3-25-2020 10:00 AM

# Practical Applications and Future Directions of Genetic Code Expansion: Validation of Novel Akt1 Substrates and the Design of a Synthetic Auxotroph Strain of B. subtilis

McShane M. McKenna, *The University of Western Ontario*

Supervisor: O'Donoghue, Patrick, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biochemistry
© McShane M. McKenna 2020

## Thesis Abstract

In Chapter 1, site-specifically phosphorylated variants of the oncogene Akt1 were made in *Escherichia coli* using the orthogonal translation system that enable genetic code expansion with phosphoserine. The differentially phosphorylated variants of Akt1 were used to validate newly predicted Akt1 substrates. The predicted target sites of the peptide substrates were synthesized and subjected to *in vitro* kinase assays to quantify the activity of each Akt1 phosphorylated variant towards the predicted peptide. A previously uncharacterized kinase-substrate interaction between Akt1 and a peptide derived from RAB11 Family Interacting Protein 2 (RAB11FIP2) was validated in *vitro*. Chapter 2 describes the preliminary development of a novel orthogonal translation system for *Bacillus subtilis*. The work presented outlines the design process: from selection of the components to the generation of an all-in-one plasmid containing the orthogonal translation system. The work demonstrates stable integration of the orthogonal translation system into the *B. subtilis* genome.

**Keywords:**

# Summary for Lay Audience

Nearly every living organism on the planet studied to date uses the same standard 20 amino acid building blocks to create proteins and enzymes. Each of these 20 amino acids contains a unique chemical functional group, including, for example, acidic/basic groups, aromatic rings, sulfur groups, that each possess distinct chemical properties. Using different combinations and sequences of these 20 building blocks, organisms can produce thousands of different proteins with widely different functions.

Recent advances in genetics and molecular biology have led to the development of technologies that can expand the repertoire of amino acids that are available for organisms beyond these natural 20. The ability to expand the genetic code has led to 100's of different unnatural or non-standard amino acids being used by a wide range of organisms from bacteria to mice. These new amino acids can revolutionize the types of proteins that organisms can produce by providing them with chemical functional groups that were previously not available. The work presented here makes use of such a system to produce active variants of a human protein called Akt1 that is commonly overactivated in cancers. These active Akt1 variants are made by introducing phosphorylation amino acids into the Akt1 protein, leading to an active Akt1 enzyme. By testing these active variants, we discovered a novel protein targets that Akt1 was able to act on. Results like these enhance our understanding of the scope of Akt1's role regulating pathways linked to cancer and provide leads to investigate potential new drug targets. In the second chapter, I outline pioneering efforts towards the creation of a novel genetic code expansion technology in a species of bacteria (*Bacillus subtilis*) that did not previously have such a system. The design process, starting with the careful selection of the components and eventual assembly into a single plasmid DNA and integration into the *B. subtilis* genome is presented.

## Co-Authorship Statement

Chapter 1. contains data produced by myself as well as a few instances of data from experiments performed by Nileeka Balasuriya, included with her permission. The experiments in Figure S1.4 were conducted by Nileeka Balasuriya and are unpublished. The data presented in Figure 1.4 labelled "preparation 2" was generated by Nileeka Balasuriya and was published in the manuscript: Balasuriya, N., McKenna, M., Liu, X., Li, S.S.C. and O'Donoghue, P. (2018) Phosphorylation-Dependent Inhibition of Akt1. Genes (Basel). All remaining experimental data was generated by McShane McKenna. The MALDI-TOF mass spectrometry analysis of the peptide substrates presented in Figure 1.3 and Figure S1.1 was performed by Kristina Jurcic.

## Acknowledgements

# Table of Contents

# List of Tables

# List of Figures

## List of Abbreviations

| | |
|---|---|
| 4-FTrp | 4-fluoro-tryptophan |
| A100E | Alanine at position 100 mutated to glutamic acid |
| aaRS | Aminoacyl-tRNA synthetase |
| AcK | $N^{\varepsilon}$-acetyl-L-lysine |
| AckRS | $N^{\varepsilon}$-acetyl-L-lysyl-tRNA synthetase |
| AGC | Protein Kinase A, G, and C Families |
| amyE | Alpha amylase gene |
| ASH1L | Histone-lysine N-methyltransferase |
| ATP | Adenosine triphosphate |
| BGSC | Bacillus Genetic Stock Center |
| BocLys | $N^{\varepsilon}$-Boc-L-lysine |
| bp | Base pairs |
| BRPF1 | Peregrin |
| BTBD11 | Ankyrin repeat and BTB/POZ |
| CAD | Canadian Dollars |
| CDCA7L | Cell division cycle-associated 7-like |
| CHCA | $\alpha$-Cyano-4-hydroxycinnamic acid |
| CMTM4 | MARVEL transmembrane 4 |
| C-term | C-terminal |
| Cys | Cysteine |
| CYSLTR1 | Cysteinyl leukotriene receptor 1 |
| ddH$_2$0 | Double distilled H$_2$0 |
| DH5$\alpha$ | *Escherichia coli* strain |
| DNA | Deoxyribonucleic acid |
| DnaX | DNA polymerase III subunit tau gene |
| dNTP | Deoxynucleoside triphosphate |
| Dpn1 | Type-2 restriction enzyme DpnI |
| DTT | Dithiothreitol |
| E18TAG | Mutation from glutamate at position 18 to TAG |
| EDTA | Ethylenediaminetetraacetic acid |

| | |
|---|---|
| EFSep21 | Elongation factor-Tu |
| EGTA | Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid |
| ERK1/2 | Extracellular signal-regulated kinases 1 and 2 |
| FDA | Food and Drug Administration |
| FMO2 | Dimethylaniline monooxygenase |
| Fmoc | Fluorenylmethyloxycarbonyl chloride |
| fmol | Femtomole |
| FOXO1A | Forkhead box protein O1 |
| FRYL | Protein furry homolog-like |
| GFP | Green Fluorescent Protein |
| (Sp)GFP | (Staphylococcus pneumonia) GFP |
| GRAMD1C | GRAM domain containing 1C |
| GRAS | Generally Regarded As Safe |
| GSK3B | Glycogen synthase kinase-3β |
| GSK-3β(S9A) | Glycogen synthase kinase-3β (S9A) |
| H62Y | Histidine at position 62 mutated to tryptophan |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HF | High Fidelity |
| His | Histidine |
| HRH1 | Histamine H1 receptor |
| IPTG | Isopropyl β- d-1-thiogalactopyranoside |
| $k_{app}$ | Apparent catalytic rate |
| kb | Kilobase |
| KCNH2 | Potassium channel H2 |
| kDa | Kilo Dalton |
| KIAA1109 | Uncharacterized protein KIAA1109 |
| Kpn1 | Type-2 restriction enzyme KpnI |
| LB | Luria-Bertani medium |
| LPHN3 | Adhesion G protein-coupled receptor L3 |
| LPS | Lipopolysaccharides |
| LspA | Lipoprotein signal peptidase |

| | |
|---|---|
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionization-Time Of Flight |
| min | Minute |
| mL | Millilitre |
| mM | Millimolar |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| mRNA | Messenger ribonucleic acid |
| MSC | Multiple cloning site |
| mTOR | Mechanistic target of rapamycin |
| mTORC2 | mTOR Complex 2 |
| MWCO | Molecular weight cut-off |
| ng | Nanogram |
| Ni-NTA | Nickle-nitrilotriacetic acid |
| nM | Nanomolar |
| nm | Nanometer |
| N-term | N-terminal |
| o-aaRS | Orthogonal aminoacyl-tRNA synthetase |
| $OD_{600}$ | Optical Density at a wavelength of 600 nm |
| OPAL | Oriented Peptide Array Library |
| o-tRNA | orthogonal transfer RNA |
| OTS | Orthogonal Translation System |
| $pAkt1^{S473}$ | Phosphorylated Akt1 at position Serine 473 |
| $pAkt1^{T308}$ | Phosphorylated Akt1 at position Threonine 308 |
| $pAkt1^{T308,S473}$ | Dual phosphorylated Akt1 at positions Threonine 308/Serine 473 |
| PCNXL3 | Pecanex-like 3 |
| PCR | Polymerase Chain Reaction |
| PD-L1 | Programed death ligand-1 |
| PDK1 | Protein 3-phosphoinositide-dependent protein kinase-1 |
| PH domain | Pleckstrin homology domain |
| ΔPH | PH domain deletion |
| $PH_{KI}$ | Pleckstrin homology domain knock-in |
| PHLPP | PH domain and Leucine rich repeat Protein Phosphatases |

| | |
|---|---|
| PI3K | Phosphoinositide 3-kinase |
| PIP2 | Phosphatidylinositol 4,5-bisphosphate |
| PIP3 | Phosphatidylinositol (3,4,5)-trisphosphate |
| PITPNM2 | Phosphatidylinositol transfer protein 2 |
| PKB | Protein kinase B |
| pmol | Picomole |
| PP2A | Protein phosphatase 2A |
| PRAS40 | Proline-rich Akt substrate of 40 kDa |
| PRM2 | Protamine-2 |
| pSer | Phosphorylated serine |
| PSSM | Position Specific Scoring Matrix |
| PylRS | Pyrrolysyl-tRNA synthetase |
| Rab11 | Ras-related protein Rab-11A |
| RAB11FIP2 | Rab11 family-interacting protein 2 |
| RANBP3L | Ran-binding protein 3-like |
| S3TAG | Mutation from serine at position 3 to TAG |
| S473 | Serine at position 473 in Akt1 |
| sarA | Transcriptional regulator SarA |
| SASH1 | SAM and SH3 domain-containing |
| SCN2A | Sodium channel protein type 2 alpha |
| SecY | Protein translocase subunit SecY |
| SEMA4G | Semaphorin-4G |
| SepRS9 | phosphoseryl-tRNA synthetase |
| Ser | Serine |
| Sf9 | *Spodoptera frugiperda* Sf21 insect cell line |
| Sin1 | Stress-activated map kinase-interacting protein 1 |
| SKP2 | S-phase kinase-associated protein 2 |
| SRRM4 | Serine/arginine repetitive matrix protein 4 |
| T308 | Threonine at position 308 in Akt1 |
| T56P | Threonine at position 56 mutated to proline |
| TD | Thermal decomposition |

| | |
|---|---|
| TFA | Trifluoroacetic acid |
| Thr | Threonine |
| TIPS | Triisopropyl silane |
| TMC7 | Transmembrane channel-like protein 7 |
| TRAPPC1 | Trafficking protein particle complex 1 |
| tRNA | Transfer ribonucleic acid |
| Trp | Tryptophan |
| TSC2 | Tuberous Sclerosis Complex 2 |
| U | Unit of enzyme |
| uAA | Unnatural amino acid |
| $v_0$ | Initial velocity |
| V31I | Valine at position 31 mutated to isoleucine |
| WNK1 | WNK Lysine Deficient Protein Kinase 1 |
| ZNF256 | Zinc finger protein 256 |
| γ-[32P]-ATP | Adenosine triphosphate (gamma phosphate labelled with 32P) |
| μCi | Microcurie |
| μg | Microgram |
| μL | Microlitre |
| μM | Micromolar |
| μm | Micrometer |

# CHAPTER 1 – Phospho-form specific substrates of protein kinase B (Akt1)

## 1.0 Abstract

Protein kinase B (Akt1) is hyper-activated in diverse human tumors. Akt1 is activated by phosphorylation at two key regulatory sites, Thr308 and Ser473. Active Akt1 phosphorylates many, perhaps hundreds, of downstream cellular targets in the cytosol and nucleus. Akt1 is well known for phosphorylating proteins that regulate cell survival and apoptosis, however, the full catalog of Akt1 substrates remains unknown. Recent work using large peptide libraries concludes that each active, phosphorylated form of Akt1 (pAkt1$^{S473}$, pAkt1$^{T308}$, ppAkt1$^{S473,T308}$) has a distinct substrate specificity, and these data were used to predict potential new Akt1 substrates. Using the most high-confidence predictions, I designed 26 target peptides that represent putative Akt1 substrates. The peptides, in addition to positive and negative control peptides, were synthesized by solid phase synthesis and their purity was confirmed by mass spectrometry. I found that most of the predicted peptides showed phosphorylation accepting activity equal to or greater than that observed with a peptide derived from a well-established Akt1 substrate, glycogen synthase kinase 3 β (GSK-3β). Some of the peptides displayed strong selectivity for either the pAkt1$^{S473}$ or the ppAkt1$^{S473,T308}$ variant. The data support the hypothesis that Ser473 phosphorylation modulates Akt1 substrates selectivity. Akt1 was most active with putative substrates including the PIP3-binding protein Rab11 family-interacting protein 2 and cysteinyl leukotriene receptor 1, indicating their potential roles in Akt1-dependent cancers.

## 1.1 Introduction

### 1.1.1 Protein kinase B (Akt) is an oncogenic kinase.

Akt belongs to the AGC family of serine-threonine kinases (1,2). In mammals, there are three *Akt* genes that encode the Akt isozymes Akt1, Akt2 and Akt3. The three isozymes display high sequence identity, similar substrate specificity and share extensive homology

in their kinase domains(1,2). All three Akt1 isozymes share the same fundamental structure consisting of four distinct domains: an N-terminal regulatory pleckstin homology (PH) domain; an unstructured linker region which serves to connect the PH domain to the serine/threonine-specific kinase domain; and a final C-terminal domain often referred to as the hydrophobic motif, which is responsible for discrete modifications to the activity of Akt1(3). Akt1 is a key regulatory kinase that transduces signals through the phosphoinositide 3-kinase (PI3K)/Akt1 cell-signaling cascade (1) that controls cell growth and survival (2). The PI3K/Akt1 pathway is one of the most commonly deregulated pathways in human cancer (4,5). In fact, Akt1 is hyper-phosphorylated and overactive in >50% of human tumors (2,6,7), and elevated Akt1 phosphorylation status is linked to poor prognosis in patients (5,8,9).
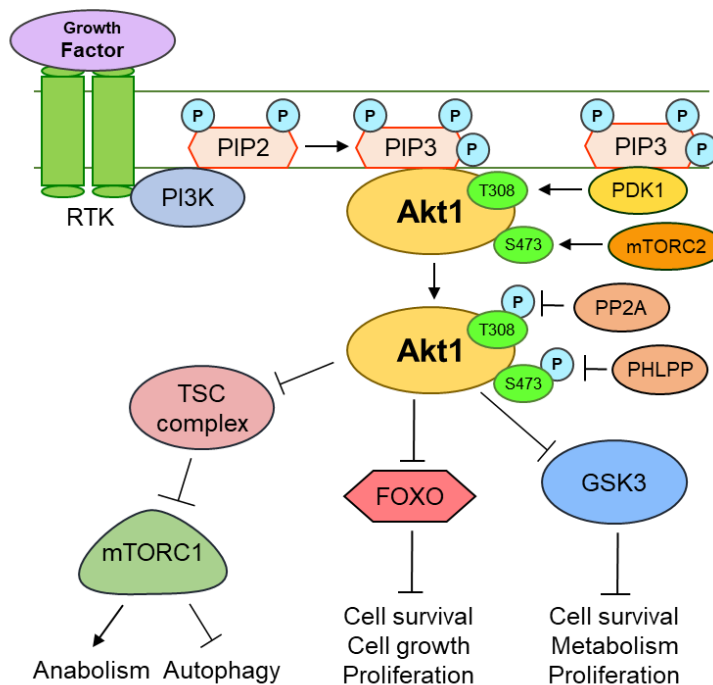


**Figure 1.1: Growth factor mediated activation of Akt1.** The activation pathway of Akt1 in response to growth factor mediated activation of the PI3K pathway is shown. Activated Akt1 promotes an oncogenic cellular phenotype that exhibits inhibited autophagy and apoptosis pathways and activated cellular growth, proliferation and anabolic pathways. Akt1 inhibits autophagy via the TSC complex/mTORC1 pathway (10), and inhibits apoptosis/promotes cellular growth through the inhibition of the tumour suppressors FOXO and GSK-3β (11,12).

**1.1.2 Akt1 activation.**

Activation of Akt1 is a multi-step process that involves a carefully orchestrated series cellular translocation (13) and post translational modification events (reviewed in (2)). Indeed, serine/threonine phosphorylation (14-16), methylation(17,18), ubiquitination (19-21) and proline hydroxylation (22) have been shown to regulate the activation or to modify the enzymatic activity of Akt1. Among these post translational modifications, serine/threonine phosphorylation has been studied most extensively (1,2). Phosphorylation of the threonine residue at position 308 (T308) in the activation loop of Akt1 is both necessary and sufficient to achieve substantial Akt1 activity (15), however, a second phosphorylation of the serine residue at position 473 (S473) in the C-terminal hydrophobic motif of Akt1 is necessary to achieve maximal Akt1 catalytic activity (15).

In cells, the phosphorylation of Akt1 at positions T308 and S473 results from insulin or growth factor stimulation of receptor tyrosine kinases and G protein coupled receptors on the cell surface (23). Growth factor stimulation promotes plasma membrane recruitment and subsequent activation of members of the class I PI3K family (Figure 1.1). Following activation by receptor kinases, PI3K phosphorylates phosphatidylinositol-4,5-bisphosphate (PIP2), converting PIP2 into phosphatidylinositol-3,4,5-triphosphate (PIP3) (2). PIP3 acts a scaffold and forms an interaction with the PH domains of various proteins, including Akt1 and phosphoinositide dependent kinase 1 (PDK1), another member of the AGC kinase family which functions as the upstream kinase responsible for the activation of Akt1. The interaction between PIP3 and the Akt1 PH domain causes Akt1 to adopt a "PH-out" conformation, relieving Akt1 of the auto-inhibitory affect of its own PH domain (24). The "PH-out" conformation of Akt1, along with its proximity to PDK1 mediated by their interactions with PIP3 allows PDK1 to activate Akt1 by phosphorylation at Thr308 (25,26). Further activation is achieved after Akt1 is phosphorylated at Ser473 by the mechanistic target of rapamycin complex 2 (mTORC2) (27). This mTORC2-mediated phosphorylation of Akt1 at S473 is itself dependant on the contribution and activity of numerous proteins (Protor (28), Sin1 (29), TSC2 (30), reviewed in (31)) which function to stabilize mTORC2. Akt1 activity is normally down regulated (32) through deactivation via

de-phosphorylation of Akt1 at Thr308 by protein phosphatase 2A (PP2A) (33) and at Ser473 by PH domain leucine-rich repeat protein phosphatases (PHLPP) (34).

### 1.1.3 Therapeutic relevance of Akt1.

The Akt1 signaling pathway is a useful prognostic marker of various cancers as well as a promising target for therapies (6-9). As of this report, the Akt1 pathway has been the subject of over 300 clinical trials (35,36). The majority of these clinical trials investigate small molecule inhibitors (37) as potential tools which can be used to inhibit or reduce the activity of the hyper-active Akt1. Unfortunately, Akt1's structural and functional similarity to its isozymes and other members of the AGC-kinase family has made it difficult to solely target Akt1 without also inhibiting additional kinases and causing unwanted side effects in clinical trials (38).

### 1.1.4 The Akt1 signalling network.

Ultimately, the cellular consequences of Akt1 activity are determined by the profile of substrates that it either activates or deactivates via phosphorylation in any given cellular event (2). In order to identify new therapeutic targets for Akt1-dependent cancers, it is important to identify the complete catalog of Akt1-dependent substrates. By expanding the scope of pathways regulated by Akt1, new routes will emerge to inhibit specific interactions between Akt1 and downstream substrates. Some Akt1 substrates are causative agents in pathologies such as cancers (39) or diabetes (40). For example, in a study investigating the rat model of the proapoptotic protein Par-4, it was found that Akt1 was directly responsible for the inactivating phosphorylation of Par-4 (at that lead to oncogenic inhibition of apoptosis (41).

As of 2019, Akt1 has over 200 reported substrates (1,2) that are involved in a wide variety of cellular events, including signal transduction, metabolism, cell-cycle regulation, transcription regulation, proliferation and angiogenesis (1,2). The level of certainty that all of these substrates are *bona fide* Akt1 targets is substantiated by differing levels of evidence. All substrates adhere to some degree to the Akt1 phosphorylation motif. Pioneering efforts with peptide arrays defined the consensus Akt1 target motif as $R_{-5}X_{-4}R_{-}$

5

$_3X_{-2}X_{-1}S/T_0\phi_1$, where X represents any amino acid, $\phi$ represents any bulky aromatic residue and subscript numbers refer to the residue's position relative to the phosphorylated residue S or T (42). As reviewed in (1), many of the now "known" Akt1 substrates were originally validated by *in vitro* kinases assays, including well-established Akt1 targets FOXO1A (43), GSK-3β (44), WNK1 (45). Traditionally, *in vivo* studies using mouse models are used to further validate reported Akt1 substrates. Homozygous knock-in mouse strains that either lack PDK1 expression, PDK1(−/−) (46), or that express a PDK1 variant that is unable to activate Akt1, which is catalytically active but unable to associate with PIP3/Akt1 (47), have been used in combination to demonstrate the ability of Akt1 to phosphorylate PRAS40, Foxo1, Foxo3, GSK-3β, TSC2 and WNK1 (47).

**1.1.5 Investigating Akt1 substrate-specificity using genetic code expansion.**
Recently, we developed a protocol that allows recombinant production of Akt1 variants that are phosphorylated at either or both regulatory sites (Thr308, Ser473) (15) (Figure 1.2). To produce recombinant Akt1 that is phosphorylated at S473, an orthogonal translation system (OTS) is used to reassign the UAG stop codon to genetically encode phosphoserine (pSer). In general, an OTS requires two key components: an orthogonal aminoacyl-tRNA synthetase (o-aaRS) and cognate orthogonal transfer RNA (o-tRNA) pair. In an OTS, the aaRS charges its cognate tRNA with an unnatural amino acid (uAA) that is usually provided to *E. coli* via direct supplementation of the growth media (Figure 1.1). Orthogonality in this case refers to the fact that the components of an OTS do not possess the ability to cross-react with the host organism's endogenous aaRSs and tRNAs.

In *E. coli*, the process of mRNA translation or protein synthesis involves over 100 discrete components, all of which must work efficiently to maintain protein synthesis and production rates, which are essential to normal cell growth (48). To expand the natural translation system, it is important that the engineered system does not interfere with or compromise the integrity of the natural components. The orthogonal aaRS must only recognize and aminoacylate its cognate orthogonal tRNA with a specific or desired unnatural amino acid. Orthogonality ensures that the engineered aaRS does not charge any of the endogenous *E. coli* tRNAs with the uAA so that the uAA is not incorporated randomly throughout the proteome. Likewise, the orthogonal tRNA must not be charged

with naturally occurring amino acids by any of the endogenous aaRSs present in *E. coli* so that the OTS does not needlessly incorporate one of the naturally occurring amino acids in place of an uAA at the UAG codon.

The pSer OTS used here (Figure 1.2A) was comprised of a phosphoseryl-tRNA synthetase (SepRS) that was derived from an archaeal pathway for Cys-tRNA$^{Cys}$ synthesis(49), a mutant tRNA$^{Sep}$ (50), and a mutant elongation factor-Tu (EF-Sep) (49). During the development of the pSer OTS, it was determined that the inclusion of EF-Sep was necessary to circumvent the quality-control functional of elongation factor-Tu, which had prevented robust pSer incorporation in initial experiments with the pSer OTS (49).

Using the pSer OTS, we are able to reassign the UAG codon to genetically encode pSer in response to a UAG codon at position 473 in the Akt1 construct (Figure 1.2 C and D). Our lab has previously validated this method using multiple-reaction monitoring MS/MS to unambiguously identify pSer at position 473 in Akt1 produced via this method with no evidence of dephosphorylation. In order to produce recombinant Akt1 that is phosphorylated at Thr308, the Akt1-upstream kinase PDK1 is co-expressed with Akt1. We have previously demonstrated using parallel reaction-monitoring MS/MS that co-expression of PDK1 with Akt1 results in undetectable levels of unphosphorylated Thr308 in the purified pAkt1$^{T308}$ product (15).

This method of Akt1 production has provided the unique opportunity to investigate the substrate specificity of individual Akt1 phosphorylated forms in a manner that was not previously possible (14). Compared to commercially available preparations of Akt1 produced using Sf9 insect cell lines that are of variable activity and contain mixtures of the active Akt1 phospho-forms (15,51), our protocol is able to produce consistent preparations of each individual phospho-form (15,16). Using these reagents, we defined the specific role that each phosphorylation has on Akt1 activation (15) and inhibition by clinically relevant compounds (16). We further showed that Akt1 phosphorylation status globally regulates substrate specificity (14).
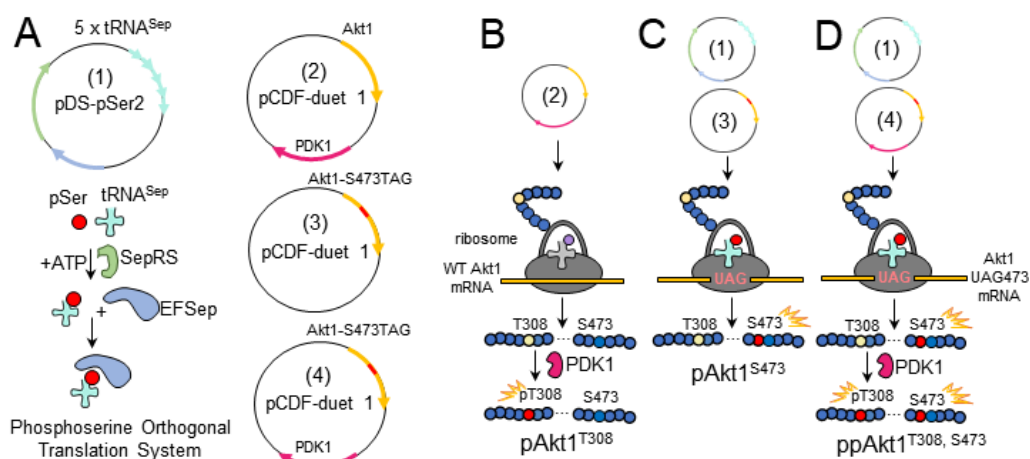
**Figure 1.2 Production of differentially phosphorylated Akt1 variants.** A) (1) pDS-pSer2 contains the phosphoserine orthogonal translation system (OTS) which was used to genetically incorporate phosphoserine at position 473 in response to a UAG nonsense codon in Akt1 mRNA transcripts. Plasmids (2-4) contain the Akt1 genes used to express Akt1. Plasmid (2) contains Akt1 (without a TAG mutation at S473) and PDK1. Plasmid (3) contains Akt1 with a TAG mutation at S473. Plasmid (4) contains both a mutated Akt1 gene (TAG at S473) and PDK1. In all, three different phospho-forms of Akt1 were produced in *E. coli*: B) (mono-phosphorylated) $pAkt1^{T308}$, C) (mono-phosphorylated) $pAkt1^{S473}$, and D) (dual-phosphorylated) $ppAkt1^{T308/S473}$. B) To create $pAkt1^{T308}$, *E. coli* was transformed with (2) pCDF-Duet1-ΔPHAkt1-PDK1 containing PDK1 (Akt1's natural upstream kinase) which is co-expressed along with Akt1 in order to phosphorylate T308. C) To create $pAkt1^{S473}$, *E. coli* was co-transformed with (1) pDS-pSer2 containing the OTS which incorporates phosphoserine into Akt1 during translation in response to TAG codons and (3) pCDF-Duet1-ΔPHAkt1-473TAG which contains an Akt1 gene bearing a TAG mutation at position S473. D) $ppAkt1^{T308,S473}$ was created by combining both methods and co-transforming *E. coli* with (1) pDS-pSer2 and (4) pCDF-Duet1-ΔPHAkt1-473TAG-PDK1. For the production of $pAkt1^{S473}$ and $ppAkt1^{T308,S473}$, *E. coli* were supplemented (2.5 mM) with pSer.

To study the substrate-specificities of individual Akt1 phospho-forms, our previous efforts (14) tested an oriented peptide array library (OPAL) which contained over 6.9 billion unique, potential Akt1 substrates. The potential substrates in the OPAL screen were partially degenerate peptides that had the following composition: Biotin-AGG-$X_{-6}X_{-5}X_{-4}R_{-3}X_{-2}X_{-1}S_0X_1X_2X_3A$; X represents any amino acid other than Ser, Thr or Cys. The OPAL screen was subjected to kinase assays using radiolabeled ATP catalyzed by each of the Akt1 phospho-forms. Data from the OPAL screen revealed that each phospho-form has

distinct preferences for specific residues at each of the degenerate positions. This is a standard method used to define the consensus substrate motif for a given kinase (42,52), but the approach was not previously used to differentiate the substrate specificity of distinct Akt1 phospho-forms. The results of the OPAL screens showed that each of the Akt1 phospho-forms had a distinct substrate specificity profile (14), demonstrating Akt1 substrate specificity is regulated by its phosphorylation status.

Here, peptides representing predicted Akt1 substrates based on the OPAL data were synthesized and characterized for activity with each Akt1 phospho-form. For each of the Akt1 phospho-forms, we were previously (14) able to determine the amino acids that were either preferred or dis-favoured according to enzyme activity at each of the degenerate (X) sites that surrounded the substrate's phosphorylatable residue ($S_0$) in the consensus motif. The OPAL data were used to create position-specific scoring matrices (PSSMs) for each Akt1 phospho-form that quantify residues preferences at each site in the consensus motif. The PSSMs where then used to search and rank the human phospho-proteome (14).

Using the PSSM scoring, we could then predict the expectation that a given known phospho-peptide sequences was a likely substrate for one or more of the Akt1 phospho-forms. In order to validate substrates predicted by the OPAL data, I synthesized 26 potential peptide substrates (Table 1.1) that represent novel putative Akt1 protein (1,2). The targeted set of 30 peptides included those ranked highest (top 1%; p-value $< 10^{-6}$) in our OPAL-derived PSSM search of the human phospho-proteome (14). The peptides were divided into three sets, with each set representing peptides that were ranked highest for each of the three phospho-forms. Kinase activity assays showed that the majority of putative substrates displayed significant activity with at least one Akt1 phospho-form. Further the data revealed novel and phospho-form-specific Akt1 substrates.

**Table 1.1. List of synthetic peptides representing predicted Akt1 substrates**

| gene | p-site | protein name | peptide sequence | MW calc (Da) | MW obs (Da) | PSSM search p-value |
|---|---|---|---|---|---|---|
| pAkt1^S473 | | | | | | |
| HRH1 | S396 | histamine H1 receptor | TWKRLRSHSRQYV | 1717.0 | 1716.0 | 4.6×10⁻⁷ |
| BRPF1 | S542 | peregrin | FMQRLHSYWTLKR | 1766.2 | 1765.0 | 8.2×10⁻⁷ |
| KIAA1109 | S4592 | uncharacterized protein | TSKRALSTWGPVP | 1399.7 | 1398.8 | 1.6×10⁻⁶ |
| TMC7 | S125 | transmembrane channel-like protein 7 | QWKRYSSKSWKRF | 1787.2 | 1786.0 | 1.6×10⁻⁶ |
| RANBP3L | S77 | Ran-binding protein 3-like | KRVRSSSFTFHIT | 1565.9 | 1564.9 | 3.1×10⁻⁶ |
| ZNF256 | T374 | zinc finger protein 256 | THQRVHTGTRPYM | 1583.7 | 1582.8 | 3.3×10⁻⁶ |
| SKP2 | S72 | S-phase kinase-associated protein 2 | PRKRLKSKGSDKD | 1515.0 | 1514.0 | 3.4×10⁻⁶ |
| LPHN3 | T816 | adhesion G protein-coupled receptor L3 | YSKRTMTGYWSTQ | 1608.9 | 1607.8 | 3.7×10⁻⁶ |
| CMTM4 | T208 | MARVEL transmembrane 4 | EIQRLDTFSYSTN | 1573.8 | N/A | 3.9×10⁻⁶ |
| PITPNM2 | S668 | phosphatidylinositol transfer protein 2 | PRKRSDSSTYELD | 1553.8 | N/A | 5.2×10⁻⁶ |
| pAkt1^T308 | | | | | | |
| CYSLTR1 | T308 | cysteinyl leukotriene receptor 1 | FRKRLSTFRKHSL | 1676.2 | 1675.1 | 1.8×10⁻⁷ |
| GRAMD1C | T515 | GRAM domain containing 1C | LRRRRRTFNRTAE | 1732.1 | 1731.1 | 5.5×10⁻⁷ |
| KCNH2 | S890 | potassium channel H2 | QRKRKLSFRRRTD | 1747.2 | 1746.2 | 1.4×10⁻⁶ |
| CDCA7L | S321 | cell division cycle-associated 7-like | RRHRISSFRPVED | 1654.9 | 1654.0 | 1.7×10⁻⁶ |
| FRYL | T1959 | protein furry homolog-like | DRRRSNTLDIMDG | 1548.9 | 1547.8 | 1.8×10⁻⁶ |
| ASH1L | S1226 | histone-lysine N-methyltransferase | QKKRRHSFEHVSL | 1651.9 | 1651.0 | 1.9×10⁻⁶ |
| SRRM4 | S125 | serine/arginine repetitive matrix protein 4 | KRRRSSSYSPSPV | 1506.8 | 1505.9 | 2.6×10⁻⁶ |
| SCN2A | S687 | sodium channel protein type 2 alpha | RKRRSSSYHVSMD | 1608.9 | 1607.9 | 2.6×10⁻⁶ |
| PRM2 | S59 | protamine-2 | YRRRHCSRRRLHR | 1852.2 | N/A | 3.3×10⁻⁶ |
| SEMA4G | S713 | semaphorin-4G | GRRRKYSLGRASR | 1563.1 | N/A | 3.4×10⁻⁶ |
| ppAkt1^T308,S473 | | | | | | |
| KCNH2 | S890 | potassium channel H2 | QRKRKLSFRRRTD | 1747.2 | 1729.1 | 7.3×10⁻⁷ |
| CYSLTR1 | T308 | cysteinyl leukotriene receptor 1 | FRKRLSTFRKHSL | 1676.2 | 1675.1 | 7.3×10⁻⁷ |
| GRAMD1C | T515 | GRAM domain containing 1C | LRRRRRTFNRTAE | 1732.1 | 1731.0 | 1.1×10⁻⁶ |
| SASH1 | S407 | SAM and SH3 domain-containing | SHGRTCSFGGFDL | 1383.7 | 1382.7 | 3.1×10⁻⁶ |
| RAB11FIP2 | S277 | Rab11 family-interacting protein 2 | PHRRTLSFDTSKM | 1575.9 | 1574.9 | 4.4×10⁻⁶ |
| PCNXL3 | S505 | pecanex-like 3 | THARVLSMDGAGG | 1271.6 | 1270.7 | 5.0×10⁻⁶ |
| FMO2 | S241 | dimethylaniline monooxygenase | FHTRFRSMLRNVL | 1677.1 | 1676.0 | 5.1×10⁻⁶ |
| ASH1L | S1226 | histone-lysine N-methyltransferase | QKKRRHSFEHVSL | 1651.9 | 1651.0 | 5.4×10⁻⁶ |
| BTBD11 | S65 | ankyrin repeat and BTB/POZ | MHSRHNSFDTVNT | 1545.6 | N/A | 5.7×10⁻⁶ |
| TRAPPC1 | S132 | trafficking protein particle complex 1 | FRSRLDSYVRSLP | 1596.0 | N/A | 5.9×10⁻⁶ |
| Known substrate | | | | | | |
| GSK3B S9A | N/A | Glycogen synthase kinase-3β (S9A) | SGRPRTTAFAESCKP | 1522.7 | N/A | N/A |
| GSK3B | S9 | Glycogen synthase kinase-3β | SGRPRTTSFAESCKP | 1538.7 | N/A | 1.1×10⁻⁴ |

The ~18 Da discrepancy between the theoretical (1747.2 Da) and observed (1729.1 Da) masses of the ppAkt1^T308,S473 peptide QRKRKLSFRRRTD can be explained by a phenomenon called mass-spec ionization induced thermal decomposition (TD) (53). TD can result in the dehydration (mass – 18 Da) of peptide molecules in cases where the peptide contains a C-terminal aspartic acid as is the case here. Due to limitations in the capacity of samples that could be analyzed by the AB Sciex 5800 TOF/TOF System, the potential peptide substrates corresponding to CMTM4, PITPNM2, PRM2, SEMA4G, BTBD11, TRAPPC1, GSK3B and GSK3B(S9A) were not analyzed.

## 1.2 Materials and Methods

### 1.2.1 Plasmids and bacterial strains.

Previously, our lab designed two expression plasmids capable of producing differentially phosphorylated forms of recombinant human Akt1 from *E. coli*. Briefly, the first expression plasmid contained a codon-optimized, PH domain–deficient, 6xHis-tagged human *AKT1* gene (residues 109-472, 45 kDa) (Δ*PH-AKT1*), which was synthesized by ATUM (Newark, CA, USA). The Akt1 gene was subcloned into an isopropyl β-D-1-thiogalactopyranoside (IPTG)–inducible T7lac promoter–driven pCDF-Duet1 vector with CloDF13-derived CDF replicon and streptomycin/spectinomycin resistance (pCDF-Duet1-ΔPH-Akt1) (15,16). In the absence of the genetic code expansion system (see below) the pCDF-Duet1-ΔPH-Akt1 vector causes *E. coli* to express an unphosphorylated and inactive Akt1. The *PDK1* gene was cloned at the second multi-cloning site (MSC2) (*Nde*I/*Kpn*I) in pCDF-Duet1-ΔPHAkt1 to create the second expression plasmid: pCDF-Duet1-ΔPHAkt1-PDK1. The human *PDK1* was purchased from the Harvard PlasmidID repository service (plasmid ID: HsCD00001584; Boston, MA, USA).

In the absence of the genetic code expansion system (Figure 1.2), expression of the pCDF-Duet1-ΔPHAkt1-PDK1 vector in *E. coli* leads to production of a mono-phosphorylated Akt1 (pAkt1$^{T308}$). The genetic code expansion system for phosphoserine (pSer) is encoded on the pDS-pSer2 plasmid (15,54,55), which contains 5 copies of tRNA$^{Sep}$ (56), phosphoseryl-tRNA synthetase (SepRS9), and elongation factor-Tu mutant (EFSep21) (57).

Incorporation of pSer also required site-directed mutagenesis of the Ser473 codon to TAG in the pCDF-Duet1-ΔPHAkt1 and pCDF-Duet1-ΔPHAkt1-PDK1 constructs to generate the following constructs (16): pCDF-Duet1-ΔPHAkt1-473TAG and pCDF-Duet1-ΔPHAkt1-PDK1-473TAG which produce the mono-phosphorylated Akt1 phospho-form (pAkt1$^{S473}$) and dual-phosphorylated (ppAkt1$^{T308,S473}$) Akt1 phospho-forms, respectively. DNA sequencing services from the London Regional Genomics Centre (London, ON, Canada) and Genewiz (Cambridge, MA, USA) were used to verify the sequences of all cloned plasmids.

### 1.2.2 Protein production in *E. coli.*

*E. coli* strain BL21(DE3) (ThermoFisher Scientific, Waltham, MA, USA) was used to express all Akt1 protein variants (Figure 1.2). The pDS-pSer2 plasmid (54) was used to incorporate pSer in response to a UAG codon at position 473 of Δ*PHAkt1* in pCDF-Duet1-ΔPHAkt1 or pCDF-Duet1-ΔPHAkt1-PDK1.

To produce un-phosphorylated Akt1, pCDF-Duet1-ΔPHAkt1 (containing Δ*PHAkt1* at MCS 1) was transformed into *E. coli* BL21(DE3) and plated on LB agar plates with 50 µg/ml streptomycin. To produce pAkt1$^{T308}$, pCDF-Duet1-ΔPHAkt1-PDK1 (containing Δ*PHAkt1* at MCS 1 and *PDK1* at MCS 2) was transformed into *E. coli* BL21(DE3) and plated on LB agar plates with 50 µg/ml streptomycin. To produce pAkt1$^{S473}$ and ppAkt1$^{T308,S473}$, pCDF-Duet1-ΔPHAkt1-473TAG or pCDF-Duet1-ΔPHAkt1-473TAG-PDK1, respectively, was co-transformed with pDS-pSer2 into *E. coli* Bl21(DE3) and plated on LB (LB) agar plates with 25 µg/ml kanamycin and 50 µg/ml streptomycin. A schematic representation of the production of each Akt1 phospho-form is shown in Figure 1.2. In all cases, a single colony was used to inoculate 50 ml of LB (with 50 µg/ml streptomycin and, if needed, 25 µg/ml kanamycin), which was grown, shaking, overnight at 37ºC. From this starter culture, a 10 ml inoculum was added to 1 l of LB with antibiotics (as above) and, for pSer473-containing variants only, *O*-phospho-L-serine (Sigma Aldrich, Oakville, ON, Canada) was added to a final concentration of 2.5 mM. The cultures were grown at 37ºC until $OD_{600} = 0.6$, at which point, for pSer473-containing variants only, 2.5 mM of additional pSer was added to the culture. Protein expression was induced by adding 300 µM of IPTG at $OD_{600} = 0.8$. Cultures were then incubated at 16ºC for 18 h. Cells were grown and pelleted at $5000 \times g$ and stored at -80ºC until further analysis.

### 1.2.3 Protein production and purification.

Nickel (Ni+) affinity column chromatography was used to purify the 6xHis-tagged Akt1 variants. *E. coli* cell pellets containing recombinant Akt1 variants were resuspended in lysis buffer (20 mM HEPES pH 7.0, 150 mM NaCl, 3 mM 2.2) β-mercaptoethanol, 3 mM DTT, 10 mM imidazole, 1 mM $Na_3VO_4$, 5 mM NaF, one cOmplete mini tablet (EDTA-free

protease inhibitor mixture, Roche Applied Science, Millipore Sigma ID: 11697498001) and 1 mM phenylmethylsulfonyl fluoride were added to the cell suspension at 10 ml/g of cell pellet. Resuspended cell pellets were treated with lysozyme (1 mg/ml) for 20 min, shaking at 4°C, and subsequently lysed using a French pressure cell press (American Instrument Co. Inc.) at 1000 psi.

Cell lysates were centrifuged at 38,000 × g for 1 h at 4°C. The supernatant was collected and filtered through a 1.2 µm filter, then mixed with 0.5 ml of Ni-nitrilotriacetic acid (Ni-NTA) resin (Thermo-Fisher Scientific). Prior to the addition of the Ni-NTA resin to the cell lysate supernatant, the resin was pre-equilibrated with lysis buffer for 1 hr. The Ni-NTA resin-supernatant mixture was shaken gently on a rocker at 4°C for 1 hr to allow the resin to bind to the 6xHis-tagged Akt1.

The Ni-NTA resin-supernatant mixture was loaded into the Ni affinity column and washed thoroughly first with 200 ml of wash buffer A (20 mM HEPES pH 7.0, 150 mM NaCl, 3 mM β-mercaptoethanol, 3 mM Dithiothreitol, 1 mM $Na_3VO_4$, 5mM NaF and 15 mM imidazole), followed by 100 ml of wash buffer B (20 mM HEPES pH 7.0, 150 mM NaCl, 3 mM β-mercaptoethanol, 3mM DTT, 1 mM $Na_3VO_4$, mM NaF and 20 mM imidazole). The 6xHis-tagged Akt1 proteins were then eluted by washing the Ni-NTA resin with 25 ml of elution buffer (20 mM HEPES pH 7.0, 150 mM NaCl, 3 mM β-mercaptoethanol, 3mM dithiothreitol, 1 mM $Na_3VO_4$, 5 mM NaF and 100 mM imidazole). The elution samples were then added to dialysis tubing (Sigma Aldrich, D6191) and dialysed overnight in 2 L of imidazole-free lysis buffer to remove excess imidazole. Finally, the Akt1-containing elution samples were concentrated down to ~1 ml using Pierce Protein concentration PES centrifuge tubes with 30K MWCO (Thermo Scientific, 88529) and protein concentration of each sample was determined using a Bradford assay.

### 1.2.4 Synthesis of potential peptide substrates.

The peptide substrates were generated using a standard solid phase peptide synthesis protocol outlined previously (58). Briefly, an automatic Intavis AG peptide synthesizer was used to synthesize free peptides using 9-fluorenylmethyl-oxycarbonyl (Fmoc) chemistry. Rink amide resin (Rink-NH$_2$, Product ID: S30132, RAPP POLYMERE, Tuebingen,

Germany) was used to couple the first amino acid during the initial synthesis cycle. In each synthesis cycle, the carboxyl group of an Fmoc-protected amino acid (FmocAA-COOH) was linked to the amine group of the previous amino acid (or the Rink-NH2 amine in the case of the initial synthesis cycle) through an amide bond. All unoccupied amine groups were then blocked (acetylated) by acetic anhydride to prevent incorrect amide bond formation in subsequent cycles. Next, the Fmoc group was removed by piperidine (de-protection) to release the free amine group for coupling to the carboxyl group of the next amino acid residue.

On completion of synthesis, the free peptide-resin was incubated in a mixture containing 95% Trifluoroacetic acid (TFA), 3% Triisopropyl silane (TIPS) and 2% $H_2O$ to deprotect side chains and cleave the peptide from Rink-resin simultaneously. The cleavage mixture containing the free peptide is then drained via vacuum from the tubes containing the Rink-resin. The cleavage mixture containing the peptide products was then washed (repeated three times) with 1 ml of pre-cooled ether followed by centrifugation at 3000 × g for 3 minutes and the supernatant was removed in order to precipitate and clean the free peptide products. After allowing the peptides to dry overnight they were resuspended in MilliQ water, aliquoted and stored at -20ºC. A list of the synthetic peptide sequences as well as calculated and measured masses are in Table 1.1.

## 1.2.5 Mass spectrometry validation of peptides.

Matrix-assisted laser desorption/ionization with a time-of-flight analyzer (MALDI-TOF) mass spectrometry was used to characterize the molecular weights of the synthesized peptide substrates (1.2.4) to validate successful synthesis. Briefly, a MALDI matrix consisting of alpha-cyano-4-hydroxycinnamic acid (CHCA) was prepared as 5 mg/ml in 6 mM ammonium phosphate monobasic, 50% acetonitrile, and 0.1% trifluoroacetic acid. An aliquot of the matrix was then mixed with the synthesized peptide samples at 1:1 ratio (v/v).

MALDI-TOF mass spectrometric data were obtained using an AB Sciex 5800 TOF/TOF System, MALDI TOF (Framingham, MA, USA). Data acquisition and data processing were respectively done using a TOF/TOF Series Explorer and Data Explorer (AB Sciex). The instrument is equipped with a 349 nm Nd:YLF OptiBeam On-Axis laser.

The laser pulse rate was 400 Hz. Reflectron positive mode was used. Reflectron mode was externally calibrated at 50 ppm mass tolerance. Each mass spectrum was collected as a sum of 1000 shots. A representative spectrum of a single synthesized peptide substrate is shown in Figure 1.3 and a complete collection of the MALDI-TOF spectrum can be found in Appendix 1, Figure S1.1.

### 1.2.6 Kinase activity assays.

The activity of each Akt1 phospho-form was characterized by performing a quantitative *in vitro* kinase assays in the presence of 200 μM of a given potential substrate peptide (Table 1.1). Assays were performed in a reaction buffer consisting of 3-(N-morpholino)-propanesulfonic acid (MOPS, 25 mM, pH 7.0), 12.5 mM β-glycerolphosphate, 25 mM MgCl$_2$, ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA, 5 mM, pH 8.0), ethylenediaminetetraacetic acid (EDTA) (2 mM), ATP (0.02 mM), and 0.4 μCi (0.033 μM) γ-[$^{32}$P]-ATP.

Each reaction (30 μL reaction volume) was performed in triplicate, representing three independent enzyme reactions. The reactions were initiated with addition of enzyme and incubated at 37ºC. The t = 0 time point represents a control that lacks enzyme. Time points were taken over a 5 min time course at t = 1 min, 2 min, 3 min and 5 min. As previously (15,16), reactions were initiated by the addition of Akt1 (1.8 pmol of pAkt1$^{S473}$, 0.36 pmol of pAkt1$^{T308}$, or 0.09 pmol of ppAkt1$^{T308,S473}$). Different concentrations of each phosphor-form were used in the *in vitro* kinase assays to accommodate the varying levels of activity of each of the phosphor-forms, as was done previously (15). The enzyme concentration was adjusted to enable linear increase in reaction product during the reaction time, which allows accurate measurement of the initial velocity. To normalize for the effect of enzyme concentration, enzyme activities were later compared as apparent rates ($k_{app}$, vo/[enzyme]).

Aliquots of each reaction were quenched at each time point by spotting 3 μl of each reaction on P81 paper (59). Following spotting, the 3 μl spots were first allowed to dry, then were washed with 1% phosphoric acid (3×10 min), then washed in 95% ethanol (1×5 min), and finally allowed to air-dry completely. Incorporation of γ-[$^{32}$P] from the γ-[$^{32}$P]-

ATP into the potential substrate peptides was detected by exposing the P81 paper to a phosphor-imaging screen. The $\gamma$-[$^{32}$P]-peptide products were imaged and quantitated using a Storm 860 Molecular Imager and ImageQuant TL software (GE Healthcare, Mississauga, ON, Canada). The phosphor-images of each reaction are in Appendix 1 (Figures S1.2 and S1.3).

Akt1 activity is reported as initial velocity ($v_0$ fmol or pmol of phospho-peptide produced per min) of the reaction. $v_0$ is the slope of linear portion of the time course as before (15,16). To compare Akt1 activity values between different Akt1 phosphorylated variants, we determined apparent catalytic rate constants $k_{app}$ ($v_0$/pmol of Akt1). Negative control assays were collected as a single time point (in triplicate) 10 minutes after the initiation of the reaction. To derive a $v_0$ from the negative control spots, the quantitated level of transferred $\gamma$-[$^{32}$P]-ATP was divided by the number of minutes that the reaction had undergone (10 minutes) to result in a $v_0$ that could be compared to the treatment groups (Figure S1.4).

## 1.3 Results

### 1.3.1 Peptide synthesis and characterization.

In total, I synthesized 26 different peptides via solid phase peptide synthesis (Table 1.1). The peptides corresponded to potential Akt1 substrate motifs. These motifs ranked highest in database searches of the human phospho-proteome. The search relied on position specific score matrices that were derived from oriented peptide arrays that we had assayed previously with each Akt1 variant (14). The oriented peptide arrays reveal specific amino acid preferences at each location in the Akt1 target consensus motif. By scoring the known human proteome (44) with these matrices, both known and putative Akt1 substrates are among the top scoring hits (14). Moreover, separate PSSM scoring systems were developed for each of the different Akt1 phospho-forms, allowed us to identify potential Akt1 substrates that may display selectivity towards a specific Akt1 phospho-form.

Database searches using a separate PSSM developed for each Akt1 phospho-form (pAkt1$^{S473}$, pAkt1$^{T308}$, ppAkt1$^{S473,T308}$) were used rank and select 10 potential substrate

peptides for each phospho-form (Table 1.1). Each phosphorylated peptide motif in the database (44) was ranked based on three separate PSSM scoring systems, which predicted the likelihood that a given phosphorylated motif the in database was a substrate of the corresponding Akt1 phospho-form. We confined our peptides to those that have been identified to be phosphorylated yet the upstream is not known or not known to include Akt1.

The peptide substrates were synthesized using solid state peptide synthesis (see Methods, Section 1.2.4). Quality of the synthetic peptides was determined using Matrix-assisted laser desorption/ionization with a time-of-flight analyzer (MALDI-TOF) mass spectrometry. A representative MALDI-TOF spectrum acquired from one of the peptide substrates of is shown in Figure 1.3. The spectrum positively identified the presence of the expected peptide sequence by the presence of a sole major peak at the expected molecular weight of the peptide. For some peptides (E.g. the peptide derived from KCNH2 [QRKRKLSFRRRTD]), there was a discrepancy of 18 Da between the theoretical and observed masses of the synthesized peptide due to mass-spec ionization induced thermal decomposition (TD) (53). TD can cause peptide dehydration (mass – 18 Da) of peptide molecules in cases where the peptide contains a C-terminal aspartic acid. The MALDI-TOF spectrum for the remaining synthetic peptides are provided in Appendix 1 (Figure S1.1).

### 1.3.2 Quantifying the activity of Akt1 phospho-forms.

Each of the Akt1 phospho-forms were purified using affinity chromatography (Figure 1.4 and S1.5). Previous investigations from our lab have determined that the deletion of the PH domain from Akt1 abolishes its auto-inhibitory activity and improves the purification of Akt1 by increasing its solubility (15). Thus, the Akt1 phospho-forms used in this present study (Akt1 residues 109-472) lacked a PH domain. We previously characterized each Akt1 phospho-form biochemically and with mass spectrometry to confirm site-specific phosphorylation for each Akt1 variant produced in this manner (15,16)In order to verify the activity of the purified Akt1 phospho-forms produced in this study, *in vitro* kinase assays were conducted using the known Akt1 substrate GSK-3β (Figure S1.2 and S1.3).

Based on these data, apparent rate constants were determined for independent Akt1 preparations (Figure 1.5).

Activity levels for all three Akt1 phospho-forms prepared in this study were consistent with previously prepared Akt1 phospho-forms (15,16) (Figure 1.5). To confirm consistent activity between the Akt1 preparations, activity values from previously purified and experimentally validated (15,16) Akt1 phospho-forms were compared to the new preparation used here (Figure 1.5). These data demonstrate the consistency of the enzyme activity that is achievable using this method and supports the identity and functionality of the Akt1 preparations used below. Both sets of activity data (Figure 1.5) were obtained using the same *in vitro* kinase assay protocol (see Methods, Section 1.2.6).
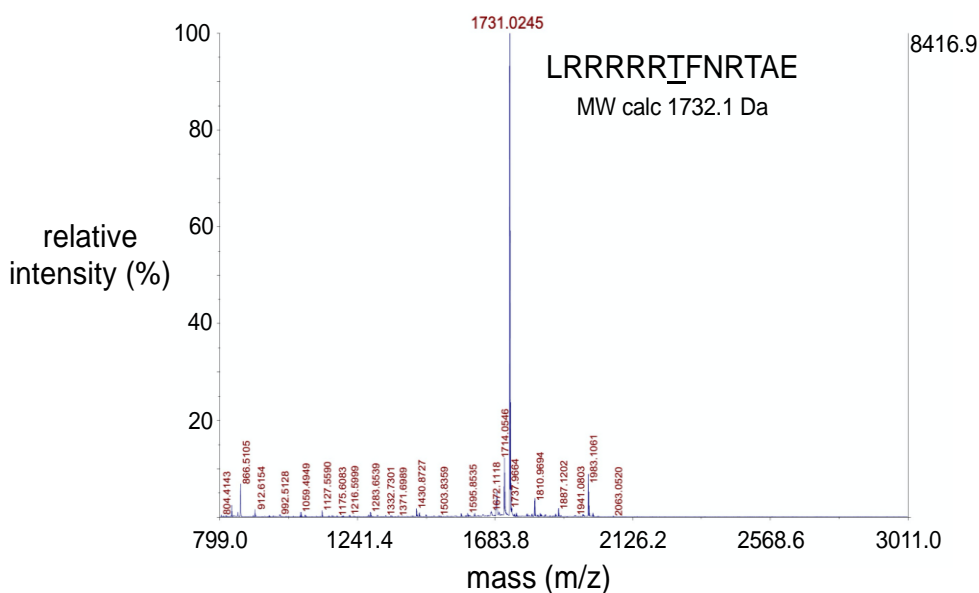


**Figure 1.3. Representative MALDI-TOF spectrum of a synthesized peptide substrate.** MALDI-TOF mass spectrometric data were obtained using an AB Sciex 5800 TOF/TOF System, MALDI TOF (Framingham, MA, USA). Data acquisition and data processing were respectively done using a TOF/TOF Series Explorer and Data Explorer (both from AB Sciex). The x-axis position of the blue peaks is measured in mass/charge number of ions (m/z). Here, the charge number of ions (z) is equal to 1 and so the (m/z) value is a measure of molecule size (m, in Daltons) which corresponds to the molecular weights of the analyzed molecules. The large peak (A) corresponding to the molecular weight of 1731.03 Da confirms the identity of the peptide substrate as well as the integrity of the

solid phase peptide synthesis protocol used to synthesize the peptides. The peptide sequence and theoretical molecular weight of the peptide is indicated.
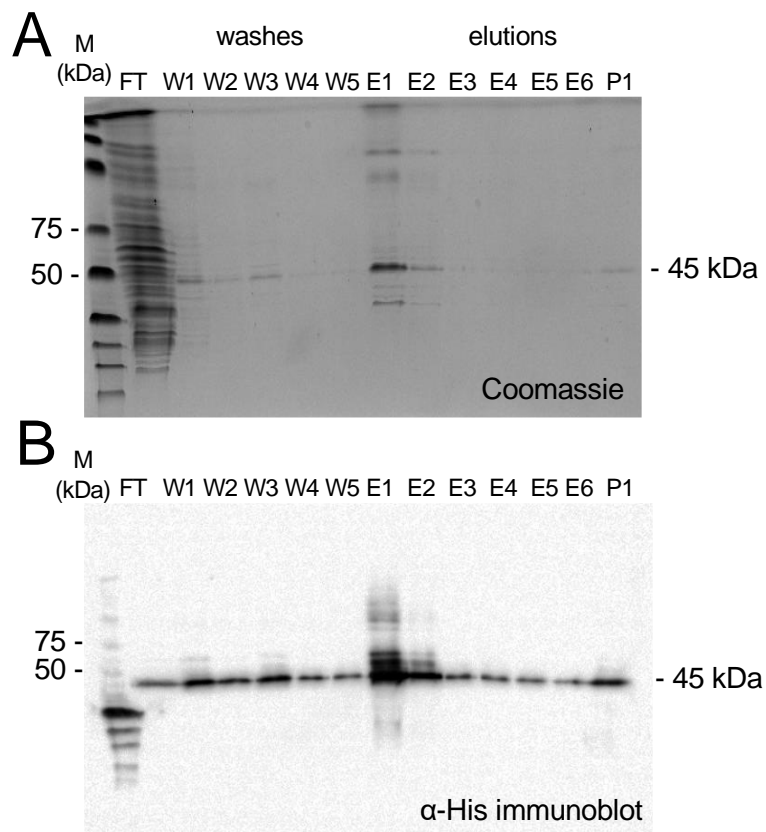


**Figure 1.4. Purification of phosphorylated Akt1 (pAkt1$^{T308}$).** (A) Coomassie stained SDS-PAGE and (B) Western blot of the flow through (FT), washes, elutions, and final pooled product (P1) of the affinity chromatography used to purify pAkt1$^{T308}$ from recombinant production in *E. coli*. The concentration of final protein product was 7.2 μM with a total yield of 32 μg/l of *E. coli* culture. ΔPH-Akt1 has an expected molecular weight of 45 kDa.
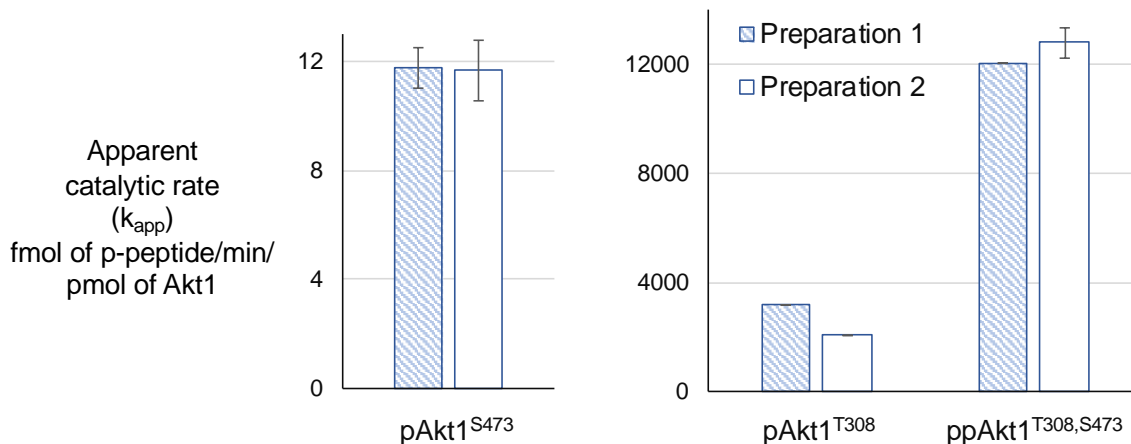
**Figure 1.5. Batch to batch consistency in Akt1 phospho-form activity across different preparations.** The apparent catalytic activity ($k_{app}$, $fmol_{substrare}/min/pmol_{enzyme}$) of the three active Akt1 phospho-forms: $pAkt1^{S473}$ (left panel), $pAkt1^{T308}$, and $ppAkt1^{T308,S473}$ (right panel) is shown here. All activity assays were conducted following the kinase assay protocol (see Methods) using 200 μM of the GSK-3β peptide (SGRPRTTSFAESCKP) as the substrate target for Akt1. The enzymes (Preparation 1) represented by the blue-striped bars were the ones purified, classified and utilized in this study (Figures 1.4, S1.5) while activity represented by the white bars (Preparation 2) were used in our previous work (15,16). Error bars represent 2 standard deviations about mean value for 3 independent enzyme reactions (n=3).

## 1.3.3 Activity of predicted Akt1 substrate peptides.

The 26 predicted Akt1 peptide substrates were then subjected to kinase assays. The activity data are based on a reaction time course during which the transfer of radiolabeled phosphate to the substrate peptide is quantitated (Figure S1.2-S1.4). The initial velocities of the enzymatic reaction ($v_0$) were determined by linear regression of the time courses. The initial velocity data (Figure S1.6) were used to compare Akt1 activity over different peptide substrates including a known Akt1 substrate GSK-3β positive control. Negative control values were derived following the same *in vitro* kinase assay protocol outlined above using a variant of the GSK-Sβ peptide that lacks the phosphorylatable Ser residue at the S9 position: GSK-3β(S9A) variant (Figure S1.4).
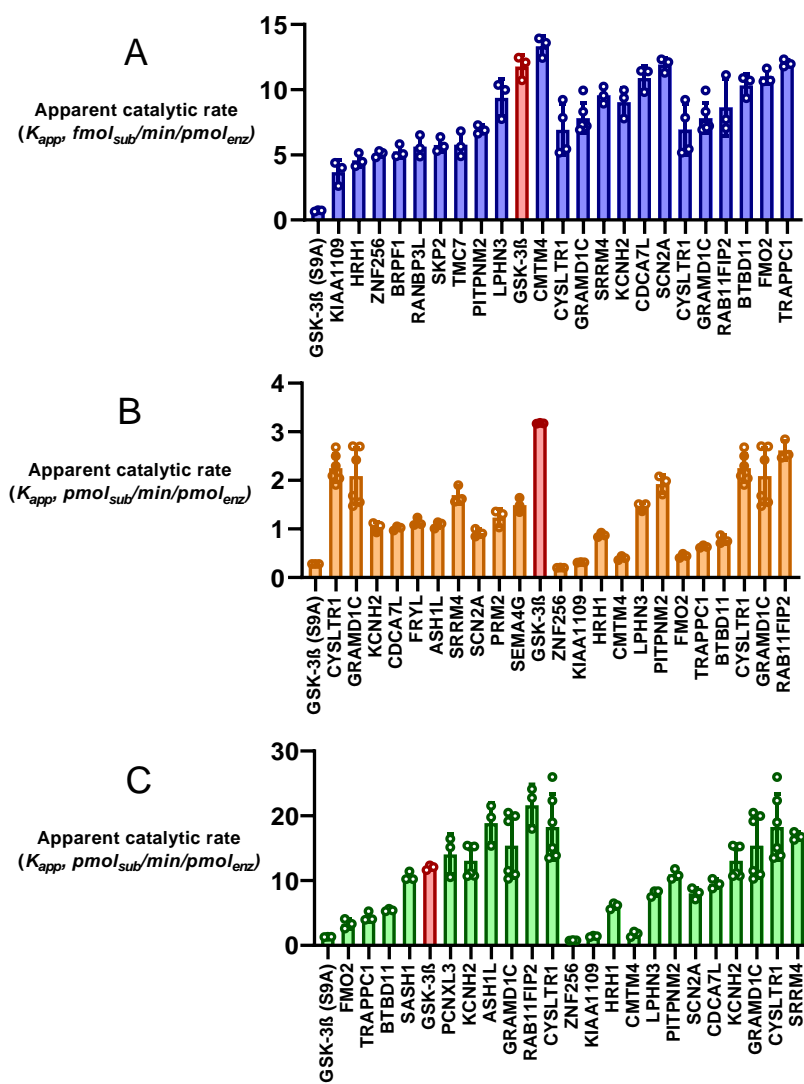
**Figure 1.6. Apparent catalytic rate (k$_{app}$) of the phosphorylation reaction catalyzed by each of the indicated Akt1 phospho-forms.** (A) pAkt1$^{S473}$, (B) pAkt1$^{T308}$, and (C) ppAkt1$^{T308,S473}$. Each bar represents the average k$_{app}$ value of 3 replicates and the value of each replicate is indicated by a closed circle. All three Akt1 phospho-forms were tested using the known Akt1 substrate GSK-3β (SGRPRTT$\underline{S}$FAESCKP, red bars) as a standard to assess the activity of the Akt1 preparations, as well as a negative control variant of GSK-3β: GSK-3β (S9A) (SGRPRTT$\underline{A}$FAESCKP, left) to obtain minimal activity values for each phospho-form. GSK-3β (S9A) contains an un-phosphorylatable alanine residue in place of the serine residue that is normally phosphorylated by Akt1. Error bars represent 2 standard deviations about mean value for 3 independent reactions (n=3).

The predicted peptide substrates were assayed using the same Akt1 phospho-form that was used to develop the PSSM scoring system that predicted the given peptide (Table 1.1). From these preliminary assays, the 3 highest and 3 lowest scoring peptides for each variant were selected for additional kinase assays conducted with the remaining two Akt1 phospho-forms. For example, from the initial assays of the 10 peptides predicted to be substrates of pAkt1$^{S473}$ (Table 1.1) the peptides with the 3 highest (CMTM4, LPHN3 and PITPNM2) and 3 lowest (KIAA1109, HRH1 and ZNF256) activity values were subsequently assayed with the remaining two Akt1 phospho-forms: pAkt1$^{T308}$ and pAkt1$^{T308,S473}$. In total, 16 of the 26 total peptides were assayed using all three Akt1 phospho-forms and the corresponding activity data ($v_0$) is outlined in Figure S1.6.

Different concentrations of each Akt1 phospho-form were used in the kinase assays (see Section 1.2.6) to permit accurate calculation of $v_0$ from the linear portion of the phosphorylation reaction time course (Figure S1.6). To account for these differences in protein concentration when comparing activity levels across the Akt1 phospho-forms, $v_0$ values were divided by the Akt1 enzyme concentration used in the given assay to arrive at the apparent catalytic rate ($k_{app}$) of each reaction (Figure 1.6).

The $k_{app}$ values obtained here agree with the previously observed trend in Akt1 activity (15,16), in which the pAkt1$^{S473}$ enzyme shows significant activity above the negative control, across all substrates tested ($p < 0.0255$ for all substrates), that is far below the activity level for the pAkt1$^{T308}$ and ppAkt1$^{T308,S473}$ enzymes. We also observed that ppAkt1$^{T308,S473}$ is more active than pAkt1$^{T308}$ on the GSK-3β peptide, as we noted before (14-16) (Figure 1.5). With respect to $k_{app}$ values, ppAkt1$^{T308,S473}$ was significantly more active ($p = 0.0053$) with the peptide derived from RAB11FIP2 than with the known Akt1 substrate GSK-3β. Pairwise comparisons of $k_{app}$ values for each peptide tested between pAkt1$^{T308}$ and ppAkt1$^{T308,S473}$ revealed that ppAkt1$^{T308,S473}$ was at least as active as pAkt1$^{T308}$ with 6 (33%) of the 18 peptides tested. Peptides representing ZNF256, CMTM4, FMO2, TRAPPC1, BTBD11 and KIAA1109 showed no significant difference in $k_{app}$ for ppAkt1 compared to pAkt1$^{T308}$. The doubly phosphorylated enzyme was more active ($p < 0.009$) with 12 (66%) of the peptides. Notably, the 6 peptides that had similar $k_{app}$ values for both pAkt1$^{T308}$ and ppAkt1$^{T308,S473}$ were also the 6 least active peptides for both phospho-forms.

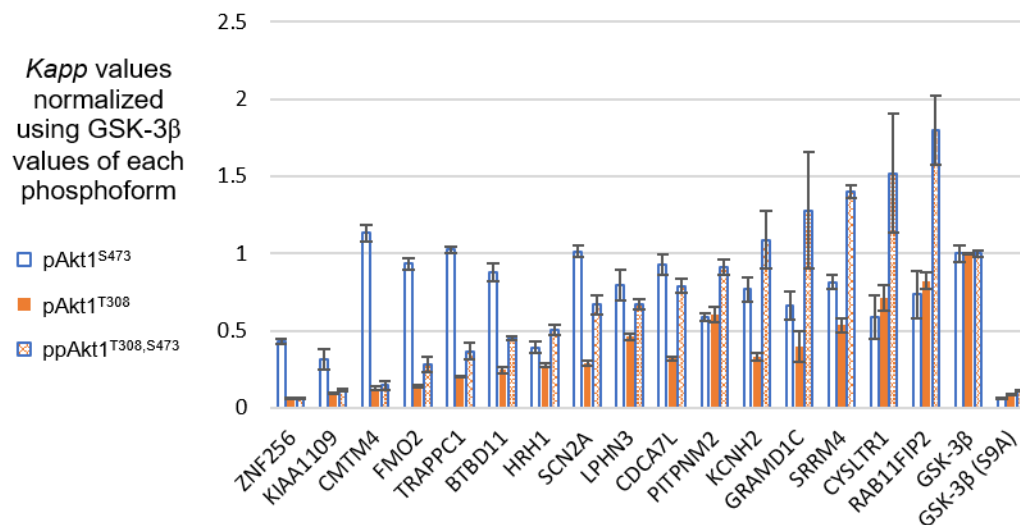**Figure 1.7. Normalized apparent catalytic rate (k_app) of the phosphorylation reaction catalyzed by different Akt1 phospho-forms on predicted Akt1 substrate peptides.** The k$_{app}$ values of each phospho-form (pAkt1$^{S473}$ (blue, open bars), pAkt1$^{T308}$ (orange, solid), and ppAkt1$^{T308,S473}$ (orange, checkered) have been normalized using each phospho-form's k$_{app}$ values for the known Akt1 substrate standard GSK-3β. Each of the peptides shown here were individually subjected to in vitro kinase assays (triplicate) with each of the three Akt1 phospho-forms. Error bars represent 2 standard deviations about mean for 3 independent reactions (n=3).

In order to estimate the relative preference or selectivity of each Akt1 phospho-form for the series of peptides, normalized $k_{app}$ values for each of the peptides tested were plotted (Figure 1.7). Normalization was based on setting the $k_{app}$ for each Akt1 variant for the positive control GSK-3β peptide to 1.0. The peptides displayed in Figure 1.7 are ordered bases on ascending $k_{app}$ values for on activity with ppAkt1$^{T308,S473}$. Figure 1.7 also contains the $k_{app}$ values of the positive control GSK-3β as well as the negative control GSK-3β (S9A) across all three Akt1 phospho-forms. Normalization of the $k_{app}$ values revealed trends in the substrate selectivity of pAkt1$^{S473}$ compared to the pAkt1$^{T308}$ and ppAkt1$^{T308,S473}$.

From the normalized $k_{app}$ values, I derived phospho-form selectivity values by dividing one Akt1 phospho-form's $k_{app}$ value for a given peptide by the average of the remaining two phospho-form's $k_{app}$ values for the same peptide. For example, to arrive at a selectivity value for pAkt1$^{S473}$ with ZNF256, I divided pAkt1$^{S473}$ normalized $k_{app}$ value for ZNF256 (0.43) by the average of pAkt1$^{T308}$ and pAkt1$^{T308,S473}$'s normalized $k_{app}$ values

for ZNF256 (0.06) to get pAkt1$^{S473}$'s selectivity value for ZNF256 (6.7 fold). A selectivity value of 1 indicates that the phospho-form is roughly as active on the given peptide as the other two phospho-forms, where as a selectivity value above or below 1 indicates that the given phospho-form is more or less active on the given peptide than the other two phospho-forms, respectively. The selectivity values are presented in Table 1.2. From these values, it is apparent that the pAkt1$^{S473}$ phospho-form has a strong selectivity (> 2-fold) towards the peptide substrates derived from ZNF256, KIAA1109, CMTM4, FMO2, TRAPPC1 and BTBD11. Moreover, the ppAkt1$^{T308,S473}$ phospho-form appears to have a strong selectivity (> 2-fold) towards a distinct set of peptide substrates derived from GRAMD1C, SRRM4, CYSLTR1, and RAB11FIP2. Taken together these results support the idea that Akt1 phosphorylation status plays an important role on determining kinase-substrate specificity.

**Table 1.2. Akt1 phospho-form selectivity values**

| gene | protein name | peptide sequence | pI | Phospho-form Relative selectivity value (x-fold) | | |
|---|---|---|---|---|---|---|
| | | | | pAkt1 $_{S473}$ | pAkt1 $_{T308}$ | ppAkt1 $_{T308,S473}$ |
| ZNF256 | zinc finger protein 256 | THQRVHTGTRPYM | 10.83 | 6.7 | 0.3 | 0.3 |
| KIAA1109 | uncharacterized protein | TSKRALSTWGPVP | 11 | 2.9 | 0.5 | 0.6 |
| CMTM4 | MARVEL transmembrane 4 | EIQRLDTFSYSTN | 4.37 | 8.3 | 0.2 | 0.2 |
| FMO2 | dimethylaniline monooxygenase | FHTRFRSMLRNVL | 12.3 | 4.4 | 0.2 | 0.5 |
| TRAPPC1 | trafficking protein particle complex 1 | FRSRLDSYVRSLP | 10.74 | 3.6 | 0.3 | 0.6 |
| BTBD11 | ankyrin repeat and BTB/POZ | MHSRHNSFDTVNT | 6.69 | 2.5 | 0.4 | 0.8 |
| HRH1 | histamine H1 receptor | TWKRLRSHSRQYV | 11.72 | 1.0 | 0.6 | 1.5 |
| SCN2A | sodium channel protein type 2 alpha | RKRRSSSYHVSMD | 10.9 | 2.1 | 0.3 | 1.0 |
| LPHN3 | adhesion G protein-coupled receptor L3 | YSKRTMTGYWSTQ | 9.7 | 1.4 | 0.6 | 1.0 |
| CDCA7L | cell division cycle-associated 7-like | RRHRISSFRPVED | 11.52 | 1.7 | 0.4 | 1.3 |
| PITPNM2 | phosphatidylinositol transfer protein 2 | PRKRSDSSTYELD | 6.54 | 0.8 | 0.8 | 1.5 |
| KCNH2 | potassium channel H2 | QRKRKLSFRRRTD | 12.18 | 1.1 | 0.4 | 2.0 |
| GRAMD1C | GRAM domain containing 1C | LRRRRRTFNRTAE | 12.3 | 0.8 | 0.4 | 2.4 |
| SRRM4 | serine/arginine repetitive matrix protein 4 | KRRRSSSYSPSPV | 11.72 | 0.8 | 0.5 | 2.1 |
| CYSLTR1 | cysteinyl leukotriene receptor 1 | FRKRLSTFRKHSL | 12.31 | 0.5 | 0.7 | 2.3 |
| RAB11FIP2 | Rab11 family-interacting protein 2 | PHRRTLSFDTSKM | 10.84 | 0.6 | 0.7 | 2.3 |

Selectivity values of each Akt1 phospho-form are presented for each of the peptides listed. Selectivity values were derived from the normalized $k_{app}$ values of each Akt1 phospho-form for each of the peptides listed.

## 1.4 Discussion

### 1.4.1 The clinical relevance of phosphorylation status-dependant substrate specificity.

Effective therapeutic targeting of Akt1 remains to be a challenge. Small molecule inhibitor-based strategies are often plagued with unwanted off target effects that lead to dose-limiting toxicities (35,60). These unwanted side effects can be attributed, at least in part, to the structural and functional similarities that are shared between Akt1 and a wide range of protein kinases in the AGC kinase family (1,2). Further complicating this issue are the Akt isozymes, Akt2 and Akt3 which likewise share exceptional similarity with Akt1. All three kinases share ~80% sequence identity in the kinase domain, 60% in the PH domain and 26% in the hinge domain that connects the kinase and PH domains across all three isozymes. Unwanted inhibition of related kinases and alternative Akt isozymes can disrupt or reverse the intended therapeutic benefits associated with targeting Akt1.

For instance, independent studies of breast cancer in transgenic mouse models have shown that in certain cellular contexts, Akt1 and Akt2 have distinct and opposing effects on tumour migration and metastatic dissemination (61,62). The suppression or deletion of Akt2 interferes with glucose homeostasis (35,60) particularly in adipose tissue (63) and can ultimately result in diabetes mellitus-like syndromes (64). Thus, small molecule inhibition of Akt1 would be ineffective in these cases, unless the inhibitor were able to selectively target Akt1. This example outlines the importance of being able to target a specific subset of Akt enzymes in order to achieve the desired therapeutic effect. Accordingly, the focus of much research has centered around the design of Akt1-isozyme specific inhibitors (65).

Yet, selectivity beyond individual Akt isozymes may prove to be necessary for future therapeutics that aim to further minimize unwanted side effects. Phosphorylation of Akt1 at position S473 has been shown previously to alter the substrate specificity of Akt1 to include members of the FOXO transcription factor family and the proline rich Akt1 substrate (PRAS40) (66). Along with the work presented here, we have elsewhere established that each differentially phosphorylated form of Akt1 (pAkt1$^{S473}$, pAkt1$^{T308}$, pAkt1$^{T308,S473}$) selectively phosphorylates different subsets of known substrates (14),

Further, our work investigating differentially phosphorylated forms of Akt1 has revealed that selective inhibition of individual Akt1 phospho-forms is possible (16).

Taken together, the results of our previous investigations into the phosphorylation-dependant activity(15), inhibitability(16) and substrate specificity(14) of Akt1 suggest that phosphorylation status constitutes a novel targetable Akt1 property that should be considered during the future development of Akt1-specific inhibitors in order to further minimize potential off-target effects. However, in order to determine which Akt1 phospho-form would be most therapeutically beneficial to target, we must first better characterize which subsets of substrates are acted on by each Akt1 phospho-form.

## 1.4.2 Comparing the genetic code expansion system to alternative methods of producing activate Akt1

In this report I utilized genetic code expansion within an *E. coli* system to incorporate the unnatural amino acid phosphoserine into Akt1 at position 473. When used in combination with co-expression of Akt1's upstream kinase PDK1, the genetic code expansion methodologies employed here allow us to produce site specifically phosphorylated variants of Akt1 that are otherwise difficult to isolate. Prior to the development of the genetic code expansion system used here, initial attempts at producing recombinant active Akt1 employed the use of phospho-mimetic mutations (67). In the case of recombinant Akt1, phospho-mimetic mutations involve replacing the serine at position 473 with glutamate as its side chain was proposed to mimic the structure and function of a phosphorylated serine residue. However, during the development of the phosphor-mimetic approach to active Akt1 production, the authors were unable to compare the activity of their phospho-mimetic S473E Akt1 construct to Akt1 that contained pSer at position 473 (67). Using the genetic code expansion system described here, our lab produced recombinant Akt1 with pSer at position 473 and was able to demonstrated that phospho-mimetic substitutions do not accurately mimic the functional role of Akt1 activation that is achieved by the phosphorylation of S473 (15). These findings call into question the validity of phospho-mimetics as a replacement for phosphorylation and also demonstrate the applicability of the Akt1 constructs produced by the genetic code expansion system described here.

**1.4.3 Validating predicted Akt1 substrates.**

Using our established method (15,16) to produce site-specifically phosphorylated variants of Akt1, we recently conducted *in vitro* kinase assays of OPALs that contained degenerate peptides representative many putative Akt1 substrates (14). By testing each separate Akt1 phospho-form (pAkt1$^{S473}$, pAkt1$^{T308}$, ppAkt1$^{T308,S473}$) using the OPAL, we were able to detect differences in substrate specificities for each of the Akt1 phospho-forms. Further, using a method derived from Scansite (68), we were able to develop a PSSM scoring system that evaluated and predicted potential substrates for each of the Akt1 phospho-forms based on the activity data generated from the *in vitro* kinase assays conducted on the OPALs. Previously, the inability to produce Akt1 in site-specifically phosphorylated forms had precluded the study of the effects of individual phosphorylation events on substrate specificity. Now, we have developed a scoring system that enables us to predict novel Akt1 substrates for the purpose of better characterizing the scope of substrates that are targeted by each Akt1 phospho-form.

Here, I synthesized and validated several potential Akt1 substrates, and some showed specificity toward individual Akt1 phospho-forms. To accomplish this, I synthesized potential peptide substrates based on the predictions derived from the peptide arrays (14). The results presented here are congruent with the findings of our initial investigations into phosphorylation-status dependent Akt1 activity (15,16) and substrate specificity (14). In particular, the general trend of Akt1 activity observed here: pAkt$^{S473}$ < pAkt1$^{T308}$ < ppAkt1$^{T308,S473}$, mirrors the trend observed across three previous studies using our methods of producing phosphorylated Akt1 variants (14-16).

Finally, Akt1 was shown to phosphorylate peptides derived from previously uncharacterized substrates *in vitro* and many of the novel peptide substrates were phosphorylated to a degree comparable to the well-established Akt1 substrate GSK-3β. The use of peptides derived from potential substrates represents only the first step towards the validation of previously uncharacterized Akt1-substrate interactions. Indeed, the results presented here validate the peptide array approach in predicting novel Akt1 substrates. However, the use of peptide substrates alone can not validate a true interaction between Akt1 and the protein from which the peptide is derived. For example, conformational

limitations may exist within the fully intact protein that preclude Akt1 phosphorylation of the peptide region that is represented by the experiments presented here. Full validation of the Akt1-substrate interaction will therefore require further testing cellular and *in vivo* model systems using fully intact substrate proteins.

### 1.4.4 Discovering novel Akt1 phospho-form-specific substrates.

The selectivity values presented in Table 1.2 showcase novel Akt1 phospho-form-specific substrates that were validated in this present study. Interestingly, none of substrates validated here were selective for the pAkt1$^{T308}$ enzyme. The lack of difference in substrate selectivity between pAkt1$^{T308}$ and ppAkt1$^{T308,S473}$ mirrors the results of our initial investigation into phosphorylation-dependant substrate specificity of Akt1 (14).

In this present study, I was able to identify a total of 11 previously unidentified Akt1 substrates that showed significant levels of phosphorylation in kinase assays with their respective Akt1 phospho-forms. In each case, significant activity was detected, i.e., $k_{app}$ substrate > $k_{app}$ negative control GSK-3β(S9A) (p < 0.0001 in all cases). Of these 11 novel substrates, 7 were relatively selective (Table 1.2 selectivity values > 2) for pAkt1$^{S473}$ (ZNF256, KIAA1109, CMTM4, FMO2, TRAPPC1, BTBD11, SCN2A) and 4 were relatively selective for ppAkt1$^{T308,S473}$ (GRAMD1C, SRRM4, CYSLTR1, RAB11FIP2) (Sequences available in Table 1.1). Relative selectivity indicates that the given substrate was highly active for the indicated Akt1 phospho-form and lowly active for the remaining two phospho-forms.

The preferred substrate consensus motif of the 11 novel Akt1 substrates observed here concur with the established consensus motif R$_{-5}$X$_{-4}$R$_{-3}$X$_{-2}$X$_{-1}$S/T$_0$ϕ$_1$ and also align with the patterns that were observed previously . The substrates that lacked an Arg residue at the -5 position were disfavoured by both pAkt1$^{T308}$ and pAkt1$^{T308,S473}$ (selectivity values < 1, Table 1.2) but were favoured by pAkt1$^{S473}$ (selectivity values > 1, Table 1.2). The same pattern was observed in all 7 of the novel peptides that were validated as pAkt1$^{S473}$-selective substrates (Tables 1.1, 1.2). Accordingly, the 4 novel peptides that were validated as being ppAkt1$^{T308,473}$-selective substrates all contained an Arg residue at the -5 position, which is likewise in agreement with our findings on the effect of phosphorylation on Akt1 substrate selectivity (14).

### 1.4.5 Novel pAkt1$^{S473}$-selective substrates.

Of the 7 novel peptides validated in this study, the peptide derived from the MARVEL transmembrane 4 (CMTM4) protein had the highest selectivity value for pAkt1$^{S473}$. Accordingly, the CMTM4 peptide also exhibited the highest $k_{app}$ of all the peptides tested with pAkt1$^{S473}$, similar to the $k_{app}$ for GSK-3β. Interestingly, the peptide derived from CMTM4 had the lowest isoelectric point of all the peptides tested, perhaps indicating the preference or higher tolerance for negatively charged substrates in pAkt1$^{S473}$. CMTM4 is a ubiquitously expressed transmembrane protein that has recently gained attention as a potential phenotypic contributor in various cancers (69-73). However, as of yet, there is minimal information on the relationship between Akt1 and CMTM4 activity, and none on the role of phosphorylation at the CMTM4 T208 site that was the source of the peptide used in this study. Topology predictions using the transmembrane topology prediction software Phobius (74) suggest that the T208 site is located within the cell cytosol and is thus would be accessible for phosphorylation by Akt1.

The function of CMTM4 with respect to its role in oncogenic phenotypes appears to be highly context dependant. In clear cell renal cell carcinomas, restoration of CMTM4 expression induces G2/M cell cycle arrest (70). In patients with hepatocellular carcinomas, the expression of CMTM4 was reduced compared to matched adjacent non-tumor tissues, and this reduced expression was associated with decreased overall survival rates (73). In colorectal adenocarcinomas cell lines, over-expression of CMTM4 (in the CMTM4-SW480 cell line) was associated with impeded cell proliferation, reduced migration and reduced levels of Akt phosphorylation. (72). In SW480 cells, CMTM4 expression was found to be inversely related to the phosphorylation of Akt1, and the tumour-suppressive functionalities observed with CMTM4 overexpression could be mimicked in the CMTM4-SW480 cell line by direct chemical inhibition of Akt using LY294002 (72). Unfortunately, due to the use of pan-Akt antibodies to detect the changes in Akt-phosphorylation, it is unclear which specific Akt1 phospho-form (or combination of phospho-form s) was affected by CMTM4 over expression.

Conversely, in some cellular contexts CMTM4 appears to promote rather than prevent oncogenic phenotypes. Along with its homolog CMTM6, CMTM4 expression has been linked to the stabilization and longevity at the cell surface of the programed death

ligand-1 (PD-L1) causing PD-L1-mediated T-cell suppression and subsequent tumor survival in multiple cell lines and primary human dendritic cells (69,75). The fact that CMTM4 has separate, seemingly divergent roles in cancer phenotypes may present similar problems with respect to the development of traditional small-molecule-based inhibitors. Just as global Akt inhibition can dys-regulated metabolic pathways (64), it may be the case that initial attempts at CMTM4-inhibition may result in unwanted side-effects stemming from the reduction of the tumour suppressor activities. Instead, approaches aimed at therapeutic modulation of CMTM4 activity should begin with an investigation into the mechanisms responsible for the CMTM4 over/under expression that has been observed in the examples of cancer. An investigation into the potential functional role of CMTM4 T208 phosphorylation is warranted and would further validation in cells and ultimately animal models, to demonstrate CMTM4 as a bona fide Akt1 target.

### 1.4.6 Novel ppAkt1$^{T308,S473}$-selective substrates.

Of all the peptides tested with ppAkt1$^{T308,S473}$, the peptide derived from RAB11FIP2 had the second highest $k_{app}$ value and was the only peptide to achieve a $k_{app}$ value that was significantly higher than the known Akt1 substrate GSK-3β (p = 0.0053). RAB11FIP2 function involves the regulation of endosomal vesicle recycling between the endosomal recycling compartment and the plasma membrane (76). RAB11FIP2 is not known to be a direct target of Akt1, though independent studies have posited functional(77) as well as spaciotemporal(76) links between the two proteins. A recent study investigating the relationship between RAB11FIP2 expression and colorectal cancer progression demonstrated that the Akt1/PI3K pathway was involved in the RAB11FIP2-mediated expression of the oncogenic Matrix metalloproteinase 7 (MMP7) (77). However, a recent investigation into subcellular localization has demonstrated that RAB11FIP2 colocalizes and indeed interacts with phosphatidylinositol-(3,4,5)-trisphosphate (PIP3) in a manner similar to Akt1 (76). Like the Akt1 PH domain, RAB11FIP2 contains a domain which promotes association with PIP3 at the cellular membrane called the C2 domain membrane. Thus, Akt1 and RAB11FIP2 may co-localize in response to the same cellular stimuli that converts PIP2 to PIP3.

The peptide derived from CysLTR1 had the highest $k_{app}$ of all the peptides tested with ppAkt1$^{T308,S473}$, on par with the $k_{app}$ value of the known ppAkt1$^{T308,S473}$ substrate: GSK-3β (p = 0.0723). CysLTR1 are cell surface receptors of the G-protein coupled receptor family that recognize a family of inflammatory lipid mediators called cysteinyl leukotrienes (78). CysLTR1 is a multifunctional mediator of allergic rhinitis, both as a mediator of bronchoconstriction and as an initiator of secondary messenger systems involved in immune cell development and immune cell response (78,79). CysLTR1 immunomodulation in allergic rhinitis was determined to involve a concurrent increase in the level of active, phosphorylated ERK1/2. Phosphorylated ERK1/2 is involved in cell adhesion, proliferation, differentiation, and cell cycle progression (80) and the inhibition of ERK1/2 has been shown to reduce allergic inflammation (81). CysLTR1 is not known to be an Akt1 target, nor have any direct interactions between CysLTR1 and Akt1 been noted in the literature. However, cross-talk between the Akt1/PI3K and ERK1/2 signalling pathways has been previously implicated to be involved in both arteriogenisis (82), as well as cell motility and invasion (83). Thus, the link between Akt1 and CysLTR1 may be direct via Akt1-mediated phosphorylation of CysLTR1's T308 residue (the motif represented here) or indirectly through the CysLTR1-ERK1/2-Akt1 axis.

**1.5 Acknowledgements.**

**1.6 References**

1.    Manning, B.D. and Cantley, L.C. (2007) AKT/PKB signaling: navigating downstream. *Cell*, **129**, 1261-1274.
2.    Manning, B.D. and Toker, A. (2017) AKT/PKB Signaling: Navigating the Network. *Cell*, **169**, 381-405.
3.    Franke, T.F. (2008) PI3K/Akt: getting it right matters. *Oncogene*, **27**, 6473-6488.
4.    Altomare, D.A. and Testa, J.R. (2005) Perturbations of the AKT signaling pathway in human cancer. *Oncogene*, **24**, 7455-7464.
5.    Dai, D.L., Martinka, M. and Li, G. (2005) Prognostic Significance of Activated Akt Expression in Melanoma: A Clinicopathologic Study of 292 Cases. *Journal of Clinical Oncology*, **23**, 1473-1482.
6.    Agarwal, E., Brattain, M.G. and Chowdhury, S. (2013) Cell survival and metastasis regulation by Akt signaling in colorectal cancer. *Cellular Signalling*, **25**, 1711-1719.
7.    Spencer, A., Yoon, S.-S., Harrison, S.J., Morris, S.R., Smith, D.A., Brigandi, R.A., Gauvin, J., Kumar, R., Opalinska, J.B. and Chen, C. (2014) The novel AKT inhibitor afuresertib shows favorable safety, pharmacokinetics, and clinical activity in multiple myeloma. *Blood*, **124**, 2190-2195.
8.    Antonelli, M., Massimino, M., Morra, I., Garrè, M.L., Gardiman, M.P., Buttarelli, F.R., Arcella, A. and Giangaspero, F. (2012) Expression of pERK and pAKT in pediatric high grade astrocytomas: Correlation with YKL40 and prognostic significance. *Neuropathology*, **32**, 133-138.
9.    Suzuki, Y., Shirai, K., Oka, K., Mobaraki, A., Yoshida, Y., Noda, S.-e., Okamoto, M., Suzuki, Y., Itoh, J., Itoh, H. *et al.* (2010) Higher pAkt Expression Predicts a Significant Worse Prognosis in Glioblastomas. *Journal of Radiation Research*, **51**, 343-348.
10.   Saxton, R.A. and Sabatini, D.M. (2017) mTOR Signaling in Growth, Metabolism, and Disease. *Cell*, **168**, 960-976.
11.   van der Vos, K.E. and Coffer, P.J. (2010) The Extending Network of FOXO Transcriptional Target Genes. *Antioxidants & Redox Signaling*, **14**, 579-592.
12.   Kaidanovich-Beilin, O. and Woodgett, J.R. (2011) GSK-3: Functional Insights from Cell Biology and Animal Models. *Front Mol Neurosci*, **4**, 40.
13.   Ebner, M., Lucic, I., Leonard, T.A. and Yudushkin, I. (2017) PI(3,4,5)P3 Engagement Restricts Akt Activity to Cellular Membranes. *Mol Cell*, **65**, 416-431.e416.
14.   Balasuriya, N., Davey, N.E., Johnson, J.L., Liu, H., Biggar, K.K., Cantley, L.C., Li, S.S.C. and O'Donoghue, P. (2020) Phosphorylation-dependent substrate selectivity of Akt1. *Journal of Biological Chemistry*, JBC/2019/012425, accepted.
15.   Balasuriya, N., Kunkel, M.T., Liu, X., Biggar, K.K., Li, S.S., Newton, A.C. and O'Donoghue, P. (2018) Genetic code expansion and live cell imaging reveal that Thr308 phosphorylation is irreplaceable and sufficient for Akt1 activity. *J Biol Chem*.
16.   Balasuriya, N., McKenna, M., Liu, X., Li, S.S.C. and O'Donoghue, P. (2018) Phosphorylation-Dependent Inhibition of Akt1. *Genes (Basel)*, **9**.

17. Guo, J., Dai, X., Laurent, B., Zheng, N., Gan, W., Zhang, J., Guo, A., Yuan, M., Liu, P., Asara, J.M. *et al.* (2019) AKT methylation by SETDB1 promotes AKT kinase activity and oncogenic functions. *Nature Cell Biology*, **21**, 226-237.

18. Wang, G., Long, J., Gao, Y., Zhang, W., Han, F., Xu, C., Sun, L., Yang, S.-C., Lan, J., Hou, Z. *et al.* (2019) SETDB1-mediated methylation of Akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis. *Nature Cell Biology*, **21**, 214-225.

19. Cederquist, C.T., Lentucci, C., Martinez-Calejman, C., Hayashi, V., Orofino, J., Guertin, D., Fried, S.K., Lee, M.J., Cardamone, M.D. and Perissi, V. (2017) Systemic insulin sensitivity is regulated by GPS2 inhibition of AKT ubiquitination and activation in adipose tissue. *Mol Metab*, **6**, 125-137.

20. Wang, Z., Liu, Y., Huang, S. and Fang, M. (2018) TRAF6 interacts with and ubiquitinates PIK3CA to enhance PI3K activation. *FEBS Lett*, **592**, 1882-1892.

21. Yang, W.L., Wang, J., Chan, C.H., Lee, S.W., Campos, A.D., Lamothe, B., Hur, L., Grabiner, B.C., Lin, X., Darnay, B.G. *et al.* (2009) The E3 ligase TRAF6 regulates Akt ubiquitination and activation. *Science*, **325**, 1134-1138.

22. Guo, J., Chakraborty, A.A., Liu, P., Gan, W., Zheng, X., Inuzuka, H., Wang, B., Zhang, J., Zhang, L., Yuan, M. *et al.* (2016) pVHL suppresses kinase activity of Akt in a proline-hydroxylation-dependent manner. *Science*, **353**, 929-932.

23. Vanhaesebroeck, B., Guillermet-Guibert, J., Graupera, M. and Bilanges, B. (2010) The emerging mechanisms of isoform-specific PI3K signalling. *Nature Reviews Molecular Cell Biology*, **11**, 329-341.

24. Parikh, C., Janakiraman, V., Wu, W.-I., Foo, C.K., Kljavin, N.M., Chaudhuri, S., Stawiski, E., Lee, B., Lin, J., Li, H. *et al.* (2012) Disruption of PH-kinase domain interactions leads to oncogenic activation of AKT in human cancers. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 19368-19373.

25. Alessi, D.R., Andjelkovic, M., Caudwell, B., Cron, P., Morrice, N., Cohen, P. and Hemmings, B.A. (1996) Mechanism of activation of protein kinase B by insulin and IGF-1. *EMBO J*, **15**, 6541-6551.

26. Stokoe, D., Stephens, L.R., Copeland, T., Gaffney, P.R.J., Reese, C.B., Painter, G.F., Holmes, A.B., McCormick, F. and Hawkins, P.T. (1997) Dual role of phosphatidylinositol-3,4,5-trisphosphate in the activation of protein kinase B. *Science*, **277**, 567-570.

27. Sarbassov, D.D., Guertin, D.A., Ali, S.M. and Sabatini, D.M. (2005) Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science*, **307**, 1098-1101.

28. Pearce, L.R., Huang, X., Boudeau, J., Pawłowski, R., Wullschleger, S., Deak, M., Ibrahim, A.F.M., Gourlay, R., Magnuson, M.A. and Alessi, D.R. (2007) Identification of Protor as a novel Rictor-binding component of mTOR complex-2. *The Biochemical journal*, **405**, 513-522.

29. Yang, Q., Inoki, K., Ikenoue, T. and Guan, K.-L. (2006) Identification of Sin1 as an essential TORC2 component required for complex formation and kinase activity. *Genes & development*, **20**, 2820-2832.

30. Huang, J. and Manning, B.D. (2009) A complex interplay between Akt, TSC2 and the two mTOR complexes. *Biochemical Society transactions*, **37**, 217-222.

31. Liao, Y. and Hung, M.-C. (2010) Physiological regulation of Akt activity and stability. *Am J Transl Res*, **2**, 19-42.

32. Kunkel, M.T., Ni, Q., Tsien, R.Y., Zhang, J. and Newton, A.C. (2005) Spatio-temporal dynamics of protein kinase B/Akt signaling revealed by a genetically encoded fluorescent reporter. *The Journal of biological chemistry*, **280**, 5581-5587.

33. Kuo, Y.C., Huang, K.Y., Yang, C.H., Yang, Y.S., Lee, W.Y. and Chiang, C.W. (2008) Regulation of phosphorylation of Thr-308 of Akt, cell proliferation, and survival by the B55alpha regulatory subunit targeting of the protein phosphatase 2A holoenzyme to Akt. *J Biol Chem*, **283**, 1882-1892.

34. Brognard, J., Sierecki, E., Gao, T. and Newton, A.C. (2007) PHLPP and a Second Isoform, PHLPP2, Differentially Attenuate the Amplitude of Akt Signaling by Regulating Distinct Akt Isoforms. *Molecular Cell*, **25**, 917-931.

35. Nitulescu, G.M., Margina, D., Juzenas, P., Peng, Q., Olaru, O.T., Saloustros, E., Fenga, C., Spandidos, D., Libra, M. and Tsatsakis, A.M. (2016) Akt inhibitors in cancer treatment: The long journey from drug discovery to clinical use (Review). *Int J Oncol*, **48**, 869-885.

36. Testas, I.J., Hu, Z.Y., Baulieuf, E.E., Robel and P. (1989) Neurosteroids: Biosynthesis of Pregnenolone and Progesterone in Primary Cultures of Rat Glial Cells*. *Endocrinology*, **125**, 2083-2091.

37. Mattmann, M.E., Stoops, S.L. and Lindsley, C.W. (2011) Inhibition of Akt with small molecules and biologics: historical perspective and current status of the patent landscape. *Expert Opin Ther Pat*, **21**, 1309-1338.

38. Cheng, J.Q., Lindsley, C.W., Cheng, G.Z., Yang, H. and Nicosia, S.V. (2005) The Akt/PKB pathway: molecular target for cancer drug discovery. *Oncogene*, **24**, 7482-7492.

39. Gonzalez, E. and McGraw, T.E. (2009) The Akt kinases: isoform specificity in metabolism and cancer. *Cell Cycle*, **8**, 2502-2508.

40. Hers, I., Vincent, E.E. and Tavare, J.M. (2011) Akt signalling in health and disease. *Cell Signal*, **23**, 1515-1527.

41. Goswami, A., Burikhanov, R., de Thonel, A., Fujita, N., Goswami, M., Zhao, Y., Eriksson, J.E., Tsuruo, T. and Rangnekar, V.M. (2005) Binding and Phosphorylation of Par-4 by Akt Is Essential for Cancer Cell Survival. *Molecular Cell*, **20**, 33-44.

42. Obata, T., Yaffe, M.B., Leparc, G.G., Piro, E.T., Maegawa, H., Kashiwagi, A., Kikkawa, R. and Cantley, L.C. (2000) Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J Biol Chem*, **275**, 36108-36115.

43. Rena, G., Guo, S., Cichy, S.C., Unterman, T.G. and Cohen, P. (1999) Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B. *J Biol Chem*, **274**, 17179-17183.

44. Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. and Zhang, B. (2004) PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, **4**, 1551-1561.

45. Vitari, A.C., Deak, M., Collins, B.J., Morrice, N., Prescott, A.R., Phelan, A., Humphreys, S. and Alessi, D.R. (2004) WNK1, the kinase mutated in an inherited

high-blood-pressure syndrome, is a novel PKB (protein kinase B)/Akt substrate. *Biochem J*, **378**, 257-268.

46.     Williams, M.R., Arthur, J.S.C., Balendran, A., van der Kaay, J., Poli, V., Cohen, P. and Alessi, D.R. (2000) The role of 3-phosphoinositide-dependent protein kinase 1 in activating AGC kinases defined in embryonic stem cells. *Current Biology*, **10**, 439-448.

47.     McManus, E.J., Collins, B.J., Ashby, P.R., Prescott, A.R., Murray-Tait, V., Armit, L.J., Arthur, J.S.C. and Alessi, D.R. (2004) The in vivo role of PtdIns(3,4,5)P3 binding to PDK1 PH domain defined by knockin mutation. *EMBO Journal*, **23**, 2071-2082.

48.     Wang, L., Xie, J. and Schultz, P.G. (2006) EXPANDING THE GENETIC CODE. *Annual Review of Biophysics and Biomolecular Structure*, **35**, 225-249.

49.     Park, H.S., Hohn, M.J., Umehara, T., Guo, L.T., Osborne, E.M., Benner, J., Noren, C.J., Rinehart, J. and Soll, D. (2011) Expanding the genetic code of Escherichia coli with phosphoserine. *Science*, **333**, 1151-1154.

50.     Hohn, M.J., Park, H.S., O'Donoghue, P., Schnitzbauer, M. and Soll, D. (2006) Emergence of the universal genetic code imprinted in an RNA record. *Proc Natl Acad Sci U S A*, **103**, 18095-18100.

51.     Fabbro, D., Batt, D., Rose, P., Schacher, B., Roberts, T.M. and Ferrari, S. (1999) Homogeneous purification of human recombinant GST-Akt/PKB from Sf9 cells. *Protein Expr Purif*, **17**, 83-88.

52.     Rodriguez, M., Li, S.S., Harper, J.W. and Songyang, Z. (2004) An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J Biol Chem*, **279**, 8802-8807.

53.     Liu, C., Topchiy, E., Lehmann, T. and Basile, F. (2015) Characterization of the dehydration products due to thermal decomposition of peptides by liquid chromatography-tandem mass spectrometry. *J Mass Spectrom*, **50**, 625-632.

54.     George, S., Aguirre, J.D., Spratt, D.E., Bi, Y., Jeffery, M., Shaw, G.S. and O'Donoghue, P. (2016) Generation of phospho-ubiquitin variants by orthogonal translation reveals codon skipping. *FEBS Lett*, **590**, 1530-1542.

55.     George, S., Wang, S.M., Bi, Y., Treidlinger, M., Barber, K.R., Shaw, G.S. and O'Donoghue, P. (2017) Ubiquitin phosphorylated at Ser57 hyper-activates parkin. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **1861**, 3038-3046.

56.     Aerni, H.R., Shifman, M.A., Rogulina, S., O'Donoghue, P. and Rinehart, J. (2015) Revealing the amino acid composition of proteins within an expanded genetic code. *Nucleic Acids Res*, **43**, e8.

57.     Lee, S., Oh, S., Yang, A., Kim, J., Soll, D., Lee, D. and Park, H.S. (2013) A facile strategy for selective incorporation of phosphoserine into histones. *Angew Chem Int Ed Engl*, **52**, 5771-5775.

58.     Wei, R., Kaneko, T., Liu, X., Liu, H., Li, L., Voss, C., Liu, E., He, N. and Li, S.S.C. (2017) Interactome mapping uncovers a general role for Numb in protein kinase regulation. *Mol Cell Proteomics*.

59.     Turowec, J.P., Duncan, J.S., French, A.C., Gyenis, L., St Denis, N.A., Vilk, G. and Litchfield, D.W. (2010) Protein kinase CK2 is a constitutively active enzyme that promotes cell survival: strategies to identify CK2 substrates and manipulate its activity in mammalian cells. *Methods Enzymol*, **484**, 471-493.

60. Nitulescu, G.M., Van De Venter, M., Nitulescu, G., Ungurianu, A., Juzenas, P., Peng, Q., Olaru, O.T., Grădinaru, D., Tsatsakis, A., Tsoukalas, D. *et al.* (2018) The Akt pathway in oncology therapy and beyond (Review). *International journal of oncology*, **53**, 2319-2331.

61. Dillon, R.L., Marcotte, R., Hennessy, B.T., Woodgett, J.R., Mills, G.B. and Muller, W.J. (2009) Akt1 and Akt2 Play Distinct Roles in the Initiation and Metastatic Phases of Mammary Tumor Progression. *Cancer Research*, **69**, 5057.

62. Hutchinson, J.N., Jin, J., Cardiff, R.D., Woodgett, J.R. and Muller, W.J. (2004) Activation of Akt-1 (PKB-α) Can Accelerate ErbB-2-Mediated Mammary Tumorigenesis but Suppresses Tumor Invasion. *Cancer Research*, **64**, 3171.

63. Czech, M.P., Tencerova, M., Pedersen, D.J. and Aouadi, M. (2013) Insulin signalling mechanisms for triacylglycerol storage. *Diabetologia*, **56**, 949-964.

64. Cho, H., Mu, J., Kim, J.K., Thorvaldsen, J.L., Chu, Q., Crenshaw, E.B., 3rd, Kaestner, K.H., Bartolomei, M.S., Shulman, G.I. and Birnbaum, M.J. (2001) Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). *Science*, **292**, 1728-1731.

65. Cherrin, C., Haskell, K., Howell, B., Jones, R., Leander, K., Robinson, R., Watkins, A., Bilodeau, M., Hoffman, J., Sanderson, P. *et al.* (2010) An allosteric Akt inhibitor effectively blocks Akt signaling and tumor growth with only transient effects on glucose and insulin levels in vivo. *Cancer Biol Ther*, **9**, 493-503.

66. Guertin, D.A., Stevens, D.M., Thoreen, C.C., Burds, A.A., Kalaany, N.Y., Moffat, J., Brown, M., Fitzgerald, K.J. and Sabatini, D.M. (2006) Ablation in mice of the mTORC components raptor, rictor, or mLST8 reveals that mTORC2 is required for signaling to Akt-FOXO and PKCalpha, but not S6K1. *Dev Cell*, **11**, 859-871.

67. Klein, S., Geiger, T., Linchevski, I., Lebendiker, M., Itkin, A., Assayag, K. and Levitzki, A. (2005) Expression and purification of active PKB kinase from Escherichia coli. *Protein Expr Purif*, **41**, 162-169.

68. Obenauer, J.C., Cantley, L.C. and Yaffe, M.B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res*, **31**, 3635-3641.

69. Mezzadra, R., Sun, C., Jae, L.T., Gomez-Eerland, R., de Vries, E., Wu, W., Logtenberg, M.E.W., Slagter, M., Rozeman, E.A., Hofland, I. *et al.* (2017) Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. *Nature*, **549**, 106-110.

70. Li, T., Cheng, Y., Wang, P., Wang, W., Hu, F., Mo, X., Lv, H., Xu, T. and Han, W. (2015) CMTM4 is frequently downregulated and functions as a tumour suppressor in clear cell renal cell carcinoma. *J Exp Clin Cancer Res*, **34**, 122-122.

71. Chrifi, I., Louzao-Martinez, L., Brandt, M.M., van Dijk, C.G.M., Burgisser, P.E., Zhu, C., Kros, J.M., Verhaar, M.C., Duncker, D.J. and Cheng, C. (2019) CMTM4 regulates angiogenesis by promoting cell surface recycling of VE-cadherin to endothelial adherens junctions. *Angiogenesis*, **22**, 75-93.

72. Xue, H., Li, T., Wang, P., Mo, X., Zhang, H., Ding, S., Ma, D., Lv, W., Zhang, J. and Han, W. (2019) CMTM4 inhibits cell proliferation and migration via AKT,

ERK1/2, and STAT3 pathway in colorectal cancer. *Acta Biochim Biophys Sin (Shanghai)*, **51**, 915-924.

73.   Bei, C., Zhang, Y., Wei, R., Zhu, X., Wang, Z., Zeng, W., Chen, Q. and Tan, S. (2017) Clinical significance of CMTM4 expression in hepatocellular carcinoma. *Onco Targets Ther*, **10**, 5439-5443.

74.   Kall, L., Krogh, A. and Sonnhammer, E.L. (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*, **338**, 1027-1036.

75.   Burr, M.L., Sparbier, C.E., Chan, Y.C., Williamson, J.C., Woods, K., Beavis, P.A., Lam, E.Y.N., Henderson, M.A., Bell, C.C., Stolzenburg, S. *et al.* (2017) CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature*, **549**, 101-105.

76.   Lindsay, A.J. and McCaffrey, M.W. (2004) The C2 domains of the class I Rab11 family of interacting proteins target recycling vesicles to the plasma membrane. *Journal of Cell Science*, **117**, 4365.

77.   Xu, C.-l., Wang, J.-z., Xia, X.-p., Pan, C.-w., Shao, X.-x., Xia, S.-l., Yang, S.-x. and Zheng, B. (2016) Rab11-FIP2 promotes colorectal cancer migration and invasion by regulating PI3K/AKT/MMP7 signaling pathway. *Biochemical and Biophysical Research Communications*, **470**, 397-404.

78.   Peters-Golden, M., Gleason, M.M. and Togias, A. (2006) Cysteinyl leukotrienes: multi-functional mediators in allergic rhinitis. *Clin Exp Allergy*, **36**, 689-703.

79.   Tsai, M.J., Wu, P.H., Sheu, C.C., Hsu, Y.L., Chang, W.A., Hung, J.Y., Yang, C.J., Yang, Y.H., Kuo, P.L. and Huang, M.S. (2016) Cysteinyl Leukotriene Receptor Antagonists Decrease Cancer Risk in Asthma Patients. *Sci Rep*, **6**, 23979.

80.   Figueroa, D.J., Breyer, R.M., Defoe, S.K., Kargman, S., Daugherty, B.L., Waldburger, K., Liu, Q., Clements, M., Zeng, Z., O'Neill, G.P. *et al.* (2001) Expression of the cysteinyl leukotriene 1 receptor in normal human lung and peripheral blood leukocytes. *Am J Respir Crit Care Med*, **163**, 226-233.

81.   El-Hashim, A.Z., Renno, W.M., Raghupathy, R., Abduo, H.T., Akhtar, S. and Benter, I.F. (2012) Angiotensin-(1-7) inhibits allergic inflammation, via the MAS1 receptor, through suppression of ERK1/2- and NF-kappaB-dependent pathways. *Br J Pharmacol*, **166**, 1964-1976.

82.   Ren, B., Deng, Y., Mukhopadhyay, A., Lanahan, A.A., Zhuang, Z.W., Moodie, K.L., Mulligan-Kehoe, M.J., Byzova, T.V., Peterson, R.T. and Simons, M. (2010) ERK1/2-Akt1 crosstalk regulates arteriogenesis in mice and zebrafish. *J Clin Invest*, **120**, 1217-1228.

83.   Gentilini, D., Busacca, M., Di Francesco, S., Vignali, M., Viganò, P. and Di Blasio, A.M. (2007) PI3K/Akt And ERK1/2 signalling pathways are involved in endometrial cell migration induced by 17β-estradiol and growth factors. *Molecular Human Reproduction*, **13**, 317-322.

## 1.7 Appendix 1

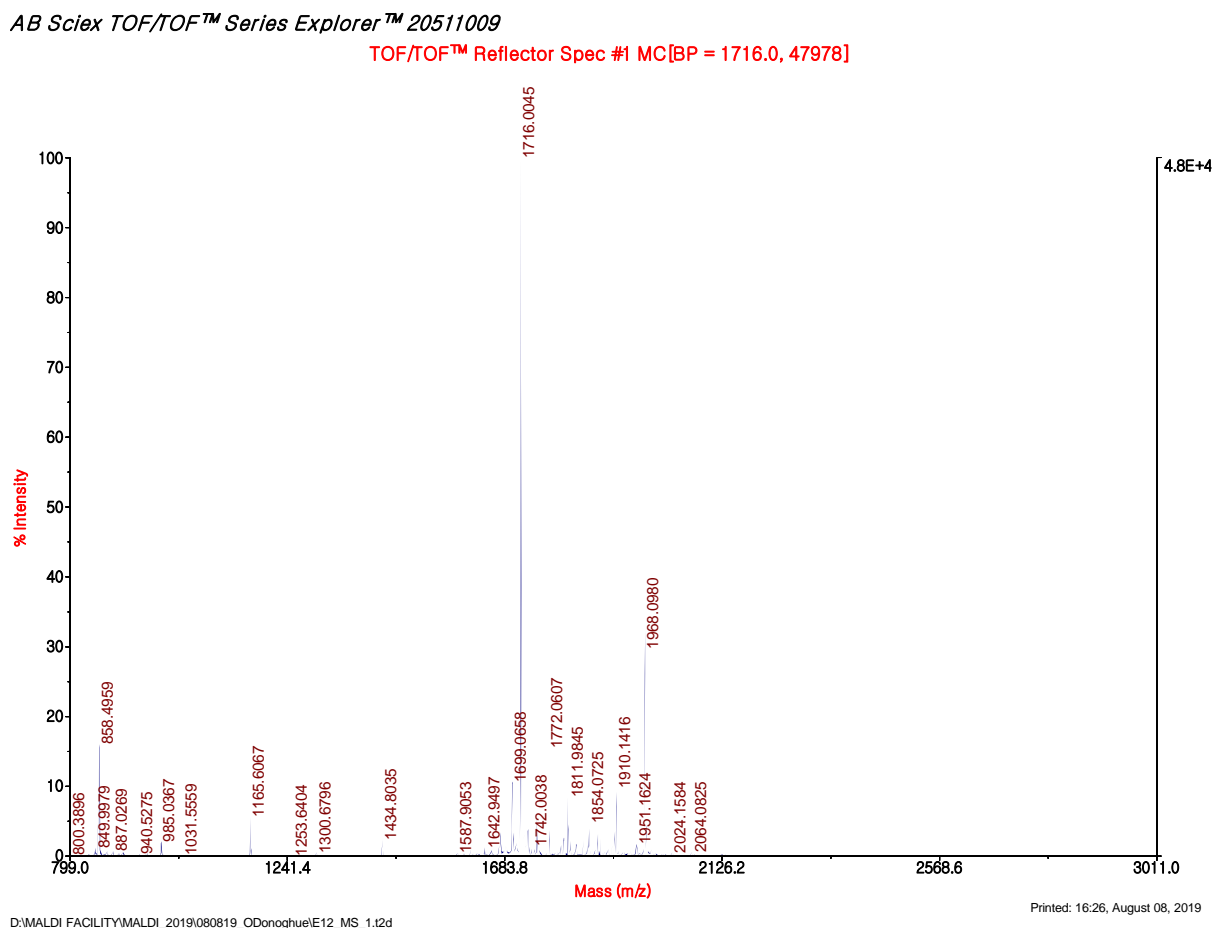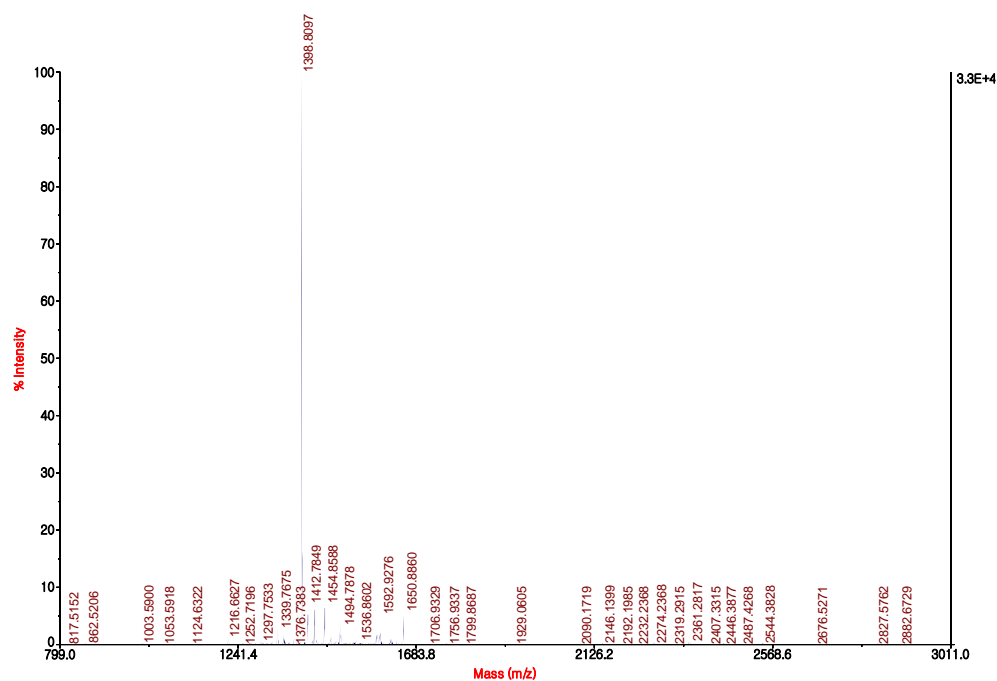## 1.7.1 Supplementary Figures



**Figure S1.1. MALDI-TOF spectrum of a synthesized peptide substrate.** MALDI-TOF mass spectrometric data were obtained using an AB Sciex 5800 TOF/TOF System, MALDI TOF (Framingham, MA, USA). Data acquisition and data processing were respectively done using a TOF/TOF Series Explorer and Data Explorer (both from AB Sciex). The x-axis scale is equivalent to mass in Daltons (Da) as the charge number of ions (z) is equal to 1. The large peak (A) corresponding to the molecular weight of 1731.03 Da confirms the identity of the peptide substrate as well as the integrity of the solid phase peptide synthesis protocol used to synthesize the peptides.

39





**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.1.** continued.

**Figure S1.2.** **Radiograph detecting the presence of γ-[³²P]-ATP in peptides phosphorylated by pAkt1^S473^ and pAkt1^T308^.** Time courses (minutes) of the in vitro kinase assay catalyzed by the indicated phospho-form of Akt1 on the indicated peptide substrate. Each spot was generated by dispensing 3 µl of kinase assay reaction mixture onto P81 paper at the indicated time point after the initiation of the reaction. Assays were performed in triplicate (R1-R3). Three spots (Standard) correspond to the γ-[³²P]-ATP standards which were used to calculate the relationship between spot intensity and [γ-[³²P]-ATP].
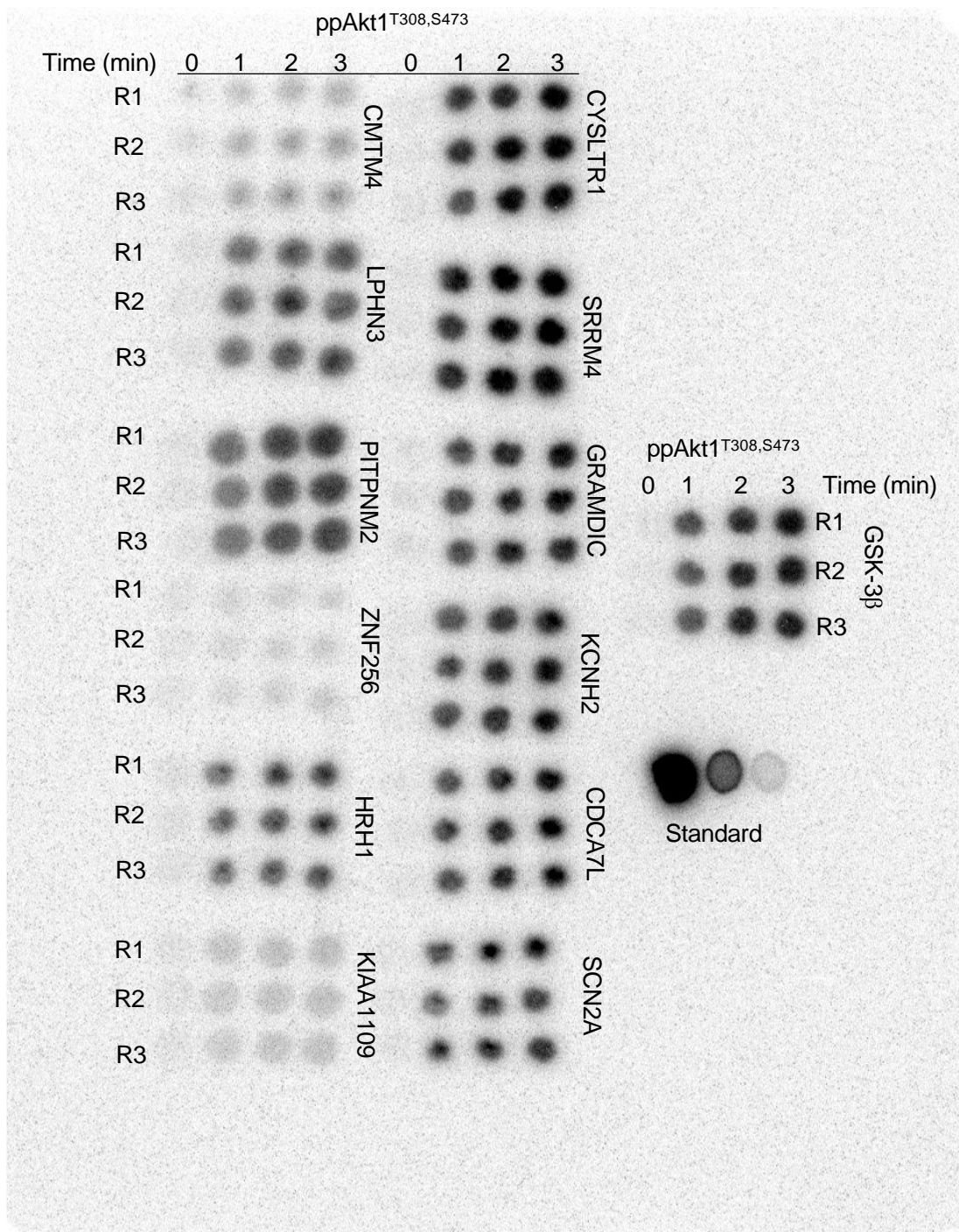
**Figure S1.3. Radiograph detecting the presence of γ-[$^{32}$P]-ATP in peptides phosphorylated by ppAkt1$^{S473,T308}$.** Time courses (minutes) of the in vitro kinase assay catalyzed by the indicated phospho-form of Akt1 on the indicated peptide substrate. Each spot was generated by dispensing 3 µl of kinase assay reaction mixture onto P81 paper at the indicated time point after the initiation of the reaction. Assays were performed in triplicate (R1-R3). Three spots (Standard) correspond to the γ-[$^{32}$P]-ATP standards which were used to calculate the relationship between spot intensity and [γ-[$^{32}$P]-ATP].

**Figure S1.4. Radiographs created by Nileeka Balasuriya that detect the presence of γ-[$^{32}$P]-ATP in control peptides phosphorylated by pAkt1$^{S473}$, pAkt1$^{T308}$, and ppAkt1$^{S473,T308}$.** Spots of the in vitro kinase assay catalyzed by the indicated phospho-form of Akt1 on the indicated control peptide substrates. Assays were performed by Nileeka Balasuryia in triplicate (R1-R3). Each spot was generated by dispensing 3 µl of kinase assay reaction mixture onto P81 paper 10 minutes after the initiation of the reaction. Three spots (Standard) correspond to the γ-[$^{32}$P]-ATP standards which were used to calculate the relationship between spot intensity and [γ-[$^{32}$P]-ATP].
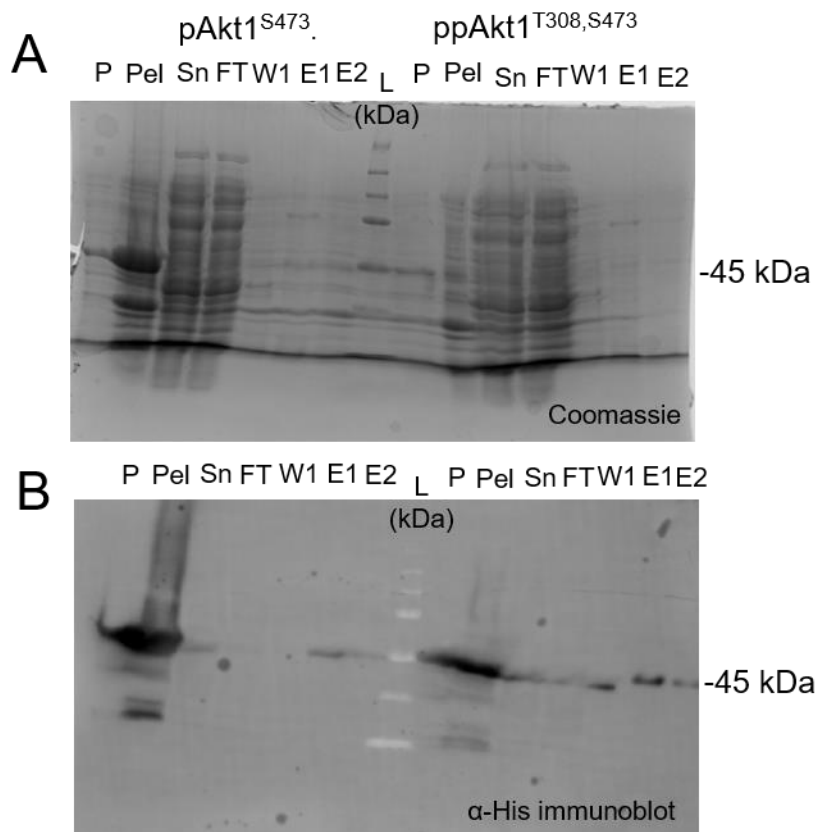
**Figure S1.5. Purification of phosphorylated Akt1 (pAkt1$^{S473}$ and ppAkt1$^{T308,S473}$).**
Coomassie stained SDS-PAGE (A) and α-His immunoblot (B) of the intermediate and final
purification products of the affinity chromatography used to purify phosphorylated Akt1
variants from recombinant production in E. coli. P denotes the final purified and
concentrated Akt1 product, Pel/Sn denote samples of the ultracentrifuge pellet/supernatant
respectively, FT denotes the initial flow-through of the purification prior to the addition of
imidazole and W1, E1 and E2 denote the washes and elution treatments used to elute Akt1
(outlined in 1.2.3). The proteins presented here were used in the in vitro kinase assays in
this study (Figure S1.2 and S1.3) and are denoted as "preparation 1" in Figure 1.5.
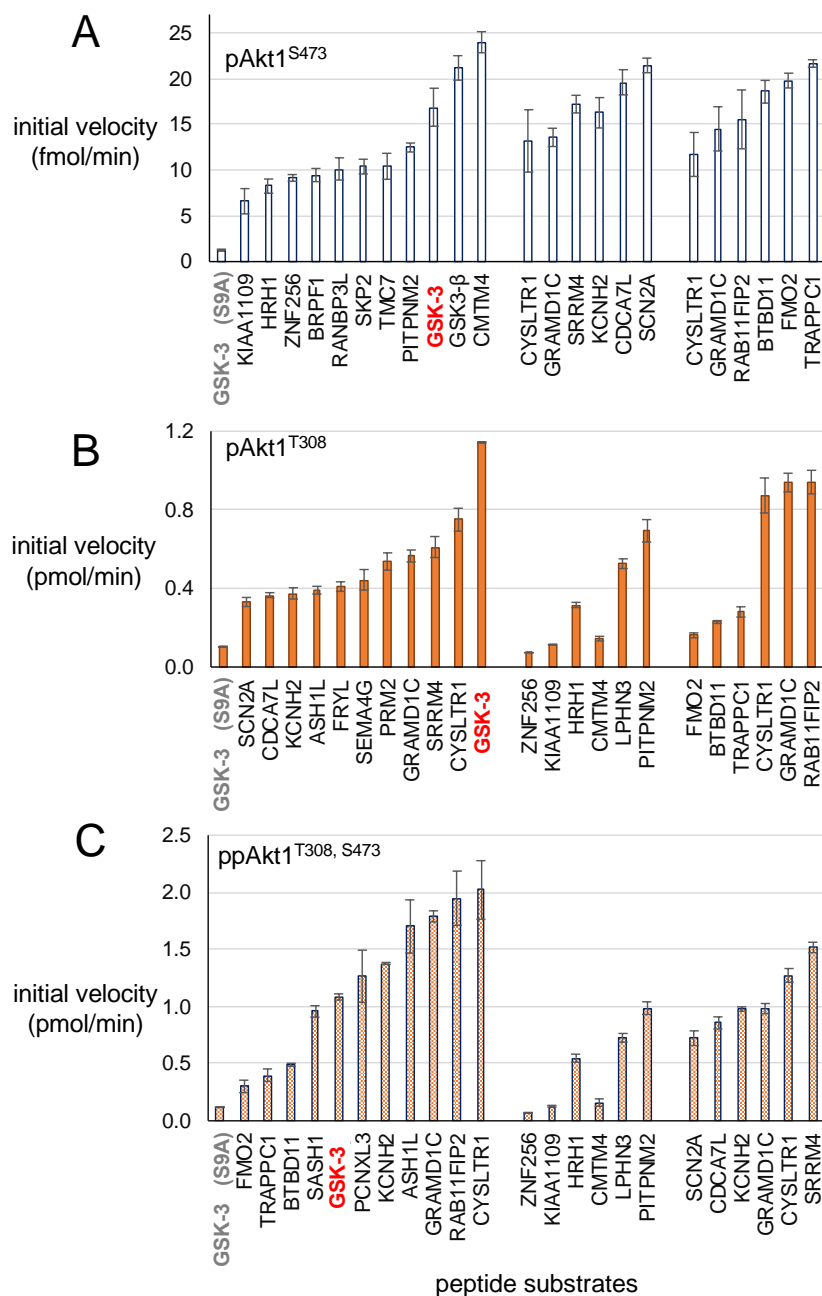
**Figure S1.6. Initial velocities of the phosphorylation reaction catalyzed by each of the indicated Akt1 phospho-forms.** pAkt1$^{S473}$ (A), pAkt1$^{T308}$ (B), ppAkt1$^{T308,S473}$ (C). Each bar represents a separate reaction containing a different peptide substrate. All three Akt1 phospho-forms were tested using the known Akt1 substrate GSK-3β (SGRPRTT$\underline{S}$FAESCKP, red bold) as a standard to assess the activity of the Akt1 preparations, as well as a negative control variant of GSK-3β: GSK-3β (S9A) (SGRPRTT$\underline{A}$FAESCKP, gray bold) to obtain minimal activity values for each phospho-form. GSK-3β (S9A) contains an un-phosphorylatable alanine residue in place of the serine residue that is normally phosphorylated by Akt1. Error bars represent 2 standard deviations about mean value for 3 independent reactions (n=3).

# CHAPTER 2 – Expanding the Genetic Code of *Bacillus subtilis*

## 2.0 Abstract

Genetic code expansion enables organisms to incorporate unnatural amino acids into proteins in response to a re-assigned codon. Commonly, the infrequently used UAG (amber) codon in *Escherichia coli* is chosen to encode unnatural or non-canonical amino acids. The co-translational incorporation of unnatural amino acids into proteins requires an orthogonal translation system consisting of an engineered tRNA that can decode the re-assigned codon, an aminoacyl synthetases to ligate the unnatural amino acids to the tRNA, and an unnatural amino acid that is not toxic and cell permeable. By inserting the recoded amber codon into the coding region of growth-essential genes, cell growth and survival becomes dependant on both the functionality of the orthogonal translation system and the provision of the otherwise unavailable unnatural amino acid. Previously, this system has been implemented as a successful means of biocontainment in *Escherichia coli*. There is yet, no such orthogonal translation system that functions in the gram-positive bacteria *Bacillus subtilis*. The work presented here represents the pioneering effort towards the development of an orthogonal translation system that functions to expand the genetic code of *B. subtilis* to include $N$ε-acetyl-L-lysine. I designed, built, and performed initial tests on an all-in-one plasmid that contains all the requisite genetic components of an orthogonal translation system along with a built-in fluorescent GFP reporter to assay UAG read-through. Stable integration of the entire orthogonal translation system in the *B. subtilis* genome was achieved.

## 2.1 Introduction

### 2.1.1 Biotechnological applications of *B. subtilis.*

*B. subtilis* is an exceptionally well-characterized gram (+) bacteria with numerous applications in both academic research and industry. *B. subtilis* has received the FDA designation of Generally Recognised As Safe (GRAS), which means that products biosynthesized by the organism are considered safe for contact and consumption by humans (1). This designation, coupled with the organism's innate ability to endure

manufacturing and storage conditions, including oxygen-rich environments, osmotic stress and heat (2,3), as well as proficiency at producing and excreting proteins and biomolecules into its environment (4), has made *B. subtilis* an important organism across a broad range of biotechnological applications.

In medicine, *B. subtilis* has garnered widespread interest as both an important probiotic (5) and as a promising bacterial vaccine candidate (6,7). Moreover, the probiotic effects of *B. subtilis* are relevant in the field of agriculture as recent studies on Broiler chickens and white shrimp have demonstrated *B. subtilis'* ability to ameliorate intestinal bacterial imbalance which promotes growth and increases the yield of harvests (8,9).

From a financial perspective, however, *B. subtilis* has an outsized role in the field of chemical and protein biomanufacturing. Nearly 900 tonnes of proteases, lipases, and amylases (active ingredients in laundry detergents) are produced annually using *B. subtilis* bio-reactors representing a yearly turnover of nearly 3 billion CAD (4). Although we are still at the advent of bacteria-based biomanufacturing, as synthetic biology applications for *B. subtilis* continue to grow, there will be increasing demands on researchers to improve the underlying technologies that allow us to use these organisms to their full potential.

## 2.1.2 Orthogonal translation systems.

The demand for bacteria with new and improved capabilities for practical industrial applications has been one of the major driving forces behind the emergence of the field of synthetic biology. In recent years, synthetic biology has grown greatly due to the invention of recombinant DNA technologies in *E. coli* (10). One such technology is the orthogonal translation system (OTS), which enables expansion of the genetic code to produce proteins with unnatural (uAA) or non-canonical amino acids (ncAA) beyond the standard 20 proteinogenic amino acids (11) (Figure 1.1).

To expand the natural translation system, it is important that the engineered system does not interfere with or compromise the integrity of the natural components. The orthogonal aaRS must only recognize and aminoacylate its cognate orthogonal tRNA with a specific or desired unnatural amino acid. Orthogonality ensures that the engineered aaRS

does not charge any of the 33 endogenous *E. coli* tRNAs with the uAA so that the uAA is not incorporated randomly throughout the proteome. Likewise, the orthogonal tRNA must not be charged with naturally occurring amino acids by any of the endogenous aaRSs present in *E. coli* so that the OTS does not needlessly incorporate one of the naturally occurring amino acids in place of an uAA at the UAG codon.

The OTS used here is composed of tRNA$^{Pyl-opt}$ (12) and an optimized version of the $N^{\varepsilon}$-acetyl-L-lysyl-tRNA synthetase (AcKRS) (13) (Figure 2.2). AcKRS and tRNA$^{Pyl-opt}$ originate from the pyrrolysyl-tRNA synthetase (PylRS) and tRNA$^{Pyl}_{CUA}$ pair derived from Methanosarcina species (14,15). The PylRS system was first used to expand the genetic code of *E. coli* in 2004 (15). To generate AcKRS, a crystal structure of PylRS in complex with pyrrolysine (PDB accession code: 2Q7H, (16)) was used to determine key PylRS residues that were within 6Å of pyrolysine's pyrroline ring. Through mutation of these residues, a variant (AcKRS) of PylRS was generated (13) that preferentially bound, and charged tRNA$_{CUA}$ with, $N\varepsilon$-acetyl-L-lysine (AcK) and to a lesser degree other similar uAAs, including $N^{\varepsilon}$-Boc-L-lysine (BocLys).

The variant of AckRS used here, contains a further 4 mutations (V31I, T56P, H62Y, and A100E) which were derived by phage assisted continuous evolution experiment that improved the yield of proteins containing unnatural amino acids by ~10-fold (17). AcKRS in the experiments presented here refers to the use of this optimized version of AcKRS. tRNA$^{Pyl-opt}$ is a rationally evolved variant of tRNA$^{Pyl}_{CUA}$ that has a 5-fold increased capacity for AcK incorporation compared to the wild type tRNA$^{Pyl}$ (12). tRNA$^{Pyl-opt}$

includes mutations in the acceptor and T stems of the tRNA that improved binding to *E. coli* elongation factor-Tu (12).
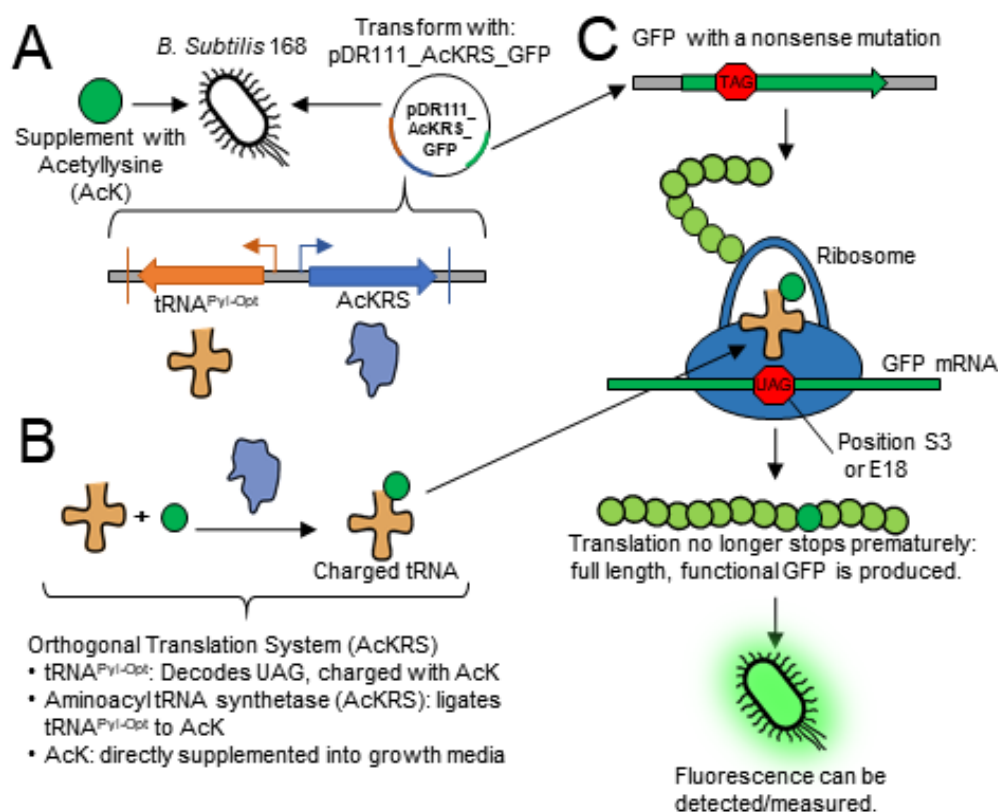


**Figure 2.1. Schematic representation of AcKRS and tRNA<sup>Pyl-opt</sup> mediated decoding of UAG codons in GFP mRNA.** A) The components of the AcKRS/tRNA<sup>Pyl-opt</sup> OTS within the pDR111_AcKRS_GFP plasmid are presented. Three key genetic components; the aminoacyl- tRNA Synthetase: AcKRS (flanked by *sarA* promoter and terminator, blue), its orthogonal cognate tRNA: tRNA<sup>Pyl-opt</sup> (flanked by *SerT* promoter and terminator, orange), and the GFP reporter system (shown in C). The unnatural amino acid *N*ε-acetyl-L-lysine (AcK) is not produced endogenously by *B. subtilis* nor are the genetic components required to synthesize *de novo* AcK present in pDR111_AcKRS_GFP. Instead, AcK must be directly supplemented (10 mM) into the growth media. B) The role of each OTS component are outlined: AcKRS charges tRNA<sup>Pyl-opt</sup> with AcK. C) The OTS functions to decode nonsense UAG codons. When the ribosome is presented with a premature UAG codon that would normally signal the end translation, it is instead supplied with AcK-tRNA<sup>Pyl-opt</sup>, which is incorporated into the growing peptide chain allowing translation elongation to continue. In the system employed here, expression of the full-length GFP protein requires a functioning OTS to decode an in-frame UAG codon. By detecting fluorescence, we can measure the degree to which the OTS is able to decode the UAG codon in the GFP mRNA and rescue the expression of full-length, active GFP.
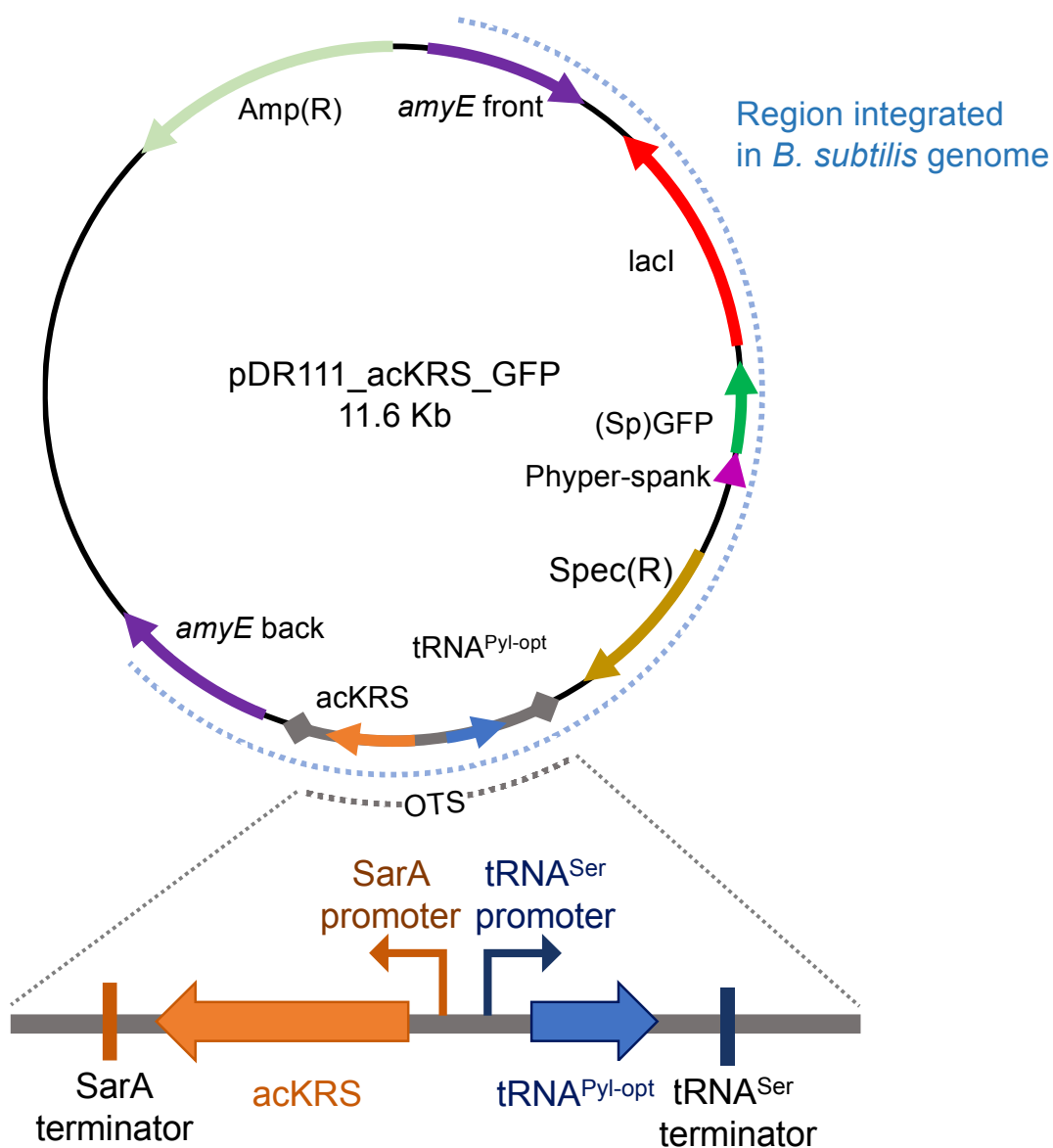
**Figure 2.2. OTS genome integration plasmid schematic.** The region flanked by *amyE* front and *amyE* back, outlined by the blue-dashed line, is integrated into the genomic *amyE* locus (BSU_03040) upon transformation into *B. subtilis*. The orthogonal translation system (expanded below) is made up of two principle components: the orthogonal aminoacyl-tRNA Synthetase (AcKRS) and its cognate tRNA$^{Pyl\text{-}opt}$, each of which is flanked by a promoter and terminator that was chosen based on a successful gram(+) OTS designed in (18).

In the years since the initial demonstration of PylRS-tRNA$^{Pyl}_{CUA}$, numerous OTS' derived from the PylRS system have been implemented in several other commonly used organisms like *Saccharomyces cerevisiae* (19), *Drosophila melanogaster* (20), mammalian

cell lines (21), and mice (22,23). OTSs are powerful tools that afford researchers the opportunity to introduce unnatural amino acids with unique chemical functional groups at specific sites in recombinant proteins. Functional groups such as ketomides (24), boronates (25), and hydroxamates (26) (which are naturally formed via post translational modification in specific cellular scenarios) are otherwise lacking from the list of functional groups in the 20 standard amino acids. These functional groups can modulate the activity of the proteins (24), act as artificial receptors for protein targeting (25), and can even provide anti-tumour activity in cyclic peptide molecules (26). OTS technology is a promising avenue to improve the functionality of bio-manufactured proteins and engineered microorganisms.

### 2.1.3 Orthogonal translation systems for synthetic auxotrophy.

When working with bacteria, especially those which contain new genetic modifications or recombinant DNA technologies like OTSs, there are safety and ethical considerations to address. One such consideration is biocontainment, i.e., the ability to control and or prevent the growth and release of engineered bacterial strains and potential spread of novel genetic material in the natural environment.

In controlled laboratory settings, a common and effective means of biocontainment is the use of auxotrophic strains, which grow if and only if they are provided with a nutrient that they are unable to synthesize on their own. While effective, this method is only applicable in settings where the availability of the requisite nutrient can be well controlled (27). Unfortunately, it is not feasible to control the levels of naturally occurring nutrients in many of the potential clinical and industrial settings in which live *B. subtilis* cultures are used. For example, the work presented here was funded in part through an industrial partnership with Germiphene Inc, a biomanufacturing company that specializes in disinfectants and infection control products containing live bacteria. One of Germiphene's flagship products (Gobble) contains a live culture of *B. subtilis* that is used to prevent the growth of pathogenic bacteria in dental suction lines. The bacteria produce a biofilm that is safe for human health while also inhibiting the growth of pathogens.

Unfortunately, Gobble reacts unfavourably in temperate climates by overgrowing and becoming burdensome to the downstream filtration systems that it is meant to disinfect. The evacuation systems in which this product is used often contain a plethora of organic matter which creates optimal conditions for *B. subtilis* thrive. To prevent the overgrowth of the active culture, we are developing a bio-contained strain of *B. subtilis* that requires the provision of an otherwise unavailable, synthetic nutrient in order to sustain growth. Though naturally occurring nutrient abundance may not be controllable, the abundance of synthetic nutrients that are not available in the natural environment can be carefully monitored and adjusted to achieve a specified growth rate. Accordingly, this project aims to address the unmet need for a method of *B. subtilis* biocontainment that can be feasibly implemented in an environment where naturally occurring nutrient levels cannot be controlled.

This rationale was similarly articulated in a recent study (27) in which a synthetic auxotroph strain of *E. coli* was developed that required supplementation of a non-naturally occurring (unnatural) amino acid in order to sustain growth. In order to generate the auxotroph strain, the researchers introduced nonsense TAG mutations in the coding region of several essential genes (*LspA*, *DnaX*, and *SecY*). Normal decoding of the corresponding UAG codon in the reading from of an mRNA then leads to prematurely terminated protein. Because these are essential genes, lack of the complete protein results in growth inhibition. To rescue growth, the strains were transformed with a plasmid containing an OTS that could decode the nonsense mutations and rescue translation of the gene transcripts by incorporating an unnatural amino acid into the nascent peptide.

Recently, Luo (18) reported success in engineering an OTS that functions in a closely related member of the *Bacillus* genus, *Bacillus cereus*. Further, genetic code expansion by means of an orthogonal translation system was recently achieved in *B. subtilis* (28). Taken together, these studies provide precedent indicating that the machinery initially developed for use in *E. coli* is both active and orthogonal across a broad range of organisms and further, that gram (+) bacteria including *B. subtilis* can tolerate genetic code expansion using orthogonal translation systems.

**2.1.4 *Goals and hypothesis*: Developing orthogonal translation for *B. subtilis*.**

Due to its robust molecular characterization and ease of use, it made sense for the initial development of the OTS to occur in *E. coli*. But looking forward, the use of these systems to improve the applications of other industrially relevant bacteria such as *B. subtilis* is necessary. Here, I present my work towards developing a plasmid that contains all the requisite genetic and molecular machinery necessary to pioneer the genetic code expansion of *B. subtilis*.

The goal of this project is to develop an orthogonal translation system that functions in *B. subtilis*. I hypothesized that co-expressing a pyrrolysyl-tRNA synthetase variant (AcKRS) in addition to a tRNA$^{\text{Pyl-opt}}$ gene in *B. subtilis* would enable reassignment of UAG codons to incorporate non-standard amino acids, including AcK and BocLys. This technology could be applied to produce a first-of-its-kind synthetic auxotroph strain of *B. subtilis* whose growth would be dependant on the supplementation of an unnatural amino acid (uAA). Not only would this technology provide a novel means of *B. subtilis* biocontainment, it would also grant the organism the ability to produce proteins that contain unnatural amino acids with novel or enhanced properties (29). By establishing this technology in *B. subtilis*, we would develop a strain that could be used for a broad range of biotechnology applications across both laboratory and clinical settings.

Using the work done in *B. cereus* (18), I designed an OTS that contains the requisite genetic components necessary for an orthogonal translation system. The OTS contains two genes: tRNA$^{\text{Pyl-opt}}$ which functions to transport the unnatural amino acid AcK to the ribosome where it then decoded the UAG amber codon present in the mRNA transcript, and AcKRS, which functions to charge or ligate tRNA$^{\text{Pyl-opt}}$ with AcK. These two genes are contained within separate gene cassettes. The expression of tRNA$^{\text{Pyl-opt}}$ within its gene cassette is controlled by the promoter and terminator pair that surround the *B. subtilis* tRNA$^{\text{Ser}}$ gene (*serT*). The expression of AcKRS is controlled by the promotor and terminator pair that surround *Staphylococcus aureus sarA* gene. The entire system is contained within a vector (pDR111) that stably integrates the entire OTS into the genome of *B. subtilis* at the *amyE* gene (genomic locus: BSU_03040). Along with the OTS, the

pDR111 vector also contains a GFP reporter gene that is likewise integrated into the *B. subtilis amyE* genomic locus.

Genetic code expansion in *B. subtilis* offers an innovative solution to the challenges associated with functional protein production and regulating the growth of an organism in a system where the availability of naturally occurring nutrients cannot be controlled. Thus, my work represents the foundational steps towards the development of such a system that can be implemented in a critical microorganism for synthetic biology applications.

## 2.2 Methods

### 2.2.1 Plasmids and bacterial strains

The plasmid pDR111_(Sp)GFP (30) was obtained from the Bacillus Genetic Stock Center (BGSC ID: ECE278). The vector was next modified for the delivery of the OTS into the *B. subtilis* genome. pDR111_(Sp)GFP is a shuttle vector, and as such is capable of replication in both *E. coli* and *B. subtilis*. (Sp)GFP refers to the presence of a variant of GFP within pDR111 that was codon optimized for expression in *Streptococcus pneumoniae* and was found to be optimally active in *B. subtilis* (30). Any mention of GFP below refers to the (Sp)GFP variant. While using pDR111_GFP, cloning and mutagenesis (Section 2.2.3) experiments were conducted in *E. coli* while GFP expression analysis (Sections 2.2.6 and 2.2.7) was conducted in *B. subtilis*.

With assistance from David Wright, we designed the 2.7 kb OTS insert containing two gene cassettes (Figure 2.2) (sequences are available in Appendix 2.7.2). The first cassette consisted of a *B. subtilis serT* promoter/terminator pair and tRNA$^{Pyl-opt}$ (12) while the second cassette consisted of a *Staphylococcus aureus sarA* promotor/terminator pair and AcKRS (17) (Figure 2.2). The insert was synthesized in pUC57 (Gene Universal, Newark, DE, USA). The full sequence is available in Appendix 2.7.2. All work involving *E. coli* was conducted with the DH5α strain. All work involving *B. subtilis* was conducted using strain 168 (BGSC ID: 1A1).

**2.2.2 Cloning OTS into pDR111_GFP**

Restriction-free cloning (31) was used to insert the 2.7 kb OTS insert into pDR111_GFP. Primers 1 and 2 (Table 2.1) were used in the primary polymerase chain reaction (PCR) step to amplify the OTS from the pUC57 vector (used by Gene Universal for shipping synthesized DNA inserts <5Kb) while simultaneously introducing flanking sequences which were homologous to pDR111_GFP. The primary PCR was conducted with the following conditions in a 50 μL PCR reaction volume: 1 × Phusion HF buffer (NEB), 200 μM dNTP (from a 2.5 mM stock solution of equimolar mixed dNTPs (NEB, N0446S)), 500 nM of each primer 1 and 2, 1 U Phusion® polymerase (NEB, Catalog# M0530S), 10 ng of pUC57_OTS template and sterile ddH$_2$O.

The thermocycler protocol had an initial denaturing for 30s at 98°C, then 35 cycles of; 8s denature at 98°C, 20s anneal at 60°C, 60s extension at 72°C and then followed by a final extension for 5 min at 72°C. The resulting PCR product was then purified via gel extraction (Figure S2.1, QIAquick Gel Extraction Kit, Qaigen, product# 28704) and used as the primer (referred to as a "mega-primer" in the original protocol outlined in (31)) for a secondary PCR to insert the OTS sequence between flanking regions of pDR111_GFP that share homology with the *B. subtilis AmyE* genomic locus (BSU_03040). The secondary PCR was likewise conducted in a 20 μL PCR reaction: 1 × Phusion HF buffer, 200 μM dNTP, a insert (PCR product from primary PCR):plasmid ratio of 20 ng to 100 ng of pDR111_GFP was used with 0.4 U Phusion® polymerase. The thermocycler protocol involving an initial denaturing for 30s at 98°C, then 15 cycles of 8s denature at 98°C, 20s anneal at 60°C, 5.5min extension at 72°C followed afterwards by a final extension for 6 min at 72°C. Once the secondary PCR was completed, 20 U of DpnI (NEB, R0176S) was added directly to the reaction mix and incubated for 2 hr at 37°C followed by 20 min at 80°C. 4 ul of the DpnI-treated secondary PCR mix was transformed into *E. coli* DH5α. Successful clones were subsequently grow overnight in 5 ml LB and plasmid DNA was prepped (Geneaid, PDH100) and sequenced (Genewiz, South Plainfield, NJ, USA) using primers 3-6 to span the entire 2.7 kb OTS insert. Sequencing confirmed the correct location and orientation of the OTS insert within the pDR111_GFP. A schematic showing the final pDR111_AcKRS_GFP plasmid map is provided (Figure 2.2).

**2.2.3 Introducing stop codons into GFP**

TAG codons were introduced into 2 different permissive sites (S3TAG and E18TAG) in the coding region of GFP in pDR111_AcKRS_GFP in order to assay the ability of the OTS to decode stop codons. In total, 3 plasmids were obtained 1) pDR111_AcKRS_GFP, 2) pDR111_AcKRS_GFP_S3TAG, and 3) pDR111_AcKRS_GFP_E18TAG. The TAG codons were introduced using a site-directed mutagenesis protocol (32).

Briefly, for each nonsense mutation two 25 µL PCR reactions were conducted simultaneously, each containing a single primer with the following parameters: PCR mix; $1 \times$ Phusion Buffer, 200µM dNTPs, 40 pmol of a single primer, 500 ng of template DNA, 0.5 U of Phusion Polymerase and sterile ddH$_2$O up to 25 µL Thermocycler protocol: 98°C for 3min, then 30 cycles of 98°C for 10s, 55°C for 30s, 72°C for 5 min 45s, followed by a final extension at 72°C for 10min. Once completed, the two reactions were added together, heated to 95°C and then cooled slowly as follows; 95°C for 5min, 90°C for 1min, 80°C for 1min, 70°C for 30s, 60°C for 30s, 50°C for 30s, 40°C for 30s and finally held at 37°C. 20 U of DpnI was then added to the cooled PCR reaction which was then incubated overnight at 37°C. The next day, 4 uL of the DpnI-treated product was transformed into *E. coli* DH5α cells. Plasmid DNA from successful colonies was purified (Geneaid, PDH100) and sequenced with primer 7 (Table 2.1) to confirm the presence and location of the TAG mutations. Primers 8 and 9 were used to replace the TCT codon (serine) at position 3 with TAG, while primers 10 and 11 were used to replace the GAA codon (glutamic acid) at position 18 with TAG.

**2.2.4 *B. subtilis* competence induction.**

*B. subtilis* minimal media included minimal salts solution (5x) per litre: (NH$_4$)$_2$SO$_4$, 10.0 g; K$_2$HPO$_4$, 74 g; KH$_2$PO$_4$, 27.0 g; trisodium citrate, 9.5 g; MgSO$_4 \bullet$ 7 H$_2$O, 1.0 g; pH 7.0; glucose 20% (w/v) solution; casamino acids 2% (w/v) solution. Minimal- growth medium per 100 ml: 20 ml minimal salts (5x); 2.5 ml glucose 20% (w/v) solution; 1 ml casamino acids; 20 µg/ml L-tryptophan. Starvation medium per 100 ml: 20 ml minimal salts; 2.5 mL glucose 20% (w/v) solution.

Competent cell preparation: On day one, *B. subtilis* 168 stock from BGSC was received as a filter disk inoculated with bacterial culture. The filter disk was aseptically transferred to a standard LB agar plate. Then, 500 μL of sterile LB was added on top of the filter disk and the plate was swirled gently to allow the culture to disperse across the plate. The plate was then incubated overnight at 37°C to allow colony growth. Day two, 10 ml of minimal growth medium in a 100 ml flask was inoculated with cells from a single colony and incubated (37°C) overnight (~16- 18 hours) with agitation. Day three, 1.4 ml of overnight culture was added to 10 ml of pre- warmed fresh minimal growth medium in a 100 ml flask. This sample was grown for 3 hours at 37°C with agitation. Then 11 ml of starvation medium was added, and the culture continued to grow for 2 hours at 37°C with agitation. At this point the culture was maximally competent (33), and 0.5 ml samples were frozen in 10% glycerol (v/v) at -80°C.

### 2.2.5 *B. subtilis* transformation

An aliquot of the competent 168 cells (Section 2.2.5) was thawed quickly in a 37°C water bath. To this thawed sample, 500-1000 ng of plasmid DNA was added and mixed with gentle pipetting. The cells were then incubated for 25 minutes at 37°C in with agitation. Then, 500 μL of LB medium was added and the cells were further incubated for 1.5 hours at 37°C with shaking to allow for the expression of antibiotic markers. After this incubation, cells were pelleted (5000 × g for 10 min), supernatant was removed, and the pellet was resuspended in 200 μL of fresh LB. 100 μL of this sample was then plated on LB agar containing selective antibiotics. For transformations using the pDR111 vector, 100 μg/ml of spectinomycin was required for selection. Transformants appeared following overnight incubation at 37°C. An aliquot of competent 168 cells was transformed with pDR111_GFP and the following day (20 hours post-transformation) transformants were imaged using a BioRad ChemiDoc imager to detect GFP fluorescence to confirm transformation of the entire pDR111_GFP plasmid (Figure S2.2). BioRad ChemiDoc imager settings for detection of GFP fluorescence; Filter: 530/38 nm, Light: Blue Epi illumination.

### 2.2.6 Integration-detection assays

pDR111_AcKRS_GFP contains two regions of DNA sequence that are homologous to the *B. subtilis amyE* gene (genomic locus: BSU_03040). Upon transformation, these regions of homology are integrated, along with the DNA in between them (Figure 2.2) into the *B. subtilis* genome at the *amyE* site. *amyE* encodes alpha-amylase, a protein which is not essential for growth but is required to metabolize starch. When pDR111_AcKRS_GFP integrates into the *B. subtilis* genome, the *amyE* gene is disrupted and starch metabolism is abolished. The loss of starch metabolism can be detected using a colorimetric iodine-staining assay as a proxy for the disruption of the *amyE* gene due to successful integration. *B. subtilis* is transformed as described in Section 2.2.5 on LB agar plates containing spectinomycin (100 µg/ml) that were supplemented with soluble starch powder (1% w/v) and incubated overnight at 37°C. To detect starch metabolism the following day, the plates containing successful transformants are flooded with 1 ml Gram's iodine stain (ThermoFisher, R40056) and transformants are observed for the presence of dye-free halos. A dye-free halo surrounding a colony indicates starch hydrolysis by the organism due to extracellular starch metabolism enzymes, i.e., alpha-amylase. Colonies that do not have dye-free halos surrounding them are unable to metabolize starch and are likely to contain the integrated DNA from the pDR111_AcKRS_GFP plasmid at their disrupted *amyE* locus.

Colonies that lack dye-free halos were then subjected to a PCR screen using primer pairs 12-13 and 14-15 which are designed to amplify the junctions (12, 13: upstream, 14, 15: downstream) between the genomic a*myE* locus and integrated region of pDR111_AcKRS_GFP in cases where pDR111_AcKRS_GFP integration into *amyE* was successful. Colony PCR screen protocol: 50 µl PCR reaction (1 × Phusion® HF buffer, 200 µM dNTP, 500 nM of each primer (either 12 and 13 or  14 and 15), 1 U Phusion® polymerase (NEB, Catalog# M0530S), sterile ddH$_2$O up to 50 µL. Template DNA which was obtained as follows: The 50 µL PCR-reaction mixture was inoculated with *B. subtilis* cells picked from a transformant colony. The cells were boiled for 5 min (98°C) in sterile ddH$_2$0 to release DNA during the first step of the PCR reaction. The thermocycler protocol was as follows: 5 min at 98°C (to release/denature DNA from transformed *B. subtilis* cells),

then 35 cycles of; 8s denature at 98°C, 20s anneal at 67°C, 60s extension at 72°C and then followed by a final extension for 5 min at 72°C.

## 2.2.7 Fluorescence measurements

For quantitative fluorescent measurements, a microplate-reader was used following the protocol established in (30). Briefly, LB medium with 100 μg/ml spectinomycin and 10 mM $N\varepsilon$-acetyl-L-lysine (AcK) was inoculated with a *B. subtilis* colony that was transformed with a GFP-bearing pDR111 plasmid (either pDR111_GFP, pDR111_AcKRS_GFP, pDR111_GFP_S3TAG, pDR111_AcKRS_GFP_S3TAG, pDR111_GFP_E18TAG pDR111_AcKRS_GFP_E18TAG, see section 2.2.3) and grown overnight at 37°C with agitation. Biological triplicate cultures started from independent colonies were maintained for each condition.

The next day, the cultures were aliquoted into 4 different LB growth medias that were supplement as follows: 1) 100 μg/ml spectinomycin, no AcK, no isopropyl-β-D-1-thiogalactopyranoside (IPTG); 2) 100 μg/ml spectinomycin, 10 mM AcK, no IPTG, 3. 100 μg/ml spectinomycin, no AcK, 0.1 mM IPTG; 4) 100 μg/ml spectinomycin, 10 mM AcK, 0.1mM IPTG. Each culture was diluted to an $OD_{600} = 0.1$ in 5 ml of the above-mentioned media. A negative control culture of untransformed WT *B. subtilis* 168 was also prepared following the same growth conditions above but in the absence of spectinomycin, AcK or IPTG. The next day, 200 μL aliquots of each culture were added to wells of a 96-well microtiter plate and fluorescence and $OD_{600}$ values were detected using a Synergy H1 Hybrid microplate reader (BioTek, Agilent), GFP signals were detected using an excitation wavelength at 485nm [20-nm width] and emission at 535 nm [25-nm width]. Relative GFP levels of each culture compared to the negative control untransformed WT *B. subtilis* 168 and determined as follows. Background values for turbidity ($OD_{600}$) and fluorescence were taken (based on 3 biological replicates) from untransformed (WT) *B. subtilis* control culture grown similarly as above. The respective averages of these background values were subtracted from the optical density and fluorescence reading taken of each culture containing a GFP variant. In order to account for potential differences in growth, final background-corrected fluorescence measurements were normalized by $OD_{600}$.

**Table 2.1. Primers used in this study.**

| Primer | Sequence | Description |
|---|---|---|
| 1 | TCTTTTCTTTTCCAGACCTGAGGGAGGCGGCTCGA GACTAGTAAGATTATGATCTTCTGAACG | pDR111_site3_fwd |
| 2 | TGGCTTTTCCCCGGGAACCTCACACCATTTCT CGAGCTGCAGAGTGCCATTAG | pDR111_site3_rev |
| 3 | GGATCATCCGTTTAGGCTGG | pDR111_site3 _seqver_fwd |
| 4 | CGAGCAAGATGCATCAAATAGGGAGG | OTS_F2 |
| 5 | ATCGCCGCAAGCCATTTCG | OTS_R2 |
| 6 | ATCTATTGACCGCAGTGATAGCC | pDR111_site3 _seqver_rev |
| 7 | AATCAAGACGTCTTATTTATACAATTCATCCATACCATG | GFP_Rev |
| 8 | CAAACATGGTTTAGAAAGGTGAAGAATTG | Fwd: S3F |
| 9 | CAATTCTTCACCTTTCTAAACCATGTTTG | Rev: S3R |
| 10 | CATCACCATCCAACTAAACCAAAATTGG | Fwd: E18F |
| 11 | CCAATTTTGGTTTAGTTGGATGGTGATG | Rev: E18R |
| 12 | CGGCTGCGAGTGCTGAAAC | Int_ver_fwd1 |
| 13 | GTTTCGGTGATGAAGATCTTCCCG | Int_ver_rev1 |
| 14 | GGATCATCCGTTTAGGCTGGG | Int_ver_fwd3 |
| 15 | CATCCTTAAACGCCTGTCGTC | Int_ver_rev3 |

## 2.3 Results

### 2.3.1 Designing a plasmid to integrate the Orthogonal Translation System (OTS) into the *B. subtilis* genome

pDR111_AcKRS_GFP is a 11.6 kb plasmid (Figure 2.2), derived from pDR111_GFP (30). The plasmid contains the OTS necessary to decode and reassign the UAG codon to incorporate AcK into proteins. Within the original pDR111_GFP (8.8 kb), there is a 4.2 kb region of DNA that is integrated into the *B. subtilis* genome upon transformation. Integration occurs via a double crossover event between the *B. subtilis* genome and sequences flanking the 4.2 kb region which share homology to the *B. subtilis* a*myE* gene (34). This 4.2 kb region also contains a copy of *GFP* that has been optimized for expression within *B. subtilis* (30) under the control of lacI that will serve as an assay for UAG-readthrough. The OTS itself, is a 2.8 kb sequence that contains the orthogonal tRNA (tRNA[Pyl-opt]) and aminoacyl-tRNA synthetase (AcKRS) pair. I also designed in 4 mutations (V31I, T56P, H62Y, and A100E; IPYE) to the sequence of AcKRS that were introduced to mirror those discovered by Bryson et al. (17) that improve the efficiency of the PylRS

system and increased the yield of proteins containing unnatural amino acids by up to 10-fold.

To promote expression of the tRNA$^{Pyl-opt}$ and AcKRS genes, the OTS insert contained two gene cassettes (all sequences are available in Appendix 2.7.2). The first cassette contained tRNA$^{Pyl-opt}$, the expression of which was controlled by the flanking *B. subtilis serT* promoter/terminator pair. The second cassette contained AcKRS, the expression of which was controlled by the flanking *Staphylococcus aureus sarA* promotor/terminator pair (Figure 2.2). The chosen *serT* promotor/terminator pair for tRNA$^{Pyl-opt}$, and *sarA* promotor/terminator pair for AcKRS were previously determined to be effective for the expression of these components in a related gram (+) bacterium, *Bacillus cereus* (18).

The 2.8 kb OTS, as described above (Appendix 2.7.2 for sequence), was synthesized by Gene Universal (pUC57_OTS) (Newark, DE, USA) and cloned (Section 2.2.2) into the 4.2 kb genome integration region of the pDR111_GFP plasmid (Section 2.2.3) to yield the final OTS construct: pDR111_AcKRS_GFP (Figure 2.2). This version of the construct contains a wild type and positive control version of GFP that will express and fluoresce upon IPTG induction without the need for OTS function. Subsequently, nonsense TAG mutations were inserted at 2 permissive sites in the coding region of GFP (S3TAG and E18TAG). These nonsense mutations will prevent full-length translation of GFP from occurring unless the OTS is able to decode the UAG codon and incorporate AcK into the growing peptide chain.

## 2.3.2 Assessing integration of pDR111_AcKRS_GFP into the *B. subtilis amyE* locus

I obtained *B. subtilis* strain 168 from the BGSC (BGSC I.D. 1A1) and made competent cell stocks using a protocol from the original paper that described *B. subtilis* transformation by Anagnostopoulos and Spizizen in 1961 (33) (Section 2.2.4). Traditional antibiotic selection with spectinomycin was used to isolate colonies which had successfully taken up the pDR111_AcKRS_GFP plasmid but further assessment was necessary to ensure that the

plasmid had integrated into the *B. subtilis* genome and that this integration had occurred at the correct locus.

An initial colorimetric (starch-iodine) screen was used to identify colonies that lacked alpha-amylase activity (starch metabolism) as a proxy assessment for the disruption of the *amyE* gene due to pDR111_AcKRS_GFP integration at the *amyE* locus. The results of this colorimetric assay are displayed in Figure 2.3. The absence of dye-free halos surrounding the *B. subtilis* colonies transformed with pDR111_AcKRS_GFP (Figure 2.3B) indicate that these bacteria are unable to digest the soluble starch present in the agar (1% starch w/v).

Final confirmation of proper integration was determined by PCR-screening (Figure 2.4). The PCR reaction used primers 12-13 (Figure 2.4 lane 2) and 14-15 (Figure 2.4 lane 3) to amplify 3 kb and 3.2 kb regions of the upstream and downstream junctions (respectively) that are created when pDR111_AcKRS_GFP is properly integrated into the genomic *amyE* locus of *B. subtilis*. The results of a positive control PCR reaction using primers 12-13 and pDR111_AcKRS_GFP as the DNA template is shown in lane 5.

## 2.3.3 Assaying OTS function by means of TAG decoding and GFP fluorescence rescue

Previously, AcK-dependant GFP fluorescence in response to UAG nonsense mutations within the GFP gene was achieved using the tRNA$^{Pyl-opt}$ and AcKRS pair in *E. coli* (13). A similar AcK-dependant GFP fluorescence system was employed here to assess the degree to which the OTS was able to decode UAG codons and rescue expression of GFP in *B. subtilis*.

Figure 2.5 contains the results of a 96-well micro plate assay conducted using a Synergy H1 microplate reader to detect GFP fluorescence in live cultures of *B. subtilis* that have been transformed with 1 of 6 pDR111_GFP plasmid variants or that remain untransformed (negative control). All 6 of the transformed *B. subtilis* cultures (each transformed with a different pDR111_GFP variant) were grown in 4 different media conditions: 1) no AcK, no IPTG, 2) 10 mM AcK, no IPTG; 3) no AcK, 0.1mM IPTG; or 4) 10 mM AcK, 0.1mM IPTG. Figure 2.5 contains the fluorescence data collected from

each of the overnight cultures presented as fluorescence intensity: measured fluorescence value/OD$_{600}$ (arbitrary units, AU).



pDR111_acKRS_GFP
integrative plasmid

pHT_(Sp)GFP
replicating plasmid
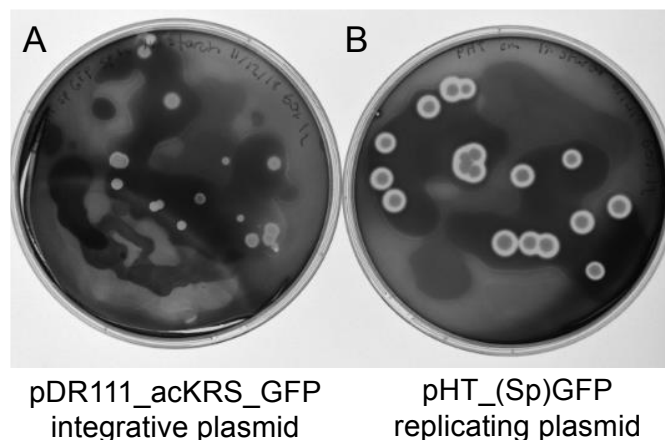
**Figure 2.3. Colorimetric assay of integration within the *B. subtilis amyE* locus.** Both plates contain 1% (w/v) soluble starch powder and were flooded with saturating amounts of Gram's iodine. The left plate (A) contains *B. subtilis* 168 (BGSC ID: 1A1) colonies that have been transformed with pDR111_AcKRS_GFP. The lack of dye-free halos surrounding the colonies in (A) is indicative of a loss of starch metabolism. The right plate (B) contains *B. subtilis* 168 (BGSC ID: 1A1) colonies that have been transformed with pHT_(Sp)GFP, a non-integrative and replicating plasmid that does not disrupt starch metabolism as indicated by the presence of white, dye-free (starch-free) halos surrounding the transformed colonies.
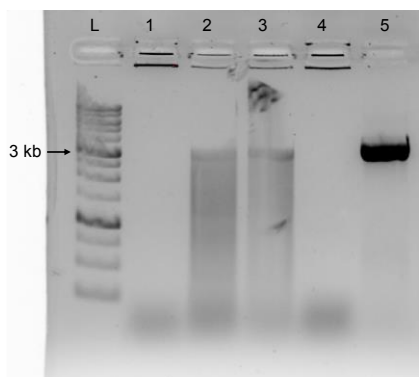


**Figure 2.4. PCR-screen of *B. subtilis* colonies with integrated OTS.** Lanes 1 and 2 used primers 12 and 13 to amplify the upstream junction of *amyE*-pDR111_AcKRS_GFP. Lane 1 used WT 168 bacterial cells as its source of template DNA and lane B used 168 cells that had been transformed with pDR111_AcKRS_GFP. Lanes 3 and 4 used primers 14 and 15 to amplify the downstream junction of *amyE*-pDR111_AcKRS_GFP. Lane 3 used 168 cells that had been transformed with pDR111_AcKRS_GFP as its source of template DNA and lane 4 used WT 168 bacterial cells. The positive control lane denoted with (lane 5) used (30ng) purified pDR111_AcKRS_OTS plasmid as the template DNA for the reaction with primers 12 and 13.

Of the 25 conditions tested, only 4 cultures had significantly increased levels of GFP fluorescence compared to the untransformed negative control culture based on a one-way ANOVA test. *B. subtilis* transformed with either pDR111_GFP or pDR111_AckRS_GFP, grown with IPTG and with or without AcK all contained significant fluorescence intensity ($p < 0.0001$) As expected, the presence or absence of AcK did not significantly affect the level of GFP fluorescence in *B. subtilis* transformed with pDR111_GFP or pDR111_ACK_GFP. Similarly, the addition of the OTS to the pDR111_GFP plasmid (pDR111_GFP vs pDR111_AcKRS_GFP) did not significantly affect the level of GFP fluorescence observed, regardless of the media. The lack of change in fluorescence values for cultures transformed with WT GFP is to be expected as the expression of these GFP variants is solely dependant on IPTG induction and not the functionality of the OTS. None of the cultures containing a GFP gene with a nonsense mutation (at either of the permissive sites S3 or E18) displayed any significant GFP fluorescence regardless of the presence or absence of the OTS, AcK or IPTG.

## 2.4 Discussion

### 2.4.1 Progress and next steps for expanding the genetic code of *B. subtilis*

In this chapter, I described the conceptualization, construction and initial testing of an orthogonal translation system designed to decode UAG codons. I stably integrated the OTS and GFP reporter variants into the genome of *B. subtilis*. The design of the system was centered around the newly improved (12,13,17), highly efficient, and broadly active orthogonal tRNA$^{Pyl-opt}$ and AcKRS pair (12). Genetic code expansion using similar orthogonal tRNA synthetase-tRNA pairs has been implemented in a number of important model organisms (19-21) including other members of the *Bacillus* genus (18). The system developed here is the first of its kind reported for *B. subtilis*.

The construction of the preliminary OTS along with its stable integration into the *B. subtilis* genome has been accomplished here, however, demonstration of the system's ability to decode UAG codons and incorporate AcK has not yet been achieved. Going forward, there are several avenues to explore in order to achieve OTS-mediated UAG decoding in *B. subtilis* using the system designed and developed here. The first step towards this goal would be to determine the aspect(s) of the system that are not functioning

properly. PCR-screens and sequencing have given us confidence that the system's genomic location, sequence integrity and ability to express wildtype GFP are satisfactory. Sequencing has confirmed that the intended TAG mutations are confined to their intended permissive locations and previous studies have demonstrated that these positions in GFP are permissible for unnatural amino acid insertion (35).



**Figure 2.5. Fluorescence quantification of GFP and GFP variants bearing nonsense mutations in _B. subtilis_.** GFP fluorescence intensities of live cultures of _B. subtilis_ represent each culture's fluorescence data/$OD_{600}$ of that culture. Cultures were either transformed with 1 of 6 pDR111_GFP plasmid variants (containing either GFP, GFP (S3TAG), or GFP (E18TAG) with or without AcKRS), or in the case of the WT negative control, remained untransformed as indicated below each bar. All 6 of the transformed _B. subtilis_ cultures were aliquoted and grown in 4 different media conditions containing neither, either, or both of 10 mM AcK and 0.1 mM IPTG. The type of GFP, presence or absence of AcKRS and media supplement conditions of each culture are indicated below each data bar. Of the 25 conditions tested, only 4 cultures had significantly increased levels of GFP fluorescence (denoted by star above data bar) compared to the untransformed negative control WT culture based on a one-way ANOVA: pDR111_GFP (AcK-/IPTG+): $p < 0.0001$, pDR111_GFP (AcK+/IPTG+): $p < 0.0001$, pDR111_AckRS_GFP (AcK-/IPTG+): $p < 0.0001$ and pDR111_AcKRS_GFP (AcK+/IPTG+), $p < 0.0001$. Each condition was completed in triplicate (n=3), the height of the bars indicate the mean value of the three replicates and the individual values that contribute to the means are shown as circles surrounding the mean. Error bars represent 1 standard deviation above and below the mean.

Further, the functionality of the tRNA$^{Pyl-opt}$-AcKRS pair, along with the ability to make use of unnatural amino acid derivatives of lysine (including, but not limited to, AcK) has been demonstrated in the closely related *B. cereus* (18) and so, while not confirmed by my own work, it is less likely that these components are incompatible with *B. subtilis* and is instead more likely that they are not being properly expressed. The problem may stem from insufficient expression of either (or both) of the tRNA$^{Pyl-opt}$ or AcKRS components of the system due to suboptimal activity of their respective promoters. In the absence of an available antibody with which to test the expression of AcKRS directly, I would next add a co-translationally cleavable mCherry tag to the N-terminus of AcKRS to monitor its expression via fluorescence and confirm via western blotting using available anti-mCherry antibodies. This approach was successfully implemented to detect the expression of PylRS in primary tissue collected from rat neocortex (23).

To test for issues related to insufficient expression of the OTS components, I could replace the current promoter and terminator pairs (*sarA* for AcKRS and *serT* for tRNA$^{Pyl-opt}$). As part of the design of our OTS, unique restriction enzyme cut sites were incorporated in between each component (*sarA* promoter, AcKRS, *sarA* terminator, *serT* promoter, tRNA$^{Pyl-opt}$, *serT* terminator) of the system to facilitate the exchange of components if necessary. Using these sites to our advantage, I would systematically replace and test a series of 5 different promotor/terminator pairs (P$_{srf}$ (36), P$_{xyl}$ (37), P$_{gsiB}$ (38), P$_{xyl}$ (39), and P$_{HpaII}$ (40)) that all have well established expression capabilities within *B. subtilis*.

## 2.4.2 Previous work involving an expanded genetic code in *B. subtilis*.

As mentioned previously (Section 2.1.3), genetic code expansion by means of recoding the UAG stop codon to code for unnatural amino acids using orthogonal translation systems has been previously established in *B. cereus*, a member of the *Bacillus* genus that is closely related to *B. subtilis* (18). However, this was not the first time that a member of the *Bacillus* genus had been the subject of research concerning the engineering the genetic code. In 1983, Wong (41) demonstrated that a Trp auxotroph strain of *B. subtilis* could become dependant instead on a Trp analog: 4-fluorotryptophan (4-FTrp). After several rounds of selection and nitrosoguanidine-mediated mutagenesis on media containing increasing

amounts of 4-FTrp and decreasing amounts of Trp, Wong and co-workers were able to isolate a strain that could grow on agar supplemented with only 4-FTrp in place of Trp. In the evolved strain, Trp supplementation could still rescue growth, but to a reduced degree compared to the original parent strain. Effectively, Wong had isolated the first ever synthetic auxotroph strain of *B. subtilis* within which all codons that once coded for Trp now coded for the unnatural amino acid 4-FTrp. These results demonstrate that the translation machinery of *B. subtilis* is tolerant and capable of utilizing unnatural amino acids for protein production.

### 2.4.3 Orthogonal translation systems within the *Bacillus* genus

More recently, the focus of molecular technologies that expand the genetic code have focused primarily on utilizing codons that are not already assigned to one of the twenty canonical amino acids, i.e., the stop codons. Of the 3 stop codons, UAG is the least frequently used stop codon in the genomes of a wide array of organisms including *E. coli* and *B. subtilis* (42), which makes UAG the most suitable to reassignment. Indeed, in their 2016 study on genetic code expansion in *B. cereus*, Luo et al. made use of a single plasmid system that contained the requisite molecular machinery to decode the UAG codon with $N^{\varepsilon}$-(tertbutoxycarbonyl)-L-lysine (18). The authors investigated a class of molecules known as thiopeptides: a subclass of ribosomally synthesized and post translationally modified cyclic peptides that have promising antibiotic (43) and antitumor (44) properties. Using their OTS, they were able to achieve increased yields of purified active thiopeptides by incorporating photo-caged amino acids that provide protection from post translational modifications that naturally inactivate the compounds. These photo caged residues could then be removed post-purification to yield active products. This is but one example of a generalizable scenario in which the expansion of the protein chemistries available in novel bacterial hosts can improve the production of medically relevant or bio-manufacturable molecules.

**2.4.4 Significance of orthogonal translation in the *B. subtilis* system**

There are circumstances in which the production of sought-after bio-manufactured molecules can be improved by means of orthogonal translation systems. As this practice inevitably becomes more commonplace, there will be a demand for its use in the manufacturing of products in the food or medical industries, intended for human contact or consumption.

In the domain of biomanufacturing, *B. subtilis* is favoured over other common biochemical 'work horses' like *E. coli* because of its lack of lipopolysaccharides (LPS). The LPS-derived endotoxins contained in the cell wall of *E. coli* necessitates additional purifications of any *E. coli*-derived products that are intended for human contact or consumption. These extra steps can be detrimental to the final yield and overall costs associated with *E. coli*-based biomanufacturing systems (45). Furthermore, the ease of final product purification from *B. subtilis* is superior to *E. coli* due to its proficiency at secreting endogenous or recombinant enzymes into its growth media (45,46). In some cases, the secretion of protein-products into the extracellular environment can facilitate proper folding if the reducing environment of the cytoplasm is incompatible with the native conditions of the given recombinant protein (47).

Aside from designer protein production, the development of an OTS for *B. subtilis* is an important step towards a novel approach to biocontainment or growth control. Previous examples of synthetic auxotroph bacterial strains have been developed using orthogonal translation systems to 'addict' *E. coli* to amino acids that do not occur in nature (27). Such strains boast nearly undetectable levels of escape mutations (27), making this method of biocontainment particularly intriguing. By introducing this method of biocontainment to an organism that already possesses the GRAS-status, we would develop a strain with an exceptionally large range of potentially applications. Such a strain would be applicable for use in a system where the risk of potential contact with humans has precluded the use of other, potentially pathogenic bacteria like *E. coli*. For example, in the case of dental evacuation system cleaners (such as Gobble from Germiphene, Inc) that use live cultures of bacteria to disinfect the accompanying plumbing system. To remain efficient, such a system requires the ability to control the growth of the active culture (such

that it does not overgrow to the point of becoming burdensome to the system) despite the plethora of nutrients present in the form of the organic material present within patient's saliva. A bio-contained, synthetic auxotroph strain of *B. subtilis* that is dependant on the provision of an unnatural amino acid not found in organic material would meet the specifications of such a situation. In conclusion, this chapter represents critical first steps toward the expansion of the genetic code of *B. subtilis* for applications in infection control and bio-manufacturing.

**2.6 References**

1. Sewalt, V., Shanahan, D., Gregg, L., La Marta, J. and Carrillo, R. (2016) The Generally Recognized as Safe (GRAS) Process for Industrial Microbial Enzymes. *Industrial Biotechnology*, **12**, 295-302.

2. Gu, Y., Xu, X., Wu, Y., Niu, T., Liu, Y., Li, J., Du, G. and Liu, L. (2018) Advances and prospects of Bacillus subtilis cellular factories: From rational design to industrial applications. *Metab Eng*, **50**, 109-121.

3. Schallmey, M., Singh, A. and Ward, O.P. (2004) Developments in the use of Bacillus species for industrial production. *Can J Microbiol*, **50**, 1-17.

4. van Dijl, J. and Hecker, M. (2013) Bacillus subtilis: from soil bacterium to super-secreting cell factory. *Microbial Cell Factories*, **12**, 3.

5. Ricca, E. and Cutting, S.M. (2003) Emerging Applications of Bacterial Spores in Nanobiotechnology. *Journal of Nanobiotechnology*, **1**, 6.

6. Ferreira, L.C., Ferreira, R.C. and Schumann, W. (2005) Bacillus subtilis as a tool for vaccine development: from antigen factories to delivery vectors. *An Acad Bras Cienc*, **77**, 113-124.

7. Sen, S., Ingale, S.L., Kim, Y.W., Kim, J.S., Kim, K.H., Lohakare, J.D., Kim, E.K., Kim, H.S., Ryu, M.H., Kwon, I.K. *et al.* (2012) Effect of supplementation of Bacillus subtilis LS 1-2 to broiler diets on growth performance, nutrient retention, caecal microbiology and small intestinal morphology. *Res Vet Sci*, **93**, 264-268.

8. Olmos, J. and Paniagua-Michel, J.d.J. (2014) Bacillus subtilis A Potential Probiotic Bacterium to Formulate FunctionalFeeds for Aquaculture. *Conference Proceedings*.

9. Tannock, G.W. (2001) Molecular assessment of intestinal microflora. *Am J Clin Nutr*, **73**, 410s-414s.

10. Cohen, S.N., Chang, A.C., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A*, **70**, 3240-3244.

11. Wang, L., Xie, J. and Schultz, P.G. (2006) EXPANDING THE GENETIC CODE. *Annual Review of Biophysics and Biomolecular Structure*, **35**, 225-249.

12. Fan, C., Xiong, H., Reynolds, N.M. and Söll, D. (2015) Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic acids research*, **43**, e156-e156.

13. Neumann, H., Peak-Chew, S.Y. and Chin, J.W. (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol*, **4**, 232-234.

14. Schimmel, P.R. and Söll, D. (1979) Aminoacyl-tRNA Synthetases: General Features and Recognition of Transfer RNAs. *Annual Review of Biochemistry*, **48**, 601-648.

15. Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K. and Krzycki, J.A. (2004) Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. *Nature*, **431**, 333-335.

16. Kavran, J.M., Gundllapalli, S., Donoghue, P., Englert, M., Söll, D. and Steitz, T.A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the National Academy of Sciences*, **104**, 11268.

17.  Bryson, D.I., Fan, C., Guo, L.-T., Miller, C., Söll, D. and Liu, D.R. (2017) Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature Chemical Biology*, **13**, 1253-1260.

18.  Luo, X., Zambaldo, C., Liu, T., Zhang, Y., Xuan, W., Wang, C., Reed, S.A., Yang, P.Y., Wang, R.E., Javahishvili, T. *et al.* (2016) Recombinant thiopeptides containing noncanonical amino acids. *Proc Natl Acad Sci U S A*, **113**, 3615-3620.

19.  Lee, H.S., Guo, J., Lemke, E.A., Dimla, R.D. and Schultz, P.G. (2009) Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in Saccharomyces cerevisiae. *J Am Chem Soc*, **131**, 12921-12923.

20.  Bianco, A., Townsley, F.M., Greiss, S., Lang, K. and Chin, J.W. (2012) Expanding the genetic code of Drosophila melanogaster. *Nature Chemical Biology*, **8**, 748-750.

21.  Kang, J.Y., Kawaguchi, D. and Wang, L. (2016) Optical Control of a Neuronal Protein Using a Genetically Encoded Unnatural Amino Acid in Neurons. *J Vis Exp*, e53818.

22.  Han, S., Yang, A., Lee, S., Lee, H.-W., Park, C.B. and Park, H.-S. (2017) Expanding the genetic code of Mus musculus. *Nature Communications*, **8**, 14568.

23.  Ernst, R.J., Krogager, T.P., Maywood, E.S., Zanchi, R., Beránek, V., Elliott, T.S., Barry, N.P., Hastings, M.H. and Chin, J.W. (2016) Genetic code expansion in the mouse brain. *Nature chemical biology*, **12**, 776-778.

24.  Lee, S.H., Kyung, H., Yokota, R., Goto, T. and Oe, T. (2014) N-Terminal α-Ketoamide Peptides: Formation and Transamination. *Chemical Research in Toxicology*, **27**, 637-648.

25.  Behnam, M.A.M., Sundermann, T.R. and Klein, C.D. (2016) Solid Phase Synthesis of C-Terminal Boronic Acid Peptides. *Organic Letters*, **18**, 2016-2019.

26.  Komatsu, Y., Tomizaki, K.Y., Tsukamoto, M., Kato, T., Nishino, N., Sato, S., Yamori, T., Tsuruo, T., Furumai, R., Yoshida, M. *et al.* (2001) Cyclic hydroxamic-acid-containing peptide 31, a potent synthetic histone deacetylase inhibitor with antitumor activity. *Cancer Res*, **61**, 4459-4466.

27.  Rovner, A.J., Haimovich, A.D., Katz, S.R., Li, Z., Grome, M.W., Gassaway, B.M., Amiram, M., Patel, J.R., Gallagher, R.R., Rinehart, J. *et al.* (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature*, **518**, 89-93.

28.  Scheidler, C.M., Vrabel, M. and Schneider, S. (2020) Genetic Code Expansion, Protein Expression, and Protein Functionalization in Bacillus subtilis. *ACS Synthetic Biology*, **9**, 486-493.

29.  Balasuriya, N., Kunkel, M.T., Liu, X., Biggar, K.K., Li, S.S.C., Newton, A.C. and O'Donoghue, P. (2018) Genetic code expansion and live cell imaging reveal that Thr-308 phosphorylation is irreplaceable and sufficient for Akt1 activity. *The Journal of biological chemistry*, **293**, 10744-10756.

30.  Overkamp, W., Beilharz, K., Detert Oude Weme, R., Solopova, A., Karsens, H., Kovacs, A., Kok, J., Kuipers, O.P. and Veening, J.W. (2013) Benchmarking various green fluorescent protein variants in Bacillus subtilis, Streptococcus pneumoniae, and Lactococcus lactis for live cell imaging. *Appl Environ Microbiol*, **79**, 6481-6490.

31.  Bond, S.R. and Naus, C.C. (2012) RF-Cloning.org: an online tool for the design of restriction-free cloning projects. *Nucleic acids research*, **40**, W209-W213.

32. Edelheit, O., Hanukoglu, A. and Hanukoglu, I. (2009) Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol*, **9**, 61-61.

33. Anagnostopoulos, C. and Spizizen, J. (1961) REQUIREMENTS FOR TRANSFORMATION IN BACILLUS SUBTILIS. *J Bacteriol*, **81**, 741-746.

34. Niaudet, B., Goze, A. and Ehrlich, S.D. (1982) Insertional mutagenesis in Bacillus subtilis: mechanism and use in gene cloning. *Gene*, **19**, 277-284.

35. Wandrey, G., Wurzel, J., Hoffmann, K., Ladner, T., Büchs, J., Meinel, L. and Lühmann, T. (2016) Probing unnatural amino acid integration into enhanced green fluorescent protein by genetic code expansion with a high-throughput screening platform. *J Biol Eng*, **10**, 11-11.

36. Guan, C., Cui, W., Cheng, J., Zhou, L., Guo, J., Hu, X., Xiao, G. and Zhou, Z. (2015) Construction and development of an auto-regulatory gene expression system in Bacillus subtilis. *Microbial Cell Factories*, **14**, 150.

37. Bhavsar, A.P., Zhao, X. and Brown, E.D. (2001) Development and Characterization of a Xylose-Dependent System for Expression of Cloned Genes in&lt;em&gt;Bacillus subtilis&lt;/em&gt;: Conditional Complementation of a Teichoic Acid Mutant. *Applied and Environmental Microbiology*, **67**, 403.

38. Maul, B., Völker, U., Riethdorf, S., Engelmann, S. and Hecker, M. (1995) σB-dependent regulation ofgsiB in response to multiple stimuli inBacillus subtilis. *Molecular and General Genetics MGG*, **248**, 114-120.

39. Kim, L., Mogk, A. and Schumann, W. (1996) A xylose-inducible Bacillus subtilis integration vector and its application. *Gene*, **181**, 71-76.

40. Zyprian, E.V.A. and Matzura, H. (1986) Characterization of Signals Promoting Gene Expression on the Staphylococcus aureus Plasmid pUB110 and Development of a Gram-Positive Expression Vector System. *DNA*, **5**, 219-225.

41. Wong, J.T. (1983) Membership mutation of the genetic code: loss of fitness by tryptophan. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 6303-6306.

42. Rocha, E.P.C., Danchin, A. and Viari, A. (1999) Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic acids research*, **27**, 3567-3576.

43. Zhang, Q. and Liu, W. (2013) Biosynthesis of thiopeptide antibiotics and their pathway engineering. *Nat Prod Rep*, **30**, 218-226.

44. Hegde, N.S., Sanders, D.A., Rodriguez, R. and Balasubramanian, S. (2011) The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. *Nat Chem*, **3**, 725-731.

45. Westers, L., Westers, H. and Quax, W.J. (2004) Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1694**, 299-310.

46. Simonen, M. and Palva, I. (1993) Protein secretion in Bacillus species. *Microbiological Reviews*, **57**, 109-137.

47. Moks, T., Abrahmsen, L., Olsson, A., Nilsson, B., Holmgren, E., Pohl, G., Sterky, C., Hultberg, H., Josephson, S., Bilich, M. *et al.* (1987) Expression of Human

Insulin-like Growth Factor I in Bacteria: Use of Optimized Gene Fusion Vectors To Facilitate Protein Purification. *Biochemistry*, **26**, 5239-5244.

## 2.7 Appendix 2

### 2.7.1 Supplementary figures.
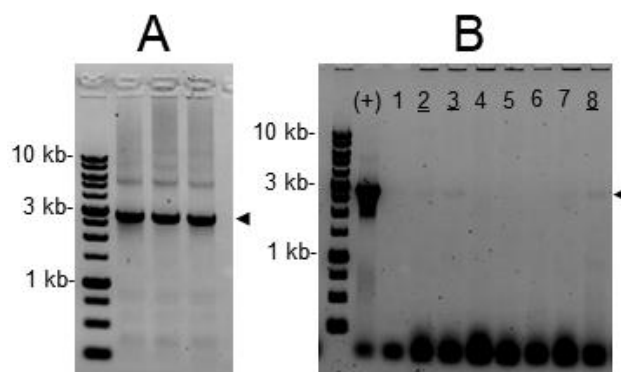


**Figure S2.1 DNA Gel image showing the progress and results of the restriction free cloning method used to insert the OTS into pDR111_GFP.** A) Restriction free cloning primers pDR111_site3_fwd and pDR111_site3_rev (Table 2.1) were used to amplify the AcKRS OTS (see Section 2.7.2) from the pUC57 vector in which the sequence was delivered (Genewiz, South Plainfield, NJ, USA). PCR reaction protocol is outlined in Methods (see Section 2.2.2). The three lanes represent 3 replicates of the same reaction. The major band, denoted with the black arrowhead, just below the 3 kb marker of the ladder corresponds to the intended AcKRS OTS product (2759 bp). These 3 bands were extracted and the DNA within was used as a primer in the secondary PCR reaction (Section 2.2.2) to insert the AcKRS OTS into pDR111_GFP. B) A PCR screen (Section 2.2.6) was conducted on *E. coli* colonies that were transformed with the final restriction free cloning product (Section 2.2.2). The lane denoted by (+) was the positive control screen that used pUC57_AcKRS as a template instead of bacterial cells. Lanes 1-8 represent colonies screened. Underlined lanes (2, 3, and 8) were positive, indicated by the presence of a bar at the expected AcKRS band size (2759 bp). Plasmid DNA was prepped from these colonies, sequenced, and found to contain pDR111_AcKRS_GFP.
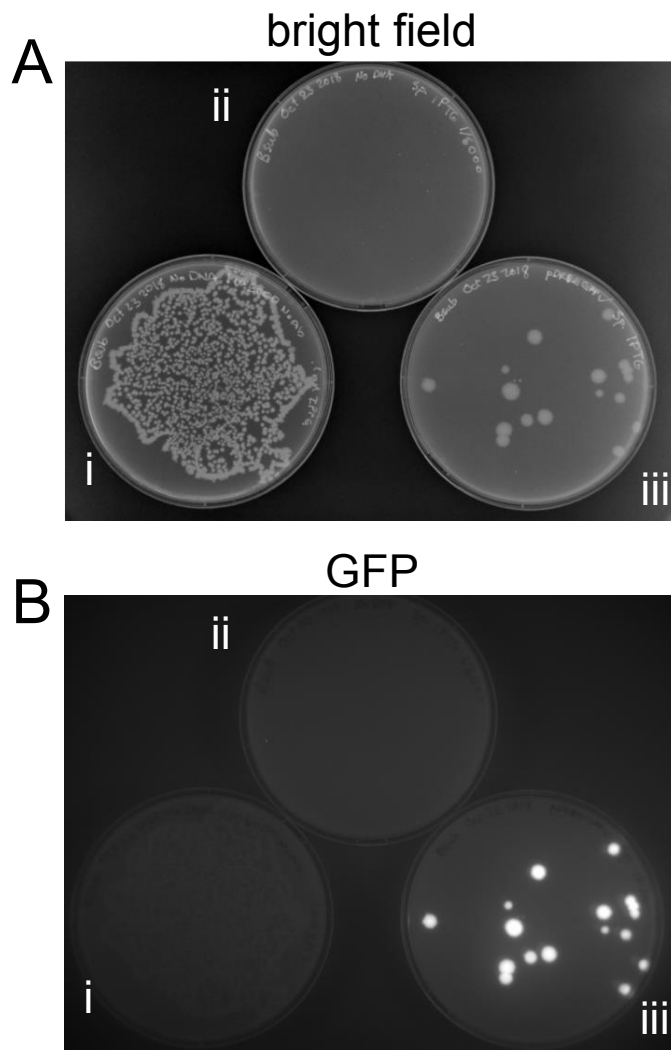
**Figure S2.2. Successful transformation and GFP expression in *B. subtilis*.** Competent cell preparations of *B. subtilis* 168 were plated and treated as follows: i) a stock of competent *B. subtilis* was diluted and plated directly onto LB agar supplemented with 0.1 mM IPTG and no antibiotics, ii) a stock of competent *B. subtilis* was diluted and plated directly onto LB agar supplemented with 0.1 mM IPTG and 100 μg/ml spectinomycin, iii) a stock of competent *B. subtilis* was transformed with pDR111_AckRS_GFP and plated on LB agar supplemented with 0.1 mM IPTG and 100 μg/ml spectinomycin. BioRad ChemiDoc imager settings for A: Filter: None, Light: White Trans illumination, White Light Conversion Screen, and for B: Filter: 530/38 nm, Light: Blue Epi illumination.

## 2.7.2 Sequences of plasmids and genes

Complete pDR111_AcKRS_GFP (11599 bp)

```
AACAAAATTCTCCAGTCTTCACATCGGTTTGAAAGGAGGAAGCGGAAGAATGAAGTAAGA
GGGATTTTTGACTCCGAAGTAAGTCTTCAAAAAATCAAATAAGGAGTGTCAAGAATGTTT
GCAAAACGATTCAAAACCTCTTTACTGCCGTTATTCGCTGGATTTTTATTGCTGTTTCAT
TTGGTTCTGGCAGGACCGGCGGCTGCGAGTGCTGAAACGGCGAACAAATCGAATGAGCTT
ACAGCACCGTCGATCAAAAGCGGAACCATTCTTCATGCATGGAATTGGTCGTTCAATACG
TTAAAACACAATATGAAGGATATTCATGATGCAGGATATACAGCCATTCAGACATCTCCG
ATTAACCAAGTAAAGGAAGGGAATCAAGGAGATAAAAGCATGTCGAACTGGTACTGGCTG
TATCAGCCGACATCGTATCAAATTGGCAACCGTTACTTAGGTACTGAACAAGAATTTAAA
GAAATGTGTGCAGCCGCTGAAGAATATGGCATAAAGGTCATTGTTGACGCGGTCATCAAT
CATACCACCAGTGATTATGCCGCGATTTCCAATGAGGTTAAGAGTATTCCAAACTGGACA
CATGGAAACACACAAATTAAAAACTGGTCTGATCGAAATAGTACATAATGGATTTCCTTA
CGCGAAATACGGGCAGACATGGCCTGCCCGGTTATTATTATTTTTGACACCAGACCAACT
GGTAATGGTAGCGACCGGCGCTCAGGATCCTAACTCACATTAATTGCGTTGCGCTCACTG
CCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCG
GGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGG
CAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCT
GGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTGACGGCGGGATATAACATGA
GCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGA
CTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGT
GGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCA
GTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCC
AGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTG
GTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAAT
AATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCA
GGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACT
GACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTC
TACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGAC
AATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTG
TTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGC
TTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAAC
GGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCATCAA
AATCGTCTCCCTCCGTTTGAATATTTGATTGATCGTAACCAGATGAAGCACTCTTTCCAC
TATCCCTACAGTGTTATGGCTTGAACAATCACGAAACAATAATTGGTACGTACGATCTTT
CAGCCGACTCAAACATCAAATCTTACAAATGTAGTCTTTGAAAGTATTACATATGTAAGA
TTTAAATGCAACCGTTTTTTCGGAAGGAAATGATGACCTCGTTTCCACCGAATTAGCTTG
CATGCAAAGAATCTTGCTTGGCAAGGTTCTTTTTTTTGTCTTTTCAAAAAAGCACGAAAA
AAGCACCCATAAGGTAAACTTTTGAGTGCTTTGCTATAAATACGCTTCACAGTTTCTTCT
TCATTTTTAACAAACAAAAAACCCAGCTCATTGAGCTGGGTTTAAGCTGGCTTGGCGGCG
TCCTAGTACTGCAGAATTGCCGACCTTGACTAGTGCTCATTATTATTTATACAATTCATC
CATACCATGTGTAATACCAGCAGCTGTAACAAATTCCAACAAAACCATGTGATCACGTTT
```

```
TTCATTTGGATCTTTAGACAATTTAGATTGTGTAGACAAATAATGATTATCTGGCAACAA
AACTGGACCATCACCAATTGGTGTATTTTGTTGATAATGATCAGCCAATTGAACAGAACC
ATCTTCAATATTATGACGAATTTTAAAATTAACTTTAATACCGTTCTTTTGTTTATCAGC
CATGATATAAACATTATGAGAATTATAATTATATTCCAATTTATGACCCAAGATATTACC
ATCTTCTTTAAAATCGATACCCTTCAATTCGATACGATTAACCAATGTATCACCTTCAAA
CTTAACCTCAGCACGTGTCTTATAATTACCATCATCCTTGAAAAGATTGTACGTTCTTG
AACATAACCTTCTGGCATAGCAGATTTAAAGAAATCATGTTGTTTCATGTGATCTGGATA
ACGAGCAAACATTGCAAACCATAAGCAAATGTTGTAACCAATGTTGGCCAAGGAACTGG
CAATTTACCAGTTGTACAAATAAATTTCAATGTCAATTTACCGTATGTAGCATCACCTTC
ACCTTCACCAGAAACAGAAATTTATGACCATTAACATCACCATCCAATTCAACCAAAAT
TGGAACAACACCTGTAAACAATTCTTCACCTTTAGAAACCATGTTTGTCCTCCTTATTAG
TTAATCAGCTAGCTGTCGACTAAGCTTAATTGTTATCCGCTCACAATTACACACATTATG
CCACACCTTGTAGATAAAGTCAACAACTTTTGCAAAATGAATTGTGAGTGCTCACATTTA
CCCTCGAGCAACGTTCTTGCCATTGCTGCATAAAAAACGCCCGGCGGCAACCGAGCGTTC
TGAATTAATTAATCATCGGGAAGATCTTCATCACCGAAACGCGGCAGGCAGCTCTAGAGT
TAACAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGATGGCCTTCTGCTTAGCTA
GAGCGGCGGATTTGTCCTACTCAGGAGAGCGTTCACCGACAAACAACAGATAAAACGAAA
GGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTTGATGCCTCAAGCTAGAGAGTCGAAT
TCCTGCAGCCCTGGCGAATGGCGATTTTCGTTCGTGAATACATGTTATAATAACTATAAC
TAATAACGTAACGTGACTGGCAAGAGATATTTTTAAAACAATGAATAGGTTTACACTTAC
TTTAGTTTTATGGAAATGAAAGATCATATCATATATAATCTAGAATAAAATTAACTAAAA
TAATTATTATCTAGATAAAAAATTTAGAAGCCAATGAAATCTATAAATAAACTAAATTAA
GTTTATTTAATTAACAACTATGGATATAAAATAGGTACTAATCAAAATAGTGAGGAGGAT
ATATTTGAATACATACGAACAAATTAATAAAGTGAAAAAAATACTTCGGAAACATTTAAA
AAATAACCTTATTGGTACTTACATGTTTGGATCAGGAGTTGAGAGTGGACTAAAACCAAA
TAGTGATCTTGACTTTTTAGTCGTCGTATCTGAACCATTGACAGATCAAAGTAAAGAAAT
ACTTATACAAAAAATTAGACCTATTTCAAAAAAAATAGGAGATAAAAGCAACTTACGATA
TATTGAATTAACAATTATTATTCAGCAAGAAATGGTACCGTGGAATCATCCTCCCAAACA
AGAATTTATTTATGGAGAATGGTTACAAGAGCTTTATGAACAAGGATACATTCCTCAGAA
GGAATTAAATTCAGATTTAACCATAATGCTTTACCAAGCAAACGAAAAAATAAAAGAAT
ATACGGAAATTATGACTTAGAGGAATTACTACCTGATATTCCATTTTCTGATGTGAGAAG
AGCCATTATGGATTCGTCAGAGGAATTAATAGATAATTATCAGGATGATGAAACCAACTC
TATATTAACTTTATGCCGTATGATTTTAACTATGGACACGGGTAAAATCATACCAAAAGA
TATTGCGGGAAATGCAGTGGCTGAATCTTCTCCATTAGAACATAGGGAGAGAATTTTGTT
AGCAGTTCGTAGTTATCTTGGAGAGAATATTGAATGGACTAATGAAAATGTAAATTTAAC
TATAAACTATTTAAATAACAGATTAAAAAAATTATAAAAAAATTGAAAAAATGGTGGAAA
CACTTTTTTCAATTTTTTTGTTTTATTATTTAATATTTGGGAAATATTCATTCTAATTGG
TAATCAGATTTTAGAAACAATAAACCCTTGCATAGGGGGATCATCCGTTTAGGCTGGGC
GGTGATAGCTTCTCGTTCAGGCAGTACGCCTCTTTTCTTTTCCAGACCTGAGGGAGGCGG
CTCGAGACTAGTAAGATTATGATCTTCTGAACGACGAAACTCGATTGAAAATATTGTAG
ATCAAATTGCTTCAGTTATTCATGATAATCAAAAAAGTGAATCTCAGTCGAGATTCACT
TTTTCTTTAAAATAAAAAATCCGCCGCACCGAATGCAACGGACAAAACAATCTCCAATAT
ATGCTTGTGTTTCATTAGAATTAAAATCAATAACCGGTTGGCGGAAACGTGGGGAATCTA
ACCCCACTGAACGGATTTAGAGTCCATTCGATCTACATGATCACGTTTCCCCTAGGGTAT
```

```
CGACAAGACTAACTATATCACGTATGTTCTCGTGTGGTCAAGCATTTTTAAAAACTTTTT
TTGTTTTTATGAAAATGCCGTGTTTACATAAGCAAAAAGGCGCACCTTGTCAGGCACGCC
TTTCCATAATTTACGGTTTCATTACTTCTCTATTCCCCATATAAGGACGAAGCACTTTCG
GAATCACAACGCTTCCGGCCGGCCAAAGCGTTGATTTGGGTAGTATGCTTTGACACAACA
AATTTTAATTTAGCAAATTCGATAGTCAACTCATTCTTAAGACCTAAATTAATGTTATTT
TTTAATAATTTACACCAAATTAATAGCAAAATTATGTTATTCGTGCTAATATTTCATAG
TTGGTTATTCAATTAATTAAAAATAAGTCAAAATGCACAACTTTTTATAATTCATTGAGT
CGAGTTTGAAAAATAAAAGTGCTTTAATGCATGATCAATTATCGTACTTTCTATTATTTG
TTACCCGTTATCAATCGGAATAACGTATAGACACTTTAACGTGCTATAGATTGGTTTTAA
TCACTAAATTAATGTGTTTTTCTTATCATTAAAACTGCACTGAGAATTACTAAATTAAAA
AAATTATAAAAATTTTTCATTTTTAGTGATAAAATTCTGAAAATGGGTATAAATAGTAG
AAGAAGTTAACTTGGAAGAGTTAAGCTATAACAAAGAATCTCTTTAGACACACATTGAAT
ATCGAAACATTTAATTGCGCTAAATCGTTTCATTAAATAAATTACCTTGTATTGTCGATT
AAATTAAGGTAAATTATAAAAAATGCTGATATTTTTGACTAAACCAAATGCTAACCCAGA
AATACAATCACTGTGTCTAATGAATAATTTGTTTTATAAACACTTTTTTGTTTACTTCTC
ATTTTTAATTAGTTATAATTAACTAAATAATAGAGCATTAAATATATTTAATAAAACTTA
TTTAATGCAAAATTATGACTAACATATCTATAATAAATAAAGATTAGATATCAATATATT
ATCGGGCAAATGTATCGAGCAAGATGCATCAAATAGGGAGGTTTTAAACTCTAGAATGGA
TAAAAAACCTCTTGACGTTCTTATCTCTGCTACAGGCCTTTGGATGTCTCGTACAGGCAC
ACTTCATAAATCAAACATTACGAAATCTCTCGTTCTAAAATCTACATCGAAATGGCTTG
CGGCGATCATCTTGTTGTTAACAACTCTCGTTCTTGCCGTCCTGCTCGTGCTTTCCGTTA
CCATAAATACCGTAAAACATGCAAACGTTGCCGTGTTTCTGGCGAAGATATCAACAACTT
CCTTACACGTTCTACAGAAGGCAAAACATCTGTTAAAGTTAAAGTTGTTTCTGAACCTAA
AGTTAAAAAAGCTATGCCTAAATCTGTTTCTCGTGCTCCTAAACCTCTTGAAAACCCTGT
TTCTGCTAAAGCTTCTACAGATACATCTCGTTCTGTTCCTTCTCCTGCTAAATCTACACC
TAACTCTCCTGTTCCTACATCTGCTTCTGCTCCTGCTCTTACAAAATCTCAAACAGATCG
TCTTGAAGTTCTTCTTAACCCTAAAGATGAAATCTCTCTTAACTCTGGCAAACCTTTCCG
TGAACTTGAATCTGAACTTCTTTCTCGTCGTAAAAAAGATCTTCAACAAATCTACGCTGA
AGAACGTGAAAACTACCTTGGCAAACTTGAACGTGAAATCACACGTTTCTTCGTTGATCG
TGGCTTCCTTGAAATCAAATCTCCTATCCTTATCCCTCTTGAATACATCGAACGTATGGG
CATCGATAACGATACAGAACTTTCTAAACAAATCTTCCGTGTTGATAAAAACTTCTGCCT
TCGTCCTATGATGGCTCCTAACATCTTCAACTACGCTCGTAAACTTGATCGTGCTCTTCC
TGATCCTATCAAAATCTTCGAAATCGGCCCTTGCTACCGTAAAGAATCTGATGGCAAAGA
ACATCTTGAAGAATTCACAATGCTTAACTTCTTCCAAATGGGCTCTGGCTGCACACGTGA
AAACCTTGAATCTATCATCACAGATTTCCTTAACCATCTTGGCATCGATTTCAAAATCGT
TGGCGATTCTTGCATGGTTTACGGCGATACACTTGATGTTATGCATGGCGATCTTGAACT
TTCTTCTGCTGTTGTTGGCCCTATCCCTCTTGATCGTGAATGGGGCATCGATAAACCTTG
GATCGGCGCTGGCTTCGGCCTTGAACGTCTTCTTAAAGTTAAACATGATTTCAAAAACAT
CAAACGTGCTGCTCGTTCTGAATCTTACTACAACGGCATCTCTACAAACCTTTAAGCGGC
CGCATCACTGAAGCAAACAACGAAATTGAACTATAATTTTGTTTAGCGCAATTTGGTGAA
GTTTGATAGATGATACATTCTATTAAACTTCCTTTTTTTATGCTCTTTTTACCTAATTGT
TAAGAGGTTTTGCACTAATGGCACTCTGCAGCTCGAGAATGGTGTGAGGTTCCCGGGGAA
AAGCCAAATAGGCGATCGCGGGAGTGCTTTATTTGAAGATCAGGCTATCACTGCGGTCAA
TAGATTTCACAATGTGATGGCTGGACAGCCTGAGGAACTCTCGAACCCGAATGGAAACAA
```

```
CCAGATATTTATGAATCAGCGCGGCTCACATGGCGTTGTGCTGGCAAATGCAGGTTCATC
CTCTGTCTCTATCAATACGGCAACAAAATTGCCTGATGGCAGGTATGACAATAAAGCTGG
AGCGGGTTCATTTCAAGTGAACGATGGTAAACTGACAGGCACGATCAATGCCAGGTCTGT
AGCTGTGCTTTATCCTGATGATATTGCAAAAGCGCCTCATGTTTTCCTTGAGAATTACAA
AACAGGTGTAACACATTCTTTCAATGATCAACTGACGATTACCTTGCGTGCAGATGCGAA
TACAACAAAAGCCGTTTATCAAATCAATAATGGACCAGACGACAGGCGTTTAAGGATGGA
GATCAATTCACAATCGGAAAAGGAGATCCAATTTGGCAAAACATACACCATCATGTTAAA
AGGAACGAACAGTGATGGTGTAACGAGGACCGAGAAATACAGTTTTGTTAAAAGAGATCC
AGCGTCGGCCAAAACCATCGGCTATCAAAATCCGAATCATTGGAGCCAGGTAAATGCTTA
TATCTATAAACATGATGGGAGCCGAGTAATTGAATTGACCGGATCTTGGCCTGGAAAACC
AATGACTAAAAATGCAGACGGAATTTACACGCTGACGCTGCCTGCGGACACGGATACAAC
CAACGCAAAAGTGATTTTTAATAATGGCAGCGCCCAAGTGCCCGGTCAGAATCAGCCTGG
CTTTGATTACGTGCTAAATGGTTTATATAATGACTCGGGCTTAAGCGGTTCTCTTCCCCA
TTGAGGGCAAGGCTAGACGGGACTTACCGAAAGAAACCATCAATGATGGTTTCTTTTTTG
TTCATAAATCAGACAAAACTTTTCTCTTGCAAAGTTTGTGAAGTGTTGCACAATATAAA
TGTGAAATACTTCACAAACAAAAGACATCAAAGAGAAACATACCCTGCAAGGATGCTGA
TATTGTCTGCATTTGCGCCGGAGCAAACCAAAAACCTGGTGAGACACGCCTTGAATTAGT
AGAAAAGAACTTGAAGATTTTCAAAGGCATCGTTAGTGAAGTCATGGCGAGCGGATTTGA
CGGCATTTTCTTAGTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGC
TTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGA
CCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGG
ACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATG
GATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAG
CCGGGCCACCTACTGAAGTGGATTTCTTTAAGAGCTCCTTTAACTTCCTCACCAGTAGTT
GTATCGGTACCATAAGTAGAAGCAGCAACCCAAGTAGCTTTACCAGCATCCGGTTCAACC
AGCATAGTAAGAATCTTACTGGACATCGGCAGTTCTTCGAACAGTGCGCCAACTACCAGC
TCTTTCTGCAGTTCATTCAGGGCACCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCA
TGCGAAACGATCCTCATCCTGTCTCTTGATCCATGGATTACGCGTTAACCCGGGCCCGCG
GATGCATATGATCAGATCTTAAGGCCTAGGTCTAGAGTCTTTGTTTTGACGCCATTAGCG
TACGTAACAATCCTCGTTAAAGGACAAGGACCTGAGCGGAAGTGTATCGTACAGTAGACG
GAGTATACTAGTATAGTCTATAGTCCGTGGAATTATTATATTTATCTCCGACGATATTCT
CATCAGTGAAATCCAGCTGGAGTTCTTTAGCAAATTTTTTTATTAGCTGAACTTAGTATT
AGTGGCCATACTCCTCCAATCCAAAGCTATTTAGAAAGATTACTATATCCTCAAACAGGC
GGTAACCGGCCTCTTCATCGGGAATGCGCGCGACCTTCAGCATCGCCGGCATGTCCCCCT
GGCGGACGGGAAGTATCCAGCTCGAGGTCGGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC
AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCC
GACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTC
TCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG
TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAG
CAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTA
CACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAG
AGTTGATAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG
```

CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC
GGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATC
AAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTC
AGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAGGCATCGTGGTGTC
ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
TGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTG
AGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGC
GCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACT
CTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA
TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTT
TCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATG
TATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAGTGCCACCTGA
CGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCC
CTTTCGTCTTCAAGAATT


(Sp)GFP (reverse strand, 720 bp)

TTATTTATACAATTCATCCATACCATGTGTAATACCAGCAGCTGTAACAAATTCCAACAA
AACCATGTGATCACGTTTTTCATTTGGATCTTTAGACAATTTAGATTGTGTAGACAAATA
ATGATTATCTGGCAACAAAACTGGACCATCACCAATTGGTGTATTTTGTTGATAATGATC
AGCCAATTGAACAGAACCATCTTCAATATTATGACGAATTTTAAAATTAACTTTAATACC
GTTCTTTTGTTTATCAGCCATGATATAAACATTATGAGAATTATAATTATATTCCAATTT
ATGACCCAAGATATTACCATCTTCTTTAAAATCGATACCCTTCAATTCGATACGATTAAC
CAATGTATCACCTTCAAACTTAACCTCAGCACGTGTCTTATAATTACCATCATCCTTGAA
AAAGATTGTACGTTCTTGAACATAACCTTCTGGCATAGCAGATTTAAAGAAATCATGTTG
TTTCATGTGATCTGGATAACGAGCAAAACATTGCAAACCATAAGCAAATGTTGTAACCAA
TGTTGGCCAAGGAACTGGCAATTTACCAGTTGTACAAATAAATTTCAATGTCAATTTACC
GTATGTAGCATCACCTTCACCTTCACCAGAAACAGAAAATTTATGACCATTAACATCACC
ATCCAA**<u>TTC</u>**AACCAAAATTGGAACAACACCTGTAAACAATTCTTCACCTTT**<u>AGA</u>**AACCAT

*Bold underlined codons (AGA and TTC) indicate the residues that were mutated to become TAG codons. AGA: S3, TTC: E18.


Complete orthogonal translation system (2797 bp)

```
CTCGAGACTAGTAAGATTATGATCTTCTGAACGACGAAACTCGATTGAAAATATTGTAG
ATCAAATTGCTTCAGTTATTCATGATAATCAAAAAAGTGAATCTCAGTCGAGATTCACT
TTTTCTTTAAAATAAAAAATCCGCCGCACCGAATGCAACGGACAAAACAATCTCCAATAT
ATGCTTGTGTTTCATTAGAATTAAAATCAATAACCGGTTGGCGGAAACGTGGGGAATCTA
ACCCCACTGAACGGATTTAGAGTCCATTCGATCTACATGATCACGTTTCCCCTAGGGTAT
CGACAAGACTAACTATATCACGTATGTTCTCGTGTGGTCAAGCATTTTTAAAAACTTTTT
TTGTTTTTATGAAAATGCCGTGTTTACATAAGCAAAAAGGCGCACCTTGTCAGGCACGCC
TTTCCATAATTTACGGTTTCATTACTTCTCTATTCCCCATATAAGGACGAAGCACTTTCG
GAATCACAACGCTTCCGGCCGGCCAAAGCGTTGATTTGGGTAGTATGCTTTGACACAACA
AATTTTAATTTAGCAAATTCGATAGTCAACTCATTCTTAAGACCTAAATTAATGTTATTT
TTTAATAATTTACACCAAATTAATAGCAAAAATTATGTTATTCGTGCTAATATTTCATAG
TTGGTTATTCAATTAATTAAAAATAAGTCAAAATGCACAACTTTTTATAATTCATTGAGT
CGAGTTTGAAAAATAAAAGTGCTTTAATGCATGATCAATTATCGTACTTTCTATTATTTG
TTACCCGTTATCAATCGGAATAACGTATAGACACTTTAACGTGCTATAGATTGGTTTTAA
TCACTAAATTAATGTGTTTTTCTTATCATTAAAACTGCACTGAGAATTACTAAATTAAAA
AAATTATAAAAATTTTTCATTTTTAGTGATAAAATTCTGAAAAATGGGTATAAATAGTAG
AAGAAGTTAACTTGGAAGAGTTAAGCTATAACAAAGAATCTCTTTAGACACACATTGAAT
ATCGAAACATTTAATTGCGCTAAATCGTTTCATTAAATAAATTACCTTGTATTGTCGATT
AAATTAAGGTAAATTATAAAAAATGCTGATATTTTTGACTAAACCAAATGCTAACCCAGA
AATACAATCACTGTGTCTAATGAATAATTTGTTTTATAAACACTTTTTTGTTTACTTCTC
ATTTTTAATTAGTTATAATTAACTAAATAATAGAGCATTAAATATATTTAATAAAACTTA
TTTAATGCAAAATTATGACTAACATATCTATAATAAATAAAGATTAGATATCAATATATT
ATCGGGCAAATGTATCGAGCAAGATGCATCAAATAGGGAGGTTTTAAACTCTAGAATGGA
TAAAAAACCTCTTGACGTTCTTATCTCTGCTACAGGCCTTTGGATGTCTCGTACAGGCAC
ACTTCATAAAATCAAACATTACGAAATCTCTCGTTCTAAAATCTACATCGAAATGGCTTG
CGGCGATCATCTTGTTGTTAACAACTCTCGTTCTTGCCGTCCTGCTCGTGCTTTCCGTTA
CCATAAATACCGTAAAACATGCAAACGTTGCCGTGTTTCTGGCGAAGATATCAACAACTT
CCTTACACGTTCTACAGAAGGCAAAACATCTGTTAAAGTTAAAGTTGTTTCTGAACCTAA
AGTTAAAAAAGCTATGCCTAAATCTGTTTCTCGTGCTCCTAAACCTCTTGAAAACCCTGT
TTCTGCTAAAGCTTCTACAGATACATCTCGTTCTGTTCCTTCTCCTGCTAAATCTACACC
TAACTCTCCTGTTCCTACATCTGCTTCTGCTCCTGCTCTTACAAAATCTCAAACAGATCG
TCTTGAAGTTCTTCTTAACCCTAAAGATGAAATCTCTCTTAACTCTGGCAAACCTTTCCG
TGAACTTGAATCTGAACTTCTTTCTCGTCGTAAAAAAGATCTTCAACAAATCTACGCTGA
AGAACGTGAAAACTACCTTGGCAAACTTGAACGTGAAATCACACGTTTCTTCGTTGATCG
TGGCTTCCTTGAAATCAAATCTCCTATCCTTATCCCTCTTGAATACATCGAACGTATGGG
CATCGATAACGATACAGAACTTTCTAAACAAATCTTCCGTGTTGATAAAACTTCTGCCT
TCGTCCTATGATGGCTCCTAACATCTTCAACTACGCTCGTAAACTTGATCGTGCTCTTCC
TGATCCTATCAAAATCTTCGAAATCGGCCCTTGCTACCGTAAAGAATCTGATGGCAAAGA
ACATCTTGAAGAATTCACAATGCTTAACTTCTTCCAAATGGGCTCTGGCTGCACACGTGA
AAACCTTGAATCTATCATCACAGATTTCCTTAACCATCTTGGCATCGATTTCAAAATCGT
TGGCGATTCTTGCATGGTTTACGGCGATACACTTGATGTTATGCATGGCGATCTTGAACT
TTCTTCTGCTGTTGTTGGCCCTATCCCTCTTGATCGTGAATGGGGCATCGATAAACCTTG
GATCGGCGCTGGCTTCGGCCTTGAACGTCTTCTTAAAGTTAAACATGATTTCAAAAACAT
CAAACGTGCTGCTCGTTCTGAATCTTACTACAACGGCATCTCTACAAACCTTTAAGCGGC
```

CGCATCACTGAAGCAAACAACGAAATTGAACTATAATTTTGTTTAGCGCAATTTGGTGAA
GTTTGATAGATGATACATTCTATTAAACTTCCTTTTTTTATGCTCTTTTTACCTAATTGT
TAAGAGGTTTTGCACTAATGGCACTCTGCAGCTCGAG


*SerT* Promoter (within the OTS sequence; reverse strand, 200 bp)

GTATCGACAAGACTAACTATATCACGTATGTTCTCGTGTGGTCAAGCATTTTTAAAAACT
TTTTTTGTTTTTATGAAAATGCCGTGTTTACATAAGCAAAAAGGCGCACCTTGTCAGGCA
CGCCTTTCCATAATTTACGGTTTCATTACTTCTCTATTCCCCATATAAGGACGAAGCACT
TTCGGAATCACAACGCTTCC

*SerT* (within the OTS sequence; reverse strand, 72 bp)

TGGCGGAAACGTGGGGAATCTAACCCCACTGAACGGATTTAGAGTCCATTCGATCTACAT
GATCACGTTTCC)

*SerT* Terminator (within the OTS sequence; reverse strand, 200 bp)

AAGATTATGATCTTCTGAACGACGAAAACTCGATTGAAAATATTGTAGATCAAATTGCTT
CAGTTATTCATGATAATCAAAAAAAGTGAATCTCAGTCGAGATTCACTTTTTCTTTAAAA
TAAAAAATCCGCCGCACCGAATGCAACGGACAAAACAATCTCCAATATATGCTTGTGTTT
CATTAGAATTAAAATCAATA


*sarA* Promoter (within the OTS sequence; forward strand, 865 bp)

AAAGCGTTGATTTGGGTAGTATGCTTTGACACAACAAATTTTAATTTAGCAAATTCGATA
GTCAACTCATTCTTAAGACCTAAATTAATGTTATTTTTTAATAATTTACACCAAATTAAT
AGCAAAAATTATGTTATTCGTGCTAATATTTCATAGTTGGTTATTCAATTAATTAAAAAT
AAGTCAAAATGCACAACTTTTTATAATTCATTGAGTCGAGTTTGAAAATAAAAGTGCTT
TAATGCATGATCAATTATCGTACTTTCTATTATTTGTTACCCGTTATCAATCGGAATAAC
GTATAGACACTTTAACGTGCTATAGATTGGTTTTAATCACTAAATTAATGTGTTTTTCTT
ATCATTAAAACTGCACTGAGAATTACTAAATTAAAAAAATTATAAAAATTTTTCATTTTT
AGTGATAAAATTCTGAAAATGGGTATAAATAGTAGAAGAAGTTAACTTGGAAGAGTTAA
GCTATAACAAGAATCTCTTTAGACACACATTGAATATCGAAACATTTAATTGCGCTAAA
TCGTTTCATTAAATAAATTACCTTGTATTGTCGATTAAATTAAGGTAAATTATAAAAAAT
GCTGATATTTTTGACTAAACCAAATGCTAACCCAGAAATACAATCACTGTGTCTAATGAA
TAATTTGTTTTATAAACACTTTTTTGTTTACTTCTCATTTTTAATTAGTTATAATTAACT
AAATAATAGAGCATTAAATATATTTAATAAAACTTATTTAATGCAAAATTATGACTAACA
TATCTATAATAAATAAGATTAGATATCAATATATTATCGGGCAAATGTATCGAGCAAGA
TGCATCAAATAGGGAGGTTTTAAAC

*sarA* (within the OTS sequence; forward strand, 1260 bp)

ATGGATAAAAAACCTCTTGACGTTCTTATCTCTGCTACAGGCCTTTGGATGTCTCGTACA
GGCACACTTCATAAAATCAAACATTACGAAATCTCTCGTTCTAAAATCTACATCGAAATG

```
GCTTGCGGCGATCATCTTGTTGTTAACAACTCTCGTTCTTGCCGTCCTGCTCGTGCTTTC
CGTTACCATAAATACCGTAAAACATGCAAACGTTGCCGTGTTTCTGGCGAAGATATCAAC
AACTTCCTTACACGTTCTACAGAAGGCAAAACATCTGTTAAAGTTAAAGTTGTTTCTGAA
CCTAAAGTTAAAAAAGCTATGCCTAAATCTGTTTCTCGTGCTCCTAAACCTCTTGAAAAC
CCTGTTTCTGCTAAAGCTTCTACAGATACATCTCGTTCTGTTCCTTCTCCTGCTAAATCT
ACACCTAACTCTCCTGTTCCTACATCTGCTTCTGCTCCTGCTCTTACAAAATCTCAAACA
GATCGTCTTGAAGTTCTTCTTAACCCTAAAGATGAAATCTCTCTTAACTCTGGCAAACCT
TTCCGTGAACTTGAATCTGAACTTCTTTCTCGTCGTAAAAAAGATCTTCAACAAATCTAC
GCTGAAGAACGTGAAAACTACCTTGGCAAACTTGAACGTGAAATCACACGTTTCTTCGTT
GATCGTGGCTTCCTTGAAATCAAATCTCCTATCCTTATCCCTCTTGAATACATCGAACGT
ATGGGCATCGATAACGATACAGAACTTTCTAAACAAATCTTCCGTGTTGATAAAAACTTC
TGCCTTCGTCCTATGATGGCTCCTAACATCTTCAACTACGCTCGTAAACTTGATCGTGCT
CTTCCTGATCCTATCAAAATCTTCGAAATCGGCCCTTGCTACCGTAAAGAATCTGATGGC
AAAGAACATCTTGAAGAATTCACAATGCTTAACTTCTTCCAAATGGGCTCTGGCTGCACA
CGTGAAAACCTTGAATCTATCATCACAGATTTCCTTAACCATCTTGGCATCGATTTCAAA
ATCGTTGGCGATTCTTGCATGGTTTACGGCGATACACTTGATGTTATGCATGGCGATCTT
GAACTTTCTTCTGCTGTTGTTGGCCCTATCCCTCTTGATCGTGAATGGGGCATCGATAAA
CCTTGGATCGGCGCTGGCTTCGGCCTTGAACGTCTTCTTAAAGTTAAACATGATTTCAAA
AACATCAAACGTGCTGCTCGTTCTGAATCTTACTACAACGGCATCTCTACAAACCTTTAA
```

*sarA* (within the OTS sequence; forward strand, 141 bp)

```
TCACTGAAGCAAACAACGAAATTGAACTATAATTTTGTTTAGCGCAATTTGGTGAAGTTT
GATAGATGATACATTCTATTAAACTTCCTTTTTTTATGCTCTTTTTACCTAATTGTTAAG
AGGTTTTGCACTAATGGCACT
```

Curriculum Vitae

| | |
|---|---|
| Name: | McShane McKenna |
| Post-Secondary Education: | Western University, London, Ontario, Canada MSc in Biochemistry Sept 2017 – May 2020 |
| | Western University, London, Ontario, Canada BMSc Honours Specialization in Biochemistry of Infection and Immunity Sept 2013 – April 2017 |
| Honours and Awards: | Dean's List Sept 2013 – April 2017 |
| | Western Scholar April 2017 |
| Related Work Experience: | Summer Research Assistant May 2016 – Sept 2017 |
| | Teaching Assistant – Biochemistry 4420A Sept 2017 – Dec 2017 Sept 2018 – Dec 2018 |
| | Teaching Assistant – Biology 1001A/2b Jan 2018 – May 2018 Sept 2018 – Dec 2019 Jan 2019 – May 2019 Sept 2019 – Dec 2019 Jan 2020 – May 2020 |
| Mentoring: | Melanie Offman Undergraduate Honours student Sept 2018 – April 2019 |
| Presentations: | 102nd Canadian Chemistry Conference and Exhibition "Pioneering the genetic code expansion of *Bacillus subtilis*" |
| Publications: | Balasuriya, N., McKenna, M., Liu, X., Li, S.S.C. and O'Donoghue, P. (2018) Phosphorylation-Dependent Inhibition of Akt1. Genes (Basel), 9. |