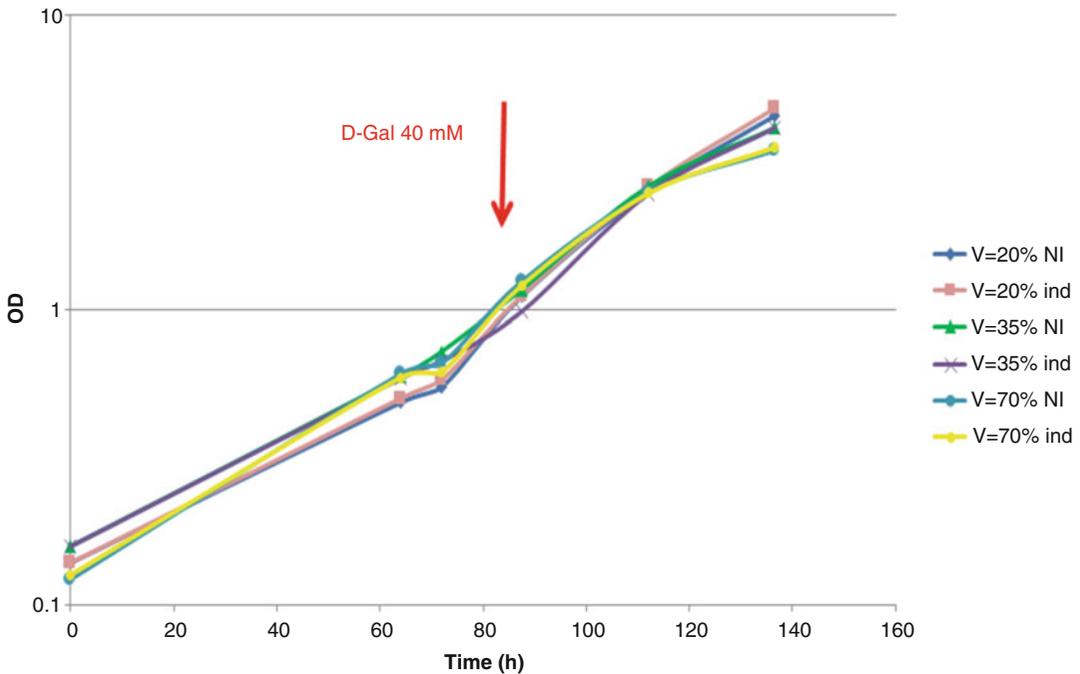


Fig. 1 Growth profiles of recombinant *P. haloplanktis* TAC125 (pMAV-*hCDKL5_5*) at 0°C in GG synthetic medium performed at different percentage of liquid gas ratio

2. Monitor the cell growth by measuring the optical density at 600 nm as described above. Register the data of at least two measurements to avoid the technical error. When the cell density reaches an OD_{600nm} of 0.8–1.2, which corresponds to the exponential phase, induce the recombinant gene expression with D-galactose addition. Add the required amount of 1 M D-galactose sterile stock solution to obtain the optimal inducer concentration of 40 mM.
3. At different times (8 h, 24 h, and 48 h) after the induction, collect two samples from each culture corresponding to an OD_{600nm} of 1 and 20 OD, respectively. Calculate the volume of each sample using the following formula:

$$\text{mL sample} = X \text{ OD}_{600nm} / \text{culture optical density } OD_{600nm}.$$

1 OD pellets will be used for total cellular expression analysis, while 20 OD pellets will be lysed for solubility screenings. Collect samples by centrifuging the calculated volume for 15 min at $4000 \times g$ at 4 °C. Discard the supernatant and store the biomass indefinitely at –80 °C.

4. Plot the optical density values versus the time of cultivation in graphs. In a typical process (Fig. 1) during the first exponential phase, the highest specific growth rate is reached ($\mu_{max} = 0.03 \text{ h}^{-1}$).

3.4 CDKL5 Production Analysis

To analyze the *b*CDKL5_5 production, a cellular lysis followed by Western blotting analysis is required.

3.4.1 Cell Lysis

1. Add 1 tab of cOMplete mini Protease Inhibitor cocktail tablet to 10 mL of Lysis buffer and mix until the tablet is completely dissolved (*see* **Note 7**).
2. Resuspend the collected bacterial pellet ($OD_{600nm} = 20$) at different fermentation time points in 2.0 mL of Lysis buffer by pipetting or vortexing.
3. Sonicate the samples for 15 min at 15% Amplitude (15 s pulse, 30 s pause). During the sonication process, place the sample in an ice/water bath to avoid sample heating.
4. Centrifuge the suspensions at $14,000 \times g$ for 20 min at 4 °C. Recover the resulting supernatant containing the total soluble protein extract for further analysis. Keep the protein extract on ice or store it at 4 °C for no longer than 2 h.

3.4.2 Western Blot Protein Detection

Protein samples were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (10%,w/v acrylamide) and transferred to a polyvinylidenedifluoride membrane (PVDF) with a Tans-Blot SD Semi-Dry Transfer Cell and immunodetection was performed with the primary anti-CDKL5 antibody.

1. For production analysis in total extracts, solubilize the 1 OD pellets of each sample prepared at Subheading 3.3.2 with 60 μ L of 4 \times SDS-PAGE loading buffer and boil the samples for 20 min at 95 °C. For the solubility analysis, mix 40 μ L of the soluble fractions collected as indicated at Subheading 3.4.1 with 13 μ L of 4 \times SDS-PAGE loading buffer and boil for 5 min at 95 °C. Load 5 μ L of the samples onto 10% SDS-PAGE gels. Run the gels for 90 min at constant 100 V.
2. Wash the gel 3 times with transfer buffer for 10 min.
3. Transfer the proteins on a 0.2 μ m PVDF membrane previously activated in methanol according to the manufacturer's instructions using a Tans-Blot SD Semi-Dry Transfer Cell for 30 min at constant 15 V.
4. Block the membrane for 1 h at room temperature (RT) in blocking buffer under shaking.
5. Dilute anti-CDKL5 antibody 1:1000 in blocking buffer by diluting 20 μ L of antibody in 20 mL. Incubate the membrane with the primary antibody solution for 1 h at RT under shaking.
6. Discard the primary antibody solution and wash the membrane 3 times with Western blot washing buffer for 10 min.

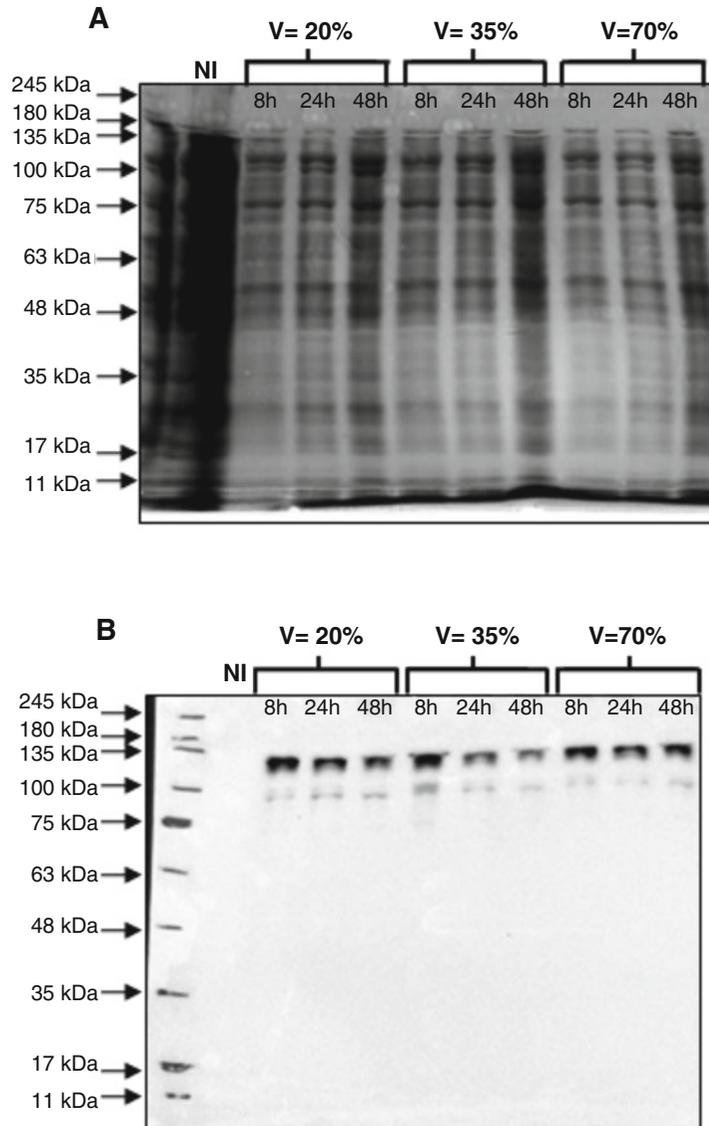


Fig. 2 *hCDKL5_5* production in *P. haloplanktis* TAC125 (pMAV-*hCDKL5_5*). **(a)** SDS-PAGE analysis of total lysates of *P. haloplanktis*TAC125 (pMAV-*hCDKL5_5*) grown at 0 °C in GG. **(b)** Analysis of *hCDKL5_5* recombinant production by Western blot using an anti-CDKL5 monoclonal antibody

7. Dilute anti-mouse HRP-conjugated antibody 1:20,000 in blocking buffer by diluting 1 μ L of antibody in 20 mL. Incubate the membrane with the primary antibody solution for 1 h at RT under shaking.
8. Discard the secondary antibody solution and wash the membrane 5 times with Western blot washing buffer for 10 min.

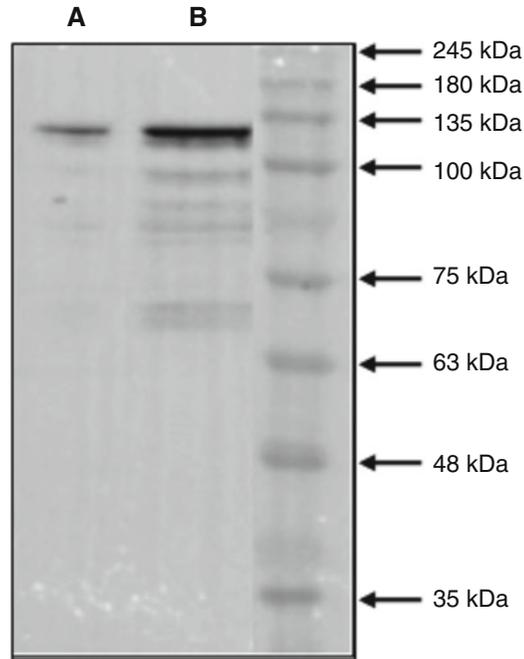


Fig. 3 Western blotting analysis of the soluble fraction of *P. haloplanktis* TAC125 (pMAV-*hCDKL5_5*) cell lysate 8 h after induction with 40 mM galactose at 0 °C. (a) 5 µg of soluble fraction. (b) 10 µg of soluble fraction

9. Develop the Western blot using the Clarity Western ECL substrate or equivalent following the manufacturer's instructions.
10. Acquire the chemiluminescent image with a ChemiDoc MP Imager and analyze with the IMAGE-lab Software.

The analysis (Fig. 2) reveals that *hCDKL5_5* is produced in the different tested conditions and that the different liquid-to-gas ratios slightly influence the production yields, while the sampling timing after induction influences the amount of recovered protein. As shown in Fig. 2b, the higher amount of protein is obtained 8 h after induction in total protein extracts. It is worth noting that the protein is present in the soluble fraction (Fig. 3) and, although a degradation pattern is present, the main product is still the larger one. In conclusion, when produced at 0 °C, the human recombinant protein CDKL5_5 resulted to be produced in soluble form. Further investigation has to be carried out to increase the production yields and to reduce proteolysis of the recombinant protein.

4 Notes

1. GelRed can be added while the gel solution is still hot.
2. The GG medium complemented with Schatz Salts can be stored one day at RT.
3. The galactose-inducible pMAV plasmid was used to produce recombinant proteins at subzero temperature [6]. pMAV deriving from the pUC18 plasmid is characterized by the presence of: (a) the pJB3-derived OriT [19], a DNA fragment responsible for the initiation of the conjugative transfer between *E. coli* S17-1 λ pir strain (donor) and the psychrophilic cells (acceptor); (b) a pUC18-derived polylinker wherein the target gene can be cloned; (c) the *E. coli* blaM gene, encoding a mesophilic β -lactamase which is used for the selection of the recombinant clones; (d) OriC, the origin of replication allowing the plasmid to replicate in *E. coli*; (e) the T/R box, a DNA fragment containing the cold-adapted origin of replication (OriR) [16]; (f) the DNA region upstream to the *Phgal* operon containing both the *PhgalTK* promoter and the gene coding for *PbGalR* [6]; (g) the *TaspC*, the transcription termination signal of the aspartate aminotransferase gene (*aspC*) isolated from *P. haloplanktis* TAC125 [20].
4. pET28a-*hCDKL5* vector was kindly provided by Prof. E. Ciani, University of Bologna.
5. Keep bacterial cultures cold while working with or transporting them using ice during the growths at 15 °C and directly sampling in the incubator during expression experiments at 0 °C.
6. The psychrophilic bacteria are able to grow at the storage temperature of 4 °C. Limiting oxygen availability can reduce growth but not avoid it.
7. The addition of a wide spectrum protease inhibitors cocktail to the lysis buffer to prevent proteolytic degradation of the recombinant product is crucial because of the intrinsic flexibility of *hCDKL5_5*, which makes it an easy target of endogenous proteases.

Acknowledgments

This work was supported by Programma Nazionale di Ricerca in Antartide Project PNRA18_00335, and by the Italian parent's association "La fabbrica dei sogni 2-New developments for Rett syndrome."

References

- Médigue C, Krin E, Pascal G et al (2005) Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* 15 (10):1325–1335. <https://doi.org/10.1101/gr.4126905>
- Parrilli E, Tutino ML (2017) Heterologous protein expression in *Pseudoalteromonas haloplanktis* TAC125. In: *Psychrophiles: from biodiversity to biotechnology*, 2nd edn. Springer, Cham, pp 513–525
- Vigentini I, Merico A, Tutino ML et al (2006) Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J Biotechnol* 127:141–150. <https://doi.org/10.1016/j.jbiotec.2006.05.019>
- Corchero JL, Gasser B, Resina D et al (2013) Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. *Biotechnol Adv* 31:140–153. <https://doi.org/10.1016/j.biotechadv.2012.09.001>
- Unzueta U, Vázquez F, Accardi G et al (2015) Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A. *Appl Microbiol Biotechnol* 99:5863–5874. <https://doi.org/10.1007/s00253-014-6328-9>
- Sannino F, Giuliani M, Salvatore U et al (2017) A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl Microbiol Biotechnol* 101(2):725–734. <https://doi.org/10.1007/s00253-016-7942-5>
- Fehr S, Wilson M, Downs J et al (2013) The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. *Eur J Hum Genet* 21:266–273. <https://doi.org/10.1038/ejhg.2012.156>
- Hector RD, Kalscheuer VM, Hennig F et al (2017) CDKL5 variants. *Neurol Genet* 3:e200. <https://doi.org/10.1212/nxg.0000000000000200>
- Alowolodu O, Johnson G, Alashwal L et al (2016) Intrinsic disorder in spondins and some of their interacting partners. *Intrinsically Disord Proteins* 4:e1255295. <https://doi.org/10.1080/21690707.2016.1255295>
- Fahmi M, Yasui G, Seki K et al (2019) In silico study of Rett syndrome treatment-related genes, MECP2, CDKL5, and FOXP1, by evolutionary classification and disordered region assessment. *Int J Mol Sci* 20:1–19. <https://doi.org/10.3390/ijms20225593>
- Kameshita I, Sekiguchi M, Hamasaki D et al (2008) Cyclin-dependent kinase-like 5 binds and phosphorylates DNA methyltransferase 1. *Biochem Biophys Res Commun* 377:1162–1167. <https://doi.org/10.1016/j.bbrc.2008.10.113>
- Katayama S, Inazu T (2019) Straightforward and rapid method for detection of cyclin-dependent kinase-like 5 activity. *Anal Biochem* 566:58–61. <https://doi.org/10.1016/j.ab.2018.11.013>
- Wilmes B, Kock H, Glagla S et al (2011) Cytoplasmic and periplasmic proteomic signatures of exponentially growing cells of the psychrophilic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Appl Environ Microbiol* 77:1276–1283. <https://doi.org/10.1128/AEM.01750-10>
- Hector RD, Dando O, Landsberger N et al (2016) Characterisation of CDKL5 transcript isoforms in human and mouse. *PLoS One* 11:1–22. <https://doi.org/10.1371/journal.pone.0157758>
- Fichou Y, Nectoux J, Bahi-Buisson N et al (2011) An isoform of the severe encephalopathy-related CDKL5 gene, including a novel exon with extremely high sequence conservation, is specifically expressed in brain. *J Hum Genet* 56:52–57. <https://doi.org/10.1038/jhg.2010.143>
- Tutino ML, Duilio A, Parrilli E et al (2001) A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles* 5 (4):257–264. <https://doi.org/10.1007/s007920100203>
- Lessard JC (2013) Molecular cloning. *Methods Enzymol* 529:85–98. <https://doi.org/10.1016/B978-0-12-418687-3.00007-0>
- Parrilli E, Duilio A, Tutino ML (2008) Heterologous protein expression in psychrophilic hosts. In: *Psychrophiles: from biodiversity to biotechnology*. Chapter 21. Springer, Cham, pp 365–379
- Blatny JM, Brautaset T, Winther-Larsen HC et al (1997) Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl Environ Microbiol* 63:370–379. <https://doi.org/10.1128/aem.63.2.370-379.1997>
- Birolo L, Tutino ML, Fontanella B et al (2000) Aspartate aminotransferase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125. Cloning, expression, properties, and molecular modelling. *Eur J Biochem* 267:2790–2802. <https://doi.org/10.1046/j.1432-1327.2000.01299.x>



ELSEVIER

Contents lists available at ScienceDirect

Research in Microbiology

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

Original Article

Conditional gene silencing in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125

Concetta Lauro^{a, b}, Andrea Colarusso^{a, b}, Marzia Calvanese^a, Ermenegilda Parrilli^a,
Maria Luisa Tutino^{a, *}

^a Dipartimento di Scienze Chimiche, Complesso Universitario Monte Sant'Angelo, Via Cintia, 80126, Napoli, Italy

^b Istituto Nazionale Biostrutture e Biosistemi—I.N.B.B., Viale Medaglie d'Oro, 305-00136, Roma, Italy



ARTICLE INFO

Article history:

Received 26 November 2021

Accepted 9 March 2022

Available online 17 March 2022

Keywords:

PTasRNA

Lon protease

Truncated hemoglobin

Gene silencing

Antarctic bacterium

Pseudoalteromonas haloplanktis (*translucida*)

TAC125

ABSTRACT

Since the release in 2005 of the genome sequence and annotation of the first Antarctic marine bacterium, the number of genomes of psychrophilic microorganisms in public databases has steadily increased. Unfortunately, the lack of effective molecular tools for the manipulation of these environmental strains still hampers our understanding of their peculiar strategies to thrive in freezing conditions, limiting the functional genomics approaches to differential analyses only. Over the past two decades, our research group established the first effective gene cloning/expression technology in the Antarctic Gram-negative marine bacterium *Pseudoalteromonas haloplanktis* TAC125. The setup of a genome mutagenesis technique (based on homologous recombination and counterselection events) further supported the use of this strain, which became an attractive model for studying microbial adaptations to freezing lifestyle. Moreover, to further extend the functional analyses to its essential genes, the set-up of a conditional gene silencing approach is desirable. In this paper, we report the development of an asRNA regulatory system in the Antarctic bacterium, testing the feasibility of Hfq-dependent and PTasRNA strategies previously developed in *Escherichia coli*. Stable and efficient silencing of two chromosomal genes was obtained by using PTasRNAs, reaching very high levels of downregulation.

© 2022 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Small RNAs (sRNAs) are regulatory RNAs involved in controlling gene expression at the post-transcriptional and/or translational level by base-pairing with their target mRNA [1]. Given the typical complementarity of their sequence with their targets, they are often denoted as antisense RNAs (asRNAs). Most bacterial asRNAs act by binding to the translation initiation region (TIR) of the transcripts, so as to interfere with the efficiency of ribosome recruiting. As a result, the translation is hindered, and the target mRNA is destabilized and susceptible to degradation [2]. Furthermore, many asRNAs can act against multiple targets explaining why they are involved in various processes such as stress responses, metabolic switches, virulence, biofilm formation, and much more [1].

In recent years, the repurposing of natural asRNAs to develop efficient gene regulation tools has widely spread. Indeed, several technologies were established in bacteria, based on engineered asRNAs, such as synthetic asRNA [3], CRISPR interference (CRISPRi) [4], CRISPR activation (CRISPRa) [5], small transcription activating RNA (STAR) [6] and riboswitch [7]. Relying on the Watson-Crick base pairing principle, such strategies hold many advantages compared to DNA or protein-based methods. They allow easy, predictable, scalable, and finely tunable control of multiple genes – even essential ones [8] – for very diverse applications, spanning from physiological studies to metabolic engineering and therapeutics [9,10].

Na and coworkers reported a strategy to design synthetic asRNAs in *Escherichia coli* (*E. coli*) with high repression capability (>90%) [3]. They identified MicC as one of the asRNA able to efficiently interact with the RNA chaperone protein Hfq in *E. coli* and developed an Hfq-dependent asRNA by the functional modularization of MicC sequence into a target binding region and a scaffold that recruits Hfq. Even if the gene silencing is strictly exerted by the target binding sequence and is strongly dependent on the efficiency

* Corresponding author. Fax: +39 081 674313.

E-mail addresses: concetta.lauro@unina.it (C. Lauro), andrea.colarusso@unina.it (A. Colarusso), marzia.calvanese@unina.it (M. Calvanese), erparrilli@unina.it (E. Parrilli), tutino@unina.it (M.L. Tutino).

of TIR masking, the interaction of synthetic asRNA with Hfq ensures a more efficient outcome.

Nakashima and collaborators designed another RNA-based regulatory tool in *E. coli*, known as Paired Termini antisense RNAs (PTasRNAs). A PTasRNA consists of a hairpin structure whose loop holds the antisense sequence able to bind the RBS of the target mRNA [11]. The complementarity of the terminal regions positively affects PTasRNAs stability, consequently enhancing their cellular concentration and silencing efficacy compared to canonical asRNA. PTasRNAs turned out to be quite successful in regulating multiple essential genes in *E. coli* [12].

Notably, the versatility and ease of such approaches can also be extended to less characterized and almost “wild” environmental microorganisms, often tricky to handle in routine genetic manipulations, given the often observed ineffectiveness of conventional mutagenesis techniques, due to the low efficiency of genome integration of exogenous DNA through homologous recombination [13–15].

In this respect, from the release of the genome sequence and annotation of the first Antarctic marine bacterium in 2005 [16], the number of genomes of psychrophilic microorganisms in public databases has steadily increased. On the contrary, the setup of molecular tools available for the manipulation of these environmental strains did not follow the same trend, limiting genetic interventions to few genera only. Among them, *Pseudoalteromonas haloplanktis* TAC125 (*PhTAC125*) [16], also named *Pseudoalteromonas translucida* TAC125, was the first Antarctic Gram-negative bacterium in which an effective gene cloning/expression technology was established [17,18], becoming a robust model for studying microbial adaptations to freezing lifestyle [19–21]. Although a genome mutagenesis technique was established in *PhTAC125* by our group [22], it relies on homologous recombination and counterselection events, often occurring at a frequency only slightly higher than spontaneous mutations. Moreover, the set-up of a conditional gene silencing approach in the psychrophilic bacterium would significantly widen the tools for its gene functional analysis, as the gene insertion/deletion technology is not applicable to essential genes.

In this work, Hfq-dependent and PTasRNA strategies were applied to the silencing of the *PhTAC125* gene coding for the Lon protease. This gene was chosen because the commercial antibodies raised against the *E. coli* Lon protease efficiently recognize the psychrophilic protease, owing to their high sequence similarity, allowing the easy evaluation of the silencing effects at the protein level by quantitative Western blotting. Furthermore, Lon is a popular target for the improvement of many common hosts for recombinant protein production, as it is reported as the main protease involved in the degradation of intracellular proteins [23,24]. The PTasRNA technology gave back the more efficient and durable silencing results, and it was applied to the downregulation of a less-intensively transcribed *PhTAC125* gene (*PhhbO*). It encodes a truncated hemoglobin (trHbO) involved in the cellular protection against oxidative stress [25]. The phenotypic characterization of the interfered strain demonstrated that *PhhbO* expression was efficiently repressed by the PTasRNA strategy, extending the design principles of these synthetic asRNAs to the application in psychrophilic bacteria.

2. Materials and methods

2.1. Bacterial strains and growth conditions

E. coli DH5 α [*supE44*, *ΔlacU169* (ϕ 80 *lacZΔM15*) *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*] was used for cloning procedures. *E. coli* S17-1 (λ *pir*) [*thi*, *pro*, *hsd* (*r- m+*) *recA::RP4-2-TCr::Mu Kmr::Tn7*,

Tpr, *Smr*, *λpir*] [26] constituted the donor strain in intergeneric conjugations for KrPL transformations [17,27]. KrPL – a *PhTAC125* strain cured of its endogenous pMtBL plasmid [27] – was used as the host for expressing antisense RNAs throughout the study. Both *E. coli* strains were grown in LB broth (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L NaCl) supplemented with 34 μg/mL chloramphenicol at 37 °C. KrPL was grown in TYP (16 g/L bacto-tryptone, 16 g/L yeast extract, 10 g/L NaCl) during conjugation experiments and in GG (10 g/L L-glutamic acid monosodium salt monohydrate, 10 g/L D-gluconic acid sodium salt, 10 g/L NaCl, 1 g/L NH₄NO₃, 1 g/L K₂HPO₄, 200 mg/L MgSO₄·7H₂O, 5 mg/L FeSO₄·7H₂O, 5 mg/L CaCl₂) [21] in recombinant production experiments. Recombinant KrPL was grown in solid media supplemented with 12.5 μg/mL chloramphenicol, while in liquid media the antibiotic was added at 25 μg/mL.

2.2. Construction of expression plasmids

The nucleotide sequences of synthetic asRNAs and oligonucleotide primers used in this work are reported in Table 1.

The DNA fragment containing the antisense sequence targeting *lon* mRNA, MicC scaffold and T1/TE terminator was synthesized by Thermo Fisher Scientific and cloned into pB40-79C, a high-copy number derivative of p79C [27] (unpublished data), using NheI and SacI restriction sites. The resulting construct was pB40-79C-asRNA-*lon* (Fig. S1A).

For the construction of pB40-79C-PTasRNA-*lon* and pB40-79C-5'-PTasRNA-*lon*, the intermediate plasmid pB40-79BsC was developed. pB40-79BsC differs from pB40-79C for the removal of HindIII, EcoRI, NheI and NdeI restriction sites present in the divergent *lacR* gene and *PlacZ* promoter, and for the introduction of BsaI and PstI sites immediately downstream of the *PlacZ* promoter (Fig. S1B). To obtain this plasmid, a synthetic fragment (Thermo Fisher Scientific) containing the *lacR* gene and *PlacZ* promoter with the above-mentioned modifications was digested in SphI and PstI sites and cloned into a high-copy number derivative of pUCC [28] (unpublished data) hydrolyzed with the same enzymes. Then, the synthetic DNA fragments (Thermo Fisher Scientific) containing PTasRNA-*lon* and 5'-PTasRNA-*lon* sequences were cloned into pB40-79BsC using BsaI and SmaI restriction sites (Fig. S1C).

For the construction of pB40-79C-PTasRNA2-*lon* and pB40-79C-PTasRNA3-*lon*, two DNA fragments of the *lon* gene (*PSHA_RS10175*) were amplified by PCR using the PT2_fw- PT2_rv and PT3_fw-PT3_rv primer pairs, respectively. The reaction was performed using 1 unit Phusion DNA Polymerase (New England Biolabs), 0.5 μM each primer, 200 μM dNTP, 1X HF Phusion buffer, and 50 ng *PhTAC125* genomic DNA as the template. The thermocycling condition was set up as follows: one cycle at 98 °C for 30 s, 25 cycles at 98 °C for 10 s, 67 °C for 30 s, 72 °C for 15 s, followed by the last cycle at 72 °C for 5 min. Then, the obtained amplicons were digested with PstI and XhoI and cloned into pB40-79C-PTasRNA-*lon* hydrolyzed with the same restriction enzymes (Fig. S1D).

pB40-79C-PTasRNA-*hbO* was constructed with the same strategy. The DNA fragment containing the TIR of *hbO* (*PSHA_RS00150*) was obtained through a PCR amplification by using hbO_fw-hbO_rv primers. The reaction was performed using 1 unit Phusion DNA Polymerase (New England Biolabs), 0.5 μM each primer, 200 μM dNTP, 1X HF Phusion buffer, and 50 ng *PhTAC125* genomic DNA as the template. The thermocycling condition was set up as follows: one cycle at 98 °C for 30 s, 20 cycles at 98 °C for 10 s, 71 °C for 30 s, 72 °C for 15 s, followed by the last cycle at 72 °C for 5 min. Then, pB40-79C-PTasRNA-*lon* and the amplicon were digested with PstI and XhoI and ligated (Fig. S1D).

The sequence of all PCR-amplified fragments was checked to rule out the occurrence of unwanted mutations.

Table 1

Nucleotide sequences of synthetic asRNAs and oligonucleotide primers used in this work. Fw: forward; Rv: reverse.

asRNA	Sequence (5' - 3')
asRNA- <i>lon</i>	TCGATCGGTTCTCTCAAGCGTCATTTC TGTTGGGCCAATGCAATGCCACTGATTTT CCAACATATAAAAAGACAAGCCGAAACAG TCGTCGGGCTTTTTTCTCGAGCCAGGC ATCAAATAAAACGAAAGGCTCAGTCGAAA GACTGGGCTTTTCGTTTTATCTGTTTTTGT CGGTGAACGCTCTCTACTAGAGTCACACTG GCTACCTTCGGGTGGGCTTTCTGCGTTTATA TATAATTGCCGTTTATGCAACGGAATAAAC AGGAGGAATTAACCATGCAAGTGGTGGTG GTGGTGCTGCAGGCGGTATACCACATCATC GCGCAGTCTAATACTGGGATTCGACTCGA TCGGTCTCTCAAGCGTCATTATATTCTCTT TGCACTTCATTTATTTTAAAGATTACTCTCG AGCACCACCACCACCACCTGCATGGTTA ATTCTCTCCCGG
5'-PTasRNA- <i>lon</i>	AGGAGGAATTAACCATGCAAGTGGTGGTG GTGGTGCTGCAGGCGGTATACCACATCATC GCGCAGTCTAATACTGGGATTCGACTCGA TCGGTCTCTCAAGCGTCATTATATTCTCTT TGCACTTCATTTATTTTAAAGATTACTCTCG AGCACCACCACCACCACCTGCATGGTTA ATTCTCTCCCGG
PTasRNA1- <i>lon</i>	AGGAGGAATTAACCATGCAAGTGGTGGTG GTGGTGCTGCAGGCGGTATACCACATCATC GCGCAGTCTAATACTGGGATTCGACTCGA TCGGTCTCTCAAGCGTCATTATATTCTCTT TGCACTTCATTTATTTTAAAGATTACTCTCG AGCACCACCACCACCACCTGCATGGTTA ATTCTCTCCCGG
PTasRNA2- <i>lon</i>	AGGAGGAATTAACCATGCAAGTGGTGGTG GTGGTGCTGCAGCATGTGCGGTATACCACATC ATCGCGCAGTCTAATACTGGGATTCGACTC GATCGGTTCTCTCAAGCGTCATTATATTCTCT TTGCACTTCATTTATTTTAAAGATTACTCG AGCACCACCACCACCACCTGCATGGTTA ATTCTCTCCCGG
PTasRNA3- <i>lon</i>	AGGAGGAATTAACCATGCAAGTGGTGGTG GTGGTGCTGCAGTATACCACATCATCGCGCA GTGCTAATACTGGGATTCGACTCGATCGGTT CTCTCAAGCGTCATTATATTCTCTTTGCATT CATTTATTTTAAAGATTACTCAAGTACTCGAG CACCACCACCACCACCACCTGCATGGTTA ATTCTCTCCCGG
PTasRNA- <i>hbO</i>	AGGAGGAATTAACCATGCAAGTGGTGGTG GTGGTGCTGCAGGTGTTTTTCAGGTGGGT GTTTGCTCTATTGTGGCGGCTTAGATTTTGA AAAAAGTCGTTTAATCATATATTATTCATCG TTAATAGGCAGTATTACTACTGCAAACTCGAG CACCACCACCACCACCACCTGCATGGTTA ATTCTCTCCCGG
Primer	Sequence (5' - 3')
PT2_fw	CAGGTCGAGATCTTAAAAATAATGAAGTCAAA
PT2_rv	CGACTGCAGCATGTGCGGTATACCACATAC
PT3_fw	CAGGTCGAGTACTTGAGTAATCTTTAAAAATAAT
PT3_rv	CGACTGCAGTATACCACATCATCGCGC
hbO_fw	CAGGTCGAGTTTGCACTAGTAATACTGCCTATTA
hbO_rv	CGACTGCAGGTGTTTTTCAGGTGGTGGG

2.3. Recombinant production procedures

Glycerol stocks (-80 °C) of KrPL recombinant strains were streaked over TYP agar plates and incubated at 15 °C for 72 h. Then, a single colony was inoculated in 2 mL of TYP at 15 °C for 1–2 days. After a first adaptation step, performed by diluting 100-fold the culture in GG medium, the inoculum was performed at 0.1 OD/mL. The recombinant production was performed at 15 °C and the induction was carried out with 10 mM IPTG when the cells reached the middle-late exponential growth phase. A Biosan PSU-20i orbital shaker was used setting the agitation at 180 rpm.

2.4. SDS-PAGE and western blotting

To quantify Lon protease intracellular levels, cell pellets corresponding to the number of cells yielding an absorbance of 1 OD₆₀₀

ml⁻¹, were collected by centrifugation at 11,000×g at 4 °C and solubilized in 60 µL of Laemmli buffer 4X. Then, the samples were boiled at 95 °C for 20 min and centrifuged at 10,000×g for 1 min at RT. 3 µL of samples were run on a TGX Stain-Free Mini-Protein 4–15% (BioRad) gel in TGS buffer for 90 min at 110 V. The gel was then Stain-Free activated for 2.5 min and imaged using the ChemiDoc MP imaging system (BioRad) and ImageLab software (version 6.0, BioRad). Then, the separated proteins were transferred to an Immuno-Blot low-fluorescence polyvinylidene fluoride (PVDF) membrane (BioRad) in 7 min using the TransBlot Turbo Transfer System (BioRad) with the mixed molecular weight setting. After the transfer, the membrane was again imaged and blocked with PBS, 0.05% Triton X-100, 5% (w/v) milk for 1 h. Then, an anti-*E*Lon antibody (ab103809) was diluted 1:5000 in the same buffer. After 1 h of incubation at RT with the primary antibody, the membrane was washed with PBS, 0.05% Triton X-100 three times (5 min each) and incubated with an HRP-conjugated anti-rabbit antibody diluted 1:3000 in PBS, 0.05% Triton X-100, 5% (w/v) milk for 1 h at RT. Then, the membrane was washed again with PBS, 0.05% Triton X-100 three times (5 min each) and the secondary antibody was detected using the ECL method (Cyanagen).

2.5. Western blotting data analysis

The signal intensities of Lon in each lane were determined by using the “Lane and Bands” tool of ImageLab software (version 6.0, BioRad). The densitometric analysis was performed by normalizing bands to total proteins in each lane detected on the blot membrane as previously described [29].

2.6. Bioinformatics analysis

The prediction of RNA secondary structures was performed using mFold website with default settings [30]. RNAPredator [31] was used for the prediction of asRNA-mRNA interaction setting NC_007481 for the selection of the *PhTAC125* genome.

2.7. H₂O₂ disk-inhibition assay

Cultures of psychrophilic non-recombinant cells, and those harboring either the pB40-79C-PTasRNA-*hbO* vector or the empty vector pB40-79BsC were diluted in 7.5 mL warm TYP soft agar (4 g/L agar) at a final concentration of 0.2 OD/mL. When required, 12.5 µg/mL chloramphenicol and 10 mM IPTG were added to the media. Then, a disk of Whatman filter paper was soaked in 256 mM H₂O₂ and placed in the center of each plate. After 24 h of incubation at 15 °C, the diameter of the killing zone was measured. Data deriving from two independent experiments, each carried out in quintuplicate, were used to calculate data reported in Table 3.

2.8. Statistics and reproducibility of results

Data were statistically validated using the t-Student test comparing the mean measurements of silenced and non-silenced samples. The significance of differences between mean values was calculated using a two-tailed Student's t-test. A P value of <0.05 was considered significant.

3. Results

3.1. Construction of Hfq-dependent asRNA-*lon* and evaluation of Phlon silencing efficiency in *PhTAC125*.

Following the established design criteria (as described in [3]), the asRNA-*lon* was obtained by fusing the Hfq-binding scaffold

Table 2

Prediction of the interactions occurring between PTasRNAs and their target *lon* mRNA in *PhTAC125*. RNAPredator tool was used to predict the total binding energy, the corresponding Z-score, the coordinates on asRNA and mRNA. Each prediction is presented with the experimental results achieved in *PhTAC125* expressed as maximum silencing activity.

PTasRNA-mRNA binding energy						
Variant	Energy [kJ/mol]	z-Score	asRNA	mRNA [Start]	mRNA [End]	Maximum silencing activity
PTasRNA1- <i>lon</i>	-41.19	-8.05	95-124	-9	20	70%
PTasRNA2- <i>lon</i>	-39.47	-7.57	105-134	-14	15	54%
PTasRNA3- <i>lon</i>	-35.20	-9.92	44-73	37	66	47%

Table 3

Diameter of the disk-inhibition zone. Bacterial growth sensitivity to H₂O₂ is directly proportional to the diameter of the disk-inhibition zone. IND, The given recombinant strain was grown and tested in the presence of 10 mM IPTG to induce the transcription of PTasRNA-*hbO* or the void vector. The reported values represented the mean of two independent experiments, each carried out in quintuplicate, and were considered significant when $p < 0.001$ according to the t-Student test.

Strain	Diameter of the disk-inhibition zone (cm)
<i>PhTAC125</i> wt	0.16 ± 0.04
<i>PhTAC125</i> -30mut	1.24 ± 0.09
KRPL-pB40-79BsC IND	0.28 ± 0.06
KRPL-pB40-79C-PTasRNA- <i>hbO</i> IND	1.32 ± 0.08

derived from *E. coli* MicC asRNA with a sequence complementary to the first 24 bp of the TIR of the *PhLon* mRNA (*PSHA_RS10175*) (see Table 1). The synthetic asRNA and the target mRNA are expected to interact with an energy of hybridization that can be predicted using RNAPredator tool. The expected hybridization energy of asRNA-*lon* with its target turned out to be $-42 \text{ kcal mol}^{-1}$, placing the interaction within the optimal binding energy range from -30 to $-40 \text{ kcal mol}^{-1}$ [3]. Finally, a strong transcription terminator T1/

TE was also added to guarantee the release of the nascent asRNA transcript (Fig. 1A, B).

Once designed, asRNA-*lon* was cloned into pB40-79C, a high copy number variant of p79C [27] (unpublished data). Then, the vector was transferred into KrPL, a *PhTAC125* strain devoid of the endogenous pMtBL plasmid [27], without any noticeable effect on the growth behavior of the recombinant cells (Fig. S2). The recombinant expression of Hfq-dependent asRNA-*lon* was performed as described in the Material and Methods section. The induction was performed during the exponential growth phase and the occurrence of gene silencing was evaluated by quantitative Western blot. In particular, the analysis was performed on samples collected 2, 4, 8, 24 and 32 h after the induction by measuring Lon levels detected in total protein extracts of cells expressing the asRNA-*lon* in comparison to cells carrying the empty vector and collected at the same time point post induction (NC). The amount of Lon protease detected in any given NC sample was set to unit and used to normalize the amount detected in the silenced cells (Fig. 1C). A statistically relevant reduction in Lon concentration was recorded only 2 h post-induction. At the following time points, the protease was consistently detected in almost comparable quantities in the two strains.

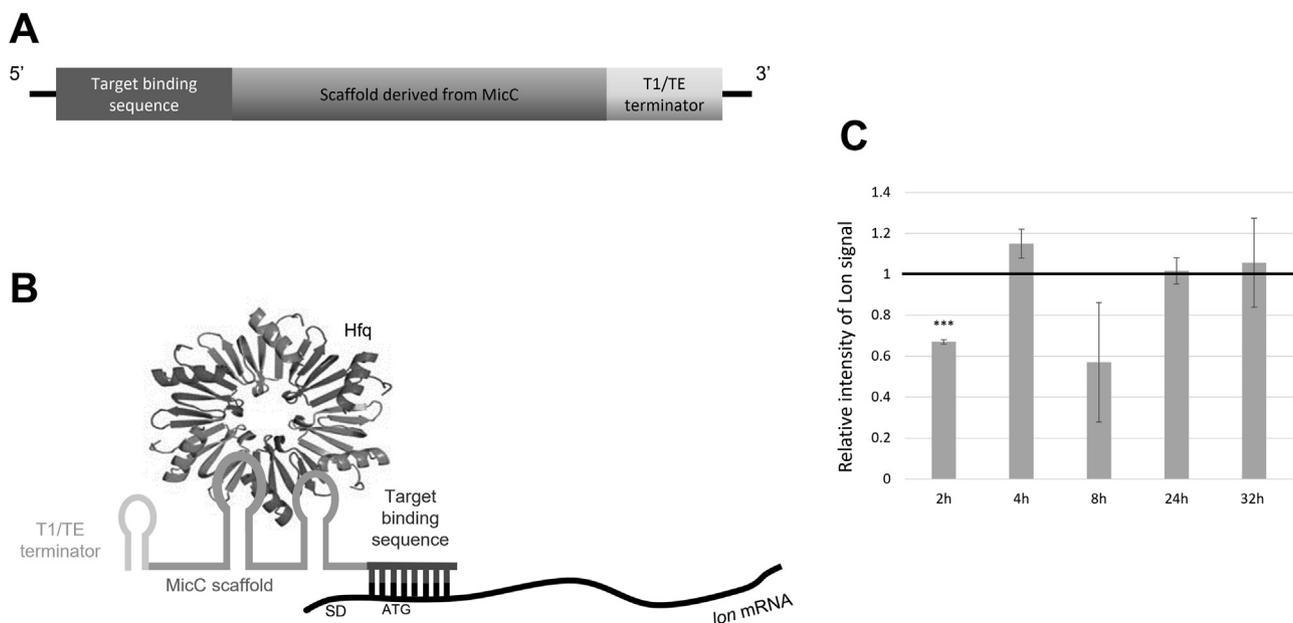


Fig. 1. Evaluation of *Phlon* gene silencing mediated by Hfq-dependent asRNA-*lon*. (A) Schematic representation of Hfq-dependent asRNA-*lon* construct. A sequence complementary to the TIR of *lon* mRNA was fused to the sequence of *E. coli* MicC recruiting Hfq proteins. To ensure the termination of the transcription, a synthetic T1/TE terminator was added. (B) Suggested mode of action of Hfq-dependent asRNA-*lon*. The antisense sequence binds the Transcription Initiation Region (TIR) of *lon* mRNA, downstream of the Shine-Dalgarno (SD) sequence, sequestering the transcript from the ribosomes. In this way, it acts by repressing the translation. The Hfq protein binds to the MicC derived scaffold, aiding the formation of the asRNA-mRNA complex resulting in a more efficient gene silencing. (C). Evaluation of relative concentration of Lon protease in cells expressing Hfq-dependent asRNA-*lon* compared to control cells, carrying the empty vector. Western blotting was performed on total protein extracts of cells recovered 2, 4, 8, 24 and 32 h after the induction of asRNA-*lon* transcription using anti-EcLon antibodies. The black bar represents the baseline amount of Lon detected in the control samples. The measurements are reported as the mean of two independent experiments whose standard deviation is indicated by the error bars. The data were considered significant when $p < 0.05$ (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$) according to t-Student test.

3.2. Construction of PTasRNAs and evaluation of Phlon silencing efficiency in PhTAC125

A second approach to achieve an efficient *Phlon* gene silencing was performed by using Paired Termini antisense RNAs (PTasRNAs). These molecules contain 38 nt-long terminal inverted repeats, characterized by a high GC content, and allowed a high silencing efficiency in *E. coli* without triggering any interference on plasmid stability [11]. mFold secondary structure prediction tool was used to arrange the antisense sequence (110 nt), harboring the *Phlon* RBS and start codon, within a loop region easily accessible for the hybridization with the target mRNA (Fig. 2A, B). To explore the effect of the specific binding region on TIR sequestering, three variants were designed, each characterized by a shift of 5 nt of the antisense sequence towards *Phlon* 3'-end, resulting in PTasRNA3-*lon*, PTasRNA1-*lon*, and PTasRNA2-*lon*, respectively (Fig. 3A). The constructs were cloned into the psychrophilic expression vector pB40-79BsC and expressed in KrPL without any noticeable effect on the growth behavior of the recombinant cells (Fig. S3A).

As previously described changes in Lon protease concentration in cells expressing the different PTasRNAs were evaluated through quantitative Western blot analysis (Fig. 3B). Again, total protein extracts of KrPL cells bearing the empty vector and subjected to the same induction protocol were used to assess the normal Lon amount in non-silenced cells. PTasRNA1-*lon* markedly reduced the concentration of Lon by about 70% after 8 h of induction, and the effect lasted up to 32 h. Both variants PTasRNA2-*lon* and PTasRNA3-*lon* achieved lower silencing efficacy than the first version, with a maximum activity of about 54% and 47%, respectively. Furthermore, they showed a less durable effect since the repression began to wane after 24 h post-induction.

Interestingly, a direct correlation between repression efficiency and interaction binding energy predicted by RNAPredator software was observed (Table 2). This finding suggested that a strong interaction is required for satisfying target suppression also in psychrophilic bacteria.

For the reasons stated above, PTasRNA1-*lon* was selected for further experiments.

As described by Nakashima and coworkers, the presence of a 5' leader sequence upstream to the hairpin structure may exert a stabilizer effect on the synthetic RNA half-life [11], potentially enhancing its repression efficiency. Hence, we designed a variant of PTasRNA1-*lon*, called 5'-PTasRNA1-*lon*, differing just for the presence of the extremely stable 5'-UTR from *PhlacZ* mRNA [27]. This asRNA was constructed and cloned into pB40-79BsC (Fig. 4A), and its recombinant production was then performed in KrPL cells as described above.

Once confirmed that the overexpression of PTasRNAs did not induce either any plasmid instability or non-specific silencing of other essential genes (Fig. S3B), their silencing capability was analyzed by quantitative Western blotting (Fig. 4B). The densitometric analysis revealed that 5'-PTasRNA1-*lon* is less effective than PTasRNA1-*lon* during the first 8 h of expression. Then, the two constructs exerted a comparable effect with about 70% reduction of Lon levels after 24 h of induction.

These results suggested that the 5'-UTR sequence did not contribute to the enhancement of the asRNA repression capability.

3.3. Validation of PTasRNA technology in PhTAC125 by silencing PhHbO encoding gene

The design of PTasRNA-*hbO* was performed considering the results obtained with the *lon* target. In detail, the synthetic asRNA was made up by the already used 38 bp long terminal inverted repeats, while the 110 nt long antisense sequence was selected following the predictions of RNAPredator tool. The sequence potentially binding to the TIR elements of *PhhbO* mRNA with the highest estimated efficiency (binding energy of -36.81 kJ/mol), and negligible non-specific off-target binding, was PCR amplified. Then, the gene coding the assembled PTasRNA-*hbO* was cloned into the psychrophilic expression vector pB40-79BsC and recombinantly expressed in KrPL. Since the *PhTAC125* truncated hemoglobin trHbO displays a notable peroxidase activity [25], the occurrence of the conditional control of its mRNA translation was evaluated by a disk-diffusion assay in the presence of 256 mM H₂O₂ (Fig. 5). As expected, *PhTAC125* wild type cells and recombinant KrPL

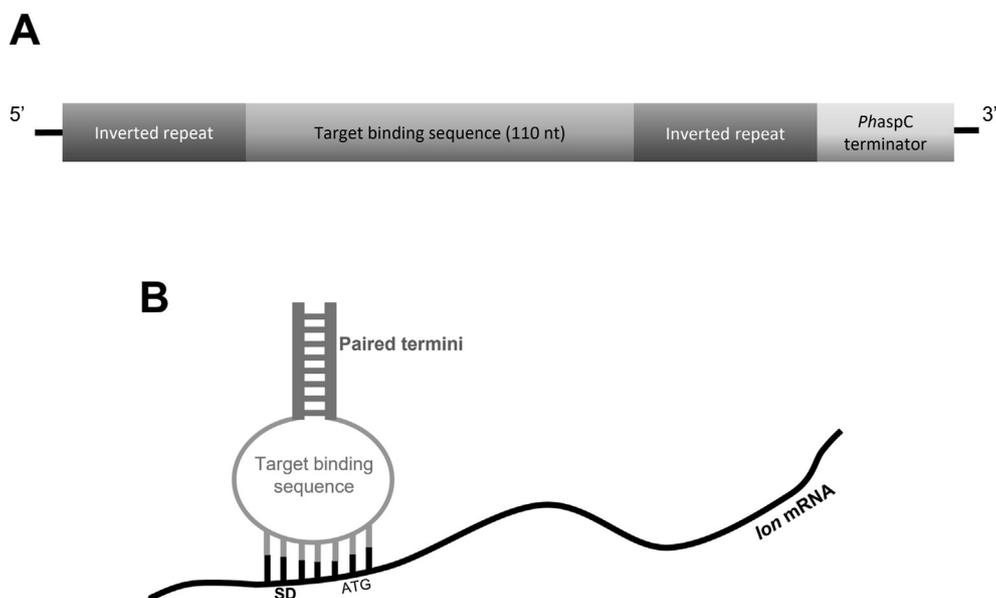


Fig. 2. Schematic representation of PTasRNA. (A) The antisense sequence, complementary to the TIR of *lon* mRNA, was included between two inverted repeats. (B) The inverted repeats at the ends of the sequence form a hairpin structure containing the antisense sequence of the target gene to be silenced into the loop. SD, Shine-Dalgarno sequence, ATG, translation start site.

A



B

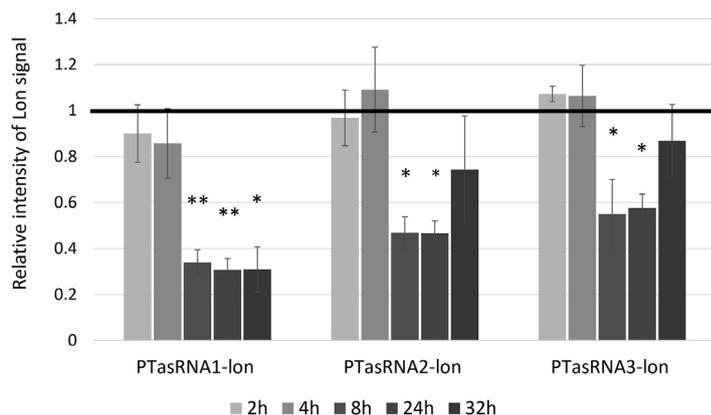


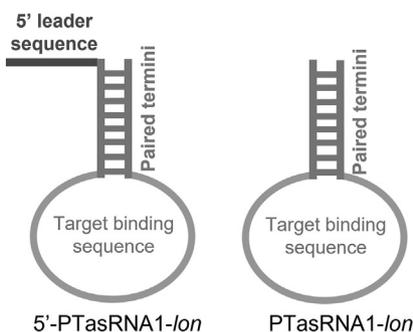
Fig. 3. Evaluation of *Phlon* gene silencing mediated by PTasRNA-*lon*. (A) Schematic illustration of the antisense fragment positions of PTasRNA-*lon* variants. The fragment positions are shown as bars covering the *Phlon* mRNA, whose predicted SD and translation start site are reported in bold and underlined bold, respectively. PTasRNA1-*lon* and PTasRNA2-*lon* were designed with a shift of 5 and 10 nt of the PTasRNA3-*lon* antisense sequence towards *Phlon* mRNA 3'-end, respectively. (B) Evaluation of relative cellular concentration of Lon protease in cells expressing the three variants of PTasRNA-*lon* in comparison to control cells, carrying the empty vector. Western blotting was performed on total protein extracts of cells recovered 2, 4, 8, 24 and 32 h after the induction of asRNA-*lon* transcription using anti-EcLon antibodies. The black bar represents the baseline amount of Lon detected in the control samples. The measurements are reported as the mean of two independent experiments whose standard deviation is indicated by the error bars. The data were considered significant when $p < 0.05$ (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$) according to t-Student test.

harboring the empty vector - used as positive controls – showed a clear peroxide-resistant phenotype, characterized by an almost negligible killing zone. In the same experimental conditions the *PhTAC125-30* knockout mutant [25] was sensitive to the oxidative stress, and its phenotype in the disk-diffusion assay was used to evaluate if the *PhhbO* conditional gene silencing was achieved (Fig. 5 and Table 3). Interestingly, the PTasRNA-*hbO* expression was coupled to an H₂O₂ growth inhibition comparable to the one observed when the *PhTAC125* knockout mutant was tested (see Table 3), suggesting that the *PSHA_RS00150* gene silencing reached a level so high to be considered as total repression.

4. Discussion

The availability of several techniques for genome-wide genetic manipulations in common bacterial hosts paved the way for targeted strain engineering in the framework of both fundamental and applied studies. However, the application of such methods in environmental strains still represents a challenging task. *P. haloplanktis* TAC125 (*PhTAC125*) is a model for environmental studies because of its notable adaptation to low temperatures lifestyle [19–21]. Despite this psychrophilic bacterium has been successfully used as a non-conventional platform for the

A



B

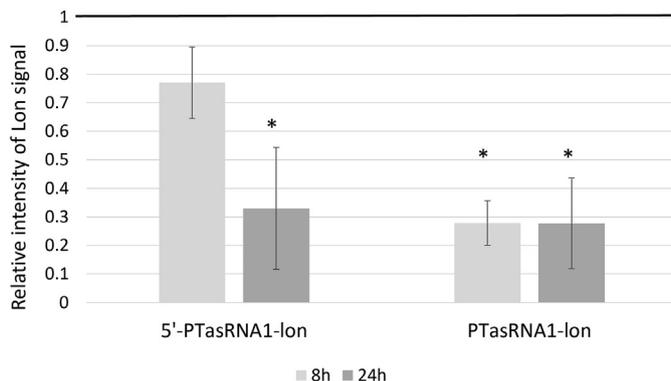


Fig. 4. Evaluation of the effect of a 5' leader sequence upstream of the hairpin structure of PTasRNA1-*lon* (A) Schematic representation of 5'-PTasRNA1-*lon* construct. The 5' leader sequence derived from *PhlacZ* was placed upstream of the inverted repeat involved in the formation of the hairpin structure, whose loop holds the antisense sequence. (B) Evaluation of the relative cellular concentration of Lon protease in cells expressing either 5'-PTasRNA1-*lon* or PTasRNA1-*lon* in comparison to control cells, carrying the empty vector. Western blotting was performed on total protein extracts of cells recovered 8 and 24 h after the induction of asRNA-*lon* transcription using anti-EcLon antibodies. The black bar represents the baseline amount of Lon detected in the control samples. The measurements are reported as the mean of two independent experiments whose standard deviation is indicated by the error bars. The data were considered significant when $p < 0.05$ (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$) according to t-Student test.

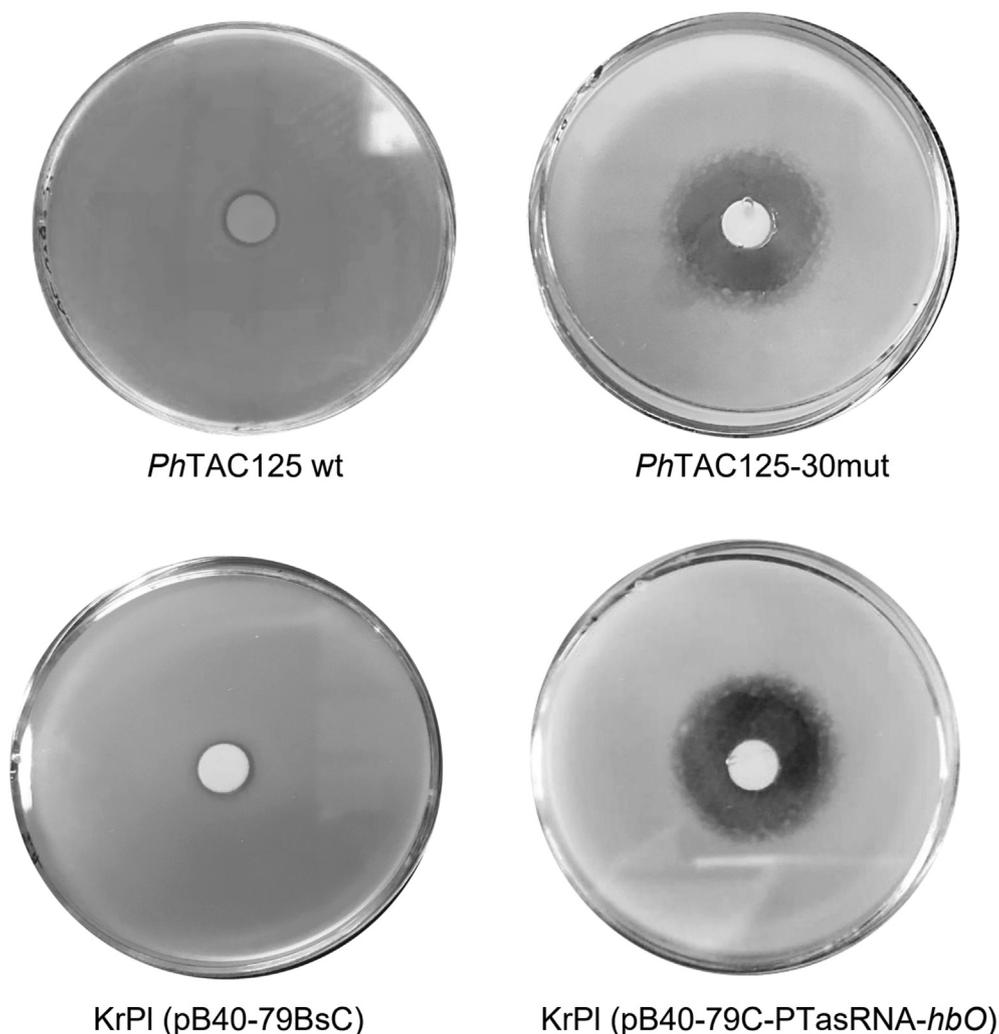


Fig. 5. Evaluation of PTasRNA-*hbO* expression on hydrogen peroxide sensitivity of recombinant *PhTAC125* cells. Disk diffusion assay in the presence of H_2O_2 (265 mM) of *PhTAC125*, *PhTAC125-30mut*, KrPL pB40-79BsC and KrPL pB40-79C-PTasRNA-*hbO* grown for 24 h at 15 °C. The peroxide sensitivity is directly proportional to the diameter of the killing zone (the dark halo) surrounding the Watmann paper disk.

recombinant production of difficult-to-produce proteins [32–36], the only consolidated genome mutagenesis technology is based on a two-steps integration-segregation procedure relying on homology recombination [22]. Nevertheless, the frequency of the second recombination event is often very low, and the “clean” gene deletion is a rare event. Furthermore, this disruption method is not suitable for the study of essential genes.

The present study intended to develop effective methods for conditional gene silencing in *PhTAC125* using synthetic antisense RNAs. Based on their mode of action, two kinds of asRNAs were evaluated: Hfq-dependent asRNA and Paired Termini asRNA. These synthetic molecules were challenged for their ability to suppress the translation of the Lon protease encoding mRNA.

Although no endogenous regulatory asRNAs of *PhTAC125* have been identified so far, the Hfq-dependent asRNA strategy was pursued relying on the observed high identity degree (85%) between the *PhTAC125* and *E. coli* Hfq amino acid sequences (Fig. S4). However, our results demonstrated that MicC scaffold-based asRNA was able to interfere with the *lon* gene expression only after 2 h post-induction, while no *lon* silencing was evident at later time points. As the cellular Hfq concentration in bacterial cells is generally limiting, it is important that Hfq leaves the mRNA-asRNA complex upon formation, to exert its RNA chaperone function on

other molecules. Most bacterial Hfqs, including the *E. coli* one, possess an intrinsically disordered region (IDR) in the C-terminal domain (CTD), which is needed to displace base-paired RNAs, so as to allow Hfq recycling [40]. *PhTAC125* Hfq is characterized by a shorter CTD, in which the predicted IDR domain is almost totally deleted (Fig. S5). Our hypothesis is that the psychrophilic Hfq binds MicC asRNA and stabilize its interaction with *lon* mRNA, but lacking the disordered region in its CTD, it could be unable to leave the complex, resulting (after only 2 h post-induction) in the reduction of the cellular concentration of the free chaperone.

We benefited from the structural features of PTasRNAs to obtain successful and efficient gene silencing in *PhTAC125*. The different efficiency of *lon* down-regulation achieved by the three PTasRNA-*lon* variants (differing in 5' and 3' interaction boundaries) confirmed that the predicted binding energy is a critical parameter to establish an efficient gene control system also at low temperatures. Our finding is in agreement with what was described by Nakashima and coworkers [11]. Indeed, they reported that slightly different targeting sequences could influence the efficiency of down-regulation of *lacZ* mRNA but a general design rule was not assessed. Furthermore, a 5' leader sequence upstream of PTasRNA-*lon* did not improve its repression capability. It can be concluded that the hairpin setup of PTasRNA is

sufficiently stable to induce the desired effect also in the psychrophilic bacterium.

However, a complete silencing of *lon* was never observed in our experimental conditions. To highlight the potential bottlenecks, if any, limiting the gene silencing outcomes, we focused our attention on the relative abundance of the target mRNA and PTasRNA. Indeed, a molar excess of asRNA is essential for efficient gene silencing, to out-compete ribosomes for binding with the target transcript [41]. Since a transcriptomic analysis of *PhTAC125* cells grown at 15 °C placed the *lon* gene among the 140 most transcribed ones (unpublished data from this laboratory), we hypothesized that our gene expression system was unable to establish a sufficient PTasRNA-*lon* molar excess. However, total silencing could be obtained targeting a less expressed gene. A good candidate for this new trial was the *PhhbO* gene, which is transcribed about 40 times less than *lon* (unpublished data from this laboratory). Interestingly, the psychrophilic cells expressing PTasRNA-*hbO* display the same peroxide-sensitive phenotype presented by the knockout mutant [25], thus supporting a compelling reduction of trHbO protein concentration likely due to the total repression of *hbO* mRNA translation.

In this work, we defined the guidelines for designing a conditional gene control system based on asRNA in *P. haloplanktis* TAC125. For the first time we demonstrated that the down-regulation of two genes (selected as transcribed at quite different levels) is feasible also in a wild type Antarctic marine bacterium, one of the widely acknowledged models for studying bacterial adaptation to freezing lifestyle.

Ethical statements

There are no ethical issues relevant to this work.

Declaration of competing interest

There are no conflicts of interest with authors.

Acknowledgements

This research was partially funded by the Italian “Programma Nazionale di Ricerca in Antartide” (PNRA project number PNRA18_00335) and by the Italian Parents' Association “La fabbrica dei sogni 2 - New developments for Rett syndrome”.

We are grateful to Prof. Suparna Sanyal (University of Uppsala, Sweden) for her precious suggestions and critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.resmic.2022.103939>.

References

- [1] Wagner EGH, Romby P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it, vol. 90. Elsevier Ltd; 2015. <https://doi.org/10.1016/bs.adgen.2015.05.001>.
- [2] Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* 2011;43:880–91. <https://doi.org/10.1016/j.molcel.2011.08.022>.
- [3] Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat Biotechnol* 2013;31:170–4. <https://doi.org/10.1038/nbt.2461>.
- [4] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013;152:1173–83. <https://doi.org/10.1016/j.cell.2013.02.022>.
- [5] Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 2013;41:7429–37. <https://doi.org/10.1093/nar/gkt520>.
- [6] Chappell J, Takahashi MK, Lucks JB. Creating small transcription activating RNAs. *Nat Chem Biol* 2015;11:214–20. <https://doi.org/10.1038/nchembio.1737>.
- [7] Jang S, Jang S, Yang J, Seo SW, Jung GY. RNA-based dynamic genetic controllers: development strategies and applications. *Curr Opin Biotechnol* 2018. <https://doi.org/10.1016/j.copbio.2017.10.005>.
- [8] Qi LS, Arkin AP. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat Rev Microbiol* 2014;12:341–54. <https://doi.org/10.1038/nrmicro3244>.
- [9] Ahmed W, Hafeez MA, Ahmed R. Advances in engineered trans-acting regulatory RNAs and their application in bacterial genome engineering. *J Ind Microbiol Biotechnol* 2019;46:819–30. <https://doi.org/10.1007/s10295-019-02160-y>.
- [10] Ji Y, Zhang B, Van Horn SF, Warren P, Woodnutt G, Burnham MKR, et al. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001;293:2266–9. <https://doi.org/10.1126/science.1063566>.
- [11] Nakashima N, Tamura T, Good L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in *Escherichia coli*. *Nucleic Acids Res* 2006;34:1–10. <https://doi.org/10.1093/nar/gkl697>.
- [12] Nakashima N, Tamura T. Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Res* 2009;37. <https://doi.org/10.1093/nar/gkp498>.
- [13] Zess EK, Begemann MB, Pflieger BF. Construction of new synthetic biology tools for the control of gene expression in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Biotechnol Bioeng* 2016;113:424–32. <https://doi.org/10.1002/bit.25713>.
- [14] Cho C, Lee SY. Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs. *Biotechnol Bioeng* 2017;114:374–83. <https://doi.org/10.1002/bit.26077>.
- [15] Sun T, Li S, Song X, Pei G, Diao J, Cui J, et al. Re-direction of carbon flux to key precursor malonyl-CoA via artificial small RNAs in photosynthetic *Synechocystis* sp. PCC 6803. *Biotechnol Biofuels* 2018;11:1–17. <https://doi.org/10.1186/s13068-018-1032-0>.
- [16] Medigue C, Krin E, Bernsel A, Bertin PN, Cheung F, Amico SD, et al. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* 2005;1–12. <https://doi.org/10.1101/gr.4126905.1>.
- [17] Tutino ML, Duilio A, Parrilli E, Remaut E, Sannia G, Marino G. A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles* 2001;5:257–64. <https://doi.org/10.1007/s007920100203>.
- [18] Parrilli E, Tutino ML. In: Margesin R, editor. Heterologous protein expression in *Pseudoalteromonas haloplanktis* TAC125. 2017 “psychrophiles from bi-divers. To biotechnol. 2nd ed. Springer; 2017. <https://doi.org/10.1007/978-1-4614-9065-4>.
- [19] Piette F, D’Amico S, Mazzucchelli G, Danchin A, Leprince P, Feller G. Life in the cold: a proteomic study of cold-repressed proteins in the antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Appl Environ Microbiol* 2011;77:3881–3. <https://doi.org/10.1128/AEM.02757-10>.
- [20] Mocali S, Chiellini C, Fabiani A, Decuzzi S, Pascale D, Parrilli E, et al. Ecology of cold environments: new insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. *Sci Rep* 2017;7:1–13. <https://doi.org/10.1038/s41598-017-00876-4>.
- [21] Sannino F, Giuliani M, Salvatore U, Apuzzo GA, de Pascale D, Fani R, et al. A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl Microbiol Biotechnol* 2017;101:725–34. <https://doi.org/10.1007/s00253-016-7942-5>.
- [22] Giuliani M, Parrilli E, Pezzella C, Ripa V, Duilio A, Marino G, et al. A novel strategy for the construction of genomic mutants of the antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Methods Mol Biol* 2012;824:219–33. https://doi.org/10.1007/978-1-61779-433-9_11.
- [23] Tsilibaris V, Maenhaut-Michel G, Van Melderen L. Biological roles of the Lon ATP-dependent protease. *Res Microbiol* 2006;157:701–13. <https://doi.org/10.1016/j.resmic.2006.05.004>.
- [24] Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 1986;189:113–30. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2).
- [25] Parrilli E, Giuliani M, Giordano D, Russo R, Marino G, Verde C, et al. The role of a 2-on-2 haemoglobin in oxidative and nitrosative stress resistance of Antarctic *Pseudoalteromonas haloplanktis* TAC125. *Biochimie* 2010;92:1003–9. <https://doi.org/10.1016/j.biochi.2010.04.018>.
- [26] Tascon RI, Rodriguez-Ferri EF, Gutierrez-Martin CB, Rodriguez-Barbosa I, Berche P, Vazquez-Boland JA. Transposon mutagenesis in *Actinobacillus pleuropneumoniae* with a Tn10 derivative. *J Bacteriol* 1993;175:5717–22. <https://doi.org/10.1128/jb.175.17.5717-5722.1993>.
- [27] Colarusso A, Lauro C, Calvanese M, Parrilli E, Tutino ML. Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a cell factory: IPTG-inducible plasmid construction and strain engineering. *Microorganisms* 2020;8:1–24. <https://doi.org/10.3390/microorganisms8101466>.

- [28] Parrilli E, Giuliani M, Tutino ML. General secretory pathway from marine antarctic *Pseudoalteromonas haloplanktis* TAC125. *Mar Genomics* 2008;1: 123–8. <https://doi.org/10.1016/j.margen.2009.01.002>.
- [29] Taylor SC, Posch A. The design of a quantitative western blot experiment. *BioMed Res Int* 2014. <https://doi.org/10.1155/2014/361590>.
- [30] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406–15. <https://doi.org/10.1093/nar/gkg595>.
- [31] Eggenhofer F, Tafer H, Stadler PF, Hofacker IL. RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res* 2011;39:149–54. <https://doi.org/10.1093/nar/gkr467>.
- [32] Vigentini I, Merico A, Tutino ML, Compagno C, Marino G. Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J Biotechnol* 2006;127:141–50. <https://doi.org/10.1016/j.jbiotec.2006.05.019>.
- [33] Giuliani M, Parrilli E, Sannino F, Apuzzo GA, Marino G, Tutino ML. Recombinant production of a single-chain antibody fragment in *Pseudoalteromonas haloplanktis* TAC125. *Appl Microbiol Biotechnol* 2014;98:4887–95. <https://doi.org/10.1007/s00253-014-5582-1>.
- [34] Unzueta U, Vázquez F, Accardi G, Mendoza R, Toledo-Rubio V, Giuliani M, et al. Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A. *Appl Microbiol Biotechnol* 2015;99:5863–74. <https://doi.org/10.1007/s00253-014-6328-9>.
- [35] Calvanese M, Colarusso A, Lauro C, Parrilli E, Tutino ML. Soluble Recombinant Protein Production in *Pseudoalteromonas Haloplanktis* TAC125: the Case Study of the Full-Length Human CDKL5 Protein. *Insoluble Proteins Methods Protoc*. 2nd ed. 2021.
- [36] Fondi M, Gonzi S, Dziurzynski M, Turano P, Ghini V, Calvanese M, et al. Modelling hCDKL5 heterologous expression in bacteria. *Metabolites* 2021;11: 491. <https://doi.org/10.3390/metabo11080491>.
- [40] Santiago-Frangos A, Kavita K, Schu DJ, Gottesman S, Woodson SA. C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA. *Proc Natl Acad Sci USA* 2016;113:E6089–96. <https://doi.org/10.1073/pnas.1613053113>.
- [41] Chen H, Ferbeyre G, Cedergren R. Efficient hammerhead ribozyme and anti-sense RNA targeting in a slow ribosome *Escherichia coli* mutant. *Nat Biotechnol* 1997;15:432–5. <https://doi.org/10.1038/nbt0597-432>.

RESEARCH

Open Access



Active human full-length CDKL5 produced in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125

Andrea Colarusso^{1,2}, Concetta Lauro^{1,2}, Marzia Calvanese¹, Ermenegilda Parrilli^{1,2} and Maria Luisa Tutino^{1*}

Abstract

Background: A significant fraction of the human proteome is still inaccessible to in vitro studies since the recombinant production of several proteins failed in conventional cell factories. Eukaryotic protein kinases are difficult-to-express in heterologous hosts due to folding issues both related to their catalytic and regulatory domains. Human CDKL5 belongs to this category. It is a serine/threonine protein kinase whose mutations are involved in CDKL5 Deficiency Disorder (CDD), a severe neurodevelopmental pathology still lacking a therapeutic intervention. The lack of successful CDKL5 manufacture hampered the exploitation of the otherwise highly promising enzyme replacement therapy. As almost two-thirds of the enzyme sequence is predicted to be intrinsically disordered, the recombinant product is either subjected to a massive proteolytic attack by host-encoded proteases or tends to form aggregates. Therefore, the use of an unconventional expression system can constitute a valid alternative to solve these issues.

Results: Using a multiparametric approach we managed to optimize the transcription of the *CDKL5* gene and the synthesis of the recombinant protein in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 applying a bicistronic expression strategy, whose generalization for recombinant expression in the cold has been here confirmed with the use of a fluorescent reporter. The recombinant protein largely accumulated as a full-length product in the soluble cell lysate. We also demonstrated for the first time that full-length CDKL5 produced in Antarctic bacteria is catalytically active by using two independent assays, making feasible its recovery in native conditions from bacterial lysates as an active product, a result unmet in other bacteria so far. Finally, the setup of an in cellulo kinase assay allowed us to measure the impact of several CDD missense mutations on the kinase activity, providing new information towards a better understanding of CDD pathophysiology.

Conclusions: Collectively, our data indicate that *P. haloplanktis* TAC125 can be a valuable platform for both the preparation of soluble active human CDKL5 and the study of structural–functional relationships in wild type and mutant CDKL5 forms. Furthermore, this paper further confirms the more general potentialities of exploitation of Antarctic bacteria to produce “intractable” proteins, especially those containing large intrinsically disordered regions.

Keyword: *Pseudoalteromonas haloplanktis* TAC125, Antarctic bacterium, Psychrophilic gene expression system, Intrinsically disordered protein (IDP), Bicistronic design, In cellulo kinase assay, Tricistronic design, Recombinant protein aggregation, Recombinant protein condensation

*Correspondence: tutino@unina.it

¹ Department of Chemical Sciences, “Federico II” University of Naples, Complesso Universitario Monte S. Angelo-Via Cintia, 80126 Naples, Italy
Full list of author information is available at the end of the article

Background

Cyclin-dependent kinase-like 5 (CDKL5) is a serine/threonine protein kinase involved in the development of the human brain. Dozens of mutations of the *CDKL5* gene are causative of CDKL5 Deficiency Disorder (CDD);



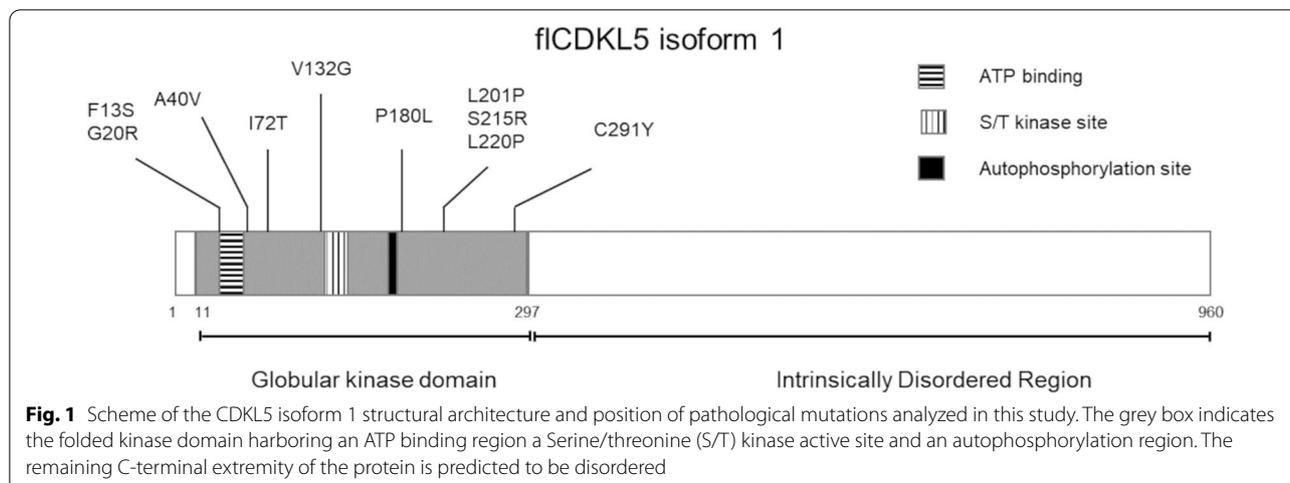
© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

OMIM 300203; 300672) [1, 2], a severe condition that is manifested with intellectual disability, autistic behavior, motor and visual impairments, infantile-onset refractory epilepsy, and many other symptoms [3]. Although no cure for CDD exists today, some studies proved that the restoration of CDKL5 activity through either protein or genetic intervention can revert CDD symptoms in mice and human models [4–6]. To clearly understand the applicability of these measures, however, a deeper knowledge of CDKL5 biology, its regulation, and the genotype-phenotype relationship of each CDD mutation should be acquired. For instance, we do not know whether some CDD mutations are dominant negative. If so, either enzyme replacement therapy (ERT) or gene addition may not be equally beneficial to all CDD patients. Another important question that should be addressed involves the levels of kinase activity that should be reached not to have a detrimental effect on CDD patients. Recently, the hyperphosphorylation of CDKL5 T169 has been demonstrated to induce the kinase hyperactivation which is a stress signal that triggers cell death during acute kidney injury [7]. This outcome would suggest that the presence and the abundance of post-translational modifications (PTMs) in CDKL5 preparations could significantly alter the therapeutic potential of a CDKL5-based ERT.

These observations indicate that multiple complementary tools to purify and study CDKL5 should be developed. Bacterial expression systems are desirable both for preparative and basic research purposes. First, the PTMs of the recombinant protein expressed in a bacterium would be limited to the phosphorylations dependent on the kinase autocatalytic activity, and a comparison with CDKL5 isolated from eukaryotic sources would make it possible to understand the role of other PTMs and their impact in an ERT. For similar reasons, a prokaryotic platform can be a valid tool to support the research in finding CDKL5 interactors, substrates, and upstream regulators. Different groups have already identified some CDKL5 substrates both in the cytoplasm [8, 9] and in the nucleus [10] using human cell cultures. Furthermore, a yeast two-hybrid screening has provided a list of proteins potentially interacting with the C-terminal extremity of CDKL5, although just two have been disclosed [11, 12]. We think that a reliable bacterial system can be useful to speed up the validation process of potential interactors and/or other putative substrates present in such lists. This aspect is crucial in the case of protein kinases since they generally can activate phosphorylation cascades and distinguishing each phosphorylation step from another can be troublesome in a eukaryotic context. In this regard, bacteria have often been useful in the study of eukaryotic kinases, from Cobb's pioneering work about the reconstruction of a MAP kinase cascade in

Escherichia coli [13] to more recent examples involving the characterization of the activation process of ciliary kinases through the comparison of protein preparations from different recombinant sources [14, 15]. The recombinant production of full-length CDKL5 (fCDKL5) is a challenging task, though. Its gene encodes a transcript that is subjected to alternative splicing leading to the production of five isoforms whose functional differences, if present, have not been explored [16]. However, the most abundant CDKL5 variant in the brain is isoform 1, followed by isoforms 2 and 3, and to a very little extent by isoform 4. Whereas isoform 5—which was the most studied one in early literature—is the bigger variant, but is exclusively expressed in the testis [16]. From a compositional point of view, all five isoforms are similar to each other since their cognate transcripts just differ for a few exons. For this reason, from now on we will generally refer to isoform 1 with the name fCDKL5. As recently highlighted by an *in silico* analysis and schematized in Fig. 1, fCDKL5 is predicted to be mainly a disordered protein, except for its N-terminal catalytic domain [17]. In support of this hypothesis, fCDKL5 has been recently demonstrated to interact with poly(ADP-ribose) with its C-terminal extremity [10], a typical feature of proteins possessing intrinsically disordered regions (IDRs) [18]. Proteins harboring extended IDRs are often problematic to be recombinantly expressed in and purified from conventional biotechnological platforms due to their propensity to either be degraded or to condensate [19]. fCDKL5 seems to fall in this category of difficult-to-express proteins. Even if its kinase catalytic domain has been successfully expressed in insect cells, yeast [20] and even *E. coli* [21], reports of recombinant production of fCDKL5 are quite scarce. Worth mentioning is the co-expression assay of fCDKL5 with two different substrates that Muñoz and co-workers established in HEK293 cells to evaluate the impact of CDD mutations [8]. Nevertheless, no examples of fCDKL5 satisfying expression in bacterial cells are available. The protein is insoluble in *E. coli* and denaturation and refolding procedures failed to achieve the active protein [22]. Regardless of this significant drawback, Katayama and Inazu managed to develop an ingenious assay to observe the fCDKL5 autophosphorylation activity on *E. coli* insoluble fractions, although such an approach would make it difficult to study the interaction of the kinase with other proteins [23]. As a result, a bacterial system that allows the soluble accumulation of fCDKL5 is desirable to allow both the validation of putative interactors with a co-expression assay and to obtain the active protein in native conditions after the cellular lysis.

To pursue this objective, we here present the genetic tools we have developed to produce fCDKL5 in the



model Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 (also named *P. translucida* TAC125). This prokaryote has already proven to be a useful alternative to conventional bacteria for the synthesis of difficult-to-express eukaryotic proteins [24–26]. The production of fICDKL5 isoform 5 has been tested in this bacterium but the overall yield was extremely low and the cultivation temperature of 0 °C was needed to preserve the protein during its synthesis [27]. However, very recently a model developed to study the *P. haloplanktis* TAC125 metabolism [28] has been specifically applied to optimize the production of fICDKL5 isoform 1, which is the most abundant in the human brain [29]. Here, we report the design of two sets of plasmids: the first set has been iteratively optimized to produce fICDKL5 isoform 1 fused with multiple tags for its detection and isolation; the second set has been designed to allow the co-expression of fICDKL5 with any one of its either putative or confirmed protein substrates. The former plasmid will be used to produce fICDKL5 for future applicative purposes, while the second will be used for basic research, e.g. either to study the effect of CDD mutations on fICDKL5 activity on known substrates or validate fICDKL5 putative interactors discovered with independent assays. As proof of concept, we here demonstrate that, by using the first set of plasmids, fICDKL5 can be produced mainly in intact and soluble form in *P. haloplanktis* TAC125 and that, after enrichment in native conditions, the kinase is catalytically active. Furthermore, by using the second set of plasmids, we co-expressed the 10 fICDKL5 CDD mutants shown in Fig. 1 with the CDKL5 bonafide protein substrate EB2 [8, 9] and demonstrated that this bacterial tool can effectively measure the impact of each mutation on the enzymatic activity. Our future goal is to use these two platforms to extend the knowledge of CDKL5 biology

and to achieve protein purification for ERT. For this reason, all the constructs presented in this work possess an N-terminal TATκ peptide that has already been used to vehiculate fICDKL5 across the blood-brain barrier [4].

Results and discussion

A multiparametric strategy to optimize full-length CDKL5 expression in *Pseudoalteromonas haloplanktis* TAC125

To test if *P. haloplanktis* TAC125 can be used to both express full-length CDKL5 (fICDKL5) for preparative purposes and as an orthogonal system for studying fICDKL5 functions and disease-related alterations, the protein kinase was produced both alone and together with one of its bonafide substrates, EB2 [8, 9] so as to carry out an *in cellulo* kinase assay. To do so, we first optimized the production of fICDKL5 alone, and then we applied the acquired knowledge to set up the co-expression assay. As schematized in Fig. 2a, six major steps were faced to progressively improve the production of intact fICDKL5 alone.

First, to express fICDKL5 we tested *P. haloplanktis* TAC125 monocistronic low copy number plasmids that were already available [30, 31]. In this phase, we focused on the selection of the best inducible promoter (Step 1 and Fig. 2b, panel (I)) and the resolution of an unforeseen abortion of mRNA translation (Step 2).

In the second part of this work, fICDKL5 synthesis was tested in bicistronic plasmids (BCDs, BiCistronic Designs), exploiting the translational coupling of the gene of interest with an optimized upstream ORF to increase the rate of translation initiation (Fig. 2b, panel (II)) [32]. This approach was pursued to increase the mRNA translation efficiency and stability. In this phase, we screened different combinations of upstream ORFs (Leader in Fig. 2b) and Shine

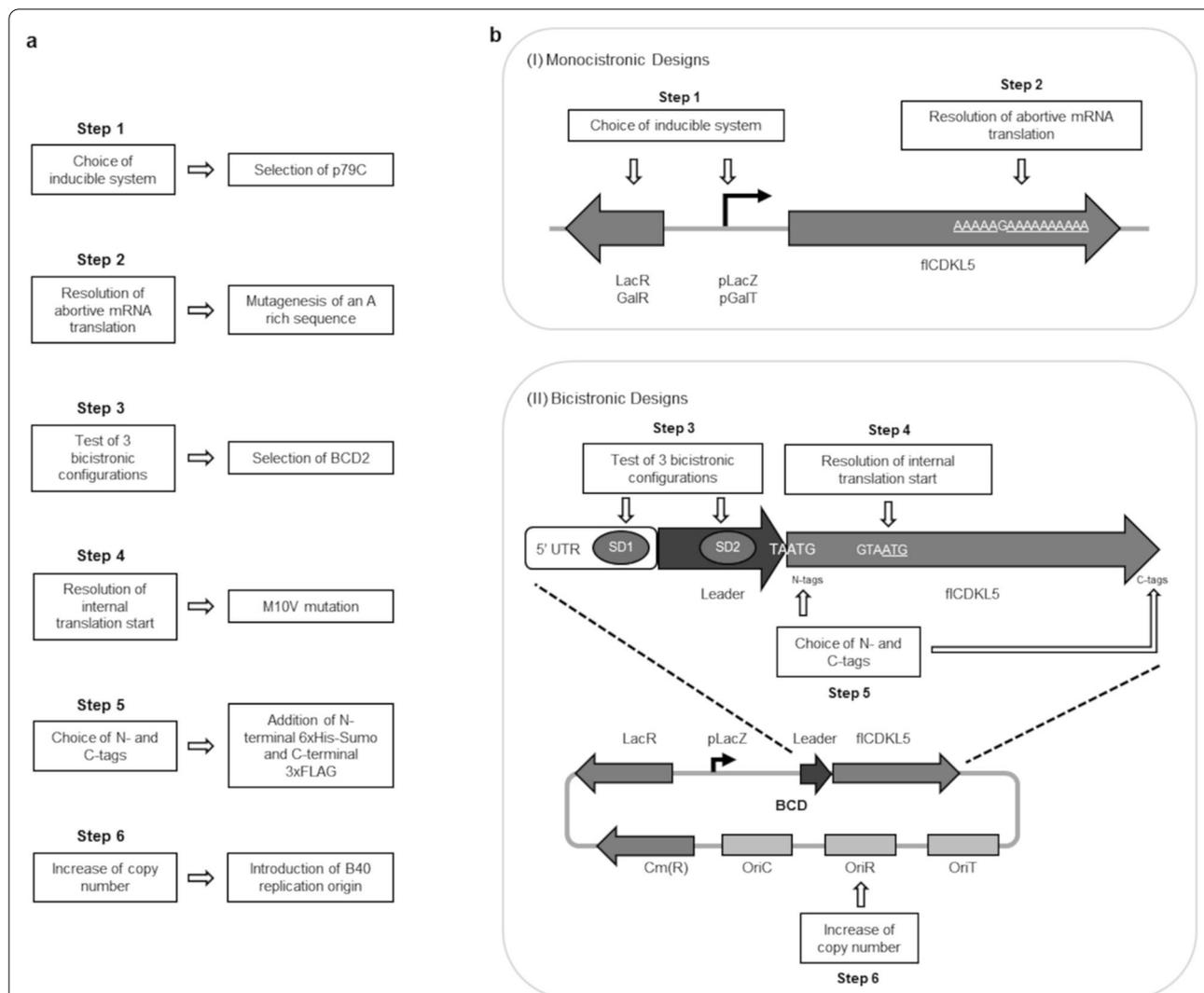


Fig. 2 Development of bicistronic plasmids for the production of fICDKL5. **a** Stepwise optimization of the production of fICDKL5. **b** Schemes of Monocistronic (I) and Bicistronic (II) plasmids for the expression of fICDKL5. (I) In the monocistronic configurations we tested two inducible systems (Step 1): p79C dependent on the LacR/pLacZ couple; pMAV dependent on the GalR/pGalT couple. Furthermore, a mutagenesis approach was used to overcome a translation abortion due to a polyA sequence (AAAAAGAAAAAAAAA, Step 2). (II) The Bicistronic Designs (BCDs) were developed by testing three different combinations of SD1, SD2 and Leader peptide encoding sequences (Step 3). Additionally, a methionine encoding sequence (GTAATG) was mutated to abolish an internal translation start (Step 4) and different permutations of N- and C-terminal tags were applied to increase the preservation and detectability of fICDKL5 extremities (Step 5). Finally, the protein yield was increased by replacing the parental plasmid replication origin with a new one named B40 (Step 6). LacR and pLacZ, the regulator and the promoter of the lacZ gene in *P. haloplanktis* TAE79, respectively; GalR and pGalT, the regulator and the promoter of the GalT gene in *P. haloplanktis* TAC125; 5' UTR, 5' untranslated region; SD1 and SD2, Shine Dalgarno 1 and 2, respectively; Leader, upstream sequence encoding a short peptide coupled to fICDKL5 translation through the overlap of a stop and a start codons (TAATG); Cm(R), chloramphenicol resistance marker; OriC, *E. coli* pMB1 replication origin; OriR, Antarctic replication origin; OriT, origin for plasmid conjugative transfer [33]. The underlined sequences were targeted for nucleotide substitution

Dalgarno sequences upstream of the Leader sequence (SD1) and the fICDKL5 encoding sequence (SD2, Step 3). We also overcame an internal translation start (Step 4) and optimized the disposition of N- and C-terminal tags to preserve fICDKL5 integrity (Step 5). Finally, the overall yield of fICDKL5 was significantly increased by the replacement of the replication origin of the plasmid

with a new one that guaranteed an increased copy number (Step 6).

fICDKL5 expression with monocistronic plasmids

To define whether fICDKL5 can be produced and is detectable in recombinant Antarctic bacteria, we started this work with a conventional monocistronic asset.

The fCDKL5 gene was designed to encode the human CDKL5 isoform 1—also called 107 variant [16]—with multiple tags. It possessed an N-terminal TATκ peptide that can be exploited for intracellular delivery [4], and tandem 3xFLAG and 6xHis C-terminal tags. Codon composition was automatically optimized for *P. haloplanktis* TAC125 by using the OPTIMIZER tool with the guided random method [34]. This CDKL5 variant was named 107 (B) and its expression was attempted in *P. haloplanktis* TAC125 cells using either a D-galactose inducible (pMAV) [30] or an IPTG inducible plasmid (p79C) [31], harboring a weak and strong regulatable promoters, respectively [31]. Regardless of their different expression mechanisms, the two plasmids share a common backbone and, most notably, harbor the same replication origin (OriR) derived from the *P. haloplanktis* TAC125 endogenous plasmid pMtBL [33]. Since pMtBL was demonstrated to be stably inherited as a single copy plasmid in Antarctic cells thanks to its partition sequences [35], we checked for the plasmid copy number (PCN) and stability of the shuttle recombinant vector p79C which is devoid of such regions. As shown in Additional file 1: Fig. S1a, the recombinant plasmid was stably kept with an average of 2–3 PCN during a 40-hour culture at 15 °C when the antibiotic selection was applied. Furthermore, IPTG induction did not seem to cause any instability (Additional file 1: Fig. S1b). Once the plasmid stability had been ascertained, 107 (B) production in the Antarctic bacterium was analyzed. After recombinant induction, 107 (B) production was not noticeable in *P. haloplanktis* TAC125 using SDS-PAGE on total lysates derived from either pMAV- or p79C-mediated expression (3a, left panel).

However, positive signals could be detected by anti-CDKL5 Western blot in both strains, proving that the target protein was synthesized, although at low levels (Fig. 3a, right panel). Furthermore, even if a signal compatible with fCDKL5 was visible in the Western blot analysis, a similarly intense band with a lower molecular weight was produced both by pMAV and p79C plasmids. Given that the used antibody targets the first half of the protein, this truncated product probably lacks the C-terminal extremity. To identify the molecular basis of this truncation, we analyzed the nucleotide sequence of the automatically codon-optimized 107 (B) gene and we noticed the occurrence of an A-rich region at about three-quarters (Fig. 2b, panel (I)). This polyA stretch corresponds to codons for six consecutive lysines, given that AAA is the most common triplet for this amino acid in *P. haloplanktis* TAC125. Since multiple authors have reported that A repetitions in mRNAs induce ribosome stalling, sliding, and accidental frameshift during protein synthesis [36–38], we modified this sequence by

replacing five AAA codons with synonymous AAG. This new fCDKL5 variant was named 107 (G) and its expression profile was compared with the one of 107 (B). As shown in Fig. 3b, the main truncation product visible in the pMAV-107 (B) bearing strain was not produced by cells hosting either pMAV-107 (G) or p79C-107 (G). Furthermore, the IPTG inducible plasmid guaranteed a higher accumulation of the target protein over time than the D-galactose-dependent vector, a result that is in agreement with the different strengths of the two promoters [31]. For these reasons, we focused on the improvement of the p79C plasmid and the 107 (G) construct for further studies.

Bicistronic cassettes and site-specific mutagenesis to overcome an internal translational start

In the second phase of this study, we wanted to address two questions: can fCDKL5 production be increased by translational regulation? Is there an ideal disposition and combination of tags to preserve fCDKL5 extremities? As fCDKL5 is a high molecular weight protein (107 kDa) with extensive disordered regions [17], we cannot exclude possible either N- or C-terminal truncations just by electrophoretic migration. This is a non-trivial issue in the case of the here presented fCDKL5 variants since they are characterized by two flexible ends, the N-terminal TATκ and the C-terminal intrinsically disordered region (IDRs). IDRs can be a target of proteolysis and impact the half-life of proteins, but their fusion with tags can change their overall stability [39].

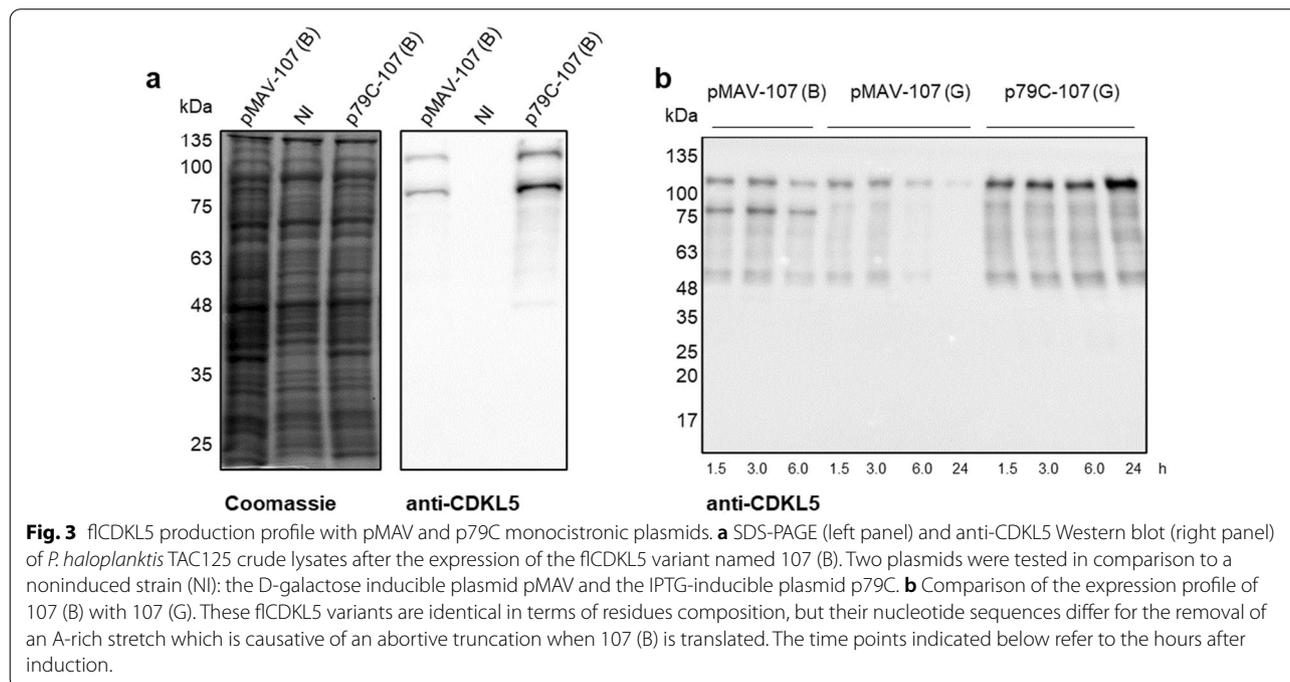
Hence, the second phase of our study aimed to examine and possibly improve these aspects contextually by using a technique that could both modulate translational efficiency and test different combinations of tags.

The tool we chose for this purpose is the bicistronic cassette (BCD) that introduces short optimized coding sequences upstream of the heterologous gene of interest to optimize the start of translation [32]. This approach has been used several times in different recombinant bacteria in the past and has demonstrated that the translation of a heterologous ORF can be modulated by the translational coupling with an upstream optimized short ORF. The application of this technology to finetune a psychrophilic translation system to produce a eukaryotic protein normally synthesized at 37 °C is reasonable given that the mRNA translation is differently regulated in bacteria and mammals [40] and that the temperature is likely to play a pivotal role in the kinetics of protein synthesis and folding. Given that protein synthesis is mainly limited by translation initiation and early elongation [41], we reasoned that the BCD strategy could serve as an insulator to optimize fCDKL5 synthesis so as to avoid too many

Table 1 Characteristics of BCD constructs for fCDKL5 expression

Name	Leader ^a	SD1 ^b	SD2 ^c	N-tags ^d	C-tags ^d	Mutations ^d
p79C-107 (G)	/	/	CAACAGGAA	TATk	3xFLAG-6xHis	/
pBCD1-107 (G)	LacZ	CAACAGGAA	CAACAGGAA	TATk	3xFLAG-6xHis	/
pBCD2-107 (G)	LacZ	CAACAGGAA	AAGGAGGTC	TATk	3xFLAG-6xHis	/
pBCD1-107 (K)	LacZ	CAACAGGAA	CAACAGGAA	6xHis-TATk	3xFLAG	/
pBCD2-107 (K)	LacZ	CAACAGGAA	AAGGAGGTC	6xHis-TATk	3xFLAG	/
pBCD3-107 (K)	TrpA	AAGGAGGTC	AAGGAGGTC	6xHis-TATk	3xFLAG	/
pBCD2-107 (G) M10V	LacZ	CAACAGGAA	AAGGAGGTC	TATk	3xFLAG-6xHis	M43V
pBCD2-107 (G) STOP	LacZ	CAACAGGAA	AAGGAGGTC	TATk	3xFLAG-6xHis	M11, G2STOP
pBCD2-107 (K) M10V	LacZ	CAACAGGAA	AAGGAGGTC	6xHis-TATk	3xFLAG	M61V
pBCD2-107 (L) M10V	LacZ	CAACAGGAA	AAGGAGGTC	6xHis-Sumo-TATk	3xFLAG	M159V

^a Leader peptide encoded upstream of fCDKL5 includes the first 19 residues of the indicated proteins. ^b Shine Dalgarno sequence upstream of the Leader peptide encoding sequence. ^c Shine Dalgarno sequence upstream of the fCDKL5 gene. Shine Dalgarno sequences which are entirely complementary to 16S rRNA are underlined. ^d N- and C-terminal tags and mutations are referred to fCDKL5 protein. M43V, M61V and M159V are the same mutation (M10V in human CDKL5, O76039, Uniprot), but the coordinates differ because of the presence of different N-terminal tags



interventions on the gene composition. Consequently, we implemented this technology in *P. haloplanktis* TAC125 using the work by Mutalik and coworkers as a model both for the design of the plasmids—as indicated in the [Methods](#) section—and for the nomenclature of the bicistronic designs [32]. Furthermore, the adopted cloning strategy allowed for the easy permutation of tags at the two extremities of fCDKL5 (see the [Methods](#) section). For these reasons BCD vectors were the most suitable tool to simultaneously analyze the effect

of tags and translational control on fCDKL5 production and preservation.

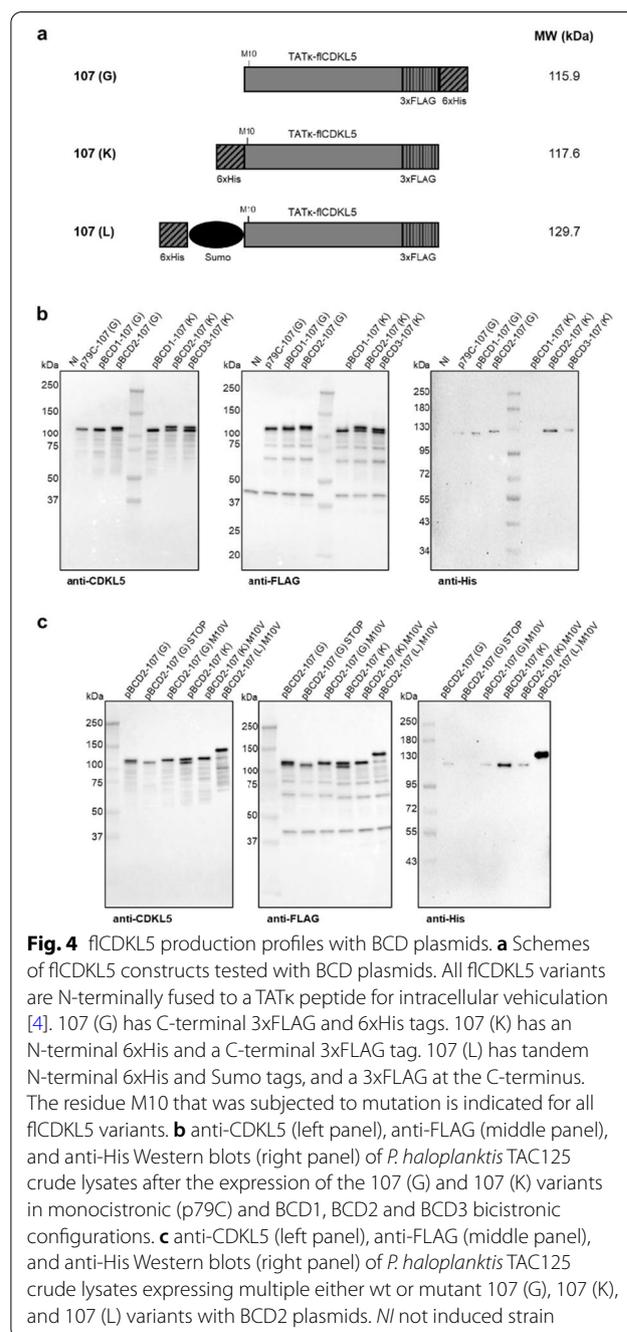
The main components of a BCD cassette are: a 5' untranslated region (5' UTR) harboring a first SD sequence (SD1); a short open reading frame (ORF) that encodes a peptide optimized for initial translation (Leader sequence) and embeds a second SD sequence (SD2); the gene of interest whose translation driven by SD2 is coupled with the translation of the Leader sequence thanks to the overlap of a Stop and Start codons

(Fig. 2b, panel (II)). The impact of variable BCD configurations on protein expression was assessed both using plasmids for the syntheses of a fluorescent reporter, pGFP [31], and a parallel set for the production of fCDKL5 (Additional file 1: Table S1 and Table 1, respectively).

In the case of pGFP, we initially tested the first 19 residues of *P. haloplanktis* TAE79 LacZ as a Leader peptide, given the high expression levels of this psychrophilic β -galactosidase in *P. haloplanktis* TAC125 [31]. The two BCD vectors that were generated differed only for the SD2 sequence. In pBCD1, SD2 was the same as SD1, that is the natural SD sequence upstream of the lacZ ORF. In pBCD2, SD2 was optimized to embed a region that is perfectly complementary to the 3'-OH end of *P. haloplanktis* TAC125 16S rRNA (Additional file 1: Table S1). When the expression levels driven by these two bicistronic vectors were compared with a monocistronic cassette (p79C-pGFP), we could detect pGFP accumulation in all the cases, but pBCD2-pGFP guaranteed a significantly higher fluorescence suggesting that the optimized SD2 sequence was pivotal to increase the pGFP translation rate (Additional file 1: Fig. S2a). We also tested a different SD1—Leader combination to produce pGFP. We aimed to replace the theoretically weak SD of lacZ in position SD1 with an SD that is perfectly complementary to 16S rRNA by using an ORF that naturally possesses an optimized SD in *P. haloplanktis* TAC125 genome. The ORF we selected is in the *trpA* gene (PSHA_RS06350) of the tryptophan operon. The new plasmid was named pBCD3-H6-pGFP (Additional file 1: Table S1) and the levels of pGFP that could be accumulated were compared to pBCD1 and pBCD2 configurations. As visible in Additional file 1: Fig. S2a, although pBCD3 performed better than pBCD1, it was worse than pBCD2. This indicates that both SD1, SD2, and the Leader sequence play a role in the accumulation of the protein encoded by the second cistron. However, while the effect of SD2 seems predictable, more subtle variables may influence the roles of SD1 and the Leader cistron, as expected. The role of BCD constructs is to optimize the expression of the second cistron independently of its codon composition thanks to the insulating effect of the first cistron. On the other hand, the translation starting from the first cistron must be optimized with thorough screening, due to the lack of any insulator [32]. The addition of N-terminal 6xHis-TATk tags did not significantly affect pGFP expression in the BCD2 configuration, indicating that multiple N-terminal tags can be added with no major detrimental effects on protein accumulation (Additional file 1: Fig. S2b). Overall, the data acquired with the fluorescent reporter indicate that BCD2 is the best bicistronic configuration we developed for recombinant expression in *P. haloplanktis* TAC125 and that the addition of N-terminal

tags does not seem to perturb the pGFP production levels when this asset is used.

In the case of fCDKL5, the achievement of protein accumulation was less straightforward than pGFP. We tested the three bicistronic configurations (BCD1, BCD2, BCD3) that were explored for pGFP expression (Table 1). Furthermore, fCDKL5 was originally produced with two different dispositions of tags. The most relevant traits are that variant 107 (G) has C-terminal 3xFLAG and 6xHis,



while 107 (K) harbors an N-terminal 6xHis and a C-terminal 3xFLAG tags (Table 1 and Fig. 4a). Anti-CDKL5 and anti-FLAG Western blots showed that pBCD2 and pBCD3 configurations gave rise to the production of a high molecular weight band that is not generated by either the monocistronic vector or pBCD1 (Fig. 4b, left and middle panels).

Apparently, p79C and pBCD1 plasmids only generated a truncated version of flCDKL5, while pBCD2 and pBCD3 allowed the synthesis of both the truncated and the putatively intact forms of flCDKL5 both as 107 (G) and 107 (K) variants. Given that such two bands are detected by the anti-FLAG antibody and that both 107 (G) and 107 (K) have a C-terminal 3xFLAG tag (Fig. 4a), the lower molecular weight band is probably N-terminally truncated. This hypothesis is also suggested by the anti-His Western blot that detected positive signals for all the strains but pBCD1-107 (K, Fig. 4b, right panel). 107 (G) was always reactive because the 6xHis tag is C-terminal in this flCDKL5 variant. On the other hand, the 107 (K) variant has an N-terminal 6xHis tag and the lack of any anti-His detection in the pBCD1-107 (K) bearing strain is indicative of the fact that in this condition flCDKL5 is exclusively synthesized in an N-terminally truncated form (detectable by anti-FLAG and anti-CDKL5 antibodies). Since the three BCD configurations just differ for elements regulating the start of flCDKL5 translation, we reasoned that such variable fragmentation patterns might be due to a translational bias. We formulated two hypotheses: (1) flCDKL5 was always N-terminally proteolyzed in *P. haloplanktis* TAC125, but BCD2 and BCD3 allowed for the accumulation of the intact form of the protein thanks to higher production levels (see Additional file 1: Fig. S2 for a comparison with pGFP accumulation); (2) the flCDKL5 encoding gene harbored an internal translation start that overrode the first translation start unless the latter is optimized like in the BCD2 and BCD3 configurations. A new analysis of the flCDKL5 encoding gene pointed toward the second hypothesis since it highlighted the possible existence of a cryptic internal translation initiation at the level of M10 according to the coordinates of the human CDKL5 isoform 1 (O76039, Uniprot). Although a strong Shine Dalgarno sequence could not be found upstream of the ATG encoding M10, a truncation at this level is compatible with the small electrophoretic difference between intact flCDKL5 and truncated flCDKL5 forms in our Western blot analysis (Fig. 4b).

To test this hypothesis of internal translation initiation, we planned multiple experiments. First, we designed a new flCDKL5 variant possessing a bulky tag between the N-terminal 6xHis and the CDKL5 wt residues to make the truncated product more easily distinguishable from

intact CDKL5. This construct was named 107 (L) and harbors N-terminal 6xHis-Sumo tags and a C-terminal 3xFLAG (Fig. 4a). This modification served also to test our first hypothesis (i.e. whether flCDKL5 was truncated due to degradation) since the fusion of flexible extremities with globular domains can change their propensity to degradation, as already mentioned [39]. Then, we introduced site-directed mutations in 107 (G), 107 (K), and 107 (L) flCDKL5 variants to reveal the putative role of M10 as an internal translation start. In particular, we abolished the translation initiation from the first residue of 107 (G) by eliminating the first ATG and introducing a STOP codon in the second position (pBCD2-107 (G) STOP in Table 1). As revealed by anti-CDKL5 and anti-FLAG Western blots (Fig. 4c, left and middle panels, respectively), the truncated form of flCDKL5 was still produced by pBCD2-107 (G) STOP bearing cells, while the upper band was absent, as expected. This outcome suggests that an internal translation start occurred. Then, we replaced M10 of human CDKL5 with a valine to have an insight into its role in flCDKL5 fragmentation. Although such residue has different coordinates depending on the protein variant (43, 61, and 159 for 107 (G), 107 (K) and 107 (L), respectively), we will always name such mutation as M10V in reference to human CDKL5 wild type coordinates. As visible in Fig. 4c, such mutations abolished the synthesis of the truncated fragment in 107 (G) M10V and 107 (K) M10V producing strains. In the case of 107 (L) M10V, a lower molecular weight fragment at the level of the untagged protein was visible in anti-CDKL5 and anti-FLAG Western blots, but the intensity was low and comparable with all the other truncated fragments in the same lane. This suggests that such a product is unlikely to be caused by the internal translation initiation. An anti-FLAG Western blot specifically comparing the production profiles of pBCD2-107 (L) and pBCD2-107 (L) M10V strains proved this point (Additional file 1: Fig. S3). This result demonstrates that the addition of the globular Sumo domain at the N-terminal extremity of flCDKL5 did not resolve its fragmentation and that such fragmentation is unlikely to be due to proteolysis. Nevertheless, the 107 (L) M10V variant guaranteed better preservation of the N-terminal 6xHis tag than the 107 (G) and 107 (K) proteins (Fig. 4c, right panel). This outcome may be since in the latter variants the 6xHis tag is directly fused to extremely flexible regions (i.e. the C-terminal intrinsically disordered region in 107 (G) and the flexible N-terminal TATκ in 107(K), Fig. 4a).

As the last proof of an internal start occurring in flCDKL5 mRNA, we wanted to obtain the N-terminal amino acid sequence of the fragmented protein by using Edman sequencing. Since flCDKL5 is a difficult-to-purify protein, we tried to isolate such a protein fragment from

a preparation of the CDKL5 catalytic domain only. We decided to produce this shortened version of CDKL5 in *Escherichia coli* rather than *P. haloplanktis* TAC125 given that previous reports demonstrate that this cell factory can be effectively used to isolate the CDKL5 catalytic domain [21]. To do so, we generated a short ORF encoding CDKL5(1-352) with a C-terminal His tag, performing a PCR on the 107 (L) encoding gene. Such construct was then expressed in *E. coli* BL21(DE3) cells using a pET system and the recombinant protein was isolated with an IMAC. As expected, the truncated fragment was co-purified and its N-terminal sequence was MNXF according to Edman degradation, where X is an undetermined amino acid. This indetermination was due to the little quantity of the protein fragment that could be recovered, but such a sequence is only compatible with a start in the M10 position in the CDKL5 protein, as expected. Hence, our results indicate that an internal translation start occurs when our engineered CDKL5 construct is expressed in bacteria which is typical of some eukaryotic cDNAs expressed in prokaryotes [42].

Based on this body of work, we selected pBCD2-107 (L) M10V as the best construct we tested for flCDKL5 expression. pBCD2 was preferred over pBCD1 because the latter only produced a truncated form of flCDKL5 (Fig. 4b). Moreover, pBCD2 was better than pBCD3 because the former guaranteed a higher accumulation of the full-length protein, as visible by anti-His Western blot (Fig. 4a, right panel) and confirmed by the pGFP results (Additional file 1: Fig. S2a). Furthermore, the pBCD2 configuration guaranteed increased stability of the flCDKL5 mRNA in comparison with the monocistronic p79C plasmid (2.73 ± 0.1 fold change), as demonstrated by qRT-PCR measurements that are in agreement with previous studies from other authors [43]. The 107 (L) variant was chosen because of the Sumo protective role on the N-terminal 6xHis tag (Fig. 4c, right panel). Finally, the M10V mutation was kept because it provided a solution to avoid the internal translational start. The presence of two different affinity tags at the two extremities of the protein constitutes a further advantage in sight of the future purification of flCDKL5. Double tagging has indeed been successfully used for the purification of both multidomain and intrinsically disordered proteins that are prone to fragmentation [19].

Overall, the multiparametric approach used to optimize flCDKL5 in *P. haloplanktis* TAC125 proved that the bicistronic strategy was pivotal to increase the yield of intact flCDKL5 from virtually absent to an amount detectable by Western blot. Nevertheless, BCDs were not sufficient to overcome an internal translational start, suggesting that the translation from the first and the internal ATGs are not in competition. This datum is corroborated

by the fact that the abolition of translation from either the first ATG (pBCD2-107 (G) STOP in Fig. 4b) or the internal ATG (M10V mutants in Fig 4b) did not increase the production levels of the truncated and intact proteins, respectively, which is in agreement with a previous study [44]. Our hypothesis of testing multiple N- and C-terminal fusions with affinity tags proved to be valid because it allowed for better discrimination of flCDKL5 fragments during protein synthesis and proved that a bulky domain like Sumo can increase the preservation of the N-terminal 6xHis tag. This approach should be extended to the flCDKL5 C-terminal IDR in the pursuit of increased stabilization. Even if the main truncated fragment of flCDKL5 during recombinant expression in *P. haloplanktis* TAC125 was due to a translational issue, our Western blots highlight the existence of other numerous low molecular weight fragments though considerably less abundant. This pattern may be ascribable to the fragility of the flCDKL5 C-terminal tail and the fusion with a terminal globular domain could lessen its impact.

Application of high-copy number plasmids for the expression of flCDKL5 and evaluation of the kinase activity

Very recently a library of replication origins with variable PCNs has been established for plasmid maintenance in *P. haloplanktis* TAC125 (Calvanese M. et al., manuscript in preparation). After we had optimized the flCDKL5 expression cassette, we moved it into a high copy number backbone named pB40 with an average PCN of ~ 100 . The comparison of the expression profiles of the Antarctic cells bearing the original low copy number monocistronic plasmid p79C-107 (G) with the ones hosting the high copy number bicistronic plasmid pB40-BCD2-107 (L) M10V revealed a strong difference.

While the protein achieved from the original construct was barely detectable by anti-CDKL5 Western blot, the optimized plasmid guaranteed a considerable higher accumulation of the target protein that was visible both by Coomassie staining after SDS-PAGE and by Western blot (Fig. 5a, left and right panels, respectively). Furthermore, when chemical-enzymatic lysis and fractionation of crude extracts were carried out in native conditions (see Materials and Methods for the details), most flCDKL5 appeared to be soluble (Fig. 5b). After a chromatographic enrichment with an anti-Flag resin, the recombinant kinase was tested for its catalytic activity by using an in vitro kinase assay with EB2 as a protein substrate. Briefly, 100 nM enzyme was incubated with 200 nM substrate in the presence of MgATP at 30 °C for 30 min. After protein inactivation with Laemmli buffer at 70 °C, the CDKL5-mediated phosphorylation of EB2 S222 was detected via Western blot with a specific antibody

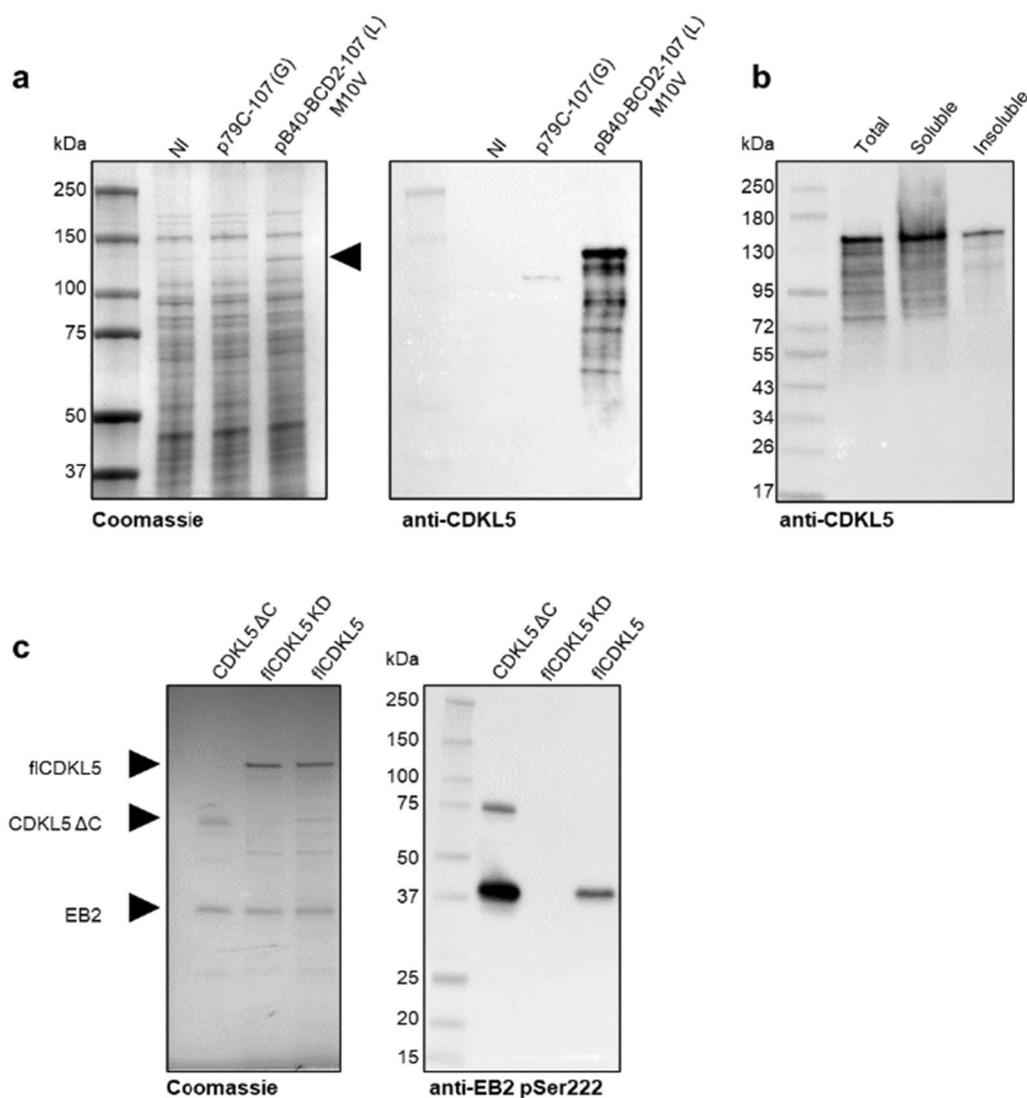


Fig. 5 fICDKL5 production with a high copy number bicistronic plasmid. **a** Comparison of fICDKL5 expression levels in *P. haloplanktis* TAC125 recombinant strains harboring either a low copy number monoistronic plasmid (p79C-107 (G)) or an optimized high copy number bicistronic plasmid (pB40-BCD2-107 (L) M10V) using Coomassie staining after SDS-PAGE (left panel) and anti-CDKL5 Western blot (right panel). The arrow indicates fICDKL5 produced with the pB40-BCD2-107 (L) M10V plasmid. NI, not induced strain. **b** Solubility analysis of 107 (L) M10V protein after cellular lysis in native conditions. Total cellular extracts and soluble and insoluble fractions achieved after centrifugation were resolved via SDS-PAGE and the target protein was detected with an anti-CDKL5 Western blot. **c** Catalytic activity of enriched 107 (L) M10V on pure EB2. After chromatographic enrichment, either 107 (L) M10V (fICDKL5), or a catalytically inactive variant (fICDKL5 KD) was incubated with EB2 and MgATP and then assayed for catalytic activity. A commercial preparation of GST-CDKL5(1–498) from insect cells (CDKL5 ΔC) was used as a positive control. The total amount of the full-length enzyme and the substrate were detected by Coomassie staining (left panel, highlighted by arrows), while phosphorylated EB2 was observed with anti-EB2 pSer222 antibody (right panel)

[9]. Remarkably, recombinant fICDKL5 enriched from *P. haloplanktis* TAC125 showed enzymatic activity on EB2, though it was less active than a commercial preparation of the CDKL5 catalytic domain from insect cells (CDKL5 ΔC, Fig. 5c). On the other hand, a catalytically inactive variant, fICDKL5 KD, did not phosphorylate EB2, as expected (Fig. 5c). The discrepancy between fICDKL5

and CDKL5 ΔC activities may be ascribable to more factors. First, the fICDKL5 C-terminal region is known to play an inhibitory effect on the enzymatic activity [7, 45, 46]. Second, the protein produced in the eukaryotic system may harbor higher phosphorylation levels that cannot be reached in a bacterial system. These two phenomena need to be dissected in the future by using

CDKL5 ΔC purified from bacteria and fCDKL5 obtained from a eukaryotic system as controls, which is currently unavailable from commercial sources.

Besides these limitations, this work demonstrated that an increased gene dosage was needed to have satisfying fCDKL5 accumulation in *P. haloplanktis* TAC125 at 15 °C, a prerequisite that allowed the recovery of enough protein to assess its activity *in vitro*. To the best of our knowledge, no other bacterial platform has guaranteed to obtain fCDKL5 in native conditions as an active enzyme so far.

Use of TCD plasmids to measure the impact of CDKL5 pathogenic mutations

As the last proof of concept study, we planned to co-express fCDKL5 with one of its substrates to demonstrate that this prokaryotic platform can be used to validate CDKL5 substrates and model CDD mutations. The plasmid configuration optimized with the BCD2 design was converted into a tricistronic plasmid for the co-expression of EB2 and fCDKL5. As schematized in Fig. 6a, the TCD plasmid triggered the production of a polycistronic mRNA encoding the LacZ leader peptide, EB2 and fCDKL5.

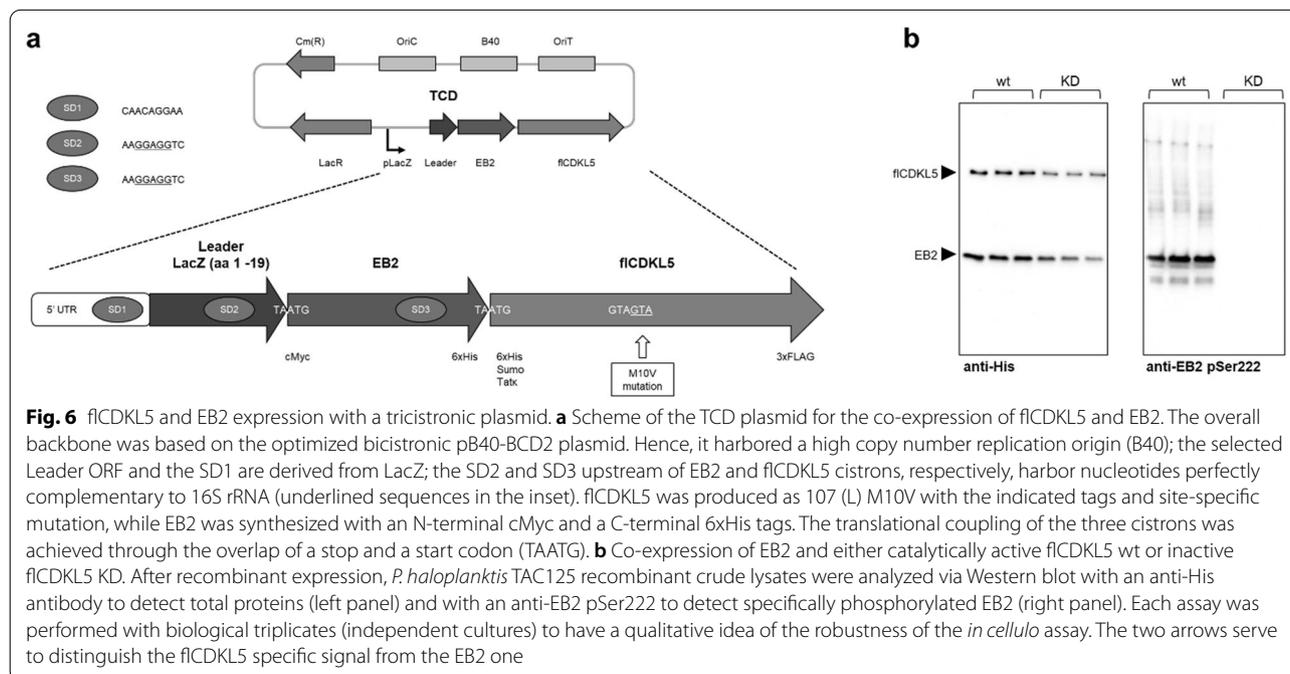
To test whether this system can be used to reliably measure fCDKL5 catalytic activity through an *in cellulo* kinase assay, EB2 was co-expressed either with fCDKL5 or fCDKL5 KD, a catalytically inactive variant. As visible in Fig. 6b (left panel), the synthesis of EB2 and fCDKL5 could be contextually detected with an anti-His antibody

and was consistent through biological triplicates. Furthermore, specific CDKL5-mediated phosphorylation of EB2 could be detected only when active fCDKL5 was produced. This result confirms that this assay can be used to undoubtedly measure CDKL5 specific activity.

Most of the patients affected by CDKL5 Deficiency Disorder (CDD) suffer from refractory epilepsy, hypotonia, intellectual and motor disabilities, and visual impairments [3]. However, the severity of these symptoms is variable, and clear genotype-phenotype correlations are scarce. For this reason, we wanted to test if our co-expression assay can be a valid tool to model CDD CDKL5 variants at the molecular level. To do so, 10 pathogenic fCDKL5 missense variants were co-expressed with EB2 using TCD plasmids, and EB2 phosphorylation levels were quantified. All fCDKL5 variants were hypoactive and the levels of EB2 cross-phosphorylation were variable, indicating a different impact of each mutation on fCDKL5 functionality (Fig. 7). This kind of study can be extended in the future to other known CDKL5 substrates to have a deeper understanding of CDD pathogenesis.

Conclusions

Our multivariate approach to optimize the production of fCDKL5 in *P. haloplanktis* TAC125 at 15 °C led to the accumulation of the target protein mainly as an intact, soluble, and active form, making feasible further developments for its purification and formulation. Furthermore, the co-expression assay with EB2 demonstrated two main points: (1) fCDKL5 is active



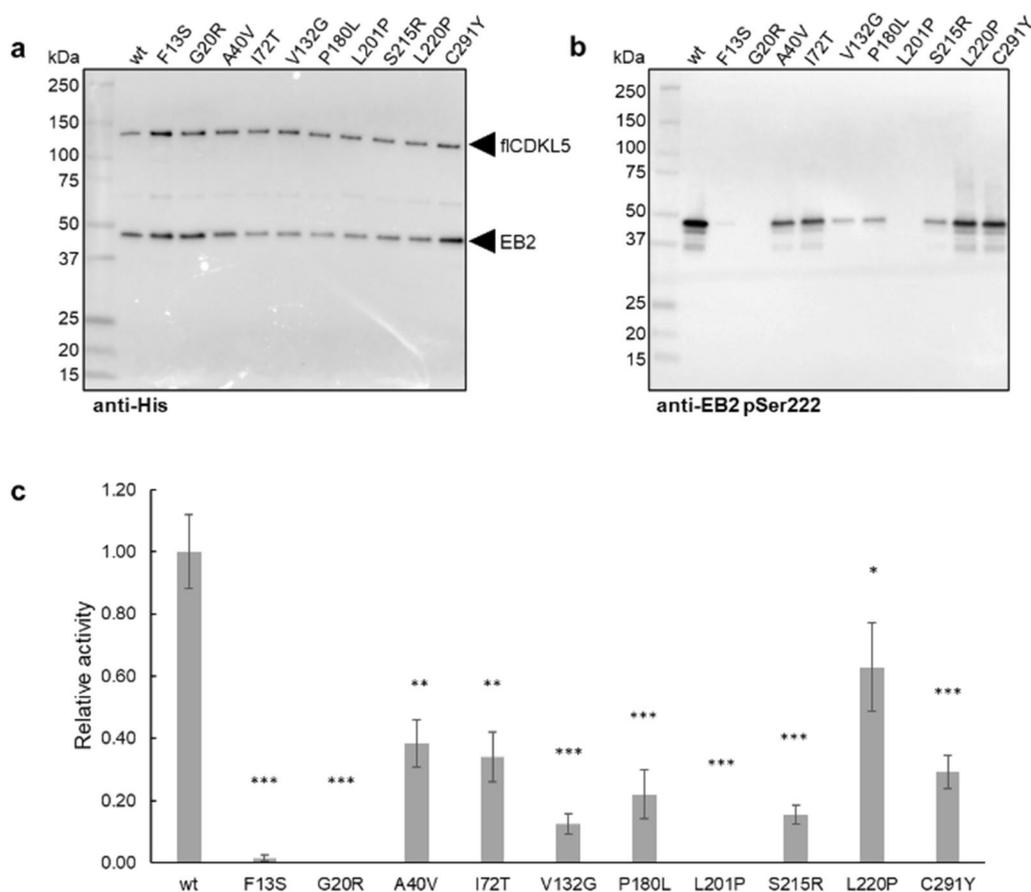


Fig. 7 Effect of pathogenic CDKL5 mutations on fCDKL5 activity. **a** fCDKL5 variants and EB2 were co-expressed in *P. haloplanktis* TAC125 using TCD plasmids. After cellular lysis, fCDKL5 mutants and EB2 were detected in each crude lysate using an anti-His antibody after Western blot, as indicated by the two arrows. **b** Phosphorylated EB2 was detected with an anti-EB2 pSer222 antibody. **c** EB2 phosphorylation levels were normalized for total EB2 and fCDKL5. The results are reported as the mean of technical triplicates and the error bars represent standard deviations. **a** and **b** show Western blots representative of one experiment. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Note: every CDKL5 construct (also the wt) harbors an M10V mutation besides the one indicated in the figure. Mutation coordinates are referred to the human CDKL5 isoform 1 (#O76039, Uniprot)

when properly produced in bacteria; and (2) CDD mis-sense variants can be modeled and discretized based on their hypoactivity. Such an approach will be useful in the future to support the reconstruction of CDKL5-mediated networks and to characterize the molecular effect of CDD mutations.

In addition, our results further support the more general potentialities of exploitation of the Antarctic bacterium to produce “intractable” proteins such as those endowed with a folded N-terminal domain and a long disordered tail, a common trait in the human kinome. Notably, all RCK kinases that are involved in cilia biology share such property, and our work demonstrate that *P. haloplanktis* TAC125 could be a suitable host to test their expression and purification for further characterization.

Methods

Bacterial strains and culture conditions

E. coli TOP10 was used for cloning purposes, while *E. coli* S17-1(λ pir) was employed in intergeneric conjugations as a donor strain for *P. haloplanktis* TAC125 transformations. *E. coli* BL21(DE3) was used for the recombinant production of the catalytic domain of human CDKL5 and for the synthesis of human EB2.

The *P. haloplanktis* TAC125 strain that was used for the expression of fCDKL5 alone is KrPI LacY⁺ which lacks its endogenous plasmid pMtBL and the *lon* gene, and constitutively expresses the *E. coli* lactose permease [31]. For the double production of EB2 and fCDKL5, the KrPI strain was used, instead [31]. The induction of recombinant expression was performed at 1 OD with either 10 mM D-galactose or 5 mM IPTG for pMAV and

p79C-derived plasmids, respectively. If not differently specified, the recombinant expression was protracted for 4 h.

The *E. coli* strains were grown in Lysogen broth (LB, 10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37 °C with 200 rpm agitation. The BL21(DE3) recombinant strains were induced in mid-exponential phase with 0.1 mM IPTG at 15 °C for 16 h. *P. haloplanktis* TAC125 was grown in a rich medium (16 g/L bacto-tryptone, 16 g/L yeast extract, 10 g/L NaCl) during the preinocula development, and in a synthetic medium named GG for recombinant expression [30]. In both cases, the bacterial cultures were carried out at 15 °C with 200 rpm agitation.

For the selection of *E. coli* recombinant strains, either 100 µg/mL ampicillin (for pMAV derived plasmids) or 34 µg/mL chloramphenicol (for p79C derived plasmids) or 50 µg/mL kanamycin (for pET40-b derivatives) was used, depending on the selection marker. In the case of recombinant *P. haloplanktis* TAC125, chloramphenicol was used at 12.5 µg/mL in solid media and 25 µg/mL in liquid media, while ampicillin was always used at 100 µg/mL concentration.

Plasmids construction

Three genes encoding different fCDKL5 variants were first cloned into two monocistronic plasmids. Then, to test fCDKL5 expression in bicistronic configurations with different N- and C-terminal tags, BCDs were developed. To do so, two fCDKL5 genes were cloned into Bicositronic Entry Clone plasmids (BECs) and then short DNA fragments harboring different Leader ORFs, Shine Dalgarno sequences and N-terminal tags were inserted between the pLacZ promoter and the gene of interest (GOI). The same procedure was followed for the preparation of the pGFP-based vectors. The list of the primers used in this work is reported in Additional file 1: Table S2.

Monocistronic designs

For the recombinant expression of fCDKL5 with monocistronic plasmids, pMAV [30], and p79C [31] vectors were used. The genes encoding fCDKL5 variants named 107 (B), 107 (G) and 107 (H) were synthesized by an external company and cloned into pMAV by using NdeI/EcoRI double digestion. The transfer of the genes of interest from pMAV into p79C was achieved with the use of NdeI and SacI allowing the isolation of the recombinant genes together with the transcriptional terminator [31]. The sequences of the three genes are reported in the Additional file 1. 107 (B) and 107 (G) have a different codon composition, but they encode the same protein with an N-terminal TATκ peptide and C-terminal

tandem 6xHis and 3xFLAG tags. 107 (H) differs from the other two variants for the lack of a His tag.

Entry clone designs

For the design of bicistronic plasmids (BCDs), Entry Clone plasmids (BECs) were prepared first. To do so a parental plasmid, p79BsC, was generated by cloning a synthetic fragment encompassing the LacR and pLacZ sequences of p79C into the pUCC backbone [47] using SphI/PstI double digestion (Additional file 1: Fig. 4, left plasmid). The aim of this step was the introduction of a BsaI site downstream of the pLacZ promoter.

The three BEC plasmids that were then generated shared the common feature of harboring two divergent BsaI sites upstream of the GOI, so as to allow the scarless cloning of DNA fragments between the pLacZ promoter and the GOI with the Golden Gate technology [48]. pBEC-pGFP was generated by isolating the pGFP gene from p79C-pGFP [31] via PCR using the *pGFP_PstIB-sal_fw* and *pGFP_KpnI_rv* primers, and cloning it into p79BsC with PstI/KpnI digestion (Additional file 1: Fig. S4, upper branch).

pBEC-107 (G) was developed by cloning the 107 (G) gene as two fragments, 5' 107 and 3' 107 (G). The former is a synthetic DNA harboring the first half of the fCDKL5 gene with left PstI and BsaI sites and a right HindIII site. 3' 107 (G)—the second half of the gene—was extracted from p79C-107 (G) using HindIII/SacI double digestion. The two fragments were cloned between PstI and SacI into p79BsC (Additional file 1: Fig. S4, middle branch).

pBEC-107 (H) was generated in the same way as pBEC-107 (G) with the only difference that 3' 107 (H) was used in place of 3' 107 (G). Such fragment was obtained from p79C-107 (H) using HindIII/SacI digestion and was cloned together with 5' 107 into p79BsC (Additional file 1: Fig. S4, bottom branch).

Bicistronic designs

To generate the various pBCD plasmids described in this work, different DNA fragments, named Strings A-F were cloned into the BEC plasmids. As depicted in Additional file 1: Fig. S5a, each String has two divergent BsaI sites at the extremities to allow Golden Gate cloning. Furthermore, every String harbors the 5'UTR of p79C (+1 in Additional file 1: Fig. S5a indicates the transcription start) and different combinations of SD1, SD2, Leader ORF and possibly N-terminal tags to be fused to the GOI. The various combinations of BEC plasmids and Strings generated the pBCDs described in this work (Additional file 1: Fig. S5b). Finally, to generate pBCD2-H6-TATκ-pGFP, a PCR on pBCD2-107 (K) was performed with *pBCD_SphI_fw* and *H6-TATκ_BsaI_rv* primers to

isolate a DNA fragment harboring the H6-TatK encoding sequence to be cloned into pBEC-pGFP upstream of the pGFP gene.

Mutagenesis reactions into flCDKL5 encoding genes were carried out with the QuikChange II XL Site-Directed Mutagenesis kit (Agilent). In particular, the elimination of the first ATG was achieved using *M11_G2stop_fw* and *M11_G2stop_rv* primers; M10V mutations were performed with *M10V_fw* and *M10V_rv* primers; Kinase Dead mutants were obtained using *Mut_KK_fw* and *Mut_KK_rv*.

High copy number bicistronic plasmids

The low copy number OriR of the recombinant plasmids was replaced by the high copy number B40 replication origin using *AscI/NotI* double digestion.

Tricistronic designs

To co-express the LacZ Leader peptide, EB2 and flCDKL5, a similar strategy as for the BCD plasmids was applied. First, pB40-BCD2-107 (L) M10V was converted into an Entry Clone plasmid (pB40-BEC-107 (L) M10V). To do so, a PCR fragment encompassing the first half of the 107 (L) M10V gene was amplified with *CDKL5_L_PstI* and *BsaI_fw* and *CDKL5_NheI_rv* primers to introduce *PstI* and *BsaI* sites upstream of the 6xHis-Sumo tag encoding sequence. Then, such DNA fragment was cloned into pB40-BEC-107 (H) using *PstI/NheI* double digestion so as to replace the 5' half of the flCDKL5 gene and generate pB40-BEC-107 (L) M10V (Additional file 1: Fig. S6a). Finally, the BEC plasmid was converted into the tricistronic plasmid (TCD) by cloning a DNA fragment harboring the 5'UTR, the LacZ (aa 1–19) encoding sequence, and the *cmv*-EB2-6xHis gene between the two *BsaI* sites (Additional file 1: Fig. S6b).

For the development of TCD plasmids expressing CDD mutants of flCDKL5, TCD was subjected to mutagenesis with the QuikChange II XL Site-Directed Mutagenesis kit (Agilent) with the primers listed in Additional file 1: Table S2.

Measurement of average plasmid copy number (PCN)

The quantification of p79C-107(B) in *P. haloplanktis* TAC125 was carried out as previously reported [35]. Briefly, total DNA was extracted from 1 OD of bacterial cells using the E.Z.N.A. Bacterial DNA kit (Omega Bio-Tek Inc). Then, five serial dilutions of total DNA were employed as substrates in qPCR reactions with either *Prom7_fw* and *Prom7_rv* primers to detect the genomic DNA, or *CDKL5_fw* and *CDKL5_rv* to quantify the plasmid. Hence, the PCN values were calculated as follows: $PCN = E_c^{Ctc} / E_p^{Ctp}$, where E_c and E_p are the efficiencies obtained from the standard curves of the amplification

of the chromosomal and plasmid genes, respectively, and C_{tc} and C_{tp} are the threshold cycles for the two amplicons (chromosomal and plasmid genes) in each sample.

Quantification of flCDKL5 mRNA

Total RNA was isolated using the Direct-zol RNA Kit (Zymo Research, Irvine, CA, USA) adopting the manufacturer's instructions, followed by treatment with RNase-free DNase I (Roche, Mannheim, Germany) to avoid genomic DNA contamination. Total RNA was reverse transcribed using SuperScript IV (Invitrogen, Carlsbad, CA, USA) according to the recommended protocol using *PSHA_RS01090_rv* and *CDKL5_rv* primers. 1 μ L cDNA from each sample was used as the template for quantitative real-time PCR by using 1X PowerUp SYBR Green Master Mix (Applied Biosystems, Foster City, CA, USA) in the presence of 400 nM of specific primers (*PSHA_RS01090_fw*, *PSHA_RS01090_rv*; *CDKL5_fw*, *CDKL5_rv*). The reactions were run by a StepOne Real-time PCR System (Applied Biosystems, Foster City, CA, USA) and three independent sets of experiments were performed.

The thermal cycling protocol was set up as follows: UDG activation for 2 min at 50 °C; initial denaturation for 10 min at 95 °C; 40 cycles of denaturation for 15 s at 95 °C alternated with annealing/extension steps for 1 min at 60 °C. At the end of each reaction, melting curves were performed to verify the presence of a specific and unique amplification product. The housekeeping gene *PSHA_RS01090* was chosen as the normalizer for variations of mRNA amounts, cDNA synthesis efficiency and plasmid DNA contamination. The expression level of the *CDKL5* gene was assayed for up-regulation in the experimental samples (induction of expression) in comparison to the calibrator sample (noninduced cells, NI). The relative quantification of mRNA was expressed as fold-change and was calculated through the standard curve method (Pfaffl method):

$$\text{Gene expression ratio} = \frac{(E_{\text{target}})^{\Delta C_{\text{t}}^{\text{target}}(\text{control-sample})}}{(E_{\text{housekeeping}})^{\Delta C_{\text{t}}^{\text{housekeeping}}(\text{control-sample})}} \quad [49]$$

Measurement of pGFP fluorescence

After recombinant production, the equivalent of 1 OD cells was harvested and washed with PBS. Then, each sample was diluted with the same buffer to achieve the best signal-to-noise ratio in the fluorescence measurements that were carried out with a JASCO FP-750 spectrofluorometer at 25 °C (excitation at 488 nm and emission at 509 nm).

Analysis of the production of flCDKL5 and EB2 proteins

To analyze total protein productions, 1 OD pellets were solubilized in Laemmli Sample buffer and heated at 90 °C

for 20 min. Then, total cellular extracts were resolved by SDS-PAGE and analyzed either by Coomassie staining or Western blot. For the solubility analysis, cells were lysed in 20 mM sodium phosphate buffer pH 7.0, supplemented with 0.5 M NaCl, 10% Glycerol, 0.1% Triton X-100, 20 U/mL DNase I, 0.1 mg/mL lysozyme, 1 mM DTT and a protease inhibitor cocktail. Then, the soluble and insoluble fractions were segregated by centrifugation. Finally, the insoluble fraction was resuspended with lysis buffer in the same volume as the soluble fraction. 10 µg of soluble extract were analyzed by SDS-PAGE and the same volume of insoluble fraction was used as a control of solubility.

After SDS-PAGE runs, proteins were electroblotted to PVDF membranes using a semidry system. After the incubation with specific antibodies, the chemiluminescent signals were developed with the ECL method.

To detect fCDKL5, the membrane was blocked with PBS, 0.05% Triton X-100, 5% w/v milk for one hour. Then, CDKL5 (D-12): sc-376314 antibody (Santa Cruz Biotechnology) was diluted 1:1000 in the same buffer. After one hour of incubation at room temperature with the primary antibody, the membrane was washed with PBS, 0.05% v/v Triton X-100 three times (5 min each) and incubated with an anti-mouse antibody diluted 1:10,000 in PBS, 0.05% v/v Triton X-100, 5% w/v milk for one hour. Then, the membrane was washed again with PBS, 0.05% v/v Triton X-100 three times and the secondary antibody was detected using the ECL method.

For anti-FLAG Western blots, the membrane was blocked with PBS, 0.2% Tween 20, 5% w/v milk, for one hour. Then, Monoclonal ANTI-FLAG M2, Clone M2 (F1804, Sigma) was diluted 1:1000 in the same buffer. After overnight incubation at 4 °C with the primary antibody, the membrane was washed with PBS, 0.2% Tween 20 three times and incubated with an anti-mouse antibody diluted 1:5000 in PBS, 0.2% Tween 20, 5% w/v milk for one hour at room temperature. Then, the membrane was washed again with PBS, 0.2% Tween 20 three times and the secondary antibody was detected using the ECL method.

In the case of anti-His Western blots, the membrane was blocked with PBS, 5% w/v milk for one hour. Then, Monoclonal Anti-polyHistidine-Peroxidase clone HIS-1 antibody (A7058, Merck) was diluted 1:2000 in PBS, 0.05% Tween 20, 5% w/v milk. After one hour of incubation at room temperature with the antibody, the membrane was washed with PBS, 0.05% Tween 20 three times and it was developed.

To measure EB2 phosphorylation of Ser222, the membrane was blocked with TBST, 5% w/v milk for one hour. Then, anti-EB2 pS222 antibody (00117739, Covalab) was

diluted 1:4000 in the same buffer. After overnight incubation at 4 °C with the primary antibody, the membrane was washed with TBST three times and incubated with an anti-rabbit antibody diluted 1:2000 in TBST and 5% w/v milk for one hour at room temperature. Then, the membrane was washed again with TBST three times and was developed.

Parallel Coomassie stained polyacrylamide gels were always used to ascertain that complex samples (i.e. total and soluble lysates) were correctly balanced in Western blot analyses.

Preparation of samples for N-terminal sequencing

To produce the catalytic domain of CDKL5 in *E. coli*, the primers named *PhSumoCDKL5_NdeI_fw* and *PhCDKL5dC_XhoI_rv* were used in a PCR on pB40-BCD2-107 (L). This gene encoding a Sumo-tagged version of CDKL5(1–352) was cloned into the pET40-b vector in frame with a C-terminal 8xHis tag using NdeI/XhoI double digestion. After recombinant expression in *E. coli* BL21(DE3), the recombinant cells were resuspended in 50 mM TrisHCl pH 8.0, 0.5 M NaCl, 20 mM imidazole and lysed by sonication in the presence of a protease inhibitor cocktail. After centrifugation (14,000 g, 4 °C, 60 min), the soluble fraction was loaded onto a HisTrap of 1 mL (Cytiva) and both the full-length protein, and its N-terminally truncated fragment were collected with a linear gradient of imidazole. A sample containing approximately 30 µg of the intact catalytic domain and 6 µg of the N-terminally truncated fragment were loaded onto SDS-PAGE and then electroblotted onto a PVDF membrane using 10 mM CAPS, 10% methanol pH 11.0 as the transfer buffer. The protein bands were made visible by Ponceau S staining and submitted to Edman sequencing at the Institute of Biosciences and Bioresources (CNR, Naples).

Enrichment of 107 (L) M10V from *P. haloplanktis* TAC125 lysate

Recombinant *P. haloplanktis* TAC125 was lysed with a chemical-enzymatic method. Briefly, the cell paste was resuspended in 20 mM sodium phosphate buffer pH 7.0, supplemented with 0.5 M NaCl, 10% glycerol, 0.1% Triton X-100, 20 U/mL DNase I, 0.1 mg/mL lysozyme, 1 mM DTT, and a protease inhibitor cocktail to reach a final concentration of 14 OD/mL. After incubation at 4 °C for 20 min, the suspension was centrifugated (14,000 g for 45 min at 4 °C) to separate the soluble fraction from the cellular debris. Then, the soluble lysate was incubated with 0.15 mL of ANTI-FLAG M2 Affinity gel

(Millipore) at 4 °C for 4 h and a gravity flow column chromatography was performed. The resin was washed with lysis buffer containing 1% Triton X-100, while the elution was performed using in 0.5 mL of the same buffer containing 175 µM 3xFLAG peptide.

EB2 purification

A codon optimized EB2 gene was synthesized by an external company and cloned into the pET40-b with NdeI/BamHI double digestion. The resulting gene encodes human EB2 with an N-terminal 6xHis tag (see the Additional file 1: Information for the nucleotide sequence). After recombinant expression, *E. coli* BL21(DE3) recombinant cells were lysed by sonication in 50 mM TrisHCl pH 8.0, 0.5 M NaCl, 5% glycerol, 20 mM imidazole supplemented with a protease inhibitor cocktail. The soluble fraction was recovered after a centrifugation (14,000 g for 45 min at 4 °C) and loaded onto a HisTrap of 1 mL (Cytiva). The target protein was eluted with 250 mM imidazole and loaded onto a a Hiload 16/600 Superdex 200 pg using 50 mM TrisHCl pH 8.0, 0.18 M NaCl as a running buffer for a final polishing step. The final preparation was stored at – 80 °C in 40 mM TrisHCl pH 8.0, 0.15 M NaCl, 1 mM DTT, 15% glycerol at 1.8 mg/mL protein concentration.

In vitro kinase assay

EB2 phosphorylation assays with enriched CDKL5 proteins were carried out using 200 nM EB2 and 100 nM of enzyme in 30 µL of 20 mM TrisHCl pH 7.7, 0.5 M NaCl, 10 mM MgCl₂, 1 mM DTT, 0.7 mM ATP, complete protease inhibitor cocktail (Roche) and Halt phosphatase inhibitor cocktail (Thermo Fisher Scientific). The reactions were stopped after 30 min with 10 µL Laemmli Sample buffer 4 × and denatured at 70 °C for 20 min. 10 µL of each reaction were analyzed via either SDS-PAGE (for total CDKL5 and EB2 detection) or anti-EB2 pSer222 Western blot (for phosphorylated EB2 detection). As a negative control, a reaction was set up with flCDKL5 KD, a catalytically inactive CDKL5 variant. As a positive control, a reaction with commercial GST-CDKL5(1–498) (ab131695, abcam) was performed.

Statistics and reproducibility of results

The Data from the *in cellulo* kinase assays were statistically validated using the t-Student test comparing the mean measurements of experimental and control samples, both carried out as technical triplicates. The significance of differences between mean values was calculated using a two-tailed Student's t-test. A p value of < 0.05 was considered significant.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-022-01939-6>.

Additional file 1: Table S1. Characteristics of BCD constructs for pGFP expression. **Table S2.** List of primers used in this work. **Fig. S1.** Average plasmid copy number (PCN) of pP79-107 (B). **Fig. S2.** Ranking of bicistronic designs (BCDs) with a fluorescent reporter. **Fig. S3.** flCDKL5 production profiles with pBCD-107 (L) plasmids. **Fig. S4.** Development of Bicistronic Entry Clones. **Fig. S5.** Development of Bicistronic Designs. **Fig. S6.** Development of Tricistronic Designs.

Acknowledgements

This work is dedicated to Elettra Yvonne, Sofia, Elena, Arianna, Filippo, Raina, Gianluigi and all the beloved CDD patients all over the world, to their extraordinary parents and relatives. We feel responsible for your hope. Open access publication fees was kindly financed by the Department of Chemical Sciences, "Federico II" University of Naples.

Author contributions

A.C., E.P. and M.L.T. designed the project, A.C. was responsible of its experimental execution, of data collection and analysis, and wrote the manuscript. C.L. developed some bicistronic constructs for the expression of CDKL5 and pGFP, set up the *in vitro* activity assay and purified flCDKL5. M.C. handled the isolation of B40 origin of replication, the production of some flCDKL5 missense mutated genes, and the setup of the first *in cellulo* co-expression assay. C.L. and M.C. established the EB2 production and purification protocol. All the authors revised the manuscript and approved the content.

Funding

This research was funded by a research grant from the University of Pennsylvania Orphan Disease Center on behalf of the Loulou Foundation to MLT (pilot award no. CDKL5-20-101-08), by the Italian Parents' Association "La fabbrica dei sogni 2—New developments for Rett syndrome" (to A.C., M.C., C.L., E.P. and M.L.T.), and by the Italian Parents' Association "CONRETT ONLUS" (to A.C. and C.L.).

Availability of data and materials

The sequences of the original flCDKL5 encoding genes (107 (B), 107 (G), 107 (H)) cloned into pUC18 have been deposited in GenBank with the accession codes ON605205, ON605206 and ON605207, respectively. The synthetic plasmid pMK-T-H6EB2 from which the EB2 encoding gene was taken for expression in *E. coli* is available in GenBank with the accession code ON605208, while the sequence for EB2 co-expression with flCDKL5 was obtained from pMK-RQ-BCD_LacZ-cmyc-EB2-H6, whose accession code in GenBank is ON605209. All the other constructs described in this work were derived from such sequences and the ones described in refs. [30, 31]. Correspondence and material requests should be addressed to MLT (tutino@unina.it) and AC (andrea.colarusso@unina.it).

Declarations

Consent for publication
not applicable.

Competing interests

Authors declare no competing interests.

Author details

¹Department of Chemical Sciences, "Federico II" University of Naples, Complesso Universitario Monte S. Angelo-Via Cintia, 80126 Naples, Italy.
²Istituto Nazionale Biostrutture e Biosistemi-I.N.B.B., Viale Medaglie d'Oro, 305-00136 Rome, Italy.

Received: 4 July 2022 Accepted: 24 September 2022
Published online: 14 October 2022

References

- Hector RD, Kalscheuer VM, Hennig F, Leonard H, Downs J, Clarke A, et al. CDKL5 variants. *Neurol Genet.* 2017;3(6):e200.
- Krishnaraj R, Ho G, Christodoulou J. RettBASE: Rett syndrome database update. *Hum Mutat.* 2017;38(8):922–31.
- Olson HE, Demarest ST, Pestana-Knight EM, Swanson LC, Iqbal S, Lal D, et al. Cyclin-dependent kinase-like 5 deficiency disorder: clinical review. *Pediatr Neurol.* 2019;97:18–25.
- Trazzi S, De Franceschi M, Fuchs C, Bastianini S, Viggiano R, Lupori L, et al. CDKL5 protein substitution therapy rescues neurological phenotypes of a mouse model of CDKL5 disorder. *Hum Mol Genet.* 2018;27(9):1572–92.
- Terzic B, Felicia Davatolhagh M, Ho Y, Tang S, Liu YT, Xia Z, et al. Temporal manipulation of Cdkl5 reveals essential postdevelopmental functions and reversible CDKL5 deficiency disorder-related deficits. *J Clin Invest.* 2021. <https://doi.org/10.1172/JCI143655>.
- Gao Y, Irvine EE, Eleftheriadou I, Naranjo CJ, Hearn-Yeates F, Bosch L, et al. Gene replacement ameliorates deficits in mouse and human models of cyclin-dependent kinase-like 5 disorder. *Brain.* 2020;143(3):811–32.
- Kim JY, Bai Y, Jayne LA, Hector RD, Persaud AK, Ong SS, et al. A kinome-wide screen identifies a CDKL5-SOX9 regulatory axis in epithelial cell death and kidney injury. *Nat Commun.* 2020. <https://doi.org/10.1038/s41467-020-15638-6>.
- Muñoz IM, Morgan ME, Peltier J, Weiland F, Gregorczyk M, Brown FC, et al. Phosphoproteomic screening identifies physiological substrates of the CDKL 5 kinase. *EMBO J.* 2018;37(24):1–19.
- Baltussen LL, Negraes PD, Silvestre M, Claxton S, Moeskops M, Christodoulou E, et al. Chemical genetic identification of CDKL 5 substrates reveals its role in neuronal microtubule dynamics. *EMBO J.* 2018;37(24):1–18.
- Khanam T, Muñoz I, Weiland F, Carroll T, Morgan M, Borsos BN, et al. CDKL5 kinase controls transcription-coupled responses to DNA damage. *EMBO J.* 2021;40(23):e108271.
- Nawaz MS, Giarda E, Bedogni F, La MP, Ricciardi S, Ciceri D, et al. CDKL5 and shootin1 interact and concur in regulating neuronal polarization. *PLoS ONE.* 2016. <https://doi.org/10.1371/journal.pone.0148634>.
- Barbiero I, Peroni D, Tramarin M, Chandola C, Rusconi L, Landsberger N, et al. The neurosteroid pregnenolone reverts microtubule derangement induced by the loss of a functional CDKL5-IQGAP1 complex. *Hum Mol Genet.* 2017;26(18):3520–30.
- Khokhlatchev A, Xu S, English J, Wu P, Schaefer E, Cobb MH. Reconstitution of mitogen-activated protein kinase phosphorylation cascades in bacteria. Efficient synthesis of active protein kinases. *J Biol Chem.* 1997;272(17):11057–62.
- Fu Z, Schroeder MJ, Shabanowitz J, Kaldis P, Togawa K, Rustgi AK, et al. Activation of a nuclear Cdc2-related kinase within a mitogen-activated protein kinase-like TDY motif by autophosphorylation and cyclin-dependent protein kinase-activating kinase. *Mol Cell Biol.* 2005;25(14):6047–64.
- Fu Z, Larson KA, Chitta RK, Parker SA, Turk BE, Lawrence MW, et al. Identification of Yin-Yang regulators and a phosphorylation consensus for male germ cell-associated kinase (MAK)-related kinase. *Mol Cell Biol.* 2006;26(22):8639–54.
- Hector RD, Dando O, Landsberger N, Kilstrup-Nielsen C, Kind PC, Bailey MES, et al. Characterisation of CDKL5 transcript isoforms in human and mouse. *PLoS ONE.* 2016;11(6):1–22.
- Fahmi M, Yasui G, Seki K, Katayama S, Kaneko-Kawano T, Inazu T, et al. In silico study of Rett Syndrome treatment-related genes, MECP2, CDKL5, and FOXG1, by evolutionary classification and disordered region assessment. *Int J Mol Sci.* 2019;20(22):1–19.
- Altmeyer M, Neelsen KJ, Teloni F, Pozdnyakova I, Pellegrino S, Grøfte M, et al. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun.* 2015;6.
- Paz A, Zeev-Ben-Mordehai T, Sussman JL, Silman I. Purification of intrinsically disordered proteins [Internet]. *Instrumental Analysis of Intrinsically Disordered Proteins.* 2010. p. 695–704. (Wiley Online Books).
- Canning P, Park K, Gonçalves J, Li C, Howard CJ, Sharpe TD, et al. CDKL family kinases have evolved distinct structural features and ciliary function. *Cell Rep.* 2018;22(4):885–94.
- Katayama S, Sueyoshi N, Kameshita I. Critical determinants of substrate recognition by cyclin-dependent kinase-like 5 (CDKL5). *Biochemistry.* 2015;54(19):2975–87.
- Kameshita I, Sekiguchi M, Hamasaki D, Sugiyama Y, Hatano N, Suetake I, et al. Cyclin-dependent kinase-like 5 binds and phosphorylates DNA methyltransferase 1. *Biochem Biophys Res Commun.* 2008;377(4):1162–7.
- Katayama S, Inazu T. Straightforward and rapid method for detection of cyclin-dependent kinase-like 5 activity. *Anal Biochem.* 2019;566(November 2018):58–61.
- Vigentini I, Merico A, Tutino ML, Compagno C, Marino G. Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J Biotechnol.* 2006;127(1):141–50.
- Papa R, Rippla V, Sannia G, Marino G, Duilio A. An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125. *J Biotechnol.* 2007;127(2):199–210.
- Unzueta U, Vázquez F, Accardi G, Mendoza R, Toledo-Rubio V, Giuliani M, et al. Strategies for the production of difficult-to-express full-length eukaryotic proteins using microbial cell factories: production of human alpha-galactosidase A. *Appl Microbiol Biotechnol.* 2015;99(14):5863–74.
- Calvanese M, Colarusso A, Lauro C, Parrilli E, Tutino ML. Soluble recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125: the case study of the full-length human CDKL5 protein. *Methods Mol Biol.* 2022;2406:219–32.
- Fondi M, Maida I, Perrin E, Mellerà A, Mocali S, Parrilli E, et al. Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Environ Microbiol.* 2015;17(3):751–66.
- Fondi M, Gonzi S, Dziurzynski M, Turano P, Ghini V, Calvanese M, et al. Modelling hCDKL5 heterologous expression in bacteria. *Metabolites.* 2021;11(8):1–16.
- Sannino F, Giuliani M, Salvatore U, Apuzzo GA, de Pascale D, Fani R, et al. A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl Microbiol Biotechnol.* 2017;101(2):725–34.
- Colarusso A, Lauro C, Calvanese M, Parrilli E, Tutino ML. Improvement of *Pseudoalteromonas haloplanktis* TAC125 as a cell factory: IPTG-Inducible plasmid construction and strain engineering. *Microorganisms.* 2020;8(10):1–24.
- Mutalik VK, Guimaraes JC, Cambay G, Lam C, Christoffersen MJ, Mai QA, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods.* 2013;10(4):354–60.
- Tutino ML, Duilio A, Parrilli E, Remaut E, Sannia G, Marino G. A novel replication element from an Antarctic plasmid as a tool for the expression of proteins at low temperature. *Extremophiles.* 2001;5(4):257–64.
- Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* 2007;35(SUPPL2):126–31.
- Qi W, Colarusso A, Olombrada M, Parrilli E, Patrignani A, Tutino ML, et al. New insights on *Pseudoalteromonas haloplanktis* TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies. *Sci Rep.* 2019;9(1):16444.
- Koutmou KS, Schuller AP, Brunelle JL, Radhakrishnan A, Djuranovic S, Green R. Ribosomes slide on lysine-encoding homopolymeric A stretches. *Elife.* 2015;2015(4):1–18.
- Arthur LL, Pavlovic-djuranovic S, Koutmou KS, Green R. Translational control by lysine-encoding A-rich sequences. *Sci Adv.* 2015. <https://doi.org/10.1126/sciadv.1500154>.
- Buskirk AR, Green R. Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos Trans R Soc B Biol Sci.* 2017. <https://doi.org/10.1098/rstb.2016.0183>.
- Suskiewicz MJ, Sussman JL, Silman I, Shaul Y. Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Sci.* 2011;20(8):1285–97.
- Zhang G, Ignatova Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr Opin Struct Biol.* 2011;21(1):25–31.
- Verma M, Choi J, Cottrell KA, Lavagnino Z, Thomas EN, Pavlovic-Djuranovic S, et al. A short translational ramp determines the efficiency of protein synthesis. *Nat Commun.* 2019;10(1):1–15.
- Leith EM, O'Dell WB, Ke N, McClung C, Berkmen M, Bergonzo C, et al. Characterization of the internal translation initiation region in

- monoclonal antibodies expressed in *Escherichia coli*. *J Biol Chem*. 2019;294(48):18046–56.
43. Zhao Z, Liu X, Zhang W, Yang Y, Dai X, Bai Z. Construction of genetic parts from the *Corynebacterium glutamicum* genome with high expression activities. *Biotechnol Lett*. 2016;38(12):2119–26.
 44. Whitaker WR, Lee H, Arkin AP, Dueber JE. Avoidance of truncated proteins from unintended ribosome binding sites within heterologous protein coding sequences. *ACS Synth Biol*. 2015;4(3):249–57.
 45. Bertani I, Rusconi L, Bolognese F, Forlani G, Conca B, De Monte L, et al. Functional consequences of mutations in CDKL5, an X-linked gene involved in infantile spasms and mental retardation. *J Biol Chem*. 2006;281(42):32048–56.
 46. Lin C, Franco B, Rosner MR. CDKL5/Stk9 kinase inactivation is associated with neuronal developmental disorders. *Hum Mol Genet*. 2005;14(24):3775–86.
 47. Parrilli E, Giuliani M, Tutino ML. General secretory pathway from marine antarctic *Pseudoalteromonas haloplanktis* TAC125. *Mar Genomics*. 2008;1(3–4):123–8.
 48. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*. 2008;3(11): e3647.
 49. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29(9):45.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

