Fall 10-1-2021

Optimization of Orthogonal Translation Systems Enhances Access to the Human Phosphoproteome

Jack Michael Moen
Yale University Graduate School of Arts and Sciences, jack.moen@yale.edu

Abstract

Optimization of Orthogonal Translation Systems Enhances Access to the Human Phosphoproteome

Jack Michael Moen

2021

Protein phosphorylation is a ubiquitous post-translational modification that governs signaling cascades and protein-protein interactions. Orthogonal translation systems, in which an orthogonal aminoacyl-tRNA synthetase and tRNA pair have been repurposed for insertion of a non-standard amino acid, have been developed for co-translational insertion of phospho-amino acids. Here, we systematically developed variants of phosphoserine and phosphothreonine orthogonal translation systems to minimize cellular toxicity and maximize translational fidelity. We characterized novel tRNA constructs to enhance cellular fitness and change decoding functionality. Multiple genetic background strains were developed to enhance the incorporation of phospho-amino acids into recombinant proteins. Utilizing large-scale DNA synthesis, we produced serine, threonine, and tyrosine libraries representative of the entire known human phosphoproteome. By pairing our libraries with a bimolecular fluorescence complementation assay, we characterized novel phosphorylation-dependent protein-protein interactions for phosphoserine, phosphothreonine, and phosphotyrosine. Genetically encoded phosphothreonine enabled the discovery of a new activation mechanism for the protein kinase CHK2 and proteome-wide surveys of its target substrates. Finally, we developed Hi-P+ to couple kinase substrate discovery with phosphorylation-dependent protein-protein interactions. This

work enables kinase-specific, proteome-wide surveys of multi-level phosphorylation-dependent interactions with phosphosite resolution.

Optimization of Orthogonal Translation Systems Enhances Access to the Human Phosphoproteome

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Jack Michael Moen

Dissertation Director: Jesse Rinehart, Ph.D.

December 2021

Dedication

This document is dedicated to my wife, Alexandra.

I would not be here today without your constant love and support in my life.

Nothing is impossible with you at my side.

Acknowledgments

For the science: The only reason I have come so far is because of the great mentors I have had in my life. I would like to thank Dr. Lisa Julian for teaching me that science requires hard work and discipline to succeed. I would also like to thank Dr. Jerry Schaack for teaching me that science is a creative endeavor, first and foremost. Dr. Ed Lakatta for cultivating scientific curiosity and mentoring me through the complexities of the world within. Dr. Jesse Rinehart for pushing me to produce the highest quality work. Without his mentorship, this document would not be possible. More than once, his enthusiasm for science helped me to reframe negative results into positive progress forward. Most importantly, I would like to thank Dr. Kyle Mohler for working with me every day down in the trenches. Without his mentorship, I would have never been pushed to do the type of science I now know I am capable of doing. I would also like to thank my thesis committee, Dr. Benjamin Turk, Dr. Farren Isaacs, and Dr. Susumu Tomita, for their guidance and help during this project. I would also like to thank everyone that has come through both Jesse Rinehart's and Farren Isaac's lab during my time at Yale. These interactions enriched my life and scientific knowledge.

For everything else: First and foremost, I would like to thank my wife Alexandra for her love and support during my graduate career. There were highs and lows over the last few years, but she was always my constant. For my father, whose intellectual curiosity and love of puzzles still resonates within me today. For my mother and all the times, she stayed up with me as a child to help me learn to read and write. For my brothers, we had many struggles throughout life, but I know I am a better person for having you both there to challenge me. Lastly, I would like to thank my extended family for their love and support. I know my grandparents would have been proud to see how far I have come.

Table of Contents

List of Figures & Tables

Chapter 1: Introduction

Protein translation

The central dogma of molecular biology states that DNA is transcribed into RNA and then translated into proteins. In this manner, these three codes are the basis for all life on Earth. DNA and RNA consist of a 4-base nucleic acid code that results in a final complexity of 20 core amino acids in cells. RNA is read as 3-nucleotide codons to address this increase in complexity, giving rise to a potential 64 possible combinations (Fig 1-1). The way the cell accomplishes the conversion of nucleic acids to amino acids is through protein translation. Protein translation is a multi-step process involving the concerted action of RNA and protein across the cell (Fig 1-2). Specifically, the process starts when an aminoacyl-tRNA synthetase (aaRS) binds to a transfer RNA (tRNA) and aminoacylates it with a cognate amino acid. The elongation factor (EF) binds the aminoacylated tRNA to facilitate its movement and entry into the ribosome. The ribosome, a complex structure of ribosomal RNA and proteins, catalyzes amide bonds between amino acids to produce proteins. This occurs when the aminoacyl-tRNA enters the aminoacyl (A) site of the ribosome, checking for the proper base pairing between the three base tRNA anticodon and the three base mRNA codon. If the tRNA effectively binds the correct sequence, it enters into the peptidyl (P) site of the ribosome. Within the P site, the ribosome enzymatically attaches the amino acid to the growing peptide chain. The tRNA then enters the exit (E) site, where it can dissociate from the ribosomal complex. This process proceeds forward until the ribosome encounters a stop codon, UAG, UGA, or UAA. When this occurs in *E. coli*, one of two main release factors (RF1 or RF2) enter the ribosome, causing termination of the protein sequence and dissociation of the ribosomal complex. Each RF has unique properties; for instance, RF1 facilitates release at UAG and UAA codons but requires the

help of RF3 to release from the ribosomal complex [1]. Conversely, RF2 terminates translation at UGA and UAA codons but does not require RF3 to facilitate its release. Through this complex process, the cell can make up the majority of protein effectors essential for biological functions.

Figure 1-1| The 64 codons of the genetic code

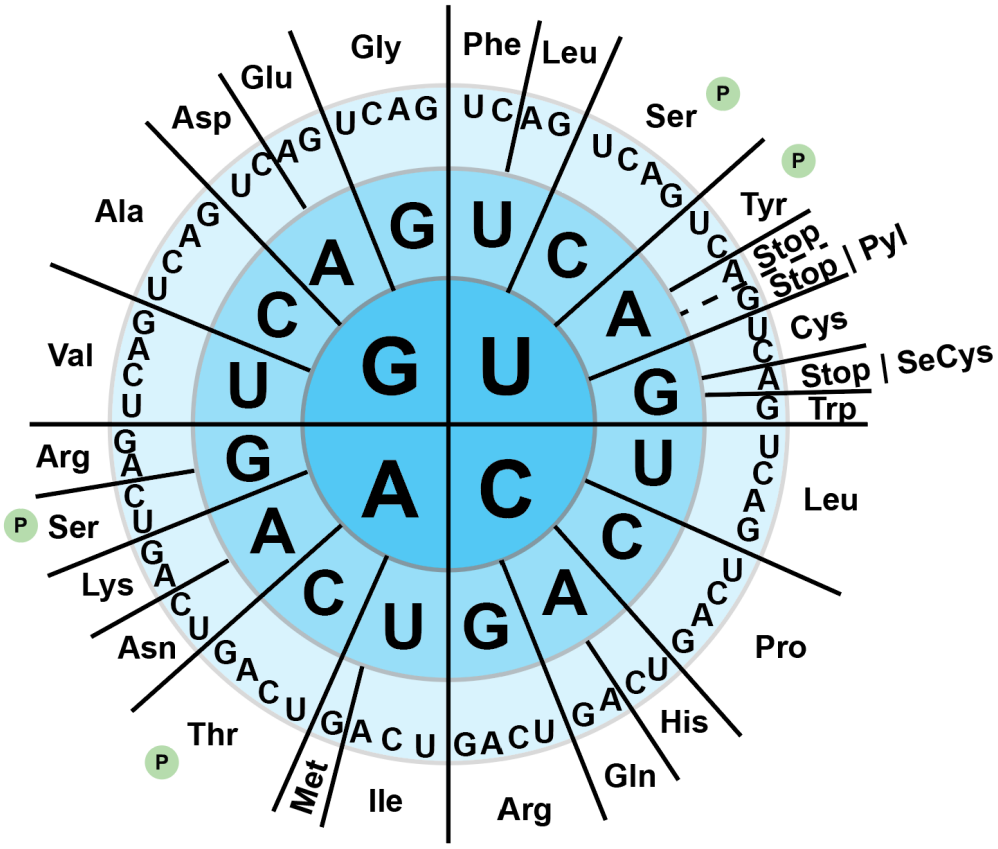

Figure 1-1| The genetic code exists as 64 possible codon positions. The twenty standard amino acids are encoded by 61 codons, and 3 stop codons. The stop codons may also encode sense codons as is the case for Pyl (UAG) and SeCys (UGA). Amino acids that are phosphorylatable are denoted with a P.

Figure 1-2| The process of protein translation

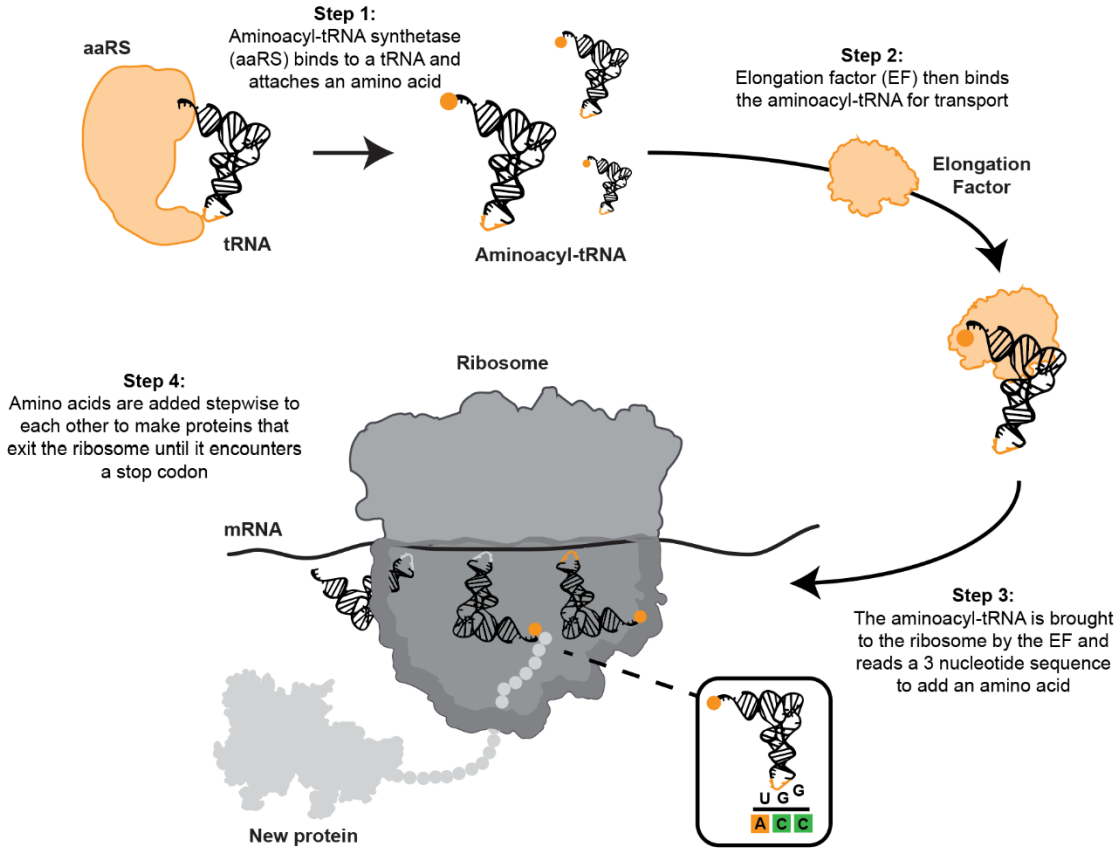

Figure 1-2| Protein translation is a concerted process requiring the interaction of several enzymes, tRNAs, and the ribosomal complex.

Most estimates for the human genome approximate the complexity around 20,000 genes [2], but the array of proteins making up the proteome is far more incredible. The vast majority of this complexity is due to alternative splicing and the addition of reversible, site-specific post-translational modifications (PTMs). The most ubiquitous reversible PTM is protein phosphorylation, with estimates reaching ~230,000 unique sites across the human proteome [3]. Mechanisms of phosphorylation-dependent regulation occur when a protein kinase attaches a phosphate molecule from ATP to the hydroxyl group of serine (Ser), threonine (Thr), or tyrosine (Tyr) to make phosphoserine (pSer), phosphothreonine (pThr), or phosphotyrosine (pTyr) within a protein (Fig. 1-3). This mechanism can be rapidly reversed by phosphatases, enzymes that remove the phosphate group by hydrolysis. Although protein kinases only encompass ~500 genes [4], they account for most characterized PTMs across the proteome. Advances in proteomics have increased our ability to identify phosphorylation sites (phosphosites), yielding an expansive map of the human phosphoproteome. However, the identification of these sites has outpaced our ability to connect them to underlying interactions. Connecting phosphorylation sites to function across the proteome has been challenging due to a lack of comprehensive information regarding the identity of upstream activating kinases, the inherently dynamic nature of kinase and phosphatase effectors, and the modulation of PPIs by non-catalytic phospho-binding proteins. Protein phosphorylation also occurs on other amino acids, such as histidine (His), aspartic acid (Asp), and cysteine (Cys), but these modifications are generally confined to prokaryotes and may only exist as a tiny fraction within eukaryotes [5-7]. The phosphorylation of proteins can facilitate many different cellular functions,

including changes in localization, complexation, and activity. For instance, many protein kinases require site-specific phosphorylation of their activation loop to facilitate conformational changes to an active state [8, 9]. Likewise, many protein-protein interactions (PPIs) are dependent on phosphorylation. This is increasingly recognized with phospho-binding proteins, such as the 14-3-3 proteins, and their role in sequestering phospho-proteins [10-12]. These regulatory complexities underscore the need for technological advances that facilitate the rapid investigation of phosphorylation-dependent signaling on a systems-wide scale.

Figure 1-3| Mechanism of protein phosphorylation

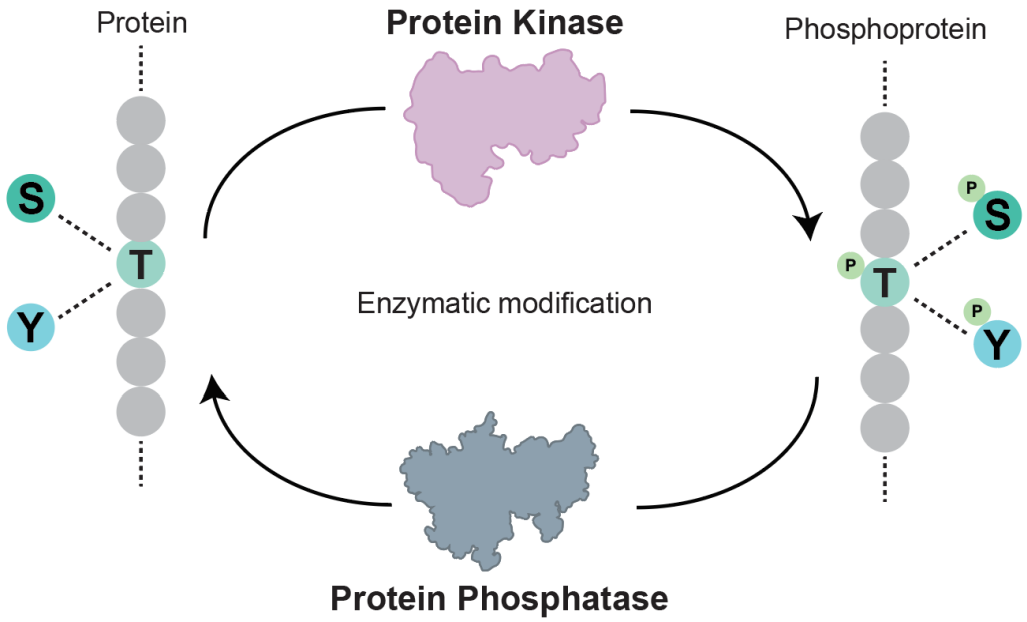

Figure 1-3| Protein phosphorylation occurs when a protein kinase attaches a phosphate molecule to Ser, Thr, or Tyr residues. The process can then be reversed in a hydrolysis dependent manner by protein phosphatases.

Suppressor tRNAs

Orthogonal translation systems (OTSs) in which an orthogonal aminoacyl-tRNA synthetase (aaRS) and tRNA pair have been repurposed for co-translational insertion of a modified amino acid is a combination of several previously validated technologies. The first discovery that enabled future OTS work was the suppressor tRNA in 1965 [13, 14]. Suppressor tRNAs occur naturally when a mutation in the anticodon causes the tRNA to recognize and decode a stop codon, leading to nonsense suppression (Fig 1-4). This finding provided the first evidence that a stop codon could be repurposed to incorporate an amino acid. Since then, several other naturally occurring *E. coli* suppressor tRNAs have been identified [15, 16]. Since the initial discovery, we have come to understand that only a particular subset of tRNAs can function as suppressors.

The ability of a tRNA to become a suppressor depends on how the identity elements of the tRNA are recognized by the aaRS, if the aaRS will still maintain aminoacylation efficiency and fidelity, and whether or not mutations in the anticodon are stable enough to persist in a population. This has been well studied with Ser suppression of UAG codons by the SupD tRNA [15]. It has since been repurposed for several applications in genetic code expansion [17]. The serine aaRS is particularly amenable to interacting with Ser suppressor tRNAs because it lacks an anticodon binding domain. In many other synthetases, such as the threonyl-tRNA synthetase, the anticodon binding domain binds to the anticodon recognizing either ACA, ACC, ACG, or ACU. If the anticodon sequence is incorrect, i.e., does not start with the sequence AC, it will not bind the tRNA. Conversely, the seryl-tRNA synthetase relies upon an enhanced variable loop (Fig 1-4) to determine tRNA identity.

This subtle difference allows for $tRNA^{Ser}$ to contain a broader range of anticodon identities, as evidenced by Ser decoding six different codons (Fig 1-1).

The variable nature of the anticodon is also due, in part, to the 'wobble' of tRNAs in decoding mRNA sequences [18]. During protein translation, the first two nucleotides of the tRNA anticodon interact with the corresponding codon of the mRNA sequence in the ribosome; as long as the first two codons pair correctly, the third codon does not have to pair perfectly. This third codon facilitated 'wobble' in the tRNA allows for decoding of mRNA codons that are not an exact match. Although still an area of study, it appears that post-transcriptional modifications to the tRNA play a key role in facilitating this wobble, or lack thereof. For instance, it is known that specific tRNAs, such as *E. coli* $tRNA^{Phe}$, have adenosine at position 37 that becomes post-transcriptionally modified to 2-methylthio-N6 isopentenyl adenosine [19]. This modification stabilizes the tRNA anticodon within the ribosome and decreases frameshifting, thereby enhancing fidelity. Insights such as these have been instrumental in improving the fidelity and function of tRNAs in OTSs.

Figure 1-4| Overview of $tRNA^{Ser}$

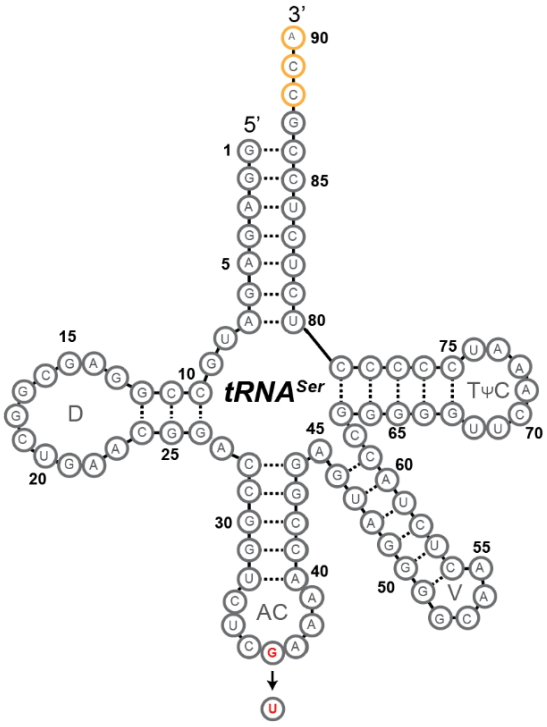

Figure 1-4| The Ser tRNA contains an enlogated variable loop (V) that serves as major recognition element by SerRS. Changes to the anticodon (red) allow for Ser incorporation at UAG codons.

Incorporation of non-standard amino acids

The other significant advancement that led to the development of OTSs was recognizing that non-standard amino acids (nsAAs) could be incorporated into proteins within a cell. The first instance of this was by Cowie and Cohen in 1957, in which it was discovered that selenomethionine, a methionine (Met) derivative, could be incorporated at Met codons in a Met auxotrophic *E. coli* strain [20]. Following this initial discovery, there was a dramatic increase in the number of nsAAs that were found to replace canonical amino acids [21]. One major limitation of this work was an inability to retain native protein function. For instance, one of the earliest experiments demonstrated that although it was possible to replace His with 2-methylhistidine in a His auxotrophic strain, isolated alkaline phosphatase was inactive [22]. Other studies performed in a similar manner were able to retain some functionality in alkaline phosphatase through misincorporation of low levels of fluorotyrosine [23]. In general, only a tiny portion of incorporated nsAAs could retain protein function and those that did generally had a deleterious effect on growth and fitness. In order to mitigate these effects, experiments on the incorporation of nsAAs occurred in organisms that could tolerate mistranslation mediated proteome stress, such as *E. coli, B. subtilis,* or bacteriophages [21, 24].

The first evidence that tRNAs could be made to incorporate different amino acids selectively occurred in 1962 when it was shown that cysteinyl-$tRNA^{Cys}$ could be chemically converted to alanyl-$tRNA^{Cys}$ and still function as the same donor tRNA during ribosomal protein synthesis *in vitro* [25]. Traditional methods involving aminoacylation of near cognate amino acids relied upon mutant or evolutionarily divergent aaRSs; however, more recent work has relied upon the flexizyme system developed by the Suga group [26].

The flexizyme system makes use of a ribozyme (flexizyme) that catalyzes tRNA aminoacylation through recognition of the 3'-CCA of tRNAs and transfer of a chemically activated amino acid [26-28]. Since its development, it has been applied to numerous standard and nsAAs in various contexts [29, 30].

The advent of orthogonal translation systems

Most organisms utilize the same genetic code; that is, 64 possible codons read in a relatively identical manner across kingdoms. However, several observations have given rise to the idea of genetic code expansion in organisms. Repurposing the genetic code occurs every day in human cells, where mitochondrial $tRNA^{Met}$ must decode both AUG and AUA codons; however, cytosolic $tRNA^{Met}$ only decodes AUG, with AUA reserved for $tRNA^{Ile}$ [31]. The more well-known instance of an altered genetic code across all kingdoms of life is the selenocysteine (Sec) system, in which Ser-$tRNA^{Sec}$ is first phosphorylated by phosphoseryl-tRNA kinase to generate pSer-$tRNA^{Sec}$ that is then converted to Sec-$tRNA^{Sec}$ by Sec synthase [32]. The aminoacylated $tRNA^{Sec}$ only decodes UGA stop codons containing mRNA structural elements that recruit the Sec-specific elongation factor. The highly regulated complex of mRNA and EFs ensures that translational insertion of Sec only occurs at a specific subset of UGA-containing codons. In light of the complex regulatory mechanisms associated with Sec, it stands to reason that it plays an essential role in physiology across prokaryotes and eukaryotes. Current work indicates that Sec is used in place of catalytic Cys residues for redox reactions, as Sec results in higher enzyme efficiencies [33]. Although Sec has particular regulatory elements required for its insertion, the 22nd amino acid, pyrrolysine (Pyl), appears to be more promiscuous. Having only been

found in some methanogenic microorganisms, it is believed that the electrophilic imine bond of Pyl is necessary for the methyl transferase activity of some enzymes [34, 35]. Pyrrolysyl-aminoacyl tRNA synthetase aminoacylates Pyl directly onto $tRNA^{Pyl}$ for incorporation at UAG codons, independent of any complex regulatory elements [36]. Although the incorporation mechanism is unclear, it appears that many of these methanogens have fewer UAG stop codons throughout their genome when compared to methanogenic archaea. In this sense, nature has performed the first genome-wide recoding to enable readthrough of UAG as a sense codon.

The Schultz group created the first OTSs to use UAG nonsense suppression and nsAAs [37-39]. It was previously established that $tRNA^{Tyr}$ could be easily converted to a suppressor tRNA by modifying the anticodon to UAG, thus enabling stop codon suppression [40]. The Schultz group built on this by finding an evolutionarily divergent organism, in this case, the archaebacterium *Methanococcus jannaschii*, that had evolved an orthogonal Tyr tRNA and aaRS pair that would have minimal cross-reactivity with native *E. coli* machinery [39]. The $tRNA^{Tyr}$ of *M. jannaschii* has a discriminator base, i.e., an identity element required for correct aminoacylation, in the acceptor stem of G:C, while *E. coli* $tRNA^{Tyr}$ has a C:G in the same position. This minor difference is a significant identity element for TyrRS that varies between *M. jannaschii* and *E. coli*. The other benefit to using *M. jannaschii* TyrRS is that it lacks an editing domain [41], meaning that aminoacylation of a non-cognate amino acid would not be removed by the aaRS. Using site-directed mutagenesis, the Schultz group was able to evolve the active site of the *M. jannaschii* TyrRS to allow for incorporation of the nsAA O-methyl-L-tyrosine at UAG codons in *E. coli*.

The incorporation of nsAAs into *E. coli* proteins were a significant step forward for the field, but it was not until this was expanded to pSer that OTSs could be used for probing phosphorylation-dependent biological questions. The pSerOTS was the first OTS capable of incorporating a nsAA to produce a recombinant post-translationally modified protein. It relied upon a phosphoseryl-tRNA synthetase (pSerRS) from *Methanococcus maripluidus* to aminoacylate pSer onto a modified UAG $tRNA^{Cys}$ suppressor from *M. jannaschii* [42]. As the $tRNA^{Cys}$ was not the native substrate for pSerRS, it required an additional mutation of C20U to enhance aminoacylation activity. Additionally, *E. coli* elongation factor Tu (EF-Tu) was evolved to accommodate the negative charge associated with pSer to generate EF-pSer. To provide a robust pool of pSer substrate, the *E. coli* phosphoserine phosphatase, serB, was deleted from the genome. It is known that pSer is produced endogenously in *E. coli* during the serine biosynthesis pathway, in which serC converts 3-phosphooxypyruvate to pSer, which is converted to Ser via serB [43]. This evolutionarily conserved pathway has been essential for expanding the genetic code to pSer in *E. coli* and human cells alike [17, 42, 44, 45]. Unfortunately, the first pSerOTS suffered from low fidelity and poor efficiency due to problems with substrate specificity and incomplete orthogonality. The first solution to these problems came through modification of the OTS, in which pSerRS, $tRNA^{pSer}$, EF-pSer were evolved for enhanced fidelity [46]. These enhanced variants (pSerRS9 and EF-pSer21) became the basis for further OTS developments in our lab [17].

OTSs have since been developed to facilitate the incorporation of phosphotyrosine (pTyr) [47-49] and phosphothreonine (pThr) [50], greatly expanding our understanding of phosphoprotein function. Iterations of pTyrOTSs utilize *M. jannaschii* TyrRS and $tRNA^{Tyr}$ pair but have proven particularly challenging to develop high fidelity OTSs for. The earliest

reported instance used a five-phosphatase knockout strain to facilitate pTyr import into the cell but resulted in low yields of pTyr containing proteins [47]. Later systems overcame the issue by using a Lys-pTyr dipeptide to enhance pTyr import [49]. By using the dipeptide ABC transporter periplasmic binding protein (DppA), it is possible to import an impermeable amino acid across the periplasmic space of *E. coli* through linkage to a standard amino acid [51, 52]. Interestingly, the pTyrRS used in these experiments was not evolved for pTyr, but rather the non-hydrolyzable analogue p-carboxymethyl-l-phenylalanine (CMF) [53]. Although pTyr mimetics cross the *E. coli* periplasmic space without modification, they are often poor substitutes for pTyr [54, 55]. Other groups have tried to circumvent these problems by chemically protecting the phospho-moiety of pTyr until incorporation into a protein of interest [48]. The major flaw in this design is that it requires an extracted protein to be treated to low pH conditions (~2) for up to 36 hours to remove the protecting group.

Despite the success of pSer, and to a lesser extent pTyr, in the production of recombinant phosphoproteins in *E. coli,* only one group, to date, has made progress towards a functional pThrOTS. The Chin group first reported on a pThrOTS in 2017 [50], with little to no follow-up from the community-at-large. The pThrOTS was enabled by two advancements: 1) an evolved pSerRS to accommodate pThr and 2) use of the *Salmonella typhimurium* enzyme pduX to catalyze the conversion Thr to pThr. Although the pThrOTS was developed several years ago, it has yet to be deployed for functional studies of protein phosphorylation.

Recoded Organisms

Although organisms with intact Pyl systems make use of the UAG codon as a sense codon, they have a built-in redundancy of an in-frame UGA or UAA stop codon to accommodate read-through [36]. However, it was believed that UAG read-through of essential genes in *E. coli* was causing an OTS-specific growth defect. To address this challenge, a genomically recoded *E. coli* strain (C321.ΔA) was developed in which all 321 instances of the TAG codon were recoded to TAA (Fig 1-5) [56]. These mutations enabled deletion of RF1 (ΔA), thereby decreasing competition for nsAA incorporation at UAG codons. Recoding efforts have further progressed, with the most ambitious efforts centering around a 57 codon *E. coli*, missing the UAG stop codon, as well as two codons for Arg, Leu, and Ser [57]. Although the 57-codon organism was never fully completed, the Chin group produced a 61-codon organism (Syn61) via large-scale DNA synthesis techniques [58]. The Syn61 organism is missing the UAG stop codon, as well as the Ser codons TCG and TCA. An evolved version of Syn61 was later demonstrated to utilize multiple OTSs simultaneously, allowing for high-fidelity site-specific incorporation of three nsAAs simultaneously [59]. Future recoding efforts are underway for more complex organisms such as yeast [60-63] but still require substantial effort for completion.

Despite these advancements, there is still a lack of knowledge surrounding the physiological impact of OTS deployment in *E. coli* that has hindered performance. To expand our understanding of the contribution of host processes to OTS function, we systematically optimized the architecture and expression of pThrOTS and pSerOTS components while monitoring the impact on cellular viability. Using previously characterized aaRS, tRNA, and EF components, we developed rationally designed and

optimized OTSs capable of facilitating high levels of phosphoamino acid incorporation. Together with several strain innovations that maximized the production of phosphoproteins, we were able to perform system-wide screens phospho-mediated protein-protein interactions (PPIs). Using these technologies, we characterized a novel mechanism of activation for checkpoint kinase 2 (CHK2). Utilizing a novel $tRNA^{Ala}$ suppressor, we demonstrated that phospho-ablative mutations may obfuscate complex phosphorylation-dependent interactions for CHK2 activation. Coupling our ability to produce active kinases with our platform to identify phosphorylation-dependent PPIs, we profiled the overlap between CHK2 substrate phosphorylation and 14-3-3β phosphorylation-dependent interactions. Overall, our work demonstrates the utility of application-driven phospho-amino acid-based OTS development and highlights how genetic code expansion to the human phosphoproteome allows for the facile characterization of multi-level kinase and phospho-binding domain-dependent PPIs.

Figure 1-5| Codon reassignment enables genetic code expansion

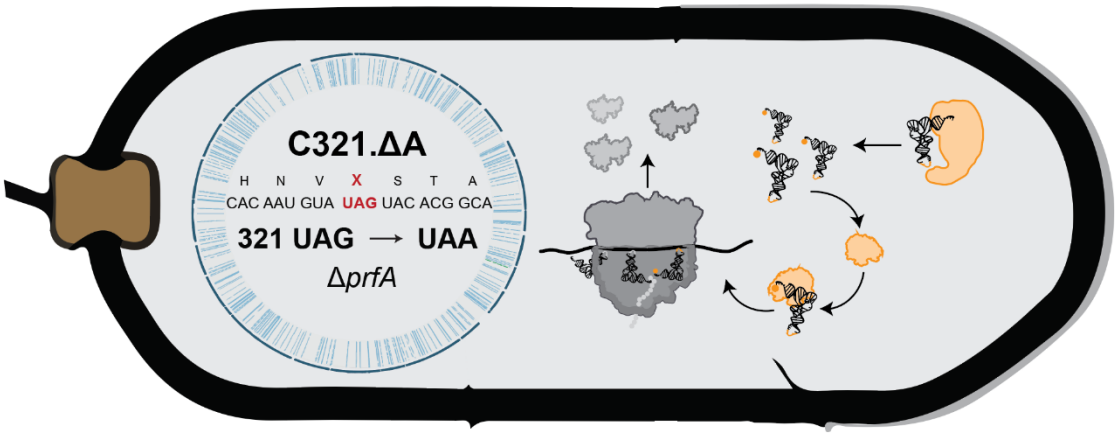

Figure 1-5| Recoding of the *E. coli* genome to remove UAG codons allows for deletion of RF1 (*prfA*), and frees UAG for sense codon reassignment.

Chapter 2: Optimizing the incorporation of phospho-amino acids in *E. coli*

Refining orthogonal translation component expression

Establishing a robust pThrOTS

To determine baseline functionality of the initial pThrOTS components, we first incorporated them into our previously developed OTS framework (pSerOTSλ) consisting of the pThr aminoacyl-tRNA synthetase (pThrRS) and pSer elongation factor 1 (EF-pSer1) under control of a high-level constitutive TRC promoter. [17, 42, 50]. EF-pSer1, along with a further optimized variant (EF-pSer21), were evolved from a bacterial elongation factor (EF-Tu) to facilitate efficient delivery of negatively charged phospho-aminoacyl-tRNA to the ribosome [42, 46]. To create a stable intracellular pool of pThr substrate, the pThrOTS system relies upon the bacterial enzyme pduX to catalyze the ATP-dependent conversion of free L-threonine to pThr [50, 64]. Our initial pThrOTS framework placed pduX under the control of the arabinose inducible promoter (pBAD). The expression of tRNA was controlled by a constitutive lpp promoter and was further modified to be insulated by surrounding terminators. We conducted our preliminary pThrOTS optimizations in a C321.ΔA strain with a deletion in *serC* (C321ΔserC), the enzyme which converts metabolic precursors to pSer during Ser biosynthesis [65]. We reasoned that limiting the intracellular levels of pSer would decrease competition for pThr at the pThrRS active site. Using this framework, we systematically modified OTS and vector components to identify their impact on the fitness of the host cell.

We used growth rate as a proxy for cellular fitness and OTS toxicity to pre-filter deleterious OTS variants from our initial pool of 41 OTSs (Table 1), selecting 12 for further analysis. Growth rate data identified antibiotic selection and promoter strength as

significant determinants of OTS toxicity (Fig 2-1A). To assess pThr incorporation facilitated by the remaining OTS variants, we utilized our Mass Spectrometry Reporter for Exact Amino acid Decoding (MS-READ) reporter protein [66]. The reporter is a 6x His tagged, TAG-containing elastin-like polypeptide, green fluorescent protein fusion that can be used to monitor relative pThr-protein expression by immunoblot and quantify amino acid incorporation at UAG codons by MS. To qualitatively assess OTS function, we expressed the reporter protein with each OTS variant and analyzed pThr-protein production by Phos-Tag-SDS-PAGE (which separates proteins based on phosphorylation status) with western blot for the 6x His tag. Samples from the same reporter proteins were then analyzed by MS to quantitatively assess pThr incorporation relative to other amino acids incorporated at the UAG codon. Combining these data with cellular fitness metrics, we observed that growth rate reduction, caused by increased pThrRS promoter strength, was inversely correlated with reporter expression and pThr incorporation (Fig 2-1A). Because promoter activity emerged as an essential mediator of pThr incorporation, we wanted to assess the impact of promoter strength variation on pduX expression and pThr production. We constructed pduX expression vectors with diverse transcriptional promoters and measured the levels of pduX protein and pThr metabolites by LC-MS/MS (Fig 2-1B). We observed that the TRC promoter (Fig 2-1B, **P4**) consistently facilitated the highest pduX expression and pThr production (Fig 2-1B). While transcriptional promoter strength was identified as a determinant of OTS toxicity and performance, additional elements of OTS architecture appeared to impact host cellular fitness. These observations were collectively used to inform design principles for our optimized pThrOTS variant. The optimized variant

(pThrOTS^{Zeus}) placed *pduX* in a polycistronic operon with *pThrRS* and *EF-pSer21* under transcriptional control of the TRC promoter (Fig 2-1C).

Table 2-1| Table of pThrOTSs tested

OTS	RS Promoter	tRNA	pduX Promoter	Antibiotic	EF-pSer Variant	Doubling Time (min)
Q66	TRC	Standard	pBAD	Tetracycline	EF-pSer1	62.68
Q77 (O1)	TRC	Insulated	pBAD	Tetracycline	EF-pSer1	67.76
R23 (O7)	glnS*	Insulated	pBAD	Tetracycline	EF-pSer1	54.90
R29 (O8)	glnS*	Insulated	OXB20	Tetracycline	EF-pSer1	46.08
S29	TRC	Standard	pBAD	Kanamycin	EF-pSer1	49.72
S30 (O2)	TRC	Insulated	pBAD	Kanamycin	EF-pSer1	47.20
S37	TRC	-	pBAD	Tetracycline	EF-pSer1	66.43
S38	TRC	-	pBAD	Kanamycin	EF-pSer1	41.98
S55	TRC	Standard	pBAD	Tetracycline	EF-pSer21	56.99
S56 (O3)	TRC	Insulated	pBAD	Tetracycline	EF-pSer21	74.46
S57	TRC	Standard	pBAD	Kanamycin	EF-pSer21	59.36
S58 (O4)	TRC	Insulated	pBAD	Kanamycin	EF-pSer21	47.83
S62	-	-	pBAD	Tetracycline	-	57.69
S63	-	Insulated	pBAD	Tetracycline	-	47.41
S64	-	-	pBAD	Kanamycin	-	46.70
S65	-	Insulated	pBAD	Kanamycin	-	43.50
S66 (O5)	glnS*	Insulated	pBAD	Tetracycline	EF-pSer21	55.25
S67 (O6)	glnS*	Insulated	OXB20	Tetracycline	EF-pSer21	47.30
S68	TRC	Standard	OXB20	Tetracycline	EF-pSer1	37.33
S69	TRC	Insulated	OXB20	Tetracycline	EF-pSer1	53.07
S70	TRC	Standard	-	Tetracycline	EF-pSer1	54.28
S71	TRC		-	Tetracycline	EF-pSer1	37.36
T1	TRC	Standard	OXB20	Kanamycin	EF-pSer1	49.36
T2	-	Insulated	-	Tetracycline	-	48.56
T3	-	Insulated	-	Kanamycin	-	37.59
T5	TRC	Insulated	-	Kanamycin	EF-pSer1	38.01
T6	-	Standard	pBAD	Tetracycline	-	51.25
T7	-	Standard	pBAD	Kanamycin	-	60.69
T8	TRC	-	-	Kanamycin	EF-pSer1	54.02
T16	TRC	-	OXB20	Kanamycin	EF-pSer1	71.59
T17	TRC	Standard	OXB20	Kanamycin	EF-pSer21	72.71
T18	TRC	Insulated	OXB20	Kanamycin	EF-pSer21	68.76
T20	TRC	Standard	OXB20	Tetracycline	EF-pSer21	43.35
T21	TRC	Insulated	OXB20	Tetracycline	EF-pSer21	73.63
T22	glnS*	Insulated	OXB20	Tetracycline	EF-pSer21	48.37
T24 (O9)	glnS*	Insulated	pBAD	Kanamycin	EF-pSer1	44.05
T25 (O10)	glnS*	Insulated	OXB20	Kanamycin	EF-pSer1	46.67
T28	TRC	-	-	Kanamycin	EF-pSer21	43.88
T29 (O11)	glnS*	Insulated	pBAD	Kanamycin	EF-pSer21	48.38
T30 (O12)	glnS*	Insulated	OXB20	Kanamycin	EF-pSer21	51.16
T31	TRC	-	-	Tetracycline	EF-pSer21	36.96

Table 2-1| List of initial OTSs tested. Plasmid name is given in the first column, with those appearing in Fig 2-1 given an additional name in parathensis. Doubling time (minutes) is provided in the last column for reference.

Figure 2-1| Screening pThrOTSs informs optimal design elements

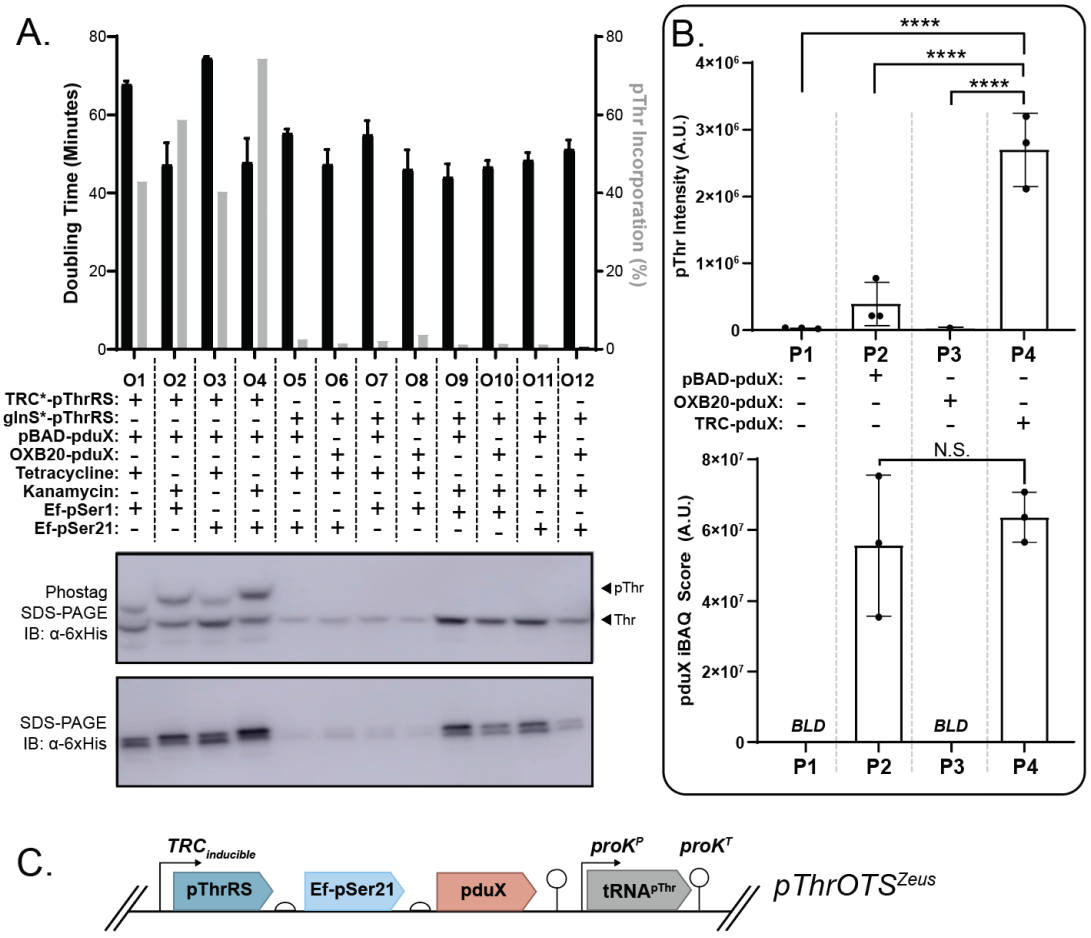

Figure 2-1| Screening pThrOTSs informs optimal design elements. (A) Top, chart displaying cell doubling time (left axis, black, N=3, error bars represent 1 STD) and percent pThr incorporation (right axis, grey, N=2) for C321ΔserC with pThrOTS variants (O1-O12) expressing the MS-READ reporter. Bottom, immunoblot for phospho-MS-READ separated by Phos-Tag SDS-PAGE (top blot), and total MS-READ protein separated by SDS-PAGE (bottom plot) visualized with α-6xHis antibody. Promoter inclusion denoted by "+" with constitutive promoter activity denoted by "*". (B) Quantification of intracellular pThr levels (top, N=3, error bars represent 1 STD, **** p-value<0.0001 by one-way ANOVA) and level of pduX expression in C321.ΔserC proteomes (bottom, N=3, error bars represent 1 STD) by LC-MS/MS (P1-P4). (C) Schematic representation of pThrOTS^{Zeus} polycistronic operon and tRNA cassette.

Expanding the stability of pSerOTS components

Our lab had previously developed pSerOTSλ [17], in which pSerRS9 and EF-pSer21 are constitutively expressed under a TRC* promoter, with four copies of a $tRNA^{pSer}$ cassette. Although functional, pSerOTSλ has been observed to cause growth defects and strain instability in some contexts. We started our initial experiments in a C321.ΔA strain with a deletion in *serB*, the phosphatase responsible for converting pSer to Ser [43], to make our experimental pSer strain $C321^{XpS}$. In these cells, the addition of pSerOTSλ caused a stark decrease in growth rate compared to cells lacking the OTS. We also found that cell size, an indicator of stress in *E. coli*, was increased ~50% in cells containing pSerOTSλ (Fig 2-2A). Transformations of pSerOTSλ into traditional BL21 protein expression cells were unsuccessful, most likely due to the toxicity associated with UAG read-through.

To identify how component expression was altered with pSerOTS, we focused on expressing individual OTS components to narrow down key features of function. We focused first on the expression of pSerRS, comparing its expression under the control of the modified TRC promoter (TRC*), which provided constitutive, high-level expression against that of the low-level constitutive glnS* promoter [67]. Compared to *E. coli* lacking an OTS (WT), cells expressing pSerRS under control of glnS* showed a slight growth defect, while cells expressing pSerRS under control of TRC* displayed a significant decrease in growth (Fig 2-2B).

Figure 2-2| pSerOTS impacts E. coli fitness

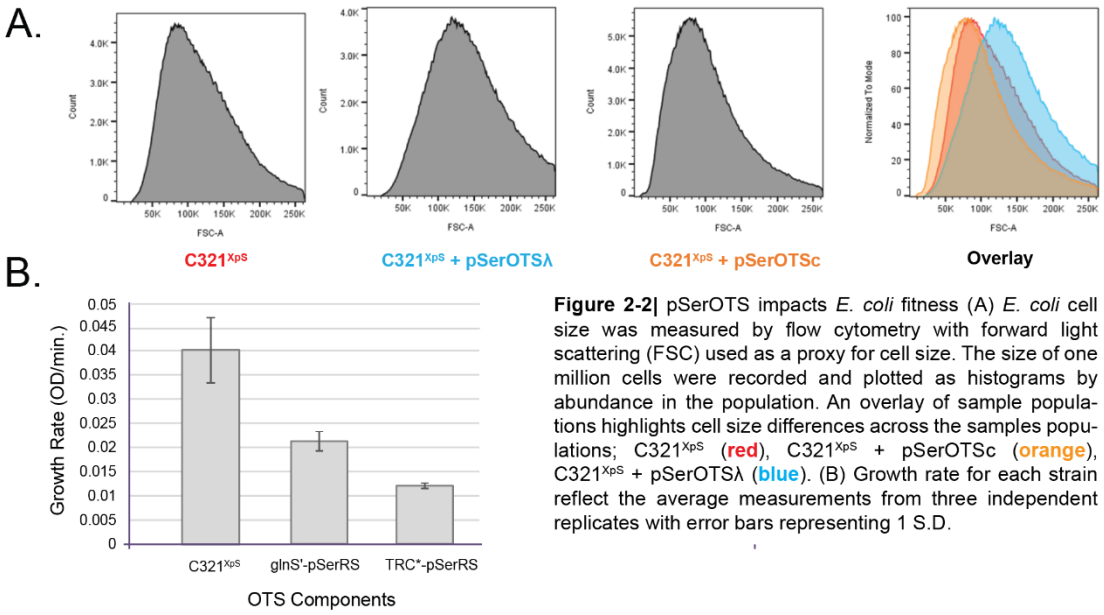

Figure 2-2| pSerOTS impacts *E. coli* fitness (A) *E. coli* cell size was measured by flow cytometry with forward light scattering (FSC) used as a proxy for cell size. The size of one million cells were recorded and plotted as histograms by abundance in the population. An overlay of sample populations highlights cell size differences across the samples populations; C321^{XpS} (**red**), C321^{XpS} + pSerOTSc (**orange**), C321^{XpS} + pSerOTSλ (**blue**). (B) Growth rate for each strain reflect the average measurements from three independent replicates with error bars representing 1 S.D.

Figure 2-3| Rank order comparison of pSerRS expression

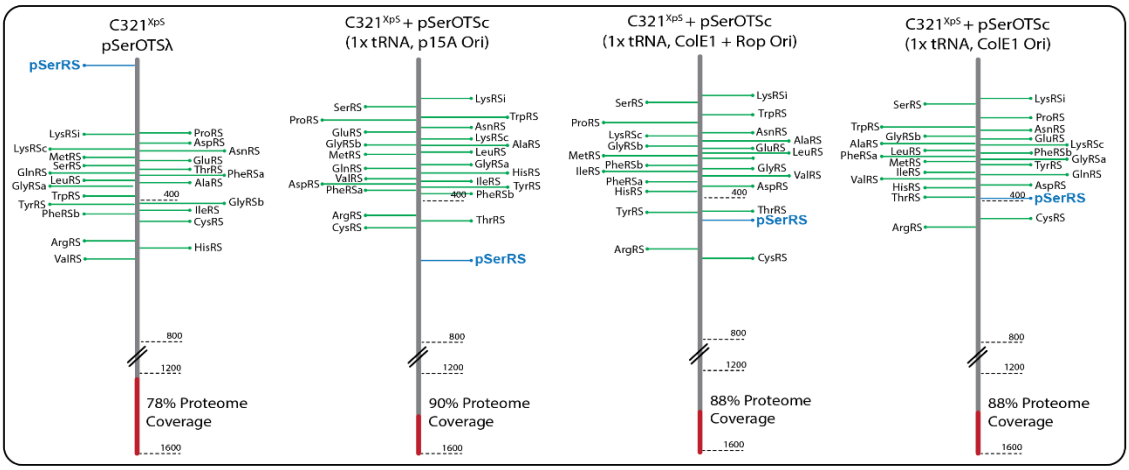

Figure 2-3| Rank order comparison of pSerRS expression. The level of pSerRS protein in cells containing various OTS constructs was measured by mass spectrometry and the position of pSerRS (blue line) relative to all proteins detected in the proteome is displayed in the context of native aaRS levels (green lines).

To better understand the implications of OTS expression on the host cell, we performed shotgun proteomics on C321^{XpS} cells expressing different pSerOTSs. Namely, we compared the relative pSerRS expression in cells transformed with pSerOTSλ against pSerOTS variants using the glnS* promoter with different copy number origins. Using the proteomic data, we constructed rank-ordered lists based on the relative abundance of individual *E. coli* and OTS proteins. pSerRS and native *E. coli* aaRSs are plotted for each experimental condition in Fig 2-3, with native *E. coli* aaRSs marked for relative comparison. While most *E. coli* aaRSs ranked in the 300-400 range, pSerRS from pSerOTSλ was one of the cell's top 10 most abundant proteins. The high level of pSerRS expression provides context for the slow growth and toxicity often seen with its expression. Presumably, pSerRS is altering the global landscape of amino acid availability through alterations in the Ser biosynthesis pathway, but a more in-depth mass spectrometry-based analysis of amino acid pools would be required for validation. Alterations to intracellular amino acid and tRNA pools have been previously demonstrated to decrease aaRS selectivity resulting in mistranslation events in *E. coli* and *S. cerevisiae* [68, 69]. Using this information, we designed our second-generation pSerOTS base plasmid, pSerOTSc, with a ColE1 origin, pSerRS:EF-pSer operon under the control of glnS*, and a single copy of tRNA^{pSer} under the *E. coli* promoter/terminator pair proK [70]. This ColE1-based OTS was used as the basis for comparison of future pSerOTS experiments, noted as C321^{EpS}.

Exploration of tRNA usage in orthogonal translation systems

tRNA copy number stability in pThrOTS design

Our experiments regarding pThrOTS identified the role of pThrRS and EF-pSer expression in meditating cellular toxicity and phosphoprotein production, thus allowing us to produce pThrOTS^{Zeus}. To simplify the expanded diversity of these initial experiments, we opted for using a single copy of tRNA^{pThr}. To enhance transcription of the single tRNA, we altered the lpp promoter and rrnC terminator system from pSerOTSλ to the native *E. coli* proK promoter/terminator pair [70]. To evaluate the role of tRNA copy number in pThrOTS^{Zeus}, we produced vectors containing 1x, 2x, 3x, and 4x copies of tRNA^{pThr}. As we had previously seen high rates of tRNA recombination with pSerOTSλ, presumably due to the four tRNA cassettes being identical, we sought to maintain high transcription of the tRNA without producing identical genetic elements. Further borrowing from native *E. coli* machinery, we used the valU tRNA linker, a genetic sequence that enhances maturation and folding of the tRNAs through RNase P meditated processing [71]. The addition of a valU linker, in theory, should minimize the rates of recombination by providing divergent genetic sequences around the tRNA while maintaining a high copy number. Attempts to increase copy number using an additional linker, valX, and tRNA resulted in multiple tRNA mutations. Instead, individual copies of the tRNA or 2x tRNA cassettes were used to make additional 3x and 4x variants (Fig 2-4A).

To assess how tRNA copy number affected phosphoprotein production, we co-transformed each variant with our MS-READ reporter and performed a qualitative analysis by Phos-tag gel and immunoblot in a modified BL21 strain. We found that generally, the level of phosphoprotein remained consistent across samples but that the increase in copy

number led to more non-phosphorylated protein being produced (Fig 2-4B). Based on these results, we determined that the 3x and 4x variants provided limited utility for quality phosphoprotein production. For a more in-depth analysis of how tRNA copy number may be affecting phosphoprotein production, we performed shotgun proteomics on our MS-READ reporter protein, looking at the fidelity of incorporation at UAG codons (Fig 2-4C). We were unable to identify significant differences in either phospho- or non-phosphoprotein protein production. As the 1x tRNA variant should be more genetically stable, we opted to use it as our basis for future experiments.

Figure 2-4| tRNA copy number variations have minimal impact on pThrOTS performance

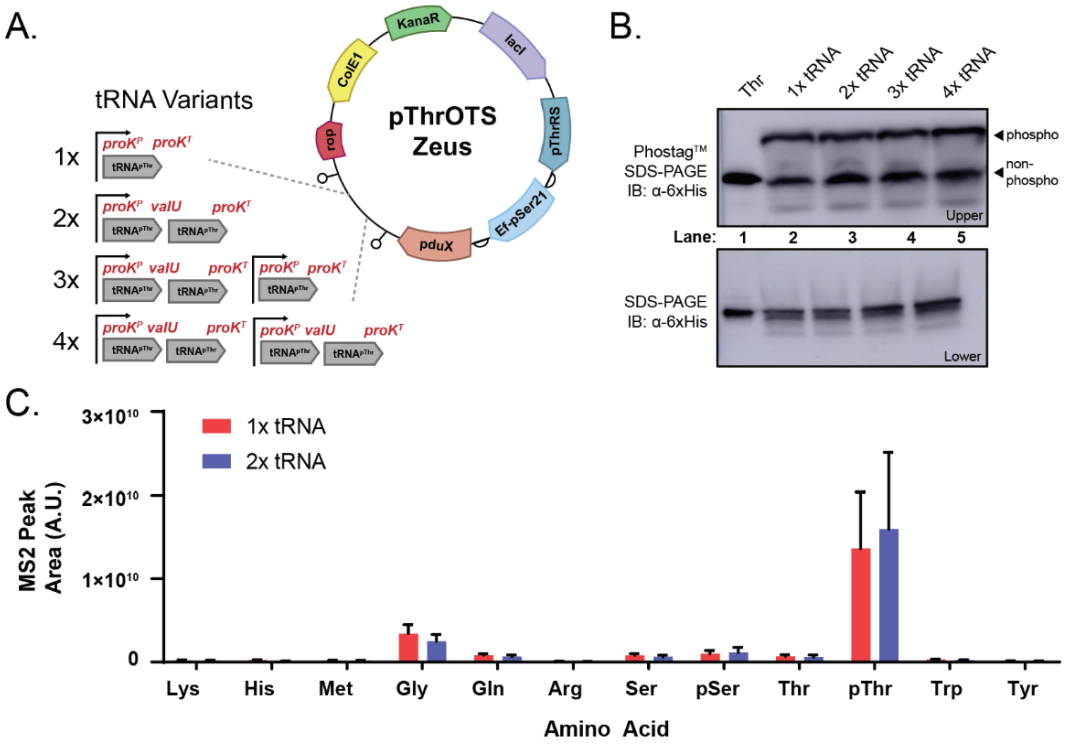

Figure 2-4| tRNA copy number variations have minimal impact on pThrOTS performance. (A). Schematic of pThrOTS tRNA copy number variants. (B) Analysis of MS-READ reporter phosphorylation and expression using Phos-Tag SDS-PAGE (Top) and SDS-PAGE (Bottom) visualized by immunoblot against α-6xHis. (C) Bar graph comparing the relative amounts of amino acid incorporation at UAG codons in MS-READ by MS2 peak area for 1x and 2x tRNA variants, error bars represent 1 S.D.

Figure 2-5| tRNA copy number variations alter growth and protein production with pSerOTSs

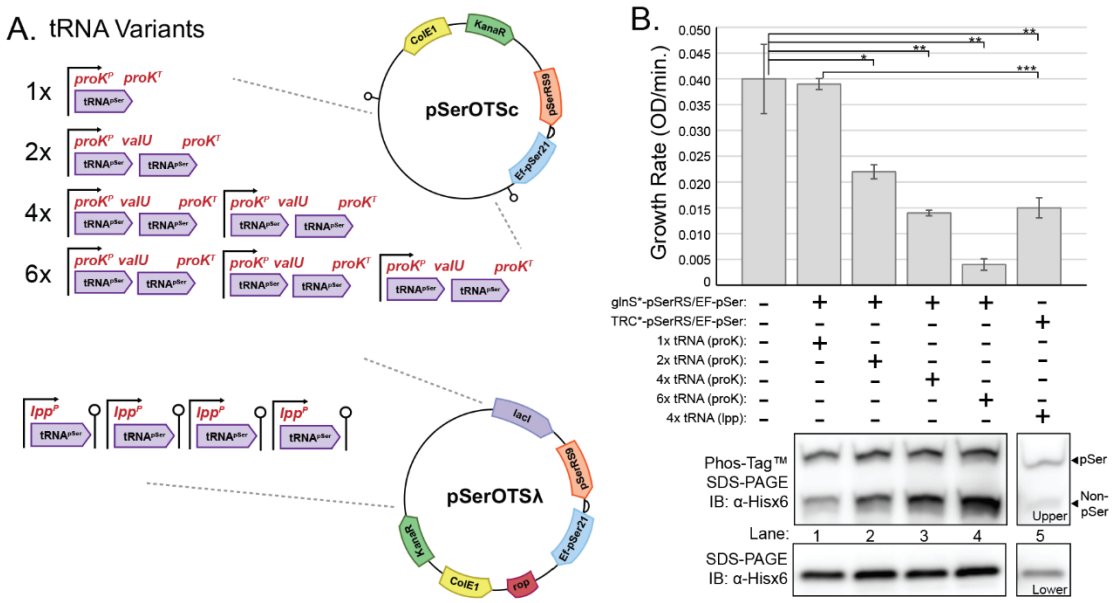

Figure 2-5| tRNA copy number variations alter growth and protein production within a pSerOTS. (A). Schematic of pSerOTS tRNA copy number variants, pSerOTSλ provided for reference. (B). The impact of tRNA copy number variation on growth and translational fidelity assessed by growth rates and Phos-Tag™ SDS-PAGE with immunoblot. Growth rate for each strain reflect the average measurements from three independent replicates with error bars representing 1 S.D. Statisticals were performed using a t-test, * equals $p < 0.05$, ** equals $p < 0.005$, *** equals $p < 0.0005$.

pSerOTS tRNA variations

One of the essential aspects of OTS design depends upon choosing the appropriate tRNA copy number. With pSerOTSc as a base, we created OTS variants with increasing tRNA copy numbers (Fig 2-5A). We transformed these OTS variants into both C321^{XpS} and BL21^{XpS}. In both cells, we observed a decrease in growth rate as the tRNA copy number was increased, with more than two copies of the tRNA being non-viable in BL21. When the OTS origin was switched to a lower copy p15a vector, cells were better able to handle increases in tRNA copy number (Fig 2-5B).

Variants of pSerOTSc with 1x, 2x, 4x, or 6x tRNA^{pSer} were co-transformed into C321^{XpS} cells with the MS-READ reporter protein. We found a copy number-dependent increase in the relative incorporation rates of Gly and Thr relative to pSer (Fig 2-6). With lower tRNA copy numbers, the rate of mis-aminoacylation stays low, but as the amount of tRNA^{pSer} rises, the misincorporation of Gly and Thr increases in a non-linear fashion. In addition to Gly and Thr, we also found appreciable levels of Ser incorporation at UAG codons. Mapping the *E. coli* SerRS recognition elements to tRNA^{pSer} identified a low degree of overlap with tRNA^{Ser}. Although it is impossible to rule out dephosphorylation as a confounding factor, the degree of overlap in identity elements provides evidence to the idea that the native SerRS is likely not interacting with the orthogonal tRNA^{pSer} (Fig 2-6) [72]. The tRNA-dependent increase in the relative ratio of Ser:pSer indicates that Ser may be mis-aminoacylated onto tRNA^{pSer} by pSerRS. Based on these findings, we surmised that the tRNA-dependent increase in non-pSer containing reporter protein results from tRNA mis-aminoacylation. These results indicate that balancing the ratio of tRNA to aaRS is essential for maintaining fidelity in OTS-dependent protein expression.

Figure 2-6| tRNA^{pSer} identity elements inform on mis-aminoacylation profiles

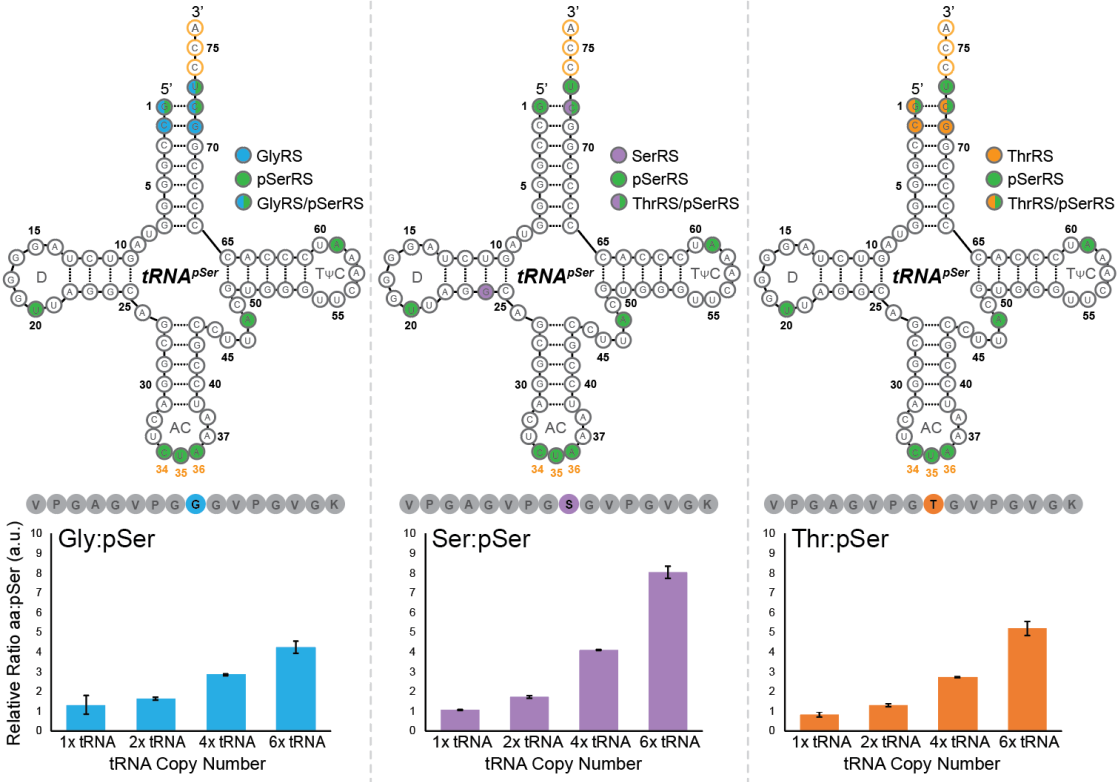

Figure 2-6| tRNA^{pSer} identity elements inform on misaminoacylation profiles. *E. coli* tRNA identity elements were overlaid onto the primary tRNA^{pSer} sequence. The effect of tRNA copy number increase on misaminoacylation was followed by mass spectrometry using a UAG containg MS-READ reporter in the presence of pSerOTSc with increasing tRNA copy number. Incorporation events from three samples were quantified using MS2 area under the curve and graphed relative to the level of pSer incorporation.

Inducing tRNA transcription with self-splicing RNA elements

One of the often-unaddressed issues of OTS expression is the role of deacylated tRNA in meditating cellular toxicity. It is known that when deacylated tRNA enters the A-site of the ribosome, it activates relA, which in turn produces the alarmone (p)ppGpp [73-75]. The presence of (p)ppGpp in the cell causes a signal cascade that activates the stringent response, where amino acid biosynthesis is upregulated, and general translation is down-regulated. Although most tRNAs within an OTS are considered 'orthogonal,' they would still be able to activate the stringent response if they enter the A-site of the ribosome. This issue is further confounded in OTSs by the inducible expression of the aaRS. If the aaRS is not induced, then the orthogonal tRNA pool will generally remain deacylated until its expression. As most protein expression protocols require cells to reach the mid-log phase before induction, there is a significant time gap in which the tRNA is most likely activating the stringent response, hampering protein expression before it has even begun.

To address this issue, we sought to create an inducible tRNA system that would be contained to the same polycistronic operon as the aaRS and EF. This would require a mechanism by which the tRNAs could be cleaved from the RNA transcript for further processing by RNase P. We found that hammerhead ribozyme could, in theory, satisfy these requirements. To test this theory, we produced pThrOTS variants in which the pThrRS:EF-pSer21:pduX operon was now linked to hammerhead ribozyme:$tRNA^{pThr}$ (Fig 2-7A), with a second variant containing an additional $tRNA^{pThr}$ with a valU linker. Modified C321 cells were then co-transformation with the MS-READ reporter to assess phosphoprotein production via Phos-tag gel and immunoblot (Fig 2-7B). As a proof of concept, we found that cells were capable of producing recombinant phosphoproteins using

the inducible ribozyme system; however, the overall yield was considerably lower than what we find with pThrOTS^{Zeus}. This indicates that although the system is functional, more copies of the tRNA are required to produce appreciable amounts of a phosphoprotein. It may be possible to incorporate more tRNA linkers, such as valX, ileT, or alaT, to produce a large operon consisting of more tRNAs.

Figure 2-7| Hammerhead ribozyme enables inducible tRNA transcription

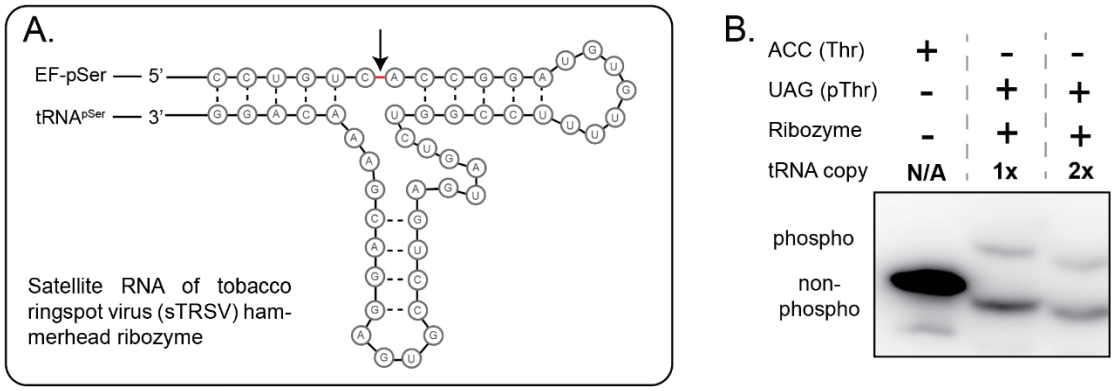

Figure 2-7| Hammerhead ribozyme enables inducible tRNA transcription. (A). Secondary structure of sTRSV hammerhead ribozyme, arrow indicates cleavage site. (B) Analysis of MS-READ reporter phosphorylation and expression using Phos-Tag SDS-PAGE.

Changes in tRNA^{pSer} recognition elements enable Ala suppression

It was previously shown that mutations in tRNA^{pSer} helped decrease mis-aminoacylation by native *E. coli* aminoacyl-tRNA synthetases, facilitating enhanced pSer incorporation [50]. In an attempt to reduce Gly mis-aminoacylation, we redesigned tRNA^{pSer} from a previously characterized pSerOTS [17] by adding recognition elements for D-aminoacyl-tRNA deacylase (DTD), an enzyme whose secondary function is to prevent Gly misincorporation by hydrolyzing mis-glycylated tRNAs (Fig 2-8A) [76]. Using MS-READ, we profiled the fidelity of our modified tRNA and, surprisingly, discovered that our modified tRNA functioned as an Ala suppressor tRNA (Fig 2-8B). The serendipitous development of an Ala suppressor tRNA provided a novel mechanism to program non-phosphorylatable residues in recombinantly expressed proteins genetically. Further investigation revealed that the major tRNA recognition element for *E. coli* AlaRS is G3:U70 [77], a mutation we had made to enhance recognition by DTD. Thus, our $tRNA^{Ala}_{CUA}$ enables the production of a non-phosphorylatable version for the protein of interest using the same UAG containing construct.

Figure 2-8| tRNA copy number variations alter growth and protein production within a pSerOTS

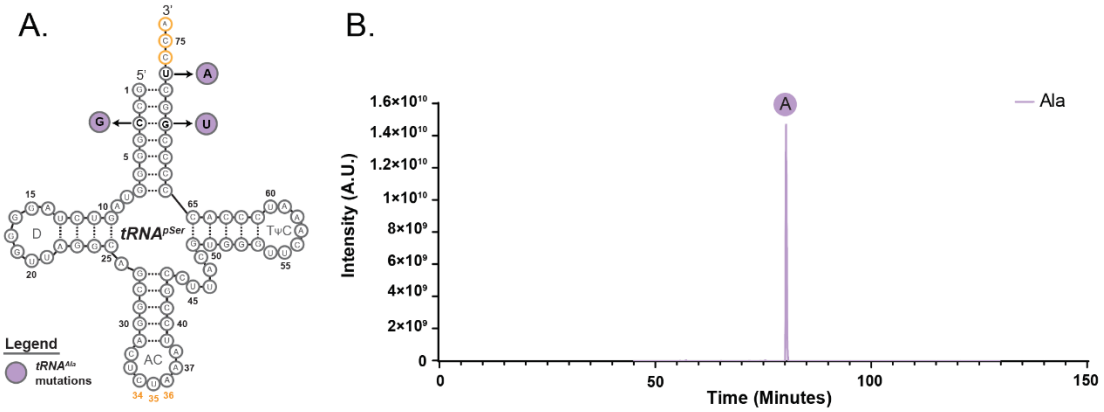

Figure 2-8| tRNA copy number variations alter growth and protein production within a pSerOTS. (A) Three substitutions (purple) in the acceptor stem of $tRNA^{pSer}$ result in suppressor activity. (B) LC-MS/MS analysis and extracted ion chromatogram of $tRNA^{Ala}$ mediated Ala incorporation at UAG codons in the MS-READ reporter.

Expanding the phosphoamino-acid genetic code to UGA codons

Although UAG codons have traditionally been used for genetic code expansion, it is not inconceivable that it could be adapted for use at additional stop codons. In order to expand the genetic code to multiple phosphoamino acids, we sought to create OTSs that could decode UGA codons and would theoretically be orthogonal to each other to facilitate the production of dually phosphorylated proteins. Many essential signaling cascades require the use of multiple phosphoamino acids for activation, for example, ERK1 & 2, which requires both a pThr and pTyr in the activation loop [78]. To this end, we created a series of pSer- and pThrOTSs containing tRNAs that would decode at UGA. For pThr, we opted only to modify the one tRNA in pThrOTS^{Zeus}, but as pSerOTSs have been evolved numerous times, we decided to begin with a larger contingent of pSerOTSs. Modifications to pSerOTSλ including replacing TRC* for an inducible TRC promoter, replacement of pSerRS9 with pSerRS1, and inclusion of tRNA cassette containing 2x $tRNA^{pSer}_{UCA}$ under control of the lpp promoter. We then co-transformed these vectors with the MS-READ reporter protein to ascertain function. Using Phos-tag gels and immunoblot for qualitative analysis, we found that only pSerRS1 was able to aminoacylate $tRNA^{pSer}_{UCA}$, EF-pSer variants did not alter phosphoprotein yields, and that the constitutive TRC* promoter enhanced phosphoprotein production (Fig 2-9A). We also found that pThrOTS^{Zeus} was able to aminoacylate $tRNA^{pThr}_{UCA}$, but not efficiently (Fig 2-9B). Due to the low protein yield produced by UGA pThrOTS^{Zeus}, we sought to characterize its decoding efficiency using our MS-READ reporter protein analyzed by MS/MS. We validated that $pThrOTS^{Zeus}_{UGA}$ was successfully incorporating pThr at UGA codons (Fig 2-9C), albeit at low levels (Fig 2-9B). Our results indicate that it is possible to incorporate phospho-amino acids at UGA codons

but with a tradeoff in efficiency. Future directions with whole-genome recoding may ease the burden by removing UGA codons and deleting RF2.

Figure 2-9| pSer- and pThrOTSs can decode UGA codons

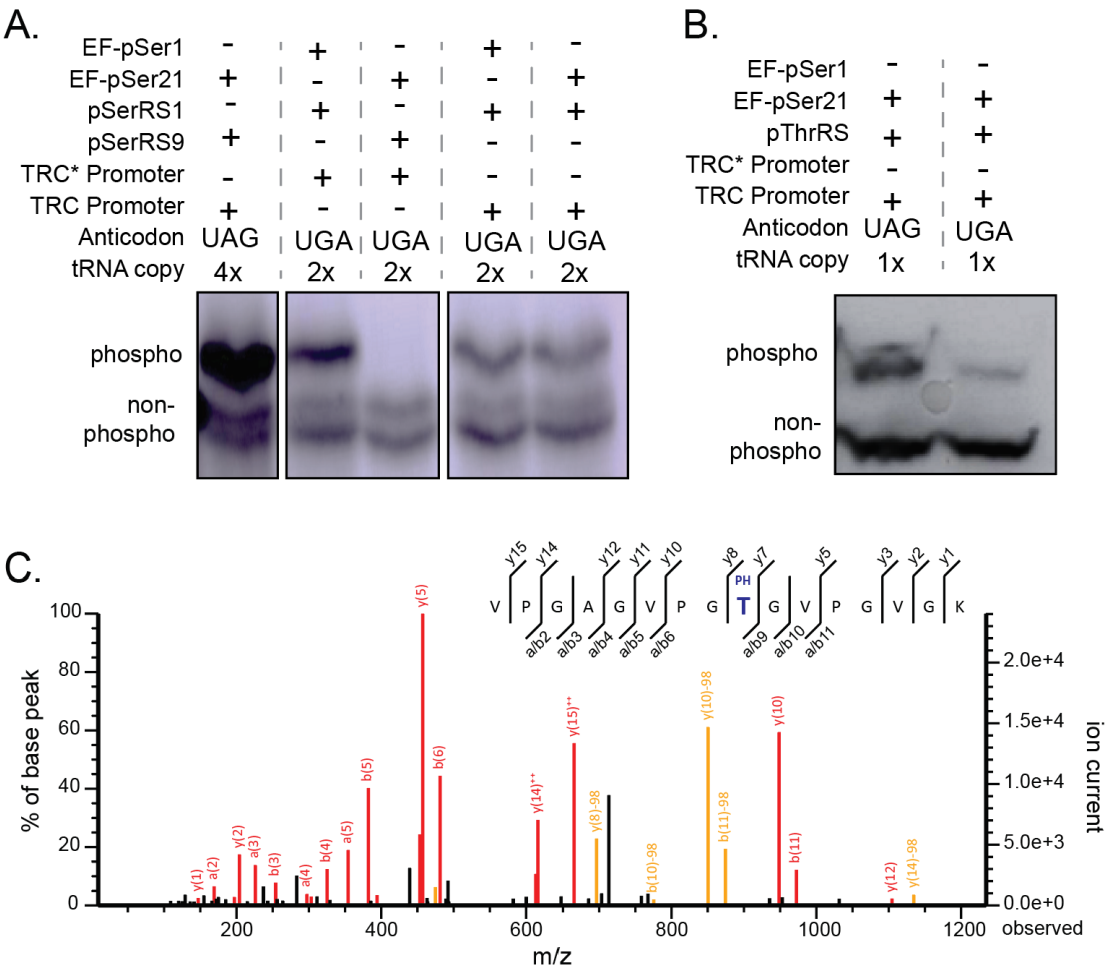

Figure 2-9| pSer- and pThrOTSs can decode UGA codons. Analysis of MS-READ reporter phosphorylation and expression using Phos-Tag SDS-PAGE for (A) UGA pSerOTS and (B) UGA pThrOTS. (C) MS2 ion spectra for pThr incorporation in UGA containing MS-READ reporter. Phosphorylated Thr is denoted in blue.

Genetic modifications to *E. coli* enhance phospho-protein production

Recombinant pThr protein production is enhanced in phosphatase deletion strains

In order to identify host contributions to OTS performance, we use targeted deletions to screen 16 candidate genes that could increase pThr protein production (Fig 2-10A). This approach showed that individual deletions of *serC*, *ycdX* and *pphA* enhanced phosphoprotein production. The impact of these candidates on steady-state pThr-protein levels was subsequently assessed both individually and combinatorically in BL21 and a modified C321 strain (C321.ΔA.exp). Strains with individual deletions for *pphA* or *ycdX* produced comparable levels of phosphoprotein reporter, while *serC* deletion strains showed significant increases in non-phosphoprotein production (Fig 2-10B&C). The triple knockout (TKO), experimental pThr strains ($BL21^{XpT}$ and $C321^{XpT}$) further improved the ratio of non-phosphorylated to phosphorylated protein reporter, mainly in $C321^{XpT}$. We next used MS-READ to quantify amino acid incorporation at UAG codons for each strain background and constructed a heat map using precursor ion intensity as a proxy for amino acid incorporation frequency (Fig 2-10D&E). Consistent with our Phos-Tag SDS-PAGE gels, the combinatorial knockouts facilitated the highest pThr-protein production. For future experiments, we designated the $C321^{XpT}$ and $BL21^{XpT}$ strains containing the episomal $pThrOTS^{Zeus}$ vector as $C321^{EpT}$ and $BL21^{EpT}$. These results highlight how host pathways can be altered to improve OTS performance.

In order to better understand the contributions of OTS architecture to cellular toxicity, we analyzed the proteome of C321ΔserC cells with and without a pThrOTS (Fig 2-1A, **O2**) and identified 463 host proteins with significantly altered expression. Protein network and GO enrichment analyses identified a cluster of metabolic sub-networks

involved in amino acid biosynthesis and catabolic processes (Fig 2-11A). The emergence of these networks suggests constitutive high-level OTS expression imparts a metabolic burden that could initiate host stress responses and ultimately diminish pThr protein expression. As noted earlier, this may, in part, be due to activation of the stringent response, either by low-affinity aminoacylation by pThrRS or excessive amounts of unaminoacylated $tRNA^{pThr}$. This was of particular concern as the stringent response decreases intracellular Thr levels [79], decreasing available substrates for pThr production via pduX. To investigate this mechanism, we knocked out the primary stringent response effector, *relA*, in $C321^{EpT}$ and $BL21^{EpT}$ strains generating $C321^{EpT}\Delta$relA and $BL21^{EpT}\Delta$relA. Using qualitative immunoblot analysis, we found no discernable difference in the ratio of phosphorylated to non-phosphorylated protein between the strains, suggesting that the stringent response was unlikely to significantly impact phosphoprotein production (Fig 2-11B). In order to expand the scope and diversity of human phosphoprotein production in our new strains, we gathered data from the PhosphoSitePlus database and constructed a codon-optimized phosphosite library DNA array corresponding to 57,536 unique pThr sites and flanking 15 amino acid sequences (resulting in 31 amino acids for each phosphosite) [80, 81]. Due to the absence of well-characterized Thr amber suppressor tRNAs, a duplicate library was synthesized by substitution of the central TAG codon for ACC to enable expression of a non-phosphorylated, Thr-containing proteome array.

As an initial experiment, we began by expressing the entire pThr phosphosite library in each of the four strains. We found that in regards to overall protein expression, there was minimal difference between genetic backgrounds (Fig 2-11C). However, many of our high-throughput experiments require a complex array of phosphosite expression *in*

vivo; in other words, we require a healthy cell population to perform most of our phosphosite-related experiments. To test this functionality, we used a bimolecular fluorescence complementation assay in which a protein-binding domain (PBD) is fused to the C-terminal half of mCherry to facilitate interactions with a phosphosite fused to the N-terminal half of mCherry for reconstitution of the fluorescent protein. We utilize this technique to produce High-throughput Interactomes of Phosphoproteins (HI-P) for systems-wide phospho-interactomics. As a measure of quality control in our assay, we wanted to understand how the relA KO background may be affecting the production of our phosphosite library. To do this, we made use of 14-3-3β, a PBD our lab had extensively validated with the pSer phosphosite library [80]. We found that only the C321 strains produced fluorescent cells, i.e., enrichment for phosphosites, but only the C321^{EpT}ΔrelA strain could produce a significant enrichment across multiple rounds (Fig 2-11D). The evidence indicates a toxicity effect from pThrOTS^{Zeus} meditated productive phosphosite expression that is mitigated in the C321^{EpT}ΔrelA strain. Conversely, the BL21 strains never experience a significant shift in fluorescence across cell populations. Whether this is due to OTS-mediated toxicity or limitations imposed by competition with the native release factor requires further investigation.

Figure 2-10| Gene deletions enhance pThr production

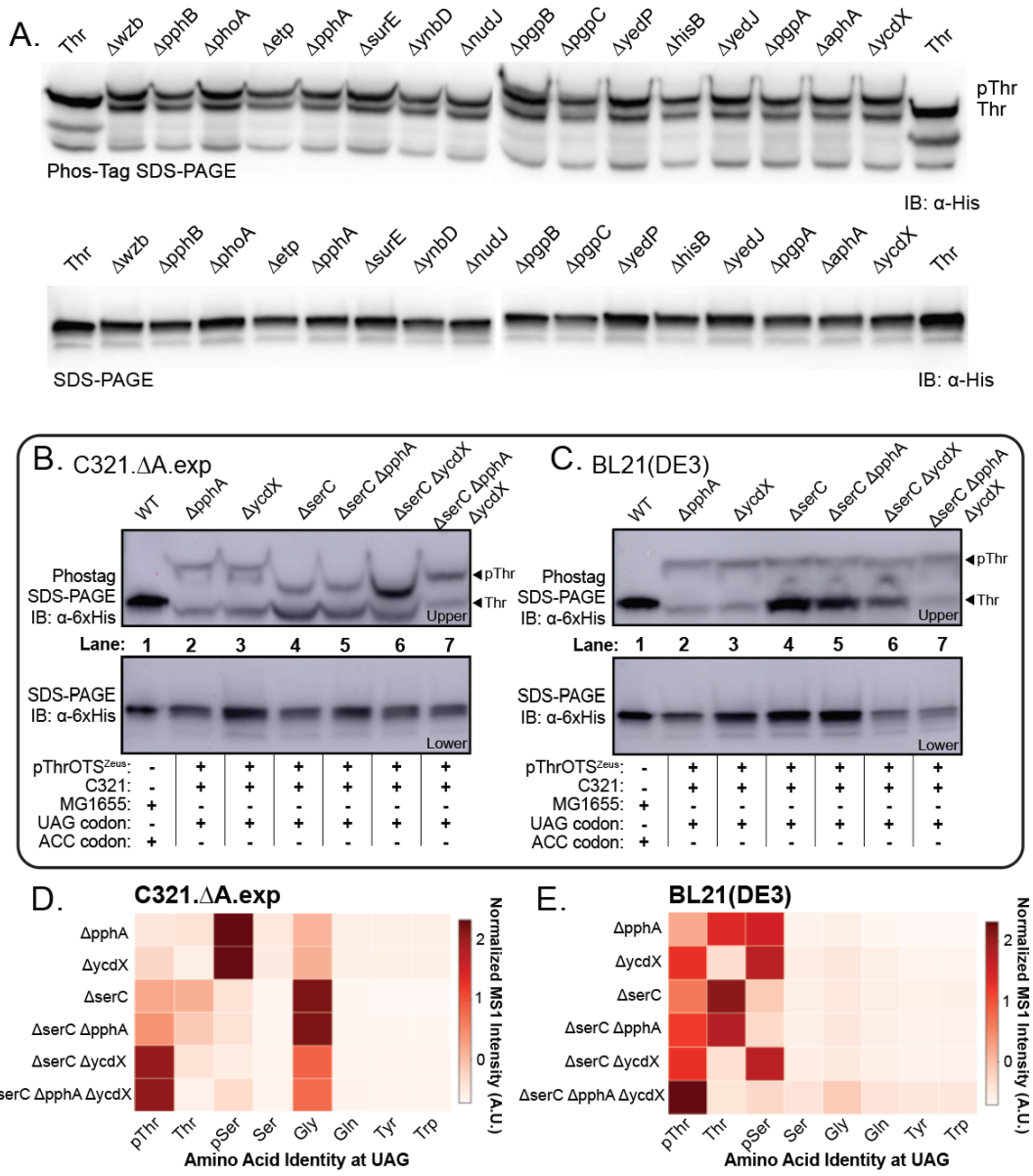

Figure 2-10| Gene deletions enhance pThr production. (A) Screen of phosphatase deletions to enhance phospho-MS-READ reporter expression using Phos-Tag SDS-PAGE (Top) and SDS-PAGE (Bottom) visualized by immunoblot against α-6xHis. (B) C321.ΔA.exp and (C) BL21 (DE3) Analysis of MS-READ reporter phosphorylation and expression using Phos-Tag SDS-PAGE (Top) and SDS-PAGE (Bottom) visualized by immunoblot against α-6xHis. (D) C321.ΔA.exp and (E) BL21 (DE3) Heatmaps comparing the relative amounts of amino acids incorporation at UAG codons in MS-READ (x-axis) plotted against modified genetic backgrounds (y-axis). Values are scaled within strains, N=3.

Figure 2-11| pThrOTS expression displays hallmarks of stringent response activation

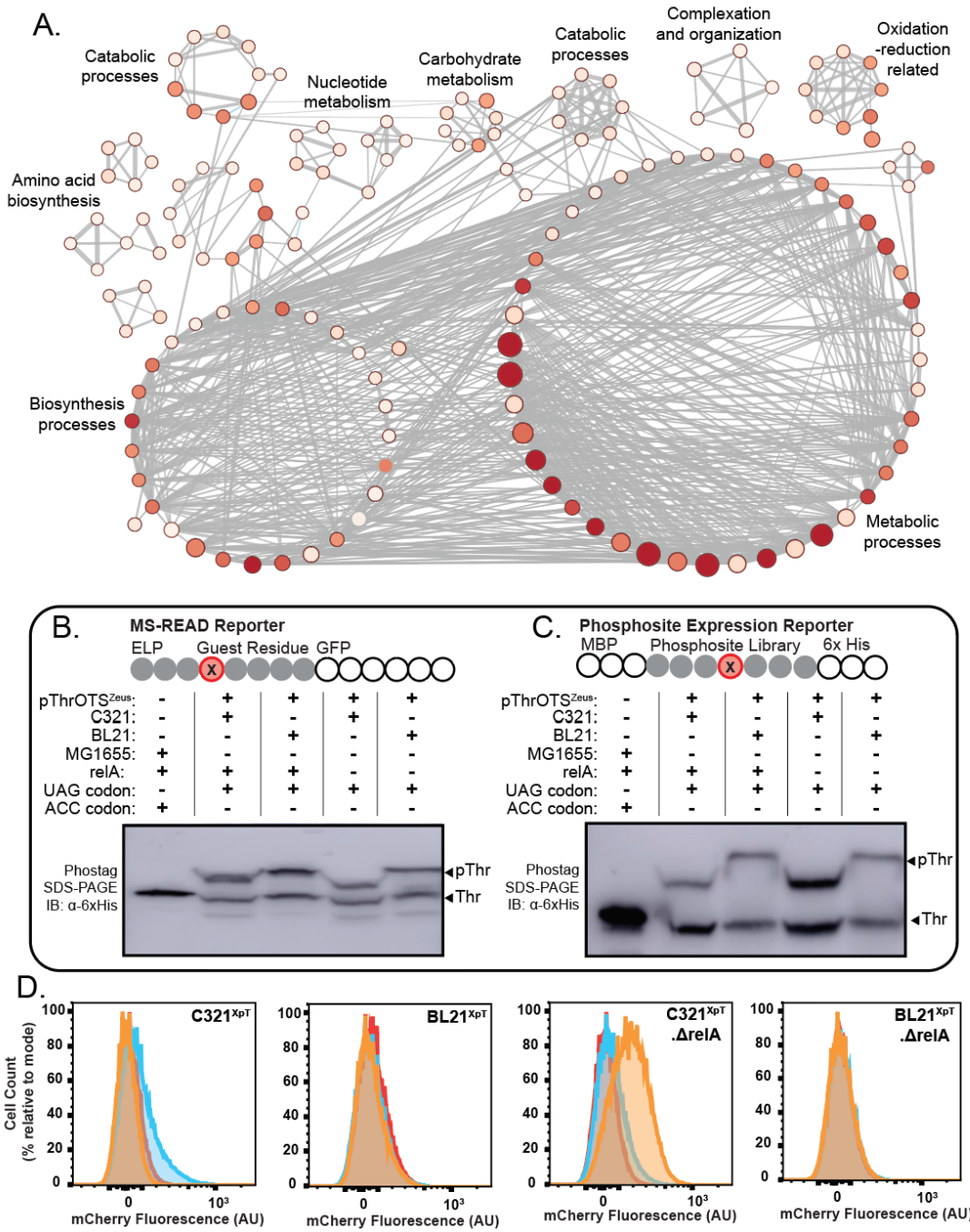

Figure 2-11| pThrOTS expression displays hallmarks of stringent response activation. (A) STRING analysis of GO enrichment comparing C321ΔserC cells with and without pThrOTS. Edges represent term relationship, size corresponds to number of enriched genes in each GO biological process, $p < 0.05$. (B) MS-READ and (C) pThr phosphosite library pThr incorporation at UAG codons in both relA+ (XpT) and relA- (XpT.ΔrelA) backgrounds assessed by Phos-Tag SDS-PAGE separation and visualized by immunoblot against α-6xHis. (D) FACS enrichment of pThr phosphosites by 14-3-3β in XpT strains. Starting population denoted by **red**, first enrichment by **blue**, and final enriched population in **orange**. One million cells were recorded and plotted as histograms representing individual mCherry fluorescence values by abundance in the population.

The optimized pSerOTSc is minimally toxic in C321

As seen with pThr, shotgun proteomics provides insights into cellular processes that may be altered by OTS expression. In order to understand the effects of pSerOTS components on C321^{XpS} cellular processes, we expressed episomal vectors containing either the tRNA alone or pSerRS alone (derived from pSerOTSλ). Comparing this dataset to that of pSerOTSc, and cells devoid of pSerOTS should enable us to understand better what metabolic processes are altered in a component-specific manner. We created rank-ordered protein abundance lists and performed statistical analyses to identify significant changes. The lists of proteins with significantly up-regulated and down-regulated protein expression compared to the C321^{XpS} control were cross-referenced to identify overlap between each experimental condition (Fig 2-12A). We found that pSerOTSc had 57 up-regulated proteins and 67 down-regulated proteins compared to the C321^{XpS} control, while pSerOTSλ and the λ aaRS control had 2-3x as many proteins upregulated and downregulated (Fig 2-12A&B). The fact that pSerOTSλ has dramatically altered the host proteome highlights the importance of understanding OTS-mediated toxicity.

To better understand OTS-mediated changes to the proteome, we constructed volcano plots comparing the up- and down-regulated proteins in pSerOTSλ and pSerOTSc. Points outside the curved boundaries represent significant changes, with those to the right representing up-regulated proteins and those to the left down-regulated proteins. As we had previously seen that expression of pSerOTSλ resulted in pSerRS being one of the most abundant cellular proteins, we theorized that the cell might have an increased translation demand. In order to test this hypothesis, we examined the expression of ribosomal subunit proteins (red) as a surrogate for translational demand (Fig 2-12C). We found that C321^{XpS}

cells expressing pSerOTSλ have a significant increase in the levels of ribosomal proteins compared to cells containing pSerOTSc (Fig 2-12C). This indicates that pSerRS expression mediated by pSerOTSλ results in a high translational burden even before expressing a recombinant phosphoprotein. Further analysis of the proteomics data revealed the up-regulation of RNA polymerase subunits and nucleotide biosynthesis, as well as the downregulation of nucleotide degradation, presumably due to the high-level transcription facilitated by TRC* and five copies of $tRNA^{pSer}$ (Fig 2-13). We also found that the tRNA-only OTS resulted in up-regulation of amino acid biosynthesis, an indicator of stringent response activation by unaminoacylated tRNA pools (Fig 2-13) [82]. Although pSerOTSλ produces numerous perturbations to the host proteome, pSerOTSc produced relatively few changes. Based on this information, we designated pSerOTSc in $C321^{XpS}$ as our episomal pSer strain ($C321^{EpS}$) for future work. This data highlights the utility of proteomics in understanding cellular mediated toxicity in OTS expression and provides a template for further studies.

Figure 2-12| C321^{XpS} cellular proteomes vary with OTS expression

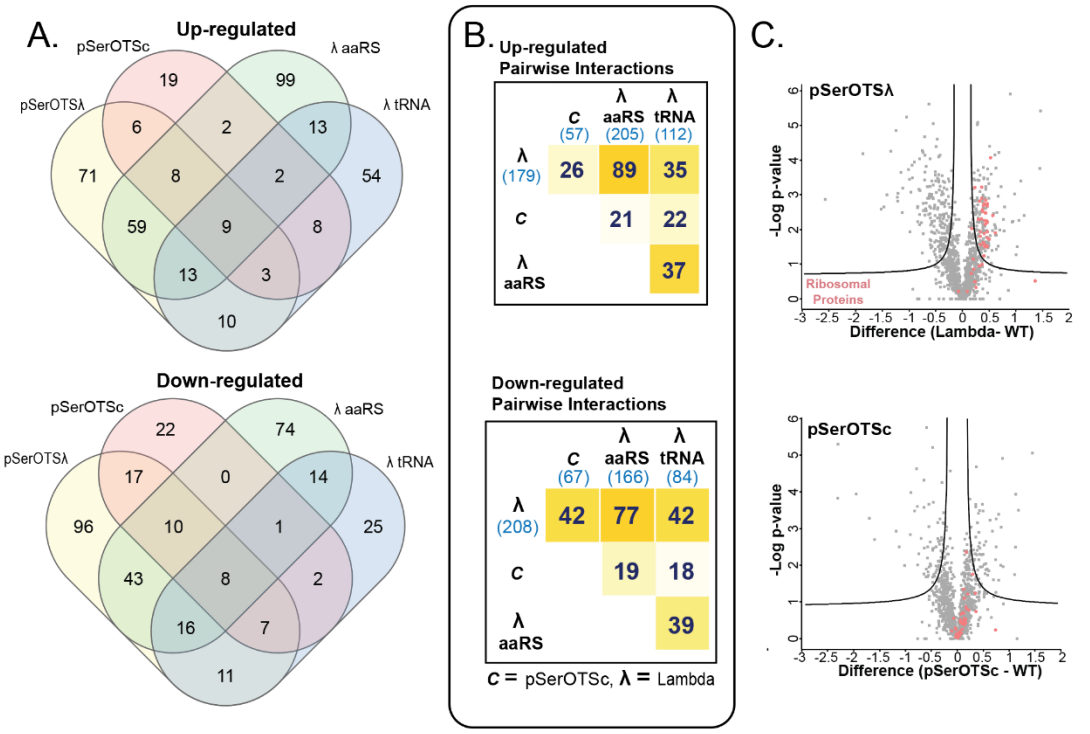

Figure 2-12| C321^{XpS} cells display irregular proteomes as determined by MS. (A) Summary of the overlap between significantly upregulated and downregulated proteins. (B) Pairwise overlap across experimental conditions. (C) Alteration of the proteomes from cells containing pSerOTSλ and pSerOTSc illustrated as volcano plots. Significant proteins are indicated as those outside the black lines with the expression of ribosomal proteins highlighted in pink.

Figure 2-13| pSerOTSλ displays markers of the stringent response

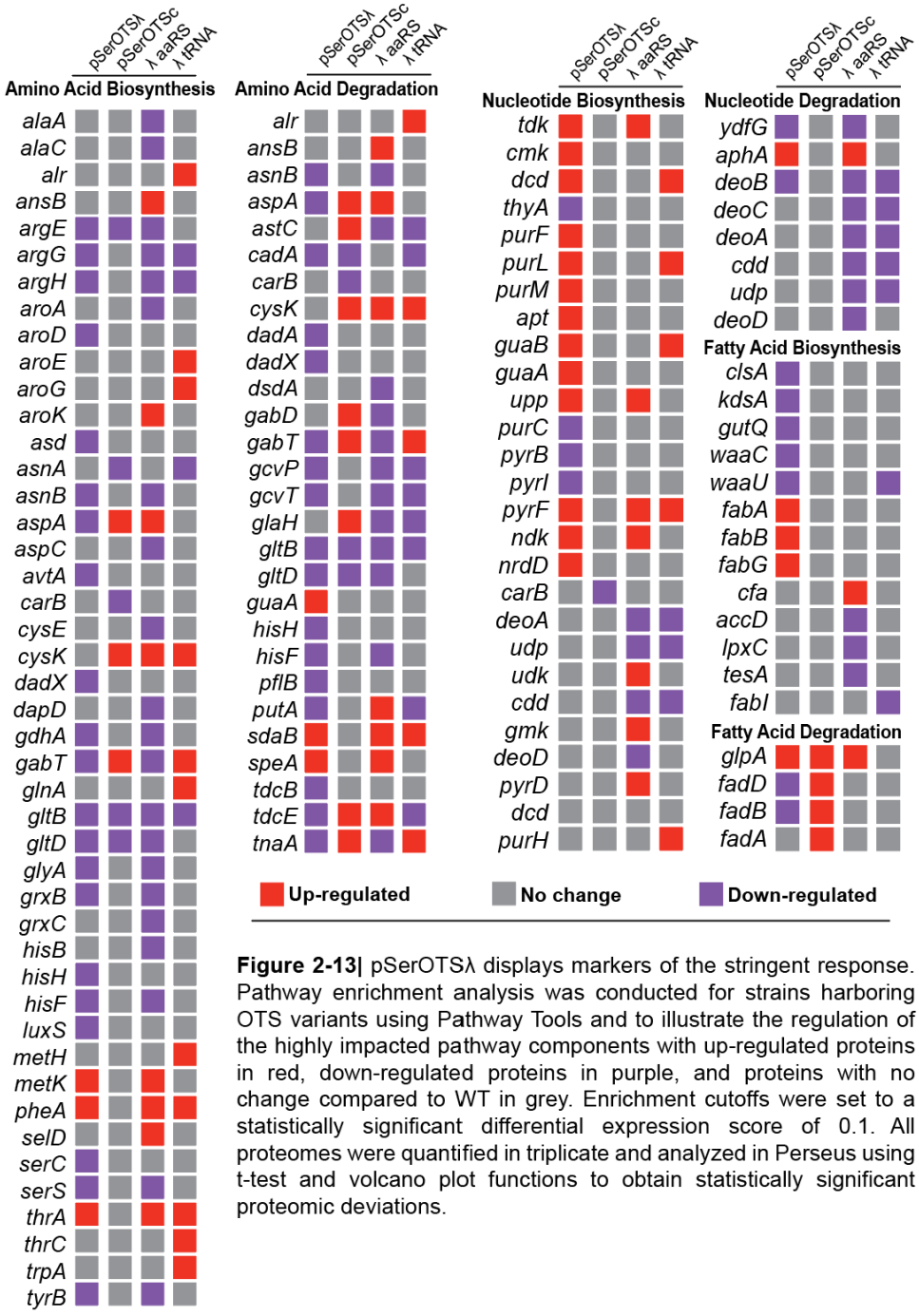

Figure 2-13| pSerOTSλ displays markers of the stringent response. Pathway enrichment analysis was conducted for strains harboring OTS variants using Pathway Tools and to illustrate the regulation of the highly impacted pathway components with up-regulated proteins in red, down-regulated proteins in purple, and proteins with no change compared to WT in grey. Enrichment cutoffs were set to a statistically significant differential expression score of 0.1. All proteomes were quantified in triplicate and analyzed in Perseus using t-test and volcano plot functions to obtain statistically significant proteomic deviations.

Incorporation of pTyr is limited by intracellular pTyr pools

Early experiments in the development of pTyrOTSs indicated that pTyr might be entering the periplasm but experiencing dephosphorylation before entering the cell [47]. Further variations circumvented this through the use of pTyr analogues or pTyr containing dipeptides to enter the cell [48, 49]. Building on this data, we produced a pTyrOTS consisting of the Schultz group's pTyrOTS with the addition of the Söll group's evolved elongation factor (EF-pTyr). Using our MS-READ reporter, we initially validated the experimental pTyrOTS in C321 cells ($C321^{EpY}$) with the pTyr-Lys dipeptide (Fig 2-14A). We found that $C321^{EpY}$ was able to incorporate pTyr, but at relatively low levels.

Due to the success of driving intracellular phospho-amino acid production for pSer and pThr, we attempted to find an enzyme that could produce intracellular pTyr. A bioinformatic search revealed that phenyl-phosphate synthase, an enzyme isolated from the soil microbe *Thauera aromatica*, was shown to readily phosphorylate phenol, the R group, of Tyr [83, 84]. Although we could express the enzyme complex successfully, we failed to identify any appreciable incorporation of pTyr with our MS-READ reporter (Fig 2-14B). As it appeared there was no clear pathway for either producing intracellular pTyr or driving import of pTyr, we decided to pursue a systems biology approach to understanding pTyr import. We first generated a Tyr auxotroph by deleting tyrA, a gene responsible for making an essential Tyr intermediate in *E. coli* (Fig 2-15A). Cells were then grown in a defined media lacking Tyr, with or without the addition of pTyr (Fig 2-15B). C321ΔtyrA cells were grown only in defined media lacking Tyr but containing pTyr. Samples were collected in triplicate and analyzed via shotgun proteomics. Statistical analyses of relative protein

abundance revealed C321ΔtyrA cells upregulated aromatic amino acid transporters, dipeptide transport proteins, and periplasmic phosphatases (Fig 2-15C&D).

Figure 2-14| pTyrOTS is capable of incorporating pTyr at UAG codons

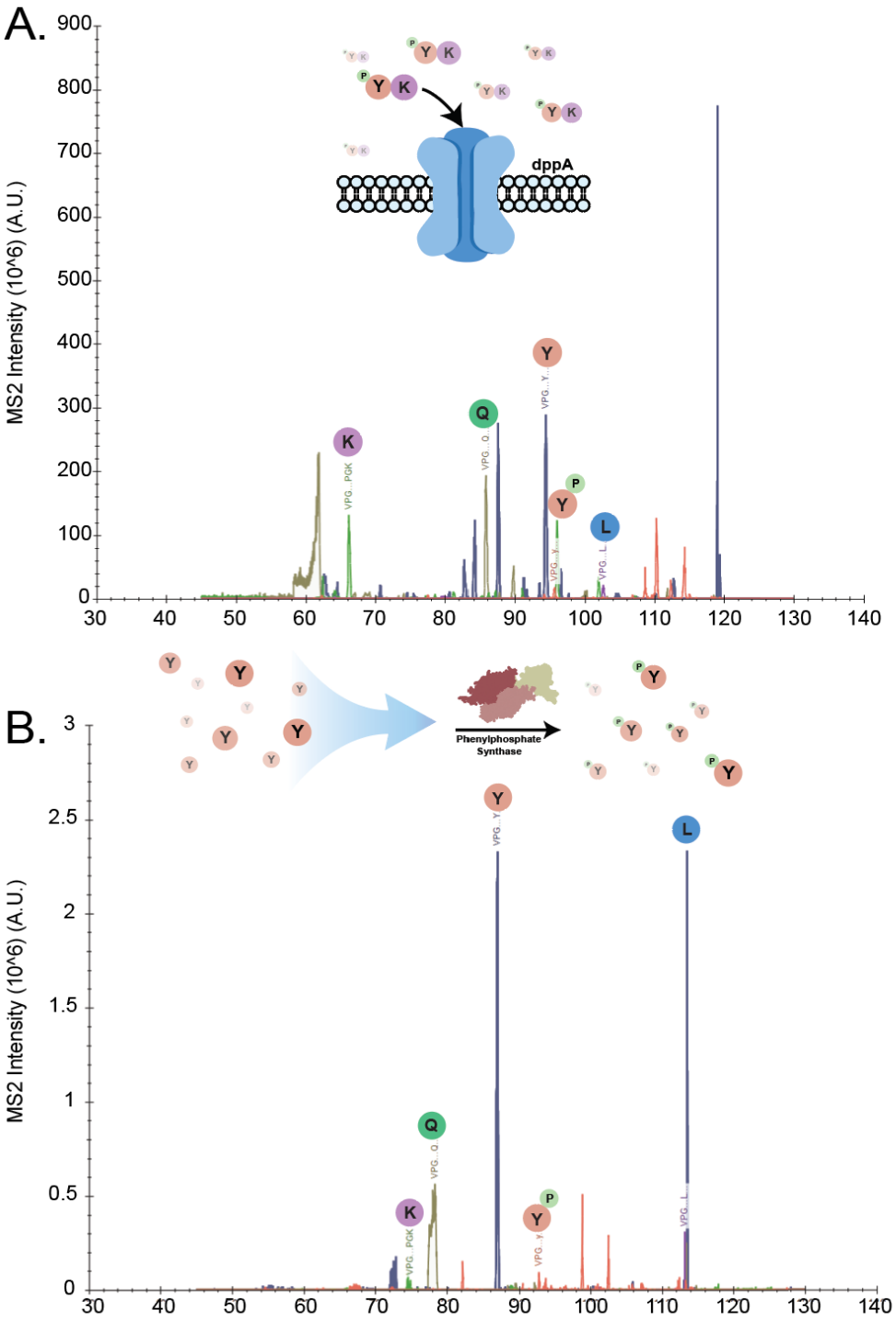

Figure 2-14| pTyrOTS is capable of incorporating pTyr. Total ion chromatograms of the UAG containing MS-READ reporter peptide. MS2 peaks are denoted by their corresponding amino acid. Intracellular pTyr was driven by either (A) Lys-pTyr dipeptide or (B) phenylphosphate synthase.

Figure 2-15| Tyr auxotrophy reveals potential targets for enhanced pTyr import

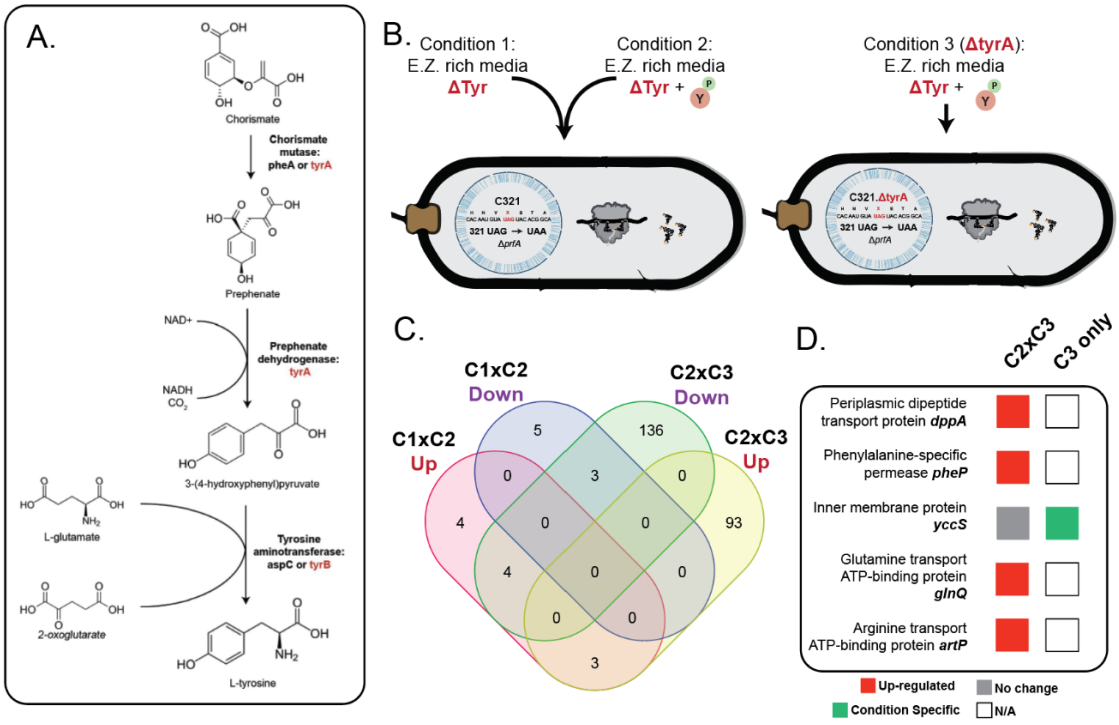

Figure 2-15| Tyr auxotrophy reveals potential targets for enhanced pTyr import. (A) Schematic of Tyr biosynthesis in *E. Coli*. (B) Experimental overview for comparative shotgun proteomics experiments. (C). Summary of the overlap between significantly upregulated and downregulated proteins across conditions. (D) Identified amino acid specific transport proteins associated Tyr auxotrophy.

Chapter 3: Applications of the incorporation of phospho-amino acids

Systems biology approaches to understanding protein-protein interactions

Mapping the phospho-interactome of 14-3-3β

Although phosphorylation is generally recognized as a regulatory medium for kinase-phosphatase-dependent processes, it also facilitates non-enzymatic PPIs via PBDs. We previously developed the HI-P method for screening individual pSer library members in parallel with a single PBD (Fig 3-1) [80]. Previously, our work had been limited to pSer, but the advancement of our pThrOTS technology allowed us, for the first time, to characterize pThr dependent interactions. Following enrichment of 14-3-3β in our HI-P system, we sequenced the phosphosites of two replicates from the final pThr population. We found that 67 phosphosites were shared between the samples, with each sample having ~200 sites. Motif analysis of enriched pThr phosphosites showed that pThr HI-P identified the canonical 14-3-3 binding motif (-3 Arg, -2 Ser, and +2 Pro; Fig 3-2A) that was also identified with pSer HI-P [80]. We found a +1 Trp motif unique to pThr, that has been previously reported in some 14-3-3β studies [85]. To better understand how this may relate to pSer, we repeated our HI-P experiment with 14-3-3β against the pSer library using the newly developed C321^{EpS} strain. We were able to identify ~220 pSer-specific 14-3-3β phosphosites using C321^{EpS}. Motif analysis of enriched pSer phosphosites also identified the canonical 14-3-3 binding motif (-3 Arg, -2 Ser, and +2 Pro; Fig 3-2A). We then performed a GO enrichment analysis of biological processes between the pThr and pSer dataset (Fig 3-2B). There was a low degree of overlap between pSer and pThr biological processes, with only terms related to cytoskeleton organization being enriched for both

sets. Our data indicates that there may be different alternative roles for pThr and pSer in phospho-binding mechanisms but requires further study.

Figure 3-1| High-throughput Interactome of Phosphoproteins (HI-P) overview

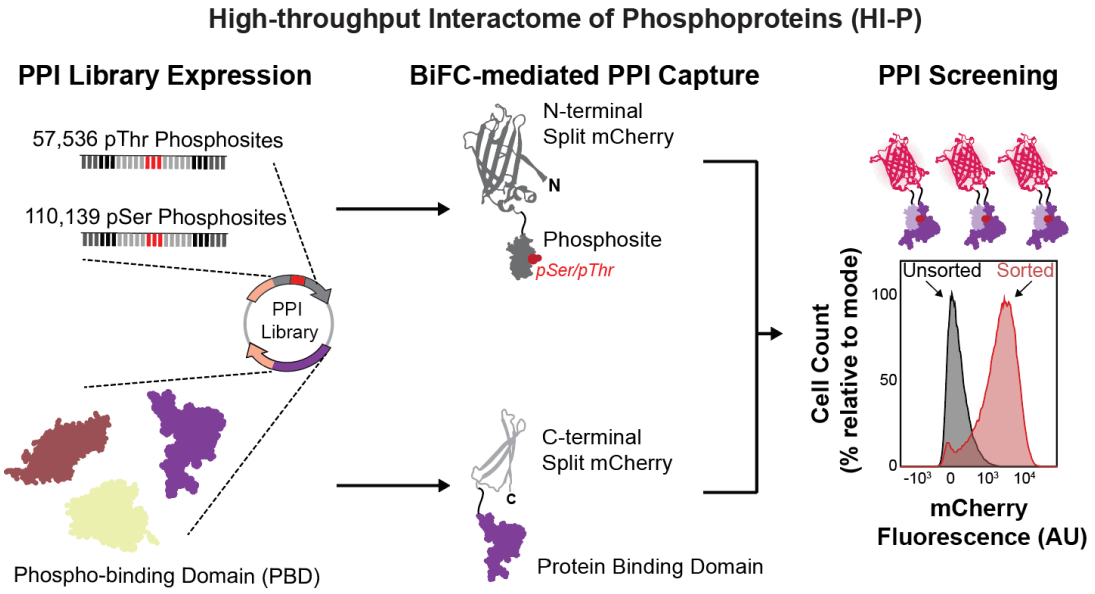

Figure 3-1| High-throughput Interactome of Phosphoproteins (HI-P) overview. The phospho-binding domain of interest is cloned into the destination vector as a fusion protein to the C-terminal half of a split mCherry protein. The phosphosite library of interest (either pSer or pThr) is then cloned in as a fusion to the N-terminal half of the split mCherry protein. Using BiFC-mediated PPI capture it is then possible to enrich for fluorescent cells via FACS.

Figure 3-2| 14-3-3β HI-P recapitulates known motifs and functions

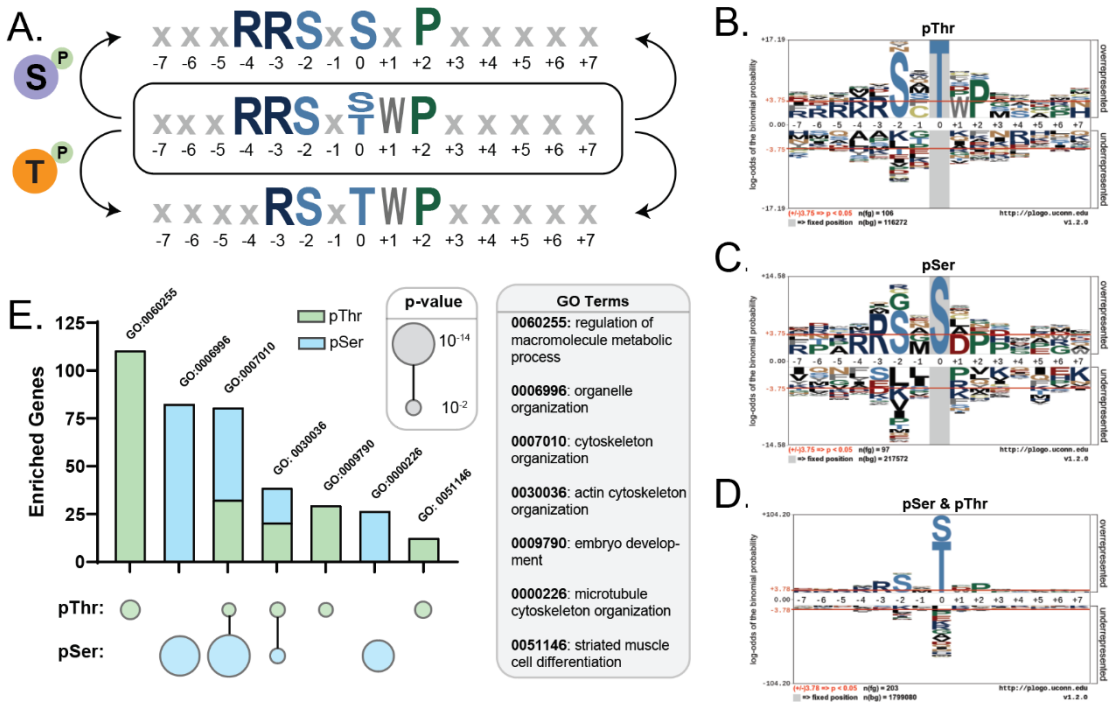

Figure 3-2| 14-3-3β HI-P recapitulates known motifs and functions. (A) Significant positive motif analysis (pLogo) for pSer, pThr, and combined motifs. Phosphosites were selected with 50x enrichment, labelled residues p<0.05. (B-D) Traditional pLogo motifs for (B) pThr, (C) pSer, and (D) combined pSer/pThr for the same data from (A). (E) Upset graph comparing PANTHER overrepresentation analysis for GO biological processes across pThr (**green**) and pSer (**blue**) datasets. For each GO term, the number of genes (chart) is compared to term inclusion (circle) and significance (circle size, p-value). GO terms listed above each bar are detailed in the grey callout to the right.

Evolutionarily conserved PBDs interact with human phosphosites

As a further proof-of-concept for our pThr based HI-P system, we decided to test a forkhead-associated (FHA) phospho-binding domain. FHA domains are known as specific binders of pThr, which has precluded using the pSer-based HI-P system to characterize its binding profile. To that end, investigated RAD53, which has one of the most well studied FHA domains from the yeast *S. cerevisiae* [86-89]. Deployment of RAD53 in a HI-P context was highly functional with BL21^{EpT} (Fig 3-3A). Analysis of individual phosphosite-specific interactions in a 96-well format pre- and post-FACS enrichment showed a striking shift in the fluorescent population distribution. Unfortunately, analysis of biological replicates failed to produce a motif, let alone recapitulate a RAD53 motif. In this instance, we could only identify 111 phosphosites in total, with 71 being shared between the two sets. The lower count may be due to limitations of yeast binding domains in a HI-P context or the use of the BL21^{EpT} strain. Although it is not possible to rule out a loss of specificity due to a mismatch of human phosphosites with a yeast binding domain, the sequence space covered by our libraries represents a vast array of potential interaction partners. Functional enrichment for GO biological processes mirror some known processes as CHK2, the direct human homologue, such as regulation of cycle progression (Fig 3-3B). Although some of the data was inconclusive, it does provide evidence that it may be possible to screen evolutionarily divergent PBDs against known human phosphosites in a high-throughput manner. Future screening with PBDs from other organisms may provide more insight into the feasibility of such an approach.

Figure 3-3| Evolutionarily conserved PBDs functionally enrich with HI-P

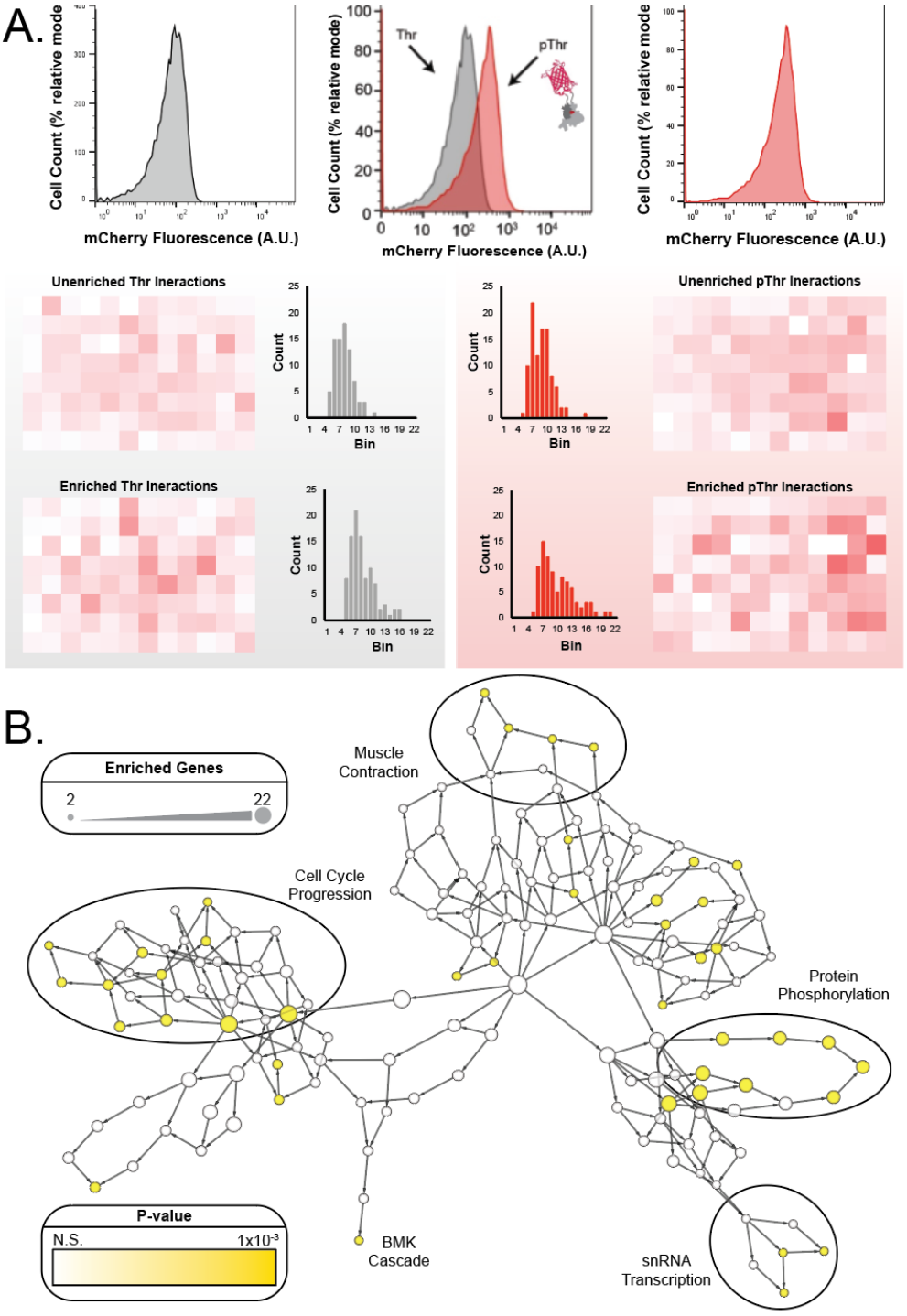

Figure 3-3| Evolutionarily conserved PBDs functionally enrich with HI-P. (A) Individual BL21^{EpT} RAD53 HI-P colonies were measured for fluorescence before enrichment (top) and post enrichment (bottom) comparing phosphosite binding for Thr (grey) and pThr (red). (B) Enrichment map for GO biological processes with RAD53 phosphosites.

The pTyr SH2 PBDs are functional in the HI-P framework

SHP2 is a tyrosine-protein phosphatase containing both an N-terminal and C-terminal Src Homology 2 (SH2) domains associated with cell growth, differentiation, and oncogenesis [90-94]. The N-terminal SH2 domain of SHP2 has many functions, but the canonical mechanism allows for *cis* binding to the SHP2 phosphatase domain, thereby blocking phosphatase activity [95]. Although such SH2 domains have been well characterized for their phosphobinding activity, there has never been a large-scale screen against all the known human pTyr phosphosites. To this end, we sought to characterize the phosphoproteome-dependent interactions of the N-terminal domain of SHP2 using HI-P and $C321^{EpY}$. Using the same methodology for pThr and pSer, we produced a DNA array containing 38,679 known pTyr phosphosites encoded with a central UAG residue. Using our pTyrOTS with dipeptide mediated uptake of pTyr and the native *M. jannaschii* TyrRS and tRNA pair as a negative control, we were able to test the binding of SH2 domains across the human phosphoproteome. When utilized in our HI-P system, we found that SHP2 (and other) SH2 domains display high levels of non-phospho-binding in HI-P (Fig 3-4A). This may be a limitation of $C321^{EpY}$ failing to incorporate high levels of pTyr (Fig 2-14) that could be overcome with future technological advancements. Motif analysis of the enriched $C321^{EpY}$ samples produced a motif directly opposed to those previously identified (Fig 3-4B) [96, 97]. However, the enrichment of acidic residues indicates there be some phospho-memetic binding of the SHP2. This data does not preclude the ability to use SH2 domains in HI-P but indicates that extensive work is required to adapt both systems for such use.

Figure 3-4| The SHP2 SH2 domain lacks specificity with C321^{EpY}

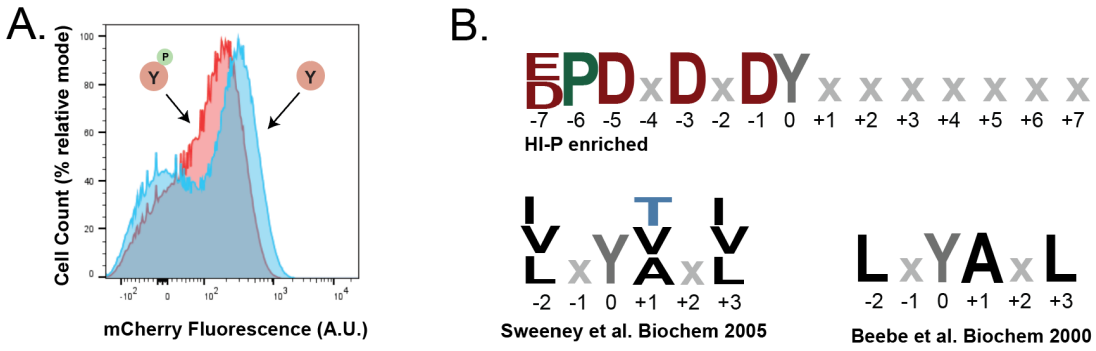

Figure 3-4| The SHP2 SH2 domain lacks specificity with C321^{EpY}. (A) Overlap of SHP2 SH2 binding using the pTyrOTS (red) and TyrOTS (blue). (B) Comparison of HI-P enriched motif (top) and those uncovered with positional peptide scanning (bottom).

Non-phosphorylated phosphosite libraries provide an ideal substrate for kinase screening

The production of active kinases and non-phosphorylated Ser/Thr phosphosite libraries provides a unique opportunity for the *in vitro* profiling of kinase substrates across the human proteome with greater coverage than previous methods [98]. The serine/threonine kinase, CHK2, is a prime example of a kinase that would provide a wealth of information by identifying its substrates involved in the modulation of cell cycle progression and DNA repair pathways [99-103]. By coupling purified Ser and Thr phosphosite libraries to *in vitro* kinase reactions and shotgun proteomics, we were able to profile the activity of both full-length active CHK2 and the CHK2 kinase domain (KD) containing pThr encoded within the activation loop (Fig 3-5A).

Overall, we identified 230 unique phosphosites when screening with full-length CHK2, of which 126 phosphosites occurred at least twice across three reactions. We found evidence for 290 unique phosphosites when screening with the CHK2-KD, with 158 identified at least twice. Most phosphosites were for pSer, corresponding to ~70% for both full-length and KD variants. Motif analysis of phosphosites identified the canonical CHK2 RxxS/T recognition motif, separation of enriched phosphosites into Ser and Thr sub-pools revealed unique motif elements that had not been characterized with traditional peptide arrays [104] (Fig 3-5B). Additional analysis identified pThr specific motifs, specifically, a +5 Tyr and +7 Cys motif. Our work demonstrates how phosphosite libraries can be repurposed for the identification of functional kinase substrates *in vitro*.

Figure 3-5| Kinase profiling of CHK2 reveals known motifs and substrates

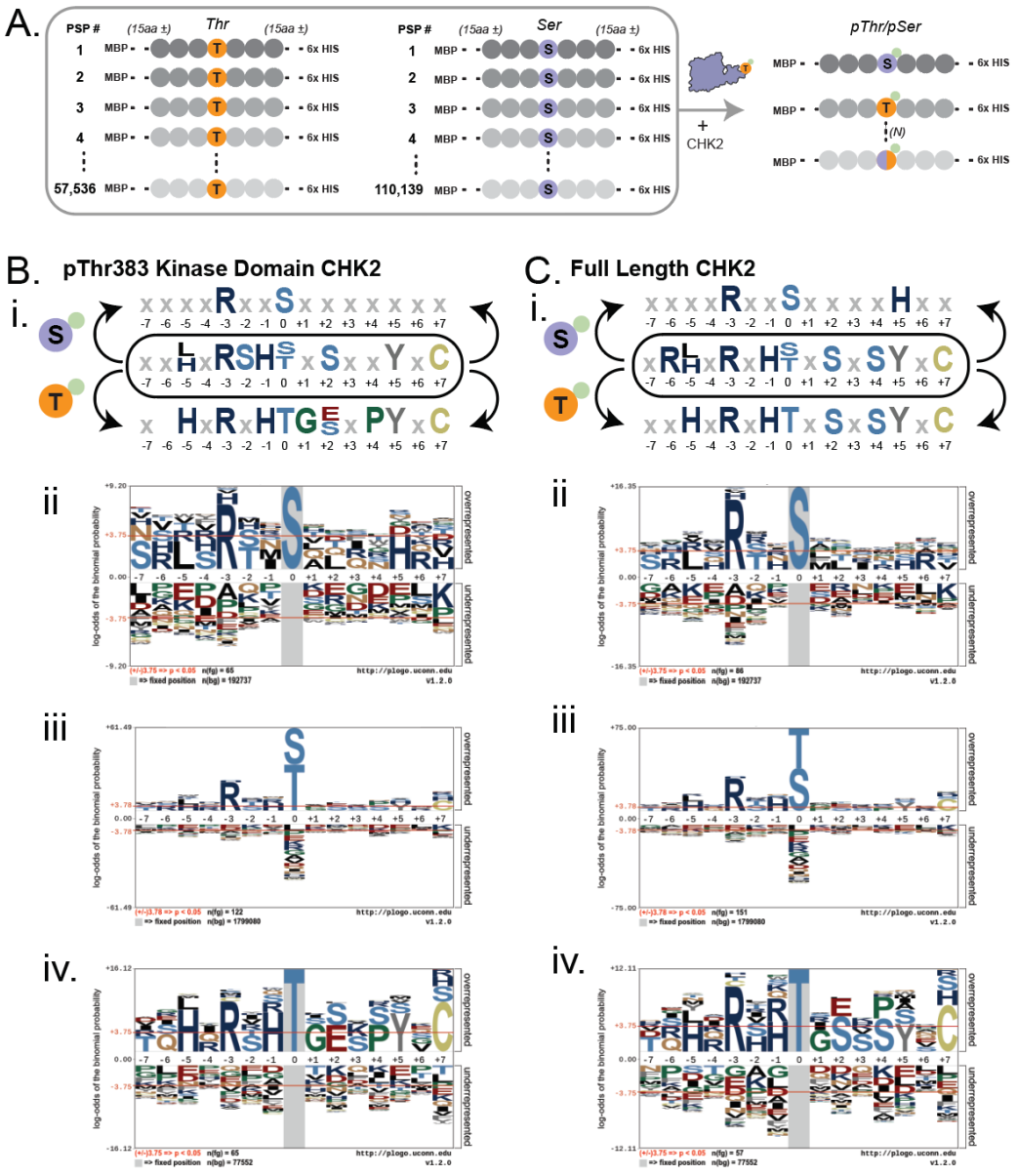

Figure 3-5| Kinase profiling of CHK2 reveals known and unknown motifs. (A) Schematic of CHK2 kinase profiling. (B) Motif analysis for pThr activated CHK2 kinase domain (left) compared to the full length motif (right). Each motif is broken down by a summary (i), pSer (ii), pSer & pThr (iii) and pThr (iv). Phosphosites were selected that occured at least twice across three replicates, labelled residues p<0.05.

HI-P+: Coupling kinases to PBDs with phosphosite resolution

A long-standing barrier to establishing mechanisms of phospho-regulation has been the need for a high-throughput method to connect kinase activity to phosphorylation-dependent PPIs while maintaining phosphosite resolution. To address this challenge, we developed a modified HI-P framework that identifies kinase-specific substrates interacting with phospho-binding domains across the phosphoproteome. Modification of the HI-P platform included utilization of the non-phosphorylated Thr-phosphosite library as the kinase substrate, leaving the TAG codon free for genetically encoded pThr to activate a kinase of interest (CHK2) (Fig 3-6A). To streamline our approach, we modified pThrOTS^{Zeus} to reduce the metabolic burden with a low copy ClodF13 origin vector [70] that constitutively expresses a kinase of interest (in this instance, CHK2). This novel reconfiguration is termed HI-P+ for HI-P "plus" a kinase of interest. Using HI-P+, we co-expressed activated CHK2-KD with the Thr-phosphosite substrate library paired with 14-3-3 β. HI-P+ enabled a coupled screen where activated CHK2 would phosphorylate substrates which would, in turn, be identified through binding by 14-3-3β (Fig 3-6A). This integrated approach to activate CHK2 and drive pThr site-specific PPIs identified 409 human phosphosites that were both CHK2 and 14-3-3β substrates. The high number of substrates may be over-represented due to the absence of human phosphatases, limiting the occurrence of low-affinity CHK2 substrates. Comparative motif analysis showed that HI-P identified CHK2/14-3-3β substrates with signature amino acids for both proteins (Fig 3-6B). Interestingly a -3 Arg is preferred by both CHK2 and 14-3-3β (Fig 3-6B), yet a -1 His was preferred in only CHK2 Thr substrates. HI-P+ demonstrated that the -1 His pThr motif was recognized by 14-3-3β and highlighted how coupling an active kinase with a

direct PPI readout could uncover kinase-specific PPI network criteria. To add functional insight to our multi-level, kinase-phospho-binding domain PPI network approach, we performed a GO enrichment analysis for biological processes on each unique dataset generated throughout this study (Fig 3-6C). Our analysis revealed a high degree of overlap for GO-terms related to the cytoskeleton and cellular organization and processes involved in gene regulation [105]. Overall, our pThrOTS enabled a unique combination of kinase profiling to establish multi-level phosphoregulatory PPI networks. HI-P+ adds unprecedented depth by directly linking kinases and phospho-binding proteins to single phosphosite PPIs across the human proteome.

Figure 3-6| HI-P+ links phospho-binding activity to a kinase of interest

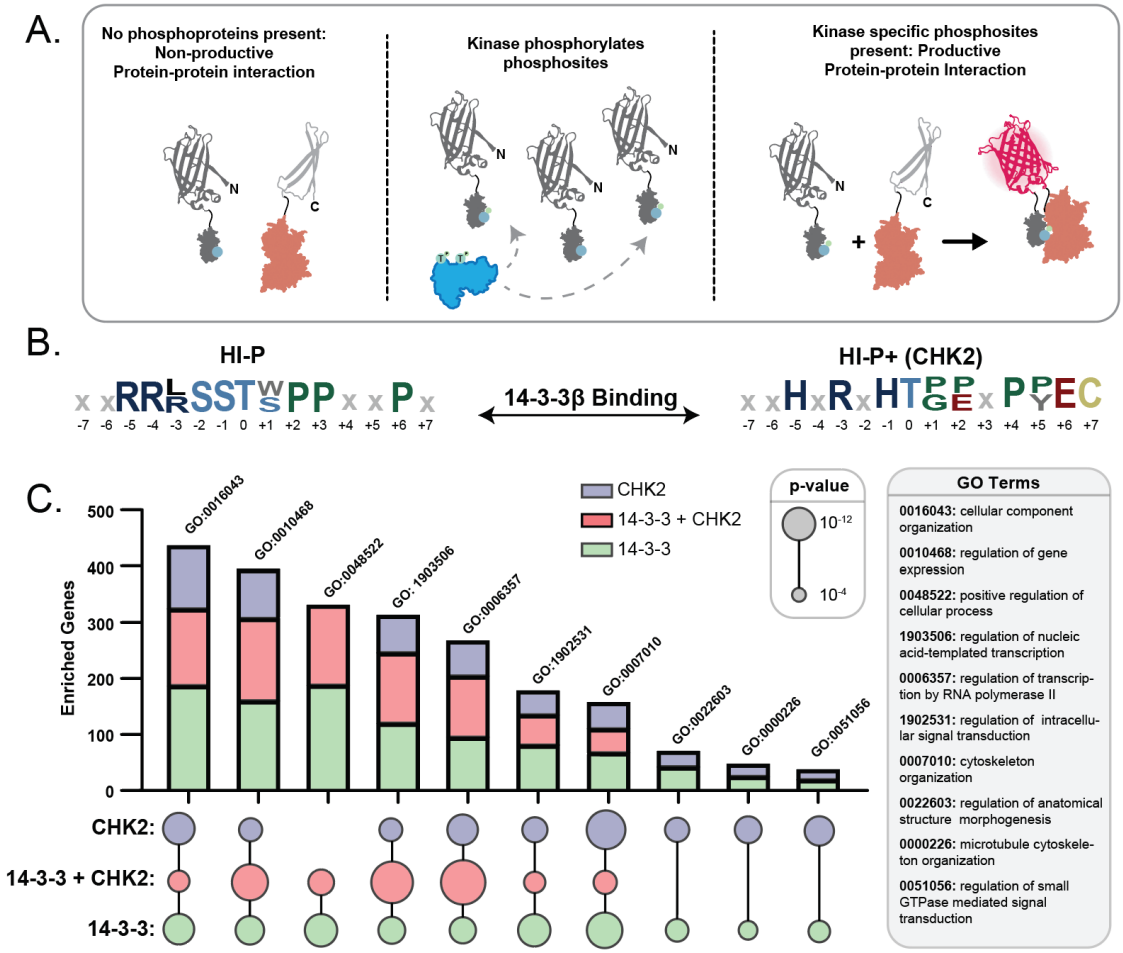

Figure 3-6| HI-P+ links phospho-binding activity to a kinase of interest. (A) HI-P+ relies upon a kinase of interest to phosphorylate Thr phosphosites to facilitate binding with a PBD. (B) Motif analysis for pThr phosphosites using 14-3-3β HI-P (left) and 14-3-3β co-expressed with active CHK2 (right). (C) PANTHER overrepresentation analysis for GO biological processes across HI-P (**green**), HI-P+ (**red**), and kinase profiling (**purple**) datasets. For each GO term, the number of genes (chart) is compared to term inclusion (circle) and significance (circle size, p-value). GO terms listed above each bar are detailed in the grey callout to the right.

Optimization of phosphoprotein expression

Based on previous work [80], we theorized that different growth conditions could enhance the expression of our pThr phosphosite library, which encompasses ~57k peptides containing various chemical properties. Using phosho-enrichment strategies coupled to MS analysis, we profiled library expression at 37°C and 20°C. After initially identifying a relatively limited number of phosphosites (~3.5k) across expression conditions (Fig 3-7A), we theorized that pThr containing peptides might have a decreased ionization efficiency compared to pSer. In order to test this hypothesis, we used an approach involving enzymatic de-phosphorylation of phosphopeptides following phospho-enrichment [106]. Phosphatase treatment resulted in the identification of an additional 4.5k phosphosites, supporting the idea that pThr hinders peptide ionization (Fig 3-7A). To characterize phosphosite expression independent of OTS function, we expressed the analogous non-phosphorylated Thr phosphosite library and identified ~25k unique phosphosite peptides (Fig 3-7B). These data illustrate the depth of our phosphosite library and highlight critical technical considerations for recombinant pThr-protein analysis by MS.

Figure 3-7| Production of phosphosite libraries varies by expression conditions

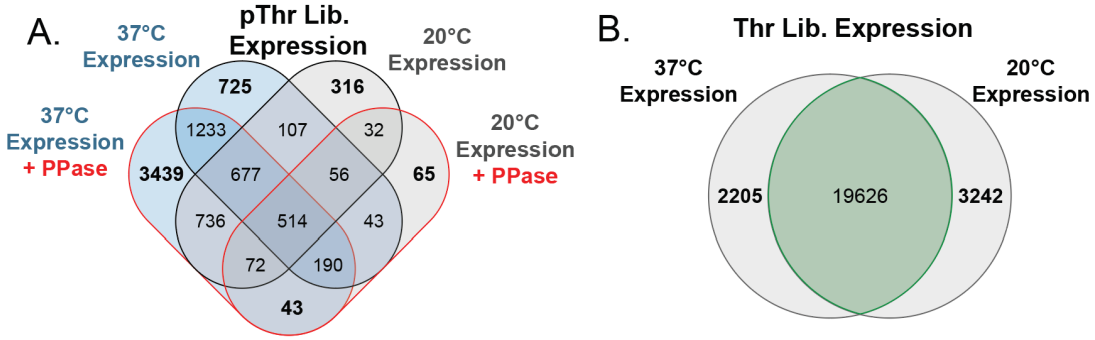

Figure 3-7| Production of phosphosite libraries varies by expression conditions. Venn diagram of phosphosites detected for (A) pThr and (B) Thr by MS/MS. Expression temperature and presence of phosphatase treatment (+ PPase) marked along the outside of each overlay.

Targeted assays for in-depth phosphoprotein studies

Genetic code expansion provides a novel toolkit for studying kinase functions

Phosphorylation by protein kinases is integral to signal transduction pathways and the regulation of cellular processes in mammalian cells. CHK2 kinase is phosphorylated at Thr68 by ATM in response to DNA damage, enabling CHK2 dimerization and trans-phosphorylation of activation loop residues Thr383 and Thr387 [107, 108]. Previous reports of recombinant CHK2 expression identified that full-length CHK2 undergoes trans-phosphorylation during protein expression and purification, yielding catalytically active kinase [109]. Trans-autoactivation at two regulatory pThr sites is difficult to dissect with traditional mutagenesis because amino acid substitutions that eliminate Thr sites can interfere with activation loop function. We aimed to use genetically encoded pThr to probe the precise mechanisms of CHK2 activation using the kinase domain (KD). Directly encoding pThr can potentially isolate the precise contributions of each pThr site without altering the primary activation loop sequence. We constructed a series of KD variants to dissect the contributions of each phosphorylation state to overall kinase activity (Fig 3-8A). Our initial analysis recapitulated previous observations of hyper-phosphorylation in full-length CHK2 and autoactivation of the KD. Double alanine substitution at Thr383 and Thr387 inactivated both the full-length kinase and the KD. Phos-Tag SDS-PAGE gels of the wildtype KD (Thr383/Thr387) suggested that only a single activation loop Thr was auto-phosphorylated, and MS analysis localized the phosphorylation site to Thr383 (Fig 3-9A&C). Direct synthesis of pThr383/Thr387 CHK2 KD appeared to result in hyperphosphorylation, while a pThr383/Ala387 mutant more closely resembled the wildtype KD shift (Fig 3-8A, lane four compared to lanes six and seven). We expected the

insertion of pThr387 into the KD to produce a phosphorylation profile similar to pThr383/Thr387 KD; however, we only observed a single phosphorylation state (Fig 3-8A, columns 8&9). MS analysis confirmed that the genetically programmed pThr387 was the only phosphorylation event within the activation loop (Fig 3-9A&D). Interestingly, encoding pThr at positions 383 and 387 produced a unique mixture of singly and doubly phosphorylated KD that was not recapitulated by any other variant. Detection of pThr383 in the activation loop of any CHK2 variant always corresponded with catalytic activity (Fig 3-9B, C&E). These results indicate that pThr383 is the sole residue responsible for CHK2 activation, while conversely, pThr387 serves to inhibit CHK2 activity.

To assess the activity of the KD variants, we performed *in vitro* phosphorylation assays using a 23 amino acid peptide of the well-characterized substrate CDC25C [100] (Fig 3-8A&B). Across all KD variants, only three phosphorylated the CDC25C substrate: full-length CHK2, wildtype KD, and pThr383 KD. We also observed that while pThr383/Thr387 was active, pThr383/Ala387 was inactive. Upon further investigation, we found that insertion of Ala at either position 383 or 387 ablated the KD activity (Fig 3-8C). However, Ala substitution at Thr225, another well-characterized CHK2 phosphosite [110], had no impact on KD activity or autophosphorylation. Our data indicate that alanine substitution at either activation loop position results in kinase inactivation but through a mechanism more complicated than phosphosite ablation. Collectivity this data shows that pThr383 is necessary and sufficient to activate CHK2 and implies that Thr383 phosphorylation may be the first step in CHK2 activation.

Figure 3-8| Genetic code expansion provides novel insights into CHK2 activity

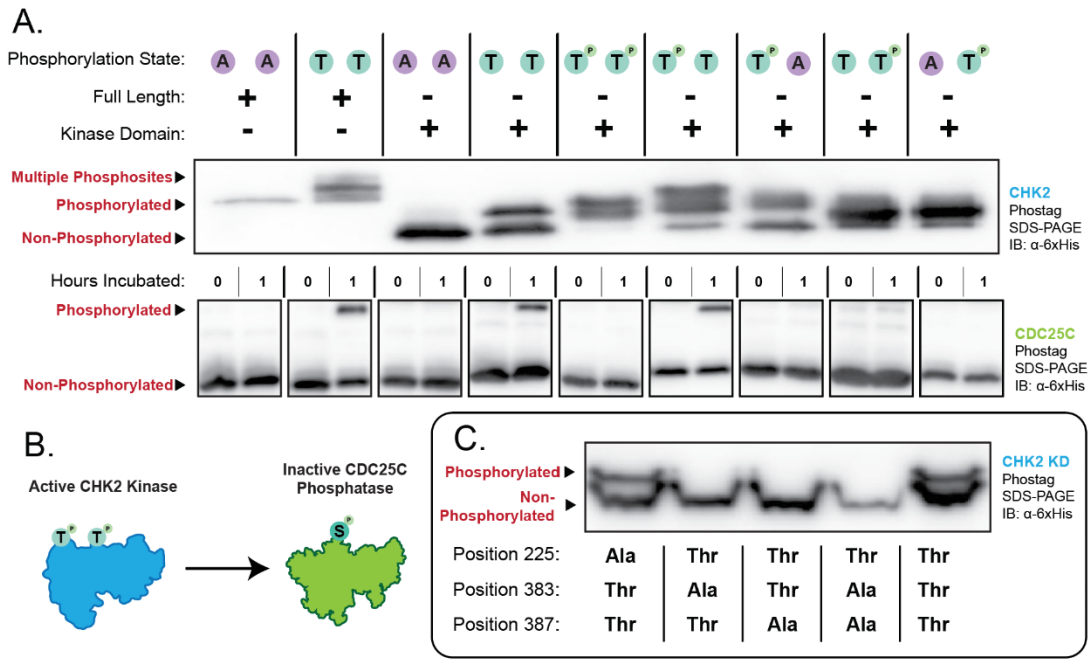

Figure 3-8| Genetic code expansion provides novel insights into CHK2 activity. (A) CHK2 programmed with either Ala, Thr, or pThr at positions 383 and 387 result in different phosphorylation patterns (top) leading to different activity with the CDC25C substrate (bottom) separated via Phos-Tag SDS-PAGE and visualized by immunoblot against 6xHis. (B) Schematic of a canonical CHK2 signaling cascade. (C) Phos-Tag SDS-PAGE gel visualized by immunoblot against 6xHis for CHK2 kinase domain variants with Ala encoded in either the activation loop (Thr383 and Thr387) or at an outlying phosphosite (Thr225).

Figure 3-9| CHK2 Analysis of CHK2 variant phosphorylation status

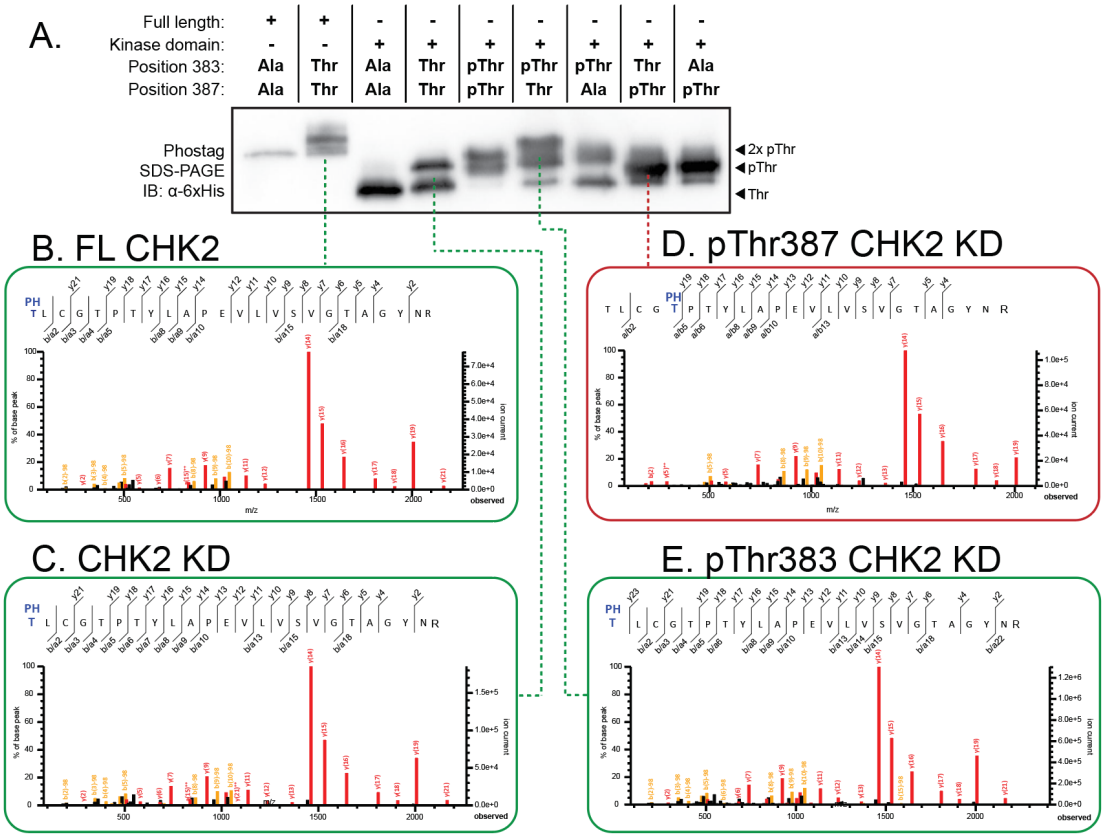

Figure 3-9| CHK2 Analysis of CHK2 variant phosphorylation status. (A) Phos-Tag SDS-PAGE gel visualized by immunoblot against 6xHis for phosphorylated CHK2 variants from Figure 3-8. (B-E) MS/MS ion spectra for CHK2 activation loop peptide fragments obtained from MASCOT. Phosphorylated amino acids denoted by PH (blue, bold) above corresponding residue. Each ion spectra corresponds to samples from a specific lane in (A), with spectra containing pThr383 (green) or pThr387 (red).

Chapter 4: Conclusion and future directions

Summary of genetic code expansion with phospho-amino acids

Since their discovery, a vast array of OTSs have been developed to incorporate diverse chemistries into recombinant proteins [111, 112]. Although genetic code expansion has been implemented across many different organisms [113-119], it is generally *E. coli*-centric. Despite the over-representation of *E. coli* in genetic code expansion, there is a lack of knowledge on how OTS expression impacts the host proteome. Many OTSs rely upon traditional protein expression strains, e.g., BL21, to enhance protein yield, opting to purify away truncated proteins [50]. A significant advancement came with the arrival of the recoded *E. coli* C321, enabling the first instance in which RF1 could be deleted. [120]. This minor change allowed for the installation of nsAAs as the 22nd amino acid at UAG codons, facilitating a wide array of new functions and activities [17, 120-122].

OTSs have provided unparalleled access to the study of PTM mediated protein-protein interactions. Many PTMs, such as protein phosphorylation, are reversible and often transient in their native environments. Protein phosphorylation is the most utilized PTM across the human proteome, but its labile nature makes it challenging to study [123, 124]. The site-specific incorporation of phospho-amino acids by OTSs has helped to address some of these challenges. Initial work centered around pSer, due to the serendipitous nature of archaea natively aminoacylating pSer to tRNA^{Cys} [42]. Further studies have expanded OTSs to pThr [50] and pTyr [48, 49, 125], enabling unprecedented access to the phosphoproteome. By far, the most well-characterized phospho-amino acid based OTS is the pSerOTS [17, 45, 46, 126], which has since been further modified to include an evolved aaRS, EF-pSer, and tRNA to enhance incorporation, and use of a non-hydrolyzable analogue to eliminate phosphatase mediated loss [45, 46]. Although considerable work has

gone into the co-translational insertion of phospho-amino acids, the field has failed to optimize OTSs in the context of cellular physiology.

Through systematic investigation of OTS components, we developed pSer-, pThr-, and pTyrOTS variants with improved efficiency for the co-translational insertion of phospho-amino acids at UAG codons (Fig 4-1). Systems-level interrogations of components demonstrated that OTS toxicity was due, in large part, to high-level aaRS expression. Reductions to plasmid copy number and regulation of component expression were central to enhancing pSer- and pThrOTS performance by reducing the metabolic burden on the host cell. Counter-intuitively, high-level expression of pSerRS results in decreased protein yield, while low-level expression enhances yield. Only in light of the cell's response to stimuli is it apparent that high-level expression activates the stringent response, thereby hampering protein output. Although this limitation was readily addressed with pSerOTS, pThrOTS designs required the deletion of *relA* to produce similar results.

Further design considerations include antibiotic resistance genes, which can alter translational burden and fidelity, highlighting an important but often overlooked, consideration for OTS design. Kanamycin, for example, interacts with the 30S ribosomal subunit to promote mistranslation, potentially enhancing non-standard amino acid incorporation, whereas tetracycline prevents aminoacylated-tRNA from entering the ribosomal A-site, undermining global translation [127, 128]. As pThr- and pTyrOTS activity are contingent upon intracellular phospho-amino acid pools for the production of aminoacylated-tRNA, the balance of tRNA and the amino acid substrate is integral to maintaining translational fidelity and regulation. For pThr, this is addressed through the expression of the *Salmonella enterica* enzyme, pduX [64], but is more limited for pTyr.

Dipeptide-mediated incorporation of pTyr, although functional, is currently technically challenging to produce and cost-prohibitive. Our work on Tyr auxotrophy provides the first insights into potential methods for increasing intracellular pTyr pools but requires further validation.

Although OTS components are generally believed to be orthogonal, our work shows that the incorporation of near-cognate amino acids is mediated by tRNA recognition elements conserved between prokaryotes and archaea. Despite the limited sequence space of tRNAs, minor substitutions can completely alter aminoacylation profiles and activity. We found that nucleotide substitutions to decrease mis-aminoacylation by Gly [76, 77] serendipitously provided a $tRNA^{Ala}$ suppressor, providing a method for incorporating phospho-ablative mutations in the same UAG containing constructs. The discovery of a novel suppressor tRNA underscores the importance of tRNA identity elements in OTS construction, representing a tRNA that has sequentially been evolved from Cys to pSer and finally to Ala with less than six nucleotide changes. Altering the anticodon, a shared recognition motif for aaRS editing domains, decreased phosphoprotein production in our hands. However, this may be mediated by RF1 competition but would require a strain with UGA recoded strain to verify.

Our findings, as a whole, demonstrate the issues surrounding incomplete orthogonality in OTS function. The field has traditionally avoided identifying limiting factors, providing a conformation bias in successive OTS production. This work highlights that better OTS design practices can decrease cellular toxicity, improve protein production, and enable more high-throughput studies of nsAAs. As a whole, our systematic approach

to OTS optimization generated improved pSer- and pThrOTSs, and notably contributed valuable insight into the rational design of OTSs by monitoring host cellular viability.

Figure 4-1| Co-translational insertion of phosphoamino acids enables recombinant phosphoprotein production

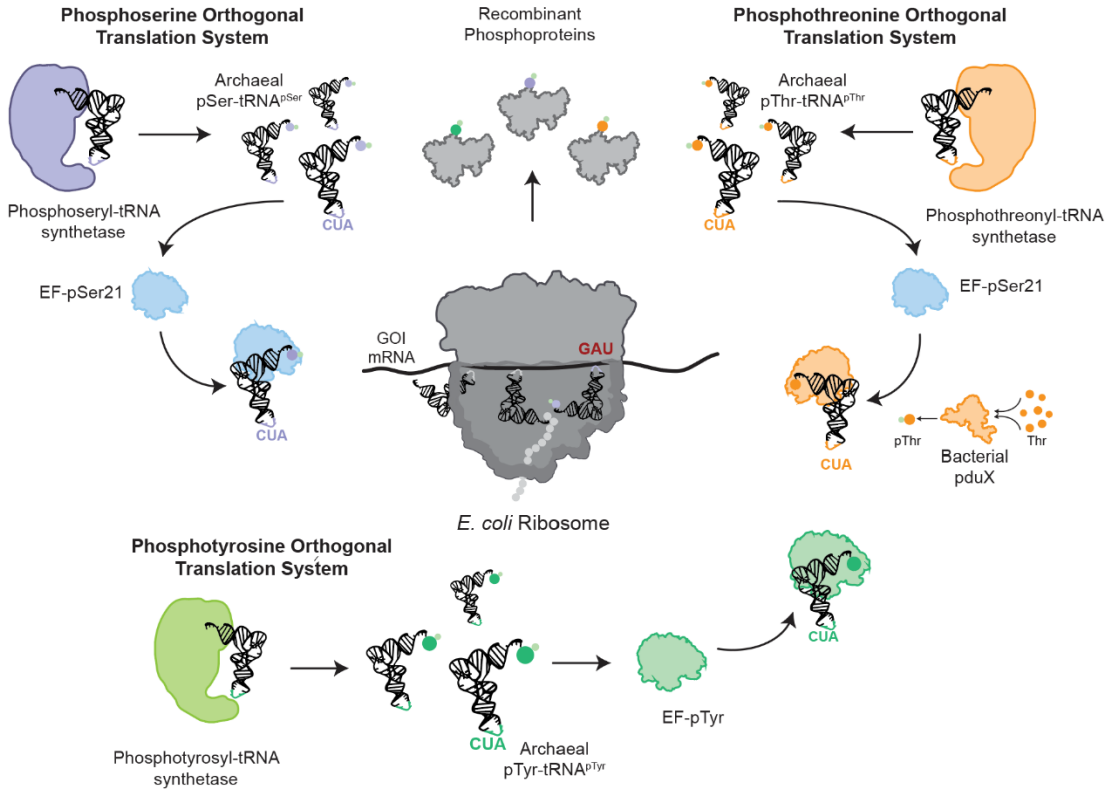

Figure 4-1| Co-translational insertion of phosphoamino acids enables recombinant phosphoprotein production. Each phospho-amino acid based OTS requires an evolved aminoacyl-tRNA synthetase, cognate tRNA, and elongation factor for high-fidelity incorporation into recombinant proteins.

Applications of co-translational insertion of phospho-amino acids

The promise of genetic code expansion to phospho-amino acids has been towards the advent of robust biological studies. To this end, we sought to characterize, for the first time, the effects of separate phosphosites on the activation loop of a kinase. CHK2, a Ser/Thr kinase involved cell cycle regulation, has been shown to have dual phosphorylation of Thr383 and Thr387 during activation by ATM [107, 108]. Our C321^{EpT} strain allowed, for the first time, identification of pThr383 within the CHK2 activation loop as the essential residue regulating kinase activity. The trans-phosphorylation activity of recombinantly expressed CHK2 made it previously impossible to dissect the independent contributions of specific phosphosites to CHK2 activity *in vitro*. Genetically encoded pThr is the only way to dissect such mechanisms. Together with our pThrOTS and the discovery of a novel Ala suppressor tRNA, we were able to evaluate the phoshpho-regulatory mechanisms of CHK2 in unprecedented detail. These advances demonstrate the general utility of our approach towards the characterization of individual kinase regulatory events, and ultimately, the expansion of our global understanding of phospho-regulatory mechanisms.

Building on these mechanisms, we sought to characterize the pThr-specific FHA binding of CHK2 and the *S. cerevisiae* homologue RAD53. Unfortunately, we were unable to recapitulate the phospho-binding activity of CHK2's FHA domain using HI-P. Utilizing the same HI-P framework (Fig 4-2B), we demonstrated how phospho-binding domain function could be maintained across evolutionarily divergent species with RAD53. Although motif analysis was inconclusive, functional enrichment for GO biological processes matched CHK2 biology [99-103]. Complications may be due to substrate motif elements being slightly altered during evolution. Validation of RAD53 interacting

substrates via AP-MS would provide clarity to the experiment. Initial work with RAD53 used a BL21^{EpT} strain, which, well functional with RAD53, failed to recapitulate results with 14-3-3β. The development of C321^{EpT}ΔrelA allowed us to characterize, for the first time, the interactome of 14-3-3β pThr sites. Utilizing our C321^{EpS} strain, we characterized the overlap of 14-3-3β binding between pSer and pThr. Our approach demonstrated the presence of unique motif elements and altered GO biological process. These enrichments detail how Ser and Thr may have differentially regulated activity with 14-3-3β and could be comparatively expanded to other Ser/Thr PBDs.

Using non-phosphorylated versions of our phosphosite libraries, we can rapidly profile the activity and specificity of Ser/Thr kinases (Fig 4-2A). Leveraging this approach, we identified the phosphosite substrates for full-length CHK2 and CHK2 pThr383-KD, allowing us to identify motif elements overlapping with canonical 14-3-3β binding sites. Until now, the determination of pSer- and pThr-mediated PPIs has been limited to single protein-substrate interactions. Hi-P+ provides the first cell-type independent platform to simultaneously link kinases and phospho-binding proteins to individual phosphosites in a single experimental approach (Fig 4-2C). Multi-level PPIs identified by HI-P+ pave the way for constructing more comprehensive protein interaction networks linking kinases, phosphosites, and phospho-binding domains across the human proteome. HI-P+ has the potential to expand our understanding of phosphorylation-dependent cellular processes.

Although considerable effort went into developing a C321^{EpY} strain, functional utilization with HI-P produced mixed results. The N-terminal SH2 domain of SHP2 produced a similar fluorescent shift with a pTyrOTS and TyrOTS in HI-P. When analyzed in conjunction with the inconsistent motif elements indicated that considerably more work

is required to make C321^{EpY} viable with HI-P. The results may also be complicated by the unique nature of SH2 domains in receptor tyrosine kinases. The localization of SH2 domains to the membrane of cells limits the number of proteins they encounter during biological processes, thus precluding the necessary specificity seen with 14-3-3 proteins. The high degree of enrichment with a TyrOTS indicates a loss of specificity; whether this is due to SH2 binding or over-expression in *E. coli* requires further study.

Figure 4-2| Summary of systems level studies enabled by genetic code expansion

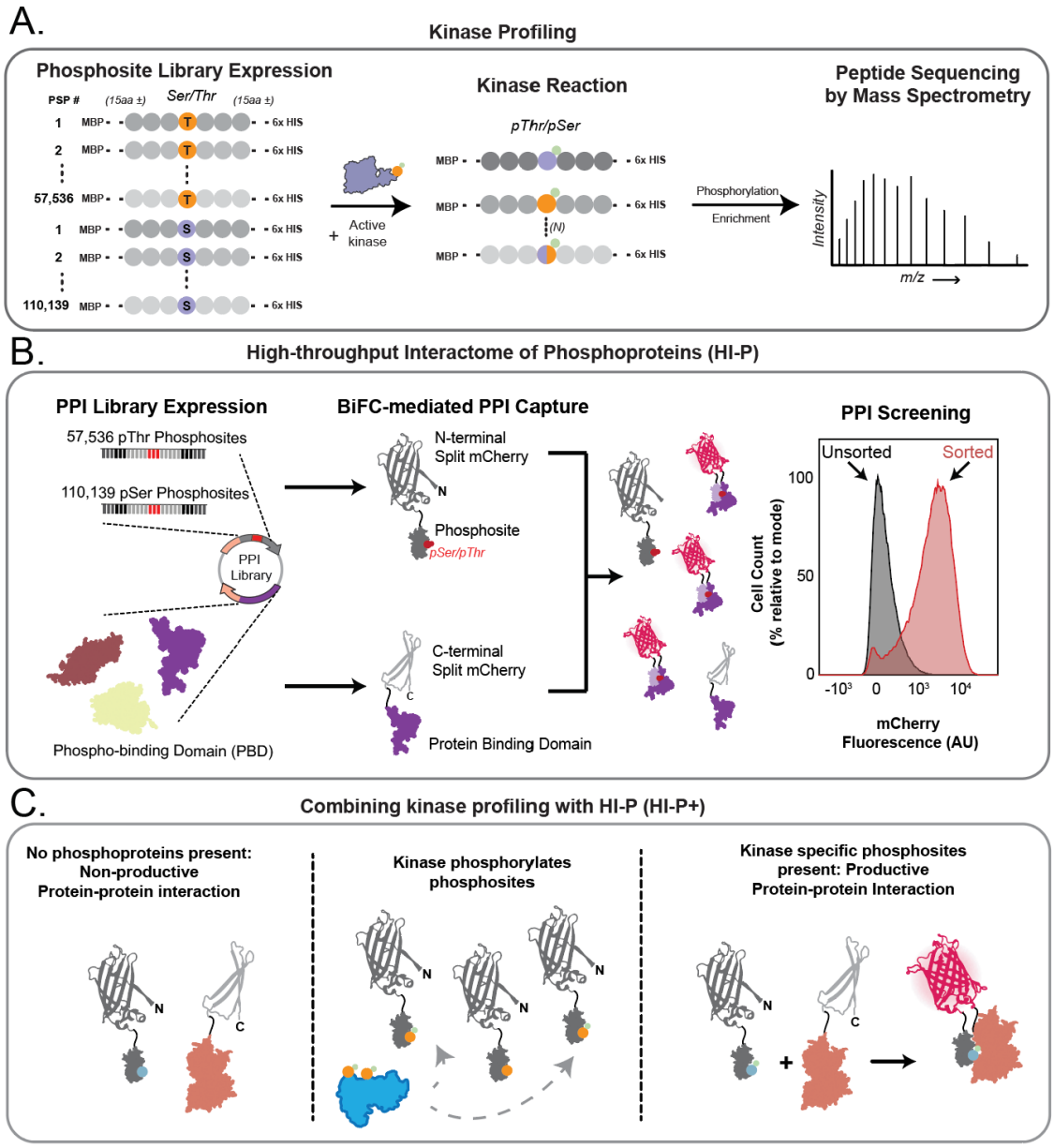

Figure 4-2| Summary of systems level studies enabled by genetic code expansion. (A) Non-phosphorylated phosphosite libraries enable screening of kinase substrates across the human proteome. (B) HI-P identifies phospho-binding domain substrates with phosphosite resolution. (C) HI-P+ couples kinase screening with HI-P to link phosphobinding-domain substrates with kinase substrates for multi-level interactomics.

Future expansion of orthogonal translation systems

In this work, OTSs have been used to probe specific biological questions and tease apart systems-wide phosphorylation-specific interactions. Although we now have robust pSer- and pThrOTSs, significant advances can still be made in both their development and application. In general, the future of OTS evolution is less in the OTS components and more in modification of the strains. Despite phospho-amino acids being incorporated as 22nd amino acids, they are still housed on episomal vectors that require constant antibiotic selection. Future developments could obviate the need for such systems through the stable integration of pSer- and pThrOTS components, as seen with other nsAAs [129]. Other modifications include the evolution of the C321 background to revert passenger mutations derived from recoding. Even though other groups have demonstrated the utility of such work [130], it has yet to be attempted in an OTS-specific context. The general problem with installing OTS components and lab-directed evolution is that the fittest population will probably disrupt the transcription and translation of said components. It is possible to address this issue by reassigning sense codons to UAG in permissible positions of proteins, enabling dependence on the OTS [131]. As *E. coli* does produce some pSer phosphosites naturally, it may be as simple as mutating known sites throughout the genome in a stably integrated strain.

This work provides significant advances for pSer- and pThrOTSs, but is still limited in scope in several regards. The pTyrOTS, featuring a functional aaRS and tRNA pair, is still hampered by phosphatases and low intracellular levels of pTyr. Target candidates identified during Tyr auxotrophy experiments provide a putative list of potential genes to enhance pTyr import. In addition to the limitations of pTyr, there are still several other

non-traditional phospho-amino acids seen throughout nature, including pHis and pAsp, that would benefit from OTS development. Canonically, non-traditional phospho-amino acids are central to bacterial two-component signaling pathways [7, 132], but more recent work has demonstrated that pHis may play a previously unknown role in mammalian signaling pathways [133-136]. Genetic code expansion to such non-standard phospho-amino acids may provide new insights into both human and bacterial physiology.

Even though we have only characterized phospho-amino acid-dependent OTSs, several other PTMs have come under the purview of genetic code expansion. One of the most well-characterized OTSs is based on the incorporation of acetyl-lysine (AcK) into recombinant proteins [137-139]. Lysine acetylation mirrors protein phosphorylation, i.e., the addition is catalyzed by acetyltransferases, removed by deacetylases, and bound by bromodomains (BRDs) [140-142]. AcKOTSs utilize the PylRS and $tRNA^{Pyl}$ pair, as they already decode UAG codons and incorporate a modified Lys [137-139]. The first AcKOTS utilized the *Methanosarcina barkeri* PylRS and $tRNA^{Pyl}$ [138], but other iterations have featured the *Methanosarcina mazei* pair [137, 139]. Despite the significant expansion to AcKOTSs, there have yet to be any high-throughput studies of acetylation-dependent interactions. Expanding the scope of HI-P to the High-throughput Interactome of AcK (HI-K) through the use of BRDs and large-scale DNA array synthesis could provide novel insights into AcK function. Although the addition of chromodomains, lysine methylation binding domains, would also prove helpful, it is currently limited by the use of methyl-lysine protected analogues in OTS variants [143, 144]. The structural similarity of methyl-lysine and Lys precludes high substrates specificity for a functional aaRS. Beyond the use

of protecting groups, the only mechanism to address this problem would be the incorporation of aaRS editing domains.

Generally, tRNA mis-aminoacylation events are corrected through the use of aaRS editing domains and trans-editing factors [145-147]. The only OTS to utilize an editing domain relied on transplantation of the PheRS editing domain (to remove Tyr) into an evolved TyrRS [148]. Naturally occurring trans editing factors, such as AlaXp, ProXp-x, ProXp-y, and ProXp-z can remove multiple amino acids that have been misacylated onto $tRNA^{Pro}$ independent of an aaRS [149, 150]. The evolution of these editing factors to current OTSs and phosphoamino-acids could significantly enhance the fidelity of incorporation. In particular, ProXp-y already removes misacylated Ser and Thr, which would be an excellent addition to the pSer- and pThrOTS repertoire [150]. Even though our work to incorporate recognition elements of DTD into $tRNA^{pSer}$ resulted in aminoacylation by AlaRS, and a serendipitous Ala suppressor, it does not preclude the evolution of DTD to recognize $tRNA^{pSer}$. If the problem is due to the presence of the G3:U70, then it should be possible to perform site-directed mutagenesis of DTD to recognize the C3:G70 of $tRNA^{pSer}$. The advent of such editing mechanisms may become more critical in implementing multiple OTSs in a single organism.

Expanding the phospho-amino acid based OTS repertoire could provide novel insights into both eukaryotic and prokaryotic signaling. As researchers expand OTSs to other PTMs, such as tyrosine sulfonation (sTyr) [151, 152] and nitrosylation (nTyr) [153, 154], it will be possible to understand more complex multi-PTM interactions. For instance, the 14-3-3 proteins are nitrosylated at Tyr residues during oxidative and nitrative stress [155, 156]. An expansion of this work would make use of encoding nTyr within 14-3-3

proteins to test for changes in interactions with pSer and pThr using HI-P. Several additional PTMs have been characterized within 14-3-3 proteins that do not have a clear function, including multiple phosphorylation sites [157] that may also be worth studying in a dual PTM dependent HI-P context. Other expansions of the dual PTM interactome may help identify how disease-relevant PTMs are regulated. A target with a high therapeutic value is CCR5, a chemokine receptor essential for the uptake of HIV-1 into cells [158, 159]. The interaction between CCR5 and HIV-1 depends on the sulfonation of multiple Tyr residues at the N-terminal domain to facilitate entry [160]. However, phosphorylation of C-terminal Ser residues may play a role in entry into endosomes [161, 162]. Combining the sTyr- and pSerOTSs with modifications necessary for GPCR expression in *E. coli* [163] could provide novel insight into the structural and functional role of PTMs in CCR5 function.

Other limitations to the field include an inability to incorporate multiple phospho-amino acids into a given protein. ERK1 & 2, significant mediators of cell growth and immune activation, rely upon both pThr and pTyr in the activation loop for proper function [78]. Our work in expanding pSer- and pThrOTSs to decode at UGA codons provides a proof-of-principle that it is possible to repurpose an additional stop codon for genetic code expansion. Based on the experiments described throughout, it is now theoretically possible to encode pThr at UGA and pTyr at UAG to make selectively phosphorylated versions of ERK1 & 2, e.g., single pThr, single pTyr, or both phosphosites. This may reveal novel insights into activation, as seen with CHK2. Expanding incorporation of dual phospho-amino acids would also provide a means to activate every member of the kinome. Combining this technology with our ability to create non-phosphorylated versions of every

known pSer, pThr, and pTyr phosphosite would theoretically allow for the screening of every kinase substrate in a cell-type independent manner, without the need for mammalian or insect cell lines.

Chapter 5: Methods

Cell growth and general techniques

All DNA oligonucleotides were obtained with standard purification and desalting from Keck Oligonucleotide Resource, Yale University. DNA sequencing services were obtained through the Keck DNA Sequencing facility, Yale University. Unless otherwise stated, all cultures were grown at 37°C in LB-Lennox medium (LB, 10 g/L bacto tryptone, 5 g/L sodium chloride, 5 g/L yeast extract). LB agar plates were LB plus 16 g/L bacto agar. Antibiotics were supplemented for selection, where appropriate (Kanamycin at 50 μg/mL, Spectinomycin at 50 μg/mL, Tetracycline at 12 μg/mL, and Ampicillin at 100 μg/mL). Lys-pTyr dipeptide was added to a concentration of 2 mM as needed. *E. coli* Top10 cells (Invitrogen, Carlsbad, CA) were used for cloning and plasmid propagation. NEBuilder HiFi Assembly Mix and restriction enzymes were obtained from New England BioLabs. Plasmid DNA preparation was carried out with the QIAprep Spin Miniprep Kit (Qiagen). Pre-cast 4–12% (wt/vol) Bis–Tris SDS–PAGE gels were purchased from Bio-Rad. Phos-Tag reagent for hand-cast protein gels was purchased from FUJIFILM Wako Chemicals U.S.A. Corporation.

All plasmids were transformed into recipient strains by electroporation. Electrocompetent cells were prepared by inoculating 20 mL of LB with 200 μL of saturated culture and growing at 37°C until reaching an OD_{600} of 0.4. Cells were harvested by centrifugation at 8,000 RPM for 2 min. at 4°C. Cells pellets were washed twice with 20 mL ice cold 10% glycerol in deionized water (dH2O). Electrocompetent pellets were re-suspended in 100 μL of 10% glycerol in dH2O. 50 ng of plasmid was mixed with 50 μL of re-suspended electrocompetent cells and transferred to 0.1 cm cuvettes, electroporated (BioRad GenePulser™, 1.78 kV, 200 Ω, 25 μF), and then immediately resuspended in 600

μL of LB. Transformed cells were recovered at 37°C for one hour and 100 μL was subsequently plated on appropriate selective medium.

Construction of OTSs and strains

$C321^{XpT}$ & $BL21^{XpT}$ strain creation

Strains were made using two different methods. The initial individual knockout strains used in 2-10 were generated used Multiplexed Automatable Genome Engineering (MAGE) [164]. Briefly, a C321.ΔA strain containing a temperature-inducible lambda red operon is grown to an OD of 0.4 in 3 mL LB at 34°C. Cells are then set to shake in a 42°C water bath for 15 minutes and are then pelleted by centrifugation at 13,000 rpm for 30s at 4°C, the supernatant is then removed by pipetting. The cells are then resuspended in 1 mL of 10% glycerol on ice. The pelleting and resuspension are repeated for a total of 3 times. Following the final pelleting, cells are resuspended in 50 ul 10% glycerol with 1 μL of 10 μM MAGE oligonucleotide. Cells are then transferred to 0.1 cm cuvettes, electroporated (BioRad GenePulser™, 1.78 kV, 200 Ω, 25 μF), and then immediately resuspended in 600 μL of LB. Transformed cells were recovered at 34°C for one hour. This process was repeated a total of three times in one day, with the final set of cells being plated on LB plates. Knockouts were confirmed using colony PCR. MAGE primers were designed to produce a frame shift mutation and generate multiple stop codons early in the ORF. MAGE primers were ordered from IDT with enhanced stability modification (denoted as *). All primers used in MAGE are listed in Table S1.

All C321.ΔA.exp and BL21 (DE3) based strains were generated by traditional recombineering methods with a FRT-kanamycin-FRT casette [165]. Primers with

homology to the gene of interest were amplified from pKD4 [165] to generate a homologous sequence with an internal FRT-kanamycin-FRT casette. Primers used in this study were those previously validated for each knockout in the Keio collection [166]. Cells were grown and transformed by electroporation as stated above, but with ~50 ng of the PCR product in place of plasmid DNA. Cells were then plated on 50 ng/µl kanamycin LB plates and grown overnight at 37°C. Individual colonies were then screened for the presence of the kanamycin-FRT casette or gene of interest. Cells that contained the kanamycin-FRT casette and were absent the gene of interest were then grown overnight in 5 mL LB containing 50 ng/µl kanamycin in a 15 mL pop-cap (Falcon) tube and set to shake at 230 RPM and 37°C. The following day cells were then transformed by electroporation, as mentioned above, using the pCP20 plasmid, containing a temperature sensitive origin and Flp recombinase. Cells were selected on 100 ng/ul ampicillin LB plates and grown overnight at 30°C. Individual colonies were then patched on separate plates containing either 100 ng/ul ampicillin, 50 ng/µl kanamycin, or no antibiotic in LB. Plates were incubated overnight at 30°C. Individual colonies that were negative for kanamycin resistance and positive for ampicillin resistance were then streaked on LB plates and incubated overnight at 42°C. The following day plates were transferred from 42°C to 30°C for additional growth. When colonies became visible to the eye they were then screened for the absence of the gene of interest as seen in Fig S3. All primers used for the generation of knockout cassette and screening for the gene of interest can be found in Table S1.

C321^{XpS} Strain construction

Strain modification using a lambda red based strategy was performed as previously

described[165]. Briefly, transformants carrying a lambda red plasmid (pKD46) were grown in 20 mL LB cultures with ampicillin and l-arabinose at 30°C to an OD_{600} of ≈0.6 and then made electrocompetent. 10–100 ng of PCR product targeting serB was transfomred into the prepared cells and recovered in 1 mL of LB for 1 h at 37°C. Following recovery, one-half was spread onto LB-agar plates with Kanamycin for selection. serB gene deletion was verified by PCR and selected colonies were purified non-selectively at 37°C remove pKD46. The KAN deletion cassette was excised from the serB locus using FLP recombinase (pCP20).

Construction of pThrOTS vectors

The starting pThr tRNA and pThrRS were generated by site-directed mutagenesis of a pSerOTS [167], gifted from the O'Donoghue lab. The pduX gene, OXB20 promoter, and proK tRNA casette were produced as separate gene gBlocks from GeneWiz. The Ef-pSer1 and Ef-pSer21 were amplified from previously reported pSerOTSs [17, 42]. Generally, OTSs were assembled by gibson assembly using NEBuilder HiFi DNA Assembly master mix (NEB). The pThrOTS backbone was made de novo by a 4 fragment Gibson assembly. tRNA cassettes were assembled into the pThrOTS backbone by Gibson assembly to create $pThrOTS^{Zeus}$. The $pThrOTS^{Hercules}$ variant was created by 3 fragment Gibson assembly consisting of the (1) pThrRS/Ef-pSer21/pduX operon, (2) ClodF13 spectinomycin pUltra backbone [49], and (3) CHK2 under control of glnS*. Modifications to $tRNA^{pThr}$ anticodon to UGA were performed using site directed mutagenesis.

Construction of pSerOTS vectors

Generally, OTS components were amplified from the progenitor OTS, pSerOTSλ (SepOTSλ, Addgene # 68292). The pSerRS (SepRS9) and EF-pSer (EF-Sep21) were amplified from pSerOTSλ and placed under the control of the glnS* promoter and assembled with the kanamycin selection marker and origin of replication from pSerOTSλ by gibson assembly using NEBuilder HiFi DNA Assembly master mix (NEB). tRNA expression cassettes (1x-, 2x-, 4x-, and 6x-tRNA) were constructed from a 2x-tRNA construct under the control of a proK tRNA promoter and terminator and separated by a valU tRNA linker that was ordered as a DNA fragment from Genewiz. tRNA cassettes were assembled into the previously assembled pSerRS-EF-pSer backbone by gibson assembly creating ColE1-based pSerOTSc OTSs. OTS variants with different origins of replication were created by assembly of Rop amplified from pSerOTSλ into the ColE1-based OTSs and p15a variants were created by replacement of ColE1 with p15a origins by gibson assembly. A 2x $tRNA^{pSer}_{UCA}$ casette was designed and synthesized by Genewiz and assembled into vectors via NEBuilder HiFi DNA Assembly master mix (NEB).

Construction of pTyrOTS and related vectors

The starting CMFOTS [53] was gifted from the Schultz lab. Site directed mutagenesis was performed on episomal housed EF-Tu to reproduce EF-pTyr [47]. The modified EF was then inserted into a polycistronic operon with CMFRS by gibson assembly using NEBuilder HiFi DNA Assembly master mix (NEB). Phenylphosphate synthase subunits were cloned from *T. aromatica* DNA and inserted into a polycistronic operon under control of pBAD promoter.

Growth characterization

Stationary phase pre-cultures were obtained by overnight growth with shaking at 37°C in 5 mL of LB supplemented with antibiotic for plasmid maintenance as appropriate. Stationary phase cultures were diluted to an OD_{600} of 0.01 in 250 μL of LB (N=3) supplemented with appropriate antibiotic and induction reagents. Growth was monitored on a Biotek Synergy H1 plate reader. OD_{600} was recorded at 10-minute intervals for 16 hours at 37°C with continuous shaking. Growth rate was determined using growthcurveR 0.3.1 [168] in R 3.5 [169] with RStudio 1.3 [170]. Data was averaged and plotted with standard deviation.

Protein production and purification

Analytical gel and immunoblotting

100 μM Phos-tag acrylamide (FUJIFILM Wako Chemicals U.S.A. Corporation) within hand-cast 12% acrylamide gels were used for separation of phosphorylated reporter proteins. SDS-PAGE gels (4–15% acrylamide, Bio-Rad) and Phos-tag gels were transferred onto PVDF membranes. All gels were visualized by immunoblot. Anti-6xHis immunoblots were performed using 1:2,500 diluted rabbit α-6xHis antibody (PA1-983B, Thermo Fisher Scientific) in 5% w/v milk TBST for 1 h and subsequently washed 3x with TBST. Secondary antibody incubations used 1:10,000 diluted donkey anti-rabbit HRP (711-035-152, Jackson ImmunoResearch) in 5% w/v milk in TBST for 1 h and were subsequently washed 3x in TBST. Protein bands were then visualized using Clarity ECL substrate (Bio-Rad) and an Amersham Imager 600 (GE Healthcare Life Sciences).

MS-READ reporter purification

Frozen *E. coli* cell pellets were thawed on ice. Pellets were lysed by sonication with a lysis buffer consisting of 50 mM Tris-HCl (pH 7.4, 23°C), 500 mM NaCl, 0.5 mM EGTA, 1mM DTT, 10 % glycerol, 50 mM NaF, and 1 mM Na_3O_4V. The extract was clarified with two rounds of centrifugation performed for 20 minutes at 4°C and 14,000 x g. Clarified lysates were applied to Ni-NTA metal affinity resin and purified according to the manufacturer's instructions. Wash buffers contained 50 mM Tris pH 7.5, 500 mM NaCl, 0.5 mM EGTA, 1mM DTT, 50 mM NaF, 1 mM Na_3VO_4 and 20 mM imidazole. Proteins were eluted with wash buffer containing 250 mM imidazole. Eluted protein was subjected to 4 rounds of buffer exchange (20mM Tris pH 8.0 and 100mM NaCl) and concentrated using an Amicon Ultra-0.5 mL 30 kDa molecular weight cutoff spin filter (Millipore Sigma #UFC503096).

CHK2, Thr, and pThr library purification

1.5L cultures of LB 50 μg/mL Kanamycin, 100 μg/mL Ampicillin were inoculated to OD 0.1 with overnight culture containing C321 cells expressing $pThrOTS^{Zeus}$ (or Ser/Ala tRNA suppressor plasmids) and the expression plasmid (see supplemental table 1) containing an MBP-6xHis fusion construct. Cultures were set to shake at 230 RPM, 37°C. At OD 0.4, 1.5L flasks were induced with IPTG and arabinose, and set back at 230 RPM, 37°C for 4 hours. For the pThr and Thr libraries, an additional round of expression was similarly inoculated, but grown to and OD of 0.8, induced with IPTG and arabinose, and set at 230 RPM, 20°C for 20 hours. Cultures were pelleted at 5,000 RPM for 5 minutes,

transferred to separate 50 mL falcon tubes and frozen at -80°C.

Thawed cell pellets were re-suspended in 30 mL of lysis buffer (50 mM Tris/HCl pH 7.4, 300 mM NaCl, 1 mM DTT, 1 mg/mL lysozyme, 50 mM NaF, 1 mM $NaVO_4$, Roche protease inhibitor tablet, 1 µL bezonase), and incubated at room temperature for 20 min, followed by sonication. The lysates were centrifuged at 45,000g, 45 min, 4°C and the clarified lysate was transferred to a 15 mL falcon tube, and centrifuged at 45,000g, 25 min, 4°C to remove all remaining insoluble material. The clarified lysate was then purified via gravity column. Briefly, 6 mLs (3 mL bed volume) of NiNTA agarose beads (Qiagen #30210) was pre-equilibrated with lysis buffer before applying cell lysate. Resin was then washed with 100 mL of NiNTA wash buffer (50 mM Tris/HCl pH 7.4, 500 mM NaCl, 500 µM DTT, 50 mM NaF, 1 mM $NaVO_4$, 5% glycerol, 10 mM imidazole) before eluting with 20 mL NiNTA elution buffer (50 mM Tris/HCl pH 7.4, 500 mM NaCl, 1 mM DTT, 50 mM NaF, 1 mM $NaVO_4$, 250 mM imidazole). Eluted protein was then diluted with 20 mL dilution buffer (20 mM Tris/HCl pH 7.4). 6 mLs (3 mL bed volume) of amylose resin (New England Biolabs #E8021L) was applied to a new column and equilibrated with 50 mL amylose wash buffer (20 mM Tris/HCl pH 7.4, 1 mM EDTA, 150 mM NaCl 50 mM NaF, 1 mM $NaVO_4$, 5% glycerol). Diluted protein elution was then applied to amylose column and washed with 100 mL amylose wash buffer before eluting with 15 mL amylose elution buffer (20 mM Tris/HCl pH 7.4, 1 mM EDTA, 150 mM NaCl, 10 mM maltose, 50 mM NaF, 1 mM $NaVO_4$, 5% glycerol). Samples were concentrated down using Amicon ultra-15 10kDa cutoff columns (Millipore Sigma #UFC901008). pThr phosphosite libraries were then frozen at -80°C. CHK2 kinases were dialyzed in SnakeSkin™ (ThermoFisher #68035) overnight at 4°C in dialysis buffer #1 (50 mM Tris/HCl pH 7.4, 150 mM NaCl, 5 mM

BME, 10% glycerol), then dialyzed for 4 hours at 4°C in dialysis buffer #2 (50 mM Tris/HCl pH 7.4, 150 mM NaCl, 5 mM BME, 50% glycerol) before being frozen at -80°C.

Ser/Thr phosphosite libraries were incubated with 20 µl (40 units) PreScission protease (GE Healthcare Life Sciences) end-over-end overnight at 4°C. The cleaved samples were then passed through an Amicon molecular weight cutoff to remove the cleaved MBP and any uncleaved protein. Samples were then concentrated using Amicon Ultra-0.5 3-kDa (Millipore Sigma #UFC500396) molecular weight cutoff columns and buffer exchanged with 10 mM Tris, pH 7.4. Eluted protein was subjected to 4 rounds of buffer exchange (10mM Tris pH 7.4 and 150mM NaCl) and concentrated using an Amicon 30 kDa molecular weight cutoff spin filter (Millipore Sigma # UFC503096).

Protein digestion and mass spectrometry

MS-READ analysis

Affinity purified, buffer exchanged protein was digested and analyzed by mass spectrometry as described previously with some modifications [17]. The concentration of protein was determined by UV280 spectroscopy, 5 µg of MS-READ reporter was dissolved in 12.5 µl solubilization buffer consisting of 10 mM Tris-HCl pH=8.5 (23°C), 10 mM DTT, 1 mM EDTA and 0.5% acid labile surfactant (ALS-101, Protea). Samples were heat denatured for 15 min at 55°C in a heat block. Alkylation of cysteines was performed with iodoacetamide (IAA) using a final IAA concentration of 24 mM. The alkylation reaction proceeded for 30 min at room temperature in the dark. Excess IAA was quenched with 200 mM DTT and the buffer concentration was adjusted using a 1 M Tris-HCl pH 8.5 resulting in a final Tris-HCl concentration of 150 mM. The reaction was then diluted with water and

1 M $CaCl_2$ solution to obtain an ALS-101 concentration of 0.045% and 2 mM $CaCl_2$ respectively. Finally, sequencing grade porcine trypsin (Promega) was added to obtain an enzyme/protein ratio of 1/5.3 and the digest was incubated for 15 h at 37°C without shaking. The digest was quenched with 20% TFA solution resulting in a sample pH of 2. Cleavage of the acid cleavable detergent proceeded for 15 min at room temperature. Digests were frozen at -80°C until further processing. Peptides were desalted on C_{18} UltraMicroSpin columns (The Nest Group Inc. #SUM SS18V). Columns were initially conditioned with 400 µL 80% ACN 0.1% TFA followed by 200 uL 0.1% TFA. Samples were pelleted at 2000g for 1 minute before having lysate applied to columns. Sample was loaded onto column and washed twice with 400 µL 0.1% TFA before eluting with twice with 200 µL 80% ACN 0.1% TFA. Peptides were dried in a vacuum centrifuge at room temperature. Dried peptides were reconstituted and analyzed by LC-MS/MS. Heatmaps were generated with heatmaply 1.2.1 [171] in R [169] using RStudio [170]

Digestion of E. coli protein extracts for shotgun proteomics

20 mL cultures were inoculated to a starting OD of 0.01 in LB media using an overnight culture to stationary phase. After reaching mid-log, cells were chilled on ice and pelleted by centrifugation for 2 min at 8000 rpm. The resulting pellet was frozen at -80°C for downstream processing. For cell lysis and protein digest, cell pellets were thawed on ice and 2 µL of cell pellet was transferred to a microcentrifuge tube containing 40 µl of lysis buffer (10 mM Tris-HCl pH 8.6, 10 mM DTT, 1 mM EDTA, and 0.5 % ALS). Cells were lysed by vortex for 30 seconds and disulfide bonds were reduced by incubating the reaction for 30 min. at 55°C. The reaction was briefly quenched on ice and 16 µl of a 60

mM IAA solution was added. Alkylation of cysteines proceeded for 30 min in the dark. Excess IAA was quenched with 14 µl of a 25 mM DTT solution and the sample was then diluted with 330 µl of 183 mM Tris-HCl buffer pH 8.0 supplemented with 2 mM $CaCl_2$. Proteins were digested overnight using 12 µg sequencing grade trypsin. Following digestion, the reaction was then quenched with 12.5 µl of a 20 % TFA solution, resulting in a sample pH ~2. Remaining ALS reagent was cleaved for 15 min at room temperature. The sample (~30 µg protein) was desalted by reverse phase clean-up using C18 MicroSpin columns (The Nest Group #SEM SS18V). Briefly, the column was conditioned twice with 400 µL 80% ACN, 0.1% TFA and twice with 400 µL 0.1% TFA by centrifugation. Samples were pelleted at 2000g for 1 minute before applying lysate to the column. Columns were then washed twice with 400 µL 0.1% TFA and eluted twice with 200 µL 80% ACN, 0.1% TFA. The desalted peptides were dried at room temperature in a rotary vacuum centrifuge and reconstituted in 30 µl 70 % formic acid 0.1 % TFA (3:8 v/v) for peptide quantitation by UV280. The sample was diluted to a final concentration of 0.2 µg/µl with 5 µl (1 µg) injected for LC-MS/MS analysis.

Phosphothreonine phosphosite library digestion, enrichment and dephosphorylation

5 mg of purified pThr phosphosite library was digested using S-Trap™ midi columns (Protifi) per manufacturer's instructions. Phosphoproteins were then enriched using Thermo Scientific High-Select™ Fe-NTA Phosphopeptide Enrichment Kit (#A32992) followed by Thermo Scientific High-Select™ TiO_2 Phosphopeptide Enrichment Kit (#A32993) following manufacturer's instructions. Samples were split into separation fractions before being dried in a vacuum centrifuge at room temperature. Pooled

Fe-NTA and TiO_2 flow through was cleaned up and desalted using C18 MicroSpin (The Nest Group #SEM SS18V) columns as described earlier. One set of phospho-enriched samples was reconstituted in 2 µl 30 % ACN 0.1 % FA and vortexed for 30s. An additional 10 µL of 3:8 70 % formic acid: 0.1% TFA was added to each sample, and vortexed for 30s before being spun down. 5 µl was injected for LC-MS/MS analysis.

The remaining dried phospho-enriched samples were resuspended in 50 µL 1x Protein MetalloPhosphatases (PMP) buffer (NEB) with 10 mM $MnCl_2$. 2,000 units of lambda phosphatase (NEB #P0753S) and 5 units calf intestinal phosphatase (NEB #M0525S) were added to each reaction, and incubated at 30°C, 600 rpm, for 1 hour in a thermo mixer. Samples were cleaned up using 200 µL StageTips similarly to what was previously described [172]. Tips consisted of 2 1.06 mM diameter punches of Empore C18 (3M #2215) fitted into 200 µL pipette tips. Columns were conditioned with 40 µl methanol and 30 µl 0.1% TFA. Sample was then acidified with 3 µl 70% FA and vortexed. Sample was applied to the column and centrifuged at 2800 rpm for 5 minutes. Columns were washed with 50 µl 0.1% TFA and centrifuged for 3 minutes at 900g. Centrifugation was repeated with 20 µl 0.1% TFA. StageTips were then transferred to 2 ml protein low bind tubes (Eppendorf #0030108450) and eluted with 30 µl of 80% ACN 0.1% TFA, centrifuged at 900g for 1 minute. Elution was repeated with 20 µL of 80% ACN 0.1% TFA, centrifuged at 900g for 2 minutes. Samples were dried down in a vacuum centrifuge at room temperature. Samples were then reconstituted in in 2 µl 30 % ACN 0.1 % FA and 10 µL of 3:8 70 % formic acid: 0.1% TFA as described earlier.

CHK2 and threonine phosphosite library digestion

24 µg of cleaved Thr phosphosite library, and 5 µg CHK2 kinase were digested using S-Trap™ mini spin columns (Protifi) per manufacturer's instructions. Following digestions samples were dried in a vacuum centrifuge at room temperature and frozen at -80°C.

Threonine phosphosite library fractionation

Dried down samples were fractionated using Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (ThermoFisher #84868) following manufacturer's instructions. Samples were then reconstituted in in 2 µl 30 % ACN 0.1 % FA and 10 µL of 3:8 70 % formic acid: 0.1% TFA as described earlier.

CHK2 phospho-enrichment

Following S-trap digestions, both full length CHK2 and CHK2 KD peptides were enriched using Thermo Scientific High-Select™ Fe-NTA Phosphopeptide Enrichment Kit (#A32992) following manufacturer's instructions. Samples were dried in a vacuum centrifuge at room temperature and reconstituted in in 2 µl 30 % ACN 0.1 % FA and 10 µL of 3:8 70 % formic acid: 0.1% TFA as described earlier.

Data acquisition and analysis

LC-MS/MS was performed using an ACQUITY UPLC M-Class (Waters) and Thermo Q Exactive Plus mass spectrometer. The analytical column employed was a 65-cm-long, 75-µm-internal-diameter PicoFrit column (New Objective) packed in-house to a

length of 50 cm with 1.9 µm ReproSil-Pur 120 Å C18-AQ (Dr. Maisch) using methanol as the packing solvent. Peptide separation was achieved using mixtures of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) with a 90-min gradient 0/1, 2/7, 60/24, 65/48, 70/80, 75/80, 80/1, 90/1; (min/%B, linear ramping between steps). Gradient was performed with a flowrate of 250 nL/min. A single blank injection (5 µl 2% B) was performed between samples to eliminate peptide carryover on the analytical column. 100 fmol of trypsin-digested BSA or 100 ng trypsin-digested wildtype K-12 MG1655 *E. coli* proteins were run periodically between samples as quality control standards. The mass spectrometer was operated with the following parameters: (MS1) 70,000 resolution, 3e6 AGC target, 300–1,700 m/z scan range; (data dependent-MS2) 17,500 resolution, 1e6 AGC target, top 10 mode, 1.6 m/z isolation window, 27 normalized collision energy, 90s dynamic exclusion, unassigned and +1 charge exclusion. Data was searched using Maxquant version 1.6.10.43 with Deamidation (NQ), Oxidation (M), and Phospho (STY) as variable modifications and Carbamidomethyl (C) or Dithiomethane (C) as a fixed modification with up to 3 missed cleavages, 5 AA minimum length, and 1% FDR against targeted libraries. *E. coli* proteome searches were run against a modified Uniprot *E. coli* database (taken on February 9, 2021) containing MS-READ reporter proteins and OTS components (supplemental data file 2). Phosphosite libraries were run against a modified FASTA file of the phosphosite library containing a partially digested MBP tag and 6xHis tag (supplemental data files 3&4). Only phosphosite peptides with a MaxQuant score greater than 40 were considered. MS-READ search results were quantified using Skyline version 20.1.0.31. Proteome search results were analyzed with Perseus version 1.6.2.2 using a two-tailed student's t-test. Ion spectra were taken from searches using

Mascot Daemon 2.5.1. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [173] partner repository.

pThr metabolite mass spectrometry

Experiments were carried out in collaboration with Dr. Xiaojian Shi and Dr. Hongying Shen at Yale University. LC-MS-based phospho-amino acid profiling experiment was performed using a Q Exactive plus benchtop orbitrap mass spectrometer equipped with an Ion Max source and a HESI II probe, coupled to a Vanquish UHPLC (Thermo Fisher Scientific, San Jose, CA, USA). The LC-MS experiment was following previously reported methods for mammalian cells [174], with 1 OD equivalent of *E. Coli* cells at mid-log phase pelleted for polar metabolite extraction. The data was analyzed using Xcalibur and GraphPad Prism.

High-throughput Interactome of Phosphoproteins

FACS

20 ml of $C321^{QKO}$ cells containing either no OTS (for Thr libraries) or $pThrOTS^{Zeus}$ were grown to an OD_{600} of 0.4 and electroporated using the method stated above with either Thr library or pThr library cloned in the split mCherry vector (see supplemental plasmid files). The cells were then resuspended in 1.2 mLs of LB and incubated for 1 h at 37°C and 230 RPM in a 15 ml culture tube. Recovered cells were directly inoculated in 20 ml of LB with 100 ng/µl ampicillin, 50 ng/µl kanamycin and grown overnight at 37°C and 230 RPM. Cells were plated at 10^{-4} and 10^{-5} serial dilutions on LB plates with antibiotic and grown at 37°C overnight. Experiments that proceeded forward required at least 20 colony-forming

units per 10^{-5} dilution. The following morning, cultures were diluted to an OD_{600} of 0.15 in 5 ml of LB containing either 100 ng/µl ampicillin or 100 ng/µl ampicillin and 50 ng/µl kanamycin grown at 37°C and 230 RPM. The cells were grown until an OD_{600} between 0.6–0.8 and set on ice. Protein expression was induced using 1 mM IPTG, 0.2% arabinose, and 100 ng/µL anhydrotetracycline, cells were then grown at 20°C and 230 RPM for 20 hours. 30 µl of cells were resuspended in 3 ml ice cold PBS in a 5 ml polystyrene tube (Falcon) prior to analysis.

Using a BD FACSAria III, cells were assessed for mCherry-based fluorescence using a 561-nm laser (Texas Red). Gating was set so that fluorescent enriched cells reached ~200 events/s with a total collection of ~50,000 cells/s. The gating was adjusted for each phospho-binding domain following successive enrichments to ensure that the most fluorescent population was isolated each round. Cells were sorted directly into 3 ml LB without antibiotic, recovered at 37°C and 230 RPM for 3 hours, and then added into 17 mL LB containing antibiotic and grown overnight at 37°C and 230 RPM (~16 hours). The following day sorted cell populations were then used as the starting culture for the second round of protein expression. The procedure for protein expression, preparation for FACS, and cell sorting was repeated, using the same sorting and gating parameters as the first round of sorting. Cells were then recovered, regrown, induced and prepared for FACS as above. Cellular mCherry fluorescence was then observed using the FACSAria III. Enriched cell populations were visualized using FlowJo 10.6.1. Following the final round of enrichment overnight cultures were pelleted, transferred to a 2 mL Eppendorf tube, and frozen at -80°C until it was possible to Miniprep the samples using a Qiagen mini-prep kit

(#27104). Hi-P experiments were repeated in triplicate during concurrent runs from the same starting population to minimize transformation bias.

Next generation sequencing analysis

Sample pellets generated from Hi-P were thawed on ice and Miniprepped (Qiagen #27104) as stated above. 100-200 ng of plasmid DNA was amplified in four reactions using a standardized set of OLS amplification primers (forward: TCTGGGTCGACTGGTGGTACC, reverse: CGTACCATGTAGCTTAATCAGCTGTTAAAGCTT) with Q5 polymerase (NEB #M0491). PCR cycle followed a 98°C denaturation for 30s, 15 cycles of 98°C for 15s, 58°C for 30s, and 72°C for 30s, with a final extension of 72°C for 2 minutes. Samples were then cleaned and concentrated with a PCR purification kit (Qiagen #28104) to one 50 µL sample. Samples were then further concentrated with a BluePippin (Sage Science) selecting for samples between 100-200 bp. Samples were cleaned and concentrated with a PCR purification kit (Qiagen #28104) to a 40 µL sample. 35 µL of sample at a concentration of 30-50 ng/µL was then sent to Massachusetts General Hospital Center for Computational & Integrative Biology DNA Core for CRISPR amplicon next-generation sequencing.

Sequencing reads were analyzed as reported previously [80]. Briefly, data was first filtered for quality using Trimmomatic, which applied a sliding window filter of width 2 bp and a Phred score cutoff of 30. If the average quality score over two consecutive bases fell below 30, the read was trimmed to remove the remaining bases. Quality trimmed read pairs were then merged using BBMerge with the stringency set to “strict”

(https://sourceforge.net/projects/bbmap/). Using custom scripts, the merged reads were then sorted and assigned to the various input libraries based on barcodes added during the PCR amplification step, libraries are provided in supplemental data files 5&6. The variable sequence region for each amplicon was then extracted, and for each input library the abundance of every unique sequence was calculated. Sequences with greater than 10x enrichment were taken forward for analysis.

Kinase screening

In vitro kinase reactions

All kinase reactions were carried out in 40 mM HEPES pH 7.5, 20 mM $MgCl_2$, 0.1 mg/mL BSA, 50 μM DTT with 200 μM ATP. CHK2 kinase variants tested against CDC25C substrate were carried out by mixing 300 ng of purified kinase with 3 μg CHKtide in a 15 μL reaction. Initial time point (0 hours) was taken before addition of ATP to reaction. Following addition of ATP, reactions were incubated at 30°C for 1 hour shaking at 600 RPM. Reactions were quenched by mixing equal amounts of 2x Laemmli buffer and reaction before freezing at -80°C. 6 μL of quenched reaction was then run on a 15 lane, 12% acrylamide, 100 μM Phos-Tag SDS-PAGE immunoblotted with α-6xHis as described earlier. CHK2 kinase profiling was carried out by mixing 250 ng purified kinase with 3.3 μg cleaved Ser library and 1.7 μg cleaved Thr library in a 40 μL reaction incubated at 30°C for 4 hours shaking at 600 RPM. After 4 hours reactions were frozen at -80°C before being digested using S-trap mini columns as described earlier.

References

1. Adio, S., et al., *Dynamics of ribosomes and release factors during translation termination in E. coli.* Elife, 2018. **7**.

2. Pertea, M., et al., *CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise.* Genome Biol, 2018. **19**(1): p. 208.

3. Vlastaridis, P., et al., *Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes.* Gigascience, 2017. **6**(2): p. 1-11.

4. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

5. Hirakawa, H., et al., *Progress Overview of Bacterial Two-Component Regulatory Systems as Potential Targets for Antimicrobial Chemotherapy.* Antibiotics (Basel), 2020. **9**(10).

6. Makwana, M.V., R. Muimo, and R.F. Jackson, *Advances in development of new tools for the study of phosphohistidine.* Lab Invest, 2018. **98**(3): p. 291-303.

7. Tiwari, S., et al., *Two-Component Signal Transduction Systems of Pathogenic Bacteria As Targets for Antimicrobial Therapy: An Overview.* Front Microbiol, 2017. **8**: p. 1878.

8. Huse, M. and J. Kuriyan, *The conformational plasticity of protein kinases.* Cell, 2002. **109**(3): p. 275-82.

9. Taylor, S.S. and E. Radzio-Andzelm, *Three protein kinase structures define a common motif.* Structure, 1994. **2**(5): p. 345-55.

10. Brunet, A., et al., *14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport.* J Cell Biol, 2002. **156**(5): p. 817-28.

11. DeYoung, M.P., et al., *Hypoxia regulates TSC1/2-mTOR signaling and tumor suppression through REDD1-mediated 14-3-3 shuttling.* Genes Dev, 2008. **22**(2): p. 239-51.

12. Margolis, S.S., et al., *Role for the PP2A/B56delta phosphatase in regulating 14-3-3 release from Cdc25 to control mitosis.* Cell, 2006. **127**(4): p. 759-73.

13. Capecchi, M.R. and G.N. Gussin, *Suppression in vitro: Identification of a Serine-sRNA as a "Nonsense" Suppressor.* Science, 1965. **149**(3682): p. 417-22.

14. Engelhardt, D.L., et al., *In vitro studies on the mechanism of suppression of a nonsense mutation.* Proc Natl Acad Sci U S A, 1965. **54**(6): p. 1791-7.

15. Steege, D.A., *A nucleotide change in the anticodon of an Escherichia coli serine transfer RNA results in supD-amber suppression.* Nucleic Acids Res, 1983. **11**(11): p. 3823-32.

16. Thorbjarnardottir, S., et al., *Leucine tRNA family of Escherichia coli: nucleotide sequence of the supP(Am) suppressor gene.* J Bacteriol, 1985. **161**(1): p. 219-22.

17. Pirman, N.L., et al., *A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation.* Nat Commun, 2015. **6**: p. 8130.

18. Agris, P.F., *Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis.* Biochimie, 1991. **73**(11): p. 1345-9.

19. Jenner, L.B., et al., *Structural aspects of messenger RNA reading frame maintenance by the ribosome.* Nat Struct Mol Biol, 2010. **17**(5): p. 555-60.

20. Cohen, G.N. and D.B. Cowie, *[Total replacement of methionine by selenomethionine in the proteins of Escherichia coli].* C R Hebd Seances Acad Sci, 1957. **244**(5): p. 680-3.

21. Budisa, N., *Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire.* Angew Chem Int Ed Engl, 2004. **43**(47): p. 6426-63.

22. Schlesinger, S. and M.J. Schlesinger, *The effect of amino acid analogues on alkaline phosphatase formation in Escherichia coli K-12. 3. Substitution of 2-methylhistidine for histidine.* J Biol Chem, 1969. **244**(14): p. 3803-9.

23. Sykes, B.D., H.I. Weingarten, and M.J. Schlesinger, *Fluorotyrosine alkaline phosphatase from Escherichia coli: preparation, properties, and fluorine-19 nuclear magnetic resonance spectrum.* Proc Natl Acad Sci U S A, 1974. **71**(2): p. 469-73.

24. Bacher, J.M., J.J. Bull, and A.D. Ellington, *Evolution of phage with chemically ambiguous proteomes.* BMC Evol Biol, 2003. **3**: p. 24.

25. Chapeville, F., et al., *On the role of soluble ribonucleic acid in coding for amino acids.* Proc Natl Acad Sci U S A, 1962. **48**: p. 1086-92.

26. Murakami, H., H. Saito, and H. Suga, *A versatile tRNA aminoacylation catalyst based on RNA.* Chem Biol, 2003. **10**(7): p. 655-62.

27. Goto, Y., T. Katoh, and H. Suga, *Flexizymes for genetic code reprogramming.* Nat Protoc, 2011. **6**(6): p. 779-90.

28. Niwa, N., et al., *A flexizyme that selectively charges amino acids activated by a water-friendly leaving group.* Bioorg Med Chem Lett, 2009. **19**(14): p. 3892-4.

29. Lee, J., et al., *Ribosome-mediated polymerization of long chain carbon and cyclic amino acids into peptides in vitro.* Nat Commun, 2020. **11**(1): p. 4304.

30. Lee, J., et al., *Expanding the limits of the second genetic code with ribozymes.* Nat Commun, 2019. **10**(1): p. 5097.

31. Takemoto, C., et al., *Unconventional decoding of the AUA codon as methionine by mitochondrial tRNAMet with the anticodon f5CAU as revealed with a mitochondrial in vitro translation system.* Nucleic Acids Res, 2009. **37**(5): p. 1616-27.

32. Labunskyy, V.M., D.L. Hatfield, and V.N. Gladyshev, *Selenoproteins: molecular pathways and physiological roles.* Physiol Rev, 2014. **94**(3): p. 739-77.

33. Arner, E.S., *Selenoproteins-What unique properties can arise with selenocysteine in place of cysteine?* Exp Cell Res, 2010. **316**(8): p. 1296-303.

34. Krzycki, J.A., *Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases.* Curr Opin Chem Biol, 2004. **8**(5): p. 484-91.

35. Paul, L., D.J. Ferguson, Jr., and J.A. Krzycki, *The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of Methanosarcina barkeri contain in-frame and read-through amber codons.* J Bacteriol, 2000. **182**(9): p. 2520-9.

36. Zhang, Y., et al., *Pyrrolysine and selenocysteine use dissimilar decoding strategies.* J Biol Chem, 2005. **280**(21): p. 20740-51.

37. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

38. Liu, D.R. and P.G. Schultz, *Progress toward the evolution of an organism with an expanded genetic code.* Proc Natl Acad Sci U S A, 1999. **96**(9): p. 4780-5.

39. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

40. Kraemer, K.H. and M.M. Seidman, *Use of supF, an Escherichia coli tyrosine suppressor tRNA gene, as a mutagenic target in shuttle-vector plasmids.* Mutat Res, 1989. **220**(2-3): p. 61-72.

41. Steer, B.A. and P. Schimmel, *Major anticodon-binding region missing from an archaebacterial tRNA synthetase.* J Biol Chem, 1999. **274**(50): p. 35601-6.

42. Park, H.S., et al., *Expanding the genetic code of Escherichia coli with phosphoserine.* Science, 2011. **333**(6046): p. 1151-4.

43. Kuznetsova, E., et al., *Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family.* J Biol Chem, 2006. **281**(47): p. 36149-61.

44. Beranek, V., et al., *Genetically Encoded Protein Phosphorylation in Mammalian Cells.* Cell Chem Biol, 2018. **25**(9): p. 1067-1074 e5.

45. Rogerson, D.T., et al., *Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog.* Nat Chem Biol, 2015. **11**(7): p. 496-503.

46. Lee, S., et al., *A facile strategy for selective incorporation of phosphoserine into histones.* Angew Chem Int Ed Engl, 2013. **52**(22): p. 5771-5.

47. Fan, C., K. Ip, and D. Soll, *Expanding the genetic code of Escherichia coli with phosphotyrosine.* FEBS Lett, 2016. **590**(17): p. 3040-7.

48. Hoppmann, C., et al., *Site-specific incorporation of phosphotyrosine using an expanded genetic code.* Nat Chem Biol, 2017. **13**(8): p. 842-844.

49. Luo, X., et al., *Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria.* Nat Chem Biol, 2017. **13**(8): p. 845-849.

50. Zhang, M.S., et al., *Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing.* Nat Methods, 2017. **14**(7): p. 729-736.

51. Ames, B.N., et al., *Illicit transport: the oligopeptide permease.* Proc Natl Acad Sci U S A, 1973. **70**(2): p. 456-8.

52. Fickel, T.E. and C. Gilvarg, *Transport of impermeant substances in E. coli by way of oligopeptide permease.* Nat New Biol, 1973. **241**(110): p. 161-3.

53. Xie, J., L. Supekova, and P.G. Schultz, *A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli.* ACS Chem Biol, 2007. **2**(7): p. 474-8.

54. Burke, T.R., Jr., et al., *Nonhydrolyzable Phosphotyrosyl Mimetics for the Preparation Of Phosphatase-Resistant SH2 Domain Inhibitors.* Biochemistry, 1994. **33**(21): p. 6490-6494.

55. Tong, L., et al., *Carboxymethyl-phenylalanine as a replacement for phosphotyrosine in SH2 domain binding.* J Biol Chem, 1998. **273**(32): p. 20238-42.

56. Isaacs, F.J., et al., *Precise manipulation of chromosomes in vivo enables genome-wide codon replacement.* Science, 2011. **333**(6040): p. 348-53.

57. Ostrov, N., et al., *Design, synthesis, and testing toward a 57-codon genome.* Science, 2016. **353**(6301): p. 819-22.

58. Fredens, J., et al., *Total synthesis of Escherichia coli with a recoded genome.* Nature, 2019. **569**(7757): p. 514-518.

59. Robertson, W.E., et al., *Sense codon reassignment enables viral resistance and encoded polymer synthesis.* Science, 2021. **372**(6546): p. 1057-1062.

60. *Building better yeast.* Nat Commun, 2018. **9**(1): p. 1939.

61. Liu, W., et al., *Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods.* Nat Commun, 2018. **9**(1): p. 1936.

62. Szymanski, E. and J. Calvert, *Designing with living systems in the synthetic yeast project.* Nat Commun, 2018. **9**(1): p. 2950.

63. Wu, Y., et al., *In vitro DNA SCRaMbLE.* Nat Commun, 2018. **9**(1): p. 1935.

64. Fan, C. and T.A. Bobik, *The PduX enzyme of Salmonella enterica is an L-threonine kinase used for coenzyme B12 synthesis.* J Biol Chem, 2008. **283**(17): p. 11322-9.

65. Drewke, C., et al., *4-O-phosphoryl-L-threonine, a substrate of the pdxC(serC) gene product involved in vitamin B6 biosynthesis.* FEBS Lett, 1996. **390**(2): p. 179-82.

66. Mohler, K., et al., *MS-READ: Quantitative measurement of amino acid incorporation.* Biochim Biophys Acta Gen Subj, 2017. **1861**(11 Pt B): p. 3081-3088.

67. Ryu, Y. and P.G. Schultz, *Efficient incorporation of unnatural amino acids into proteins in Escherichia coli.* Nat Methods, 2006. **3**(4): p. 263-5.

68. Sherman, J.M., M.J. Rogers, and D. Soll, *Competition of aminoacyl-tRNA synthetases for tRNA ensures the accuracy of aminoacylation.* Nucleic Acids Res, 1992. **20**(11): p. 2847-52.

69. Swanson, R., et al., *Accuracy of in vivo aminoacylation requires proper balance of tRNA and aminoacyl-tRNA synthetase.* Science, 1988. **242**(4885): p. 1548-51.

70. Chatterjee, A., et al., *A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli.* Biochemistry, 2013. **52**(10): p. 1828-37.

71. Agrawal, A., B.K. Mohanty, and S.R. Kushner, *Processing of the seven valine tRNAs in Escherichia coli involves novel features of RNase P.* Nucleic Acids Res, 2014. **42**(17): p. 11166-79.

72. Lenhard, B., et al., *tRNA recognition and evolution of determinants in seryl-tRNA synthesis.* Nucleic Acids Res, 1999. **27**(3): p. 721-9.

73. Cashel, M. and B. Kalbacher, *The control of ribonucleic acid synthesis in Escherichia coli. V. Characterization of a nucleotide associated with the stringent response.* J Biol Chem, 1970. **245**(9): p. 2309-18.

74. Hogg, T., et al., *Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response [corrected].* Cell, 2004. **117**(1): p. 57-68.

75. Sinha, A.K. and K.S. Winther, *The RelA hydrolase domain acts as a molecular switch for (p)ppGpp synthesis.* Commun Biol, 2021. **4**(1): p. 434.

76. Kuncha, S.K., et al., *A discriminator code-based DTD surveillance ensures faithful glycine delivery for protein biosynthesis in bacteria.* Elife, 2018. **7**.

77. Hou, Y.M. and P. Schimmel, *A simple structural feature is a major determinant of the identity of a transfer RNA.* Nature, 1988. **333**(6169): p. 140-5.

78. Canagarajah, B.J., et al., *Activation mechanism of the MAP kinase ERK2 by dual phosphorylation.* Cell, 1997. **90**(5): p. 859-69.

79. Carneiro, S., et al., *Influence of the RelA Activity on E. coli Metabolism by Metabolite Profiling of Glucose-Limited Chemostat Cultures.* Metabolites, 2012. **2**(4): p. 717-32.

80. Barber, K.W., et al., *Encoding human serine phosphopeptides in bacteria for proteome-wide identification of phosphorylation-dependent interactions.* Nat Biotechnol, 2018. **36**(7): p. 638-644.

81. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D261-70.

82. Loveland, A.B., et al., *Ribosome*RelA structures reveal the mechanism of stringent response activation.* Elife, 2016. **5**.

83. Narmandakh, A., et al., *Phosphorylation of phenol by phenylphosphate synthase: role of histidine phosphate in catalysis.* J Bacteriol, 2006. **188**(22): p. 7815-22.

84. Schmeling, S., et al., *Phenylphosphate synthase: a new phosphotransferase catalyzing the first step in anaerobic phenol metabolism in Thauera aromatica.* J Bacteriol, 2004. **186**(23): p. 8044-57.

85. Yaffe, M.B., et al., *The structural basis for 14-3-3:phosphopeptide binding specificity.* Cell, 1997. **91**(7): p. 961-71.

86. Durocher, D., et al., *The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms.* Mol Cell, 2000. **6**(5): p. 1169-82.

87. Lee, S.J., et al., *Rad53 phosphorylation site clusters are important for Rad53 regulation and signaling.* Mol Cell Biol, 2003. **23**(17): p. 6300-14.

88. Schwartz, M.F., et al., *Rad9 phosphorylation sites couple Rad53 to the Saccharomyces cerevisiae DNA damage checkpoint.* Mol Cell, 2002. **9**(5): p. 1055-65.

89. Sun, Z., et al., *Rad53 FHA domain associated with phosphorylated Rad9 in the DNA damage checkpoint.* Science, 1998. **281**(5374): p. 272-4.

90. Chen, H., et al., *SHP2 is a multifunctional therapeutic target in drug resistant metastatic breast cancer.* Oncogene, 2020. **39**(49): p. 7166-7180.

91. Eck, M.J., et al., *Spatial constraints on the recognition of phosphoproteins by the tandem SH2 domains of the phosphatase SH-PTP2.* Nature, 1996. **379**(6562): p. 277-80.

92. Hof, P., et al., *Crystal structure of the tyrosine phosphatase SHP-2.* Cell, 1998. **92**(4): p. 441-50.

93. LaRochelle, J.R., et al., *Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition.* Nat Commun, 2018. **9**(1): p. 4508.

94. Pluskey, S., et al., *Potent stimulation of SH-PTP2 phosphatase activity by simultaneous occupancy of both SH2 domains.* J Biol Chem, 1995. **270**(7): p. 2897-900.

95. Anselmi, M., et al., *Structural Determinants of Phosphopeptide Binding to the N-Terminal Src Homology 2 Domain of the SHP2 Phosphatase.* J Chem Inf Model, 2020. **60**(6): p. 3157-3171.

96. Beebe, K.D., et al., *Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library.* Biochemistry, 2000. **39**(43): p. 13251-60.

97. Sweeney, M.C., et al., *Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains.* Biochemistry, 2005. **44**(45): p. 14932-47.

98. Barber, K.W., et al., *Kinase Substrate Profiling Using a Proteome-wide Serine-Oriented Human Peptide Library.* Biochemistry, 2018. **57**(31): p. 4717-4725.

99. Bell, D.W., et al., *Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome.* Science, 1999. **286**(5449): p. 2528-31.

100. Matsuoka, S., M. Huang, and S.J. Elledge, *Linkage of ATM to cell cycle regulation by the Chk2 protein kinase.* Science, 1998. **282**(5395): p. 1893-7.

101. Cai, Z., N.H. Chehab, and N.P. Pavletich, *Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase.* Mol Cell, 2009. **35**(6): p. 818-29.

102. Li, J., et al., *Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2.* Mol Cell, 2002. **9**(5): p. 1045-54.

103. Xu, X., L.M. Tsvetkov, and D.F. Stern, *Chk2 activation and phosphorylation-dependent oligomerization.* Mol Cell Biol, 2002. **22**(12): p. 4419-32.

104. O'Neill, T., et al., *Determination of substrate motifs for human Chk1 and hCds1/Chk2 by the oriented peptide library approach.* J Biol Chem, 2002. **277**(18): p. 16102-15.

105. Chen, Y., et al., *CHK2-FOXK axis promotes transcriptional control of autophagy programs.* Sci Adv, 2020. **6**(1): p. eaax5819.

106. Dreier, R.F., et al., *Global Ion Suppression Limits the Potential of Mass Spectrometry Based Phosphoproteomics.* J Proteome Res, 2019. **18**(1): p. 493-507.

107. Ahn, J.Y., et al., *Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation.* Cancer Res, 2000. **60**(21): p. 5934-6.

108. Lee, C.H. and J.H. Chung, *The hCds1 (Chk2)-FHA domain is essential for a chain of phosphorylation events on hCds1 that is induced by ionizing radiation.* J Biol Chem, 2001. **276**(32): p. 30537-41.

109. Schwarz, J.K., C.M. Lovly, and H. Piwnica-Worms, *Regulation of the Chk2 protein kinase by oligomerization-mediated cis- and trans-phosphorylation.* Mol Cancer Res, 2003. **1**(8): p. 598-609.

110. Gabant, G., et al., *Autophosphorylated residues involved in the regulation of human chk2 in vitro.* J Mol Biol, 2008. **380**(3): p. 489-503.

111. Jin, X., O.J. Park, and S.H. Hong, *Incorporation of non-standard amino acids into proteins: challenges, recent achievements, and emerging applications.* Appl Microbiol Biotechnol, 2019. **103**(7): p. 2947-2958.

112. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annu Rev Biochem, 2010. **79**: p. 413-44.

113. Bianco, A., et al., *Expanding the genetic code of Drosophila melanogaster.* Nat Chem Biol, 2012. **8**(9): p. 748-50.

114. Brown, W. and A. Deiters, *Light-activation of Cre recombinase in zebrafish embryos through genetic code expansion.* Methods Enzymol, 2019. **624**: p. 265-281.

115. Ernst, R.J., et al., *Genetic code expansion in the mouse brain.* Nat Chem Biol, 2016. **12**(10): p. 776-778.

116. Hancock, S.M., et al., *Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair.* J Am Chem Soc, 2010. **132**(42): p. 14819-24.

117. Liu, J., et al., *Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling.* J Am Chem Soc, 2017. **139**(27): p. 9100-9103.

118. Scheidler, C.M., M. Vrabel, and S. Schneider, *Genetic Code Expansion, Protein Expression, and Protein Functionalization in Bacillus subtilis.* ACS Synth Biol, 2020. **9**(3): p. 486-493.

119. Shao, S., M. Koh, and P.G. Schultz, *Expanding the genetic code of the human hematopoietic system.* Proc Natl Acad Sci U S A, 2020. **117**(16): p. 8845-8849.

120. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* Science, 2013. **342**(6156): p. 357-60.

121. Rovner, A.J., et al., *Recoded organisms engineered to depend on synthetic amino acids.* Nature, 2015. **518**(7537): p. 89-93.

122. Schiapparelli, P., et al., *Phosphorylated WNK kinase networks in recoded bacteria recapitulate physiological function.* Cell Rep, 2021. **36**(3): p. 109416.

123. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.

124. Humphrey, S.J., D.E. James, and M. Mann, *Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation.* Trends Endocrinol Metab, 2015. **26**(12): p. 676-687.

125. Xie, J., L. Supekova, and P.G. Schultz, *A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli.* ACS chemical biology, 2007. **2**(7): p. 474-8.

126. Zhu, P., et al., *A Highly Versatile Expression System for the Production of Multiply Phosphorylated Proteins.* ACS Chem Biol, 2019. **14**(7): p. 1564-1572.

127. Chopra, I. and M. Roberts, *Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance.* Microbiol Mol Biol Rev, 2001. **65**(2): p. 232-60 ; second page, table of contents.

128. Kotra, L.P., J. Haddad, and S. Mobashery, *Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance.* Antimicrob Agents Chemother, 2000. **44**(12): p. 3249-56.

129. Amiram, M., et al., *Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids.* Nat Biotechnol, 2015. **33**(12): p. 1272-1279.

130. Wannier, T.M., et al., *Adaptive evolution of genomically recoded Escherichia coli.* Proc Natl Acad Sci U S A, 2018. **115**(12): p. 3090-3095.

131. Mandell, D.J., et al., *Biocontainment of genetically modified organisms by synthetic protein design.* Nature, 2015. **518**(7537): p. 55-60.

132. Mitrophanov, A.Y. and E.A. Groisman, *Signal integration in bacterial two-component regulatory systems.* Genes Dev, 2008. **22**(19): p. 2601-11.

133. Besant, P.G. and P.V. Attwood, *Detection of a mammalian histone H4 kinase that has yeast histidine kinase-like enzymic activity.* Int J Biochem Cell Biol, 2000. **32**(2): p. 243-53.

134. Fuhs, S.R., et al., *Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation.* Cell, 2015. **162**(1): p. 198-210.

135. Kowluru, A., *Identification and characterization of a novel protein histidine kinase in the islet beta cell: evidence for its regulation by mastoparan, an activator of G-proteins and insulin secretion.* Biochem Pharmacol, 2002. **63**(12): p. 2091-100.

136. Tan, E., et al., *Histone H4 histidine kinase displays the expression pattern of a liver oncodevelopmental marker.* Carcinogenesis, 2004. **25**(11): p. 2083-8.

137. Bryson, D.I., et al., *Continuous directed evolution of aminoacyl-tRNA synthetases.* Nat Chem Biol, 2017. **13**(12): p. 1253-1260.

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

139. Umehara, T., et al., *N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo.* FEBS Lett, 2012. **586**(6): p. 729-33.

140. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.* Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.

141. Allfrey, V.G. and A.E. Mirsky, *Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis.* Science, 1964. **144**(3618): p. 559.

142. Haynes, S.R., et al., *The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.* Nucleic Acids Res, 1992. **20**(10): p. 2603.

143. Groff, D., et al., *A genetically encoded epsilon-N-methyl lysine in mammalian cells.* Chembiochem, 2010. **11**(8): p. 1066-8.

144. Nguyen, D.P., et al., *Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones.* J Am Chem Soc, 2009. **131**(40): p. 14194-5.

145. Jakubowski, H. and A.R. Fersht, *Alternative pathways for editing non-cognate amino acids by aminoacyl-tRNA synthetases.* Nucleic Acids Res, 1981. **9**(13): p. 3105-17.

146. Martinis, S.A. and M.T. Boniecki, *The balance between pre- and post-transfer editing in tRNA synthetases.* FEBS letters., 2010. **48**(2): p. 455.

147. Roy, H., et al., *Post-transfer editing in vitro and in vivo by the beta subunit of phenylalanyl-tRNA synthetase.* The EMBO journal, 2004. **23**(23): p. 4639-48.

148. Oki, K., et al., *Transplantation of a tyrosine editing domain into a tyrosyl-tRNA synthetase variant enhances its specificity for a tyrosine analog.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13298-303.

149. Ahel, I., et al., *Trans-editing of mischarged tRNAs.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15422-7.

150. Liu, Z., et al., *Homologous trans-editing factors with broad tRNA specificity prevent mistranslation caused by serine/threonine misactivation.* Proc Natl Acad Sci U S A, 2015. **112**(19): p. 6027-32.

151. Liu, C.C., et al., *Efficient expression of tyrosine-sulfated proteins in E. coli using an expanded genetic code.* Nat Protoc, 2009. **4**(12): p. 1784-9.

152. Liu, C.C. and P.G. Schultz, *Recombinant expression of selectively sulfated proteins in Escherichia coli.* Nat Biotechnol, 2006. **24**(11): p. 1436-40.

153. Cooley, R.B., et al., *Structural basis of improved second-generation 3-nitro-tyrosine tRNA synthetases.* Biochemistry, 2014. **53**(12): p. 1916-24.

154. Neumann, H., et al., *Genetically encoding protein oxidative damage.* J Am Chem Soc, 2008. **130**(12): p. 4028-33.

155. Ghesquiere, B., et al., *In vitro and in vivo protein-bound tyrosine nitration characterized by diagonal chromatography.* Mol Cell Proteomics, 2009. **8**(12): p. 2642-52.

156. Zhao, Y., et al., *Selective Affinity Enrichment of Nitrotyrosine-Containing Peptides for Quantitative Analysis in Complex Samples.* J Proteome Res, 2017. **16**(8): p. 2983-2992.

157. Bian, Y., et al., *An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome.* J Proteomics, 2014. **96**: p. 253-62.

158. Deng, H., et al., *Identification of a major co-receptor for primary isolates of HIV-1.* Nature, 1996. **381**(6584): p. 661-6.

159. Dragic, T., et al., *HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5.* Nature, 1996. **381**(6584): p. 667-73.

160. Schnur, E., et al., *The conformation and orientation of a 27-residue CCR5 peptide in a ternary complex with HIV-1 gp120 and a CD4-mimic peptide.* J Mol Biol, 2011. **410**(5): p. 778-97.

161. Pollok-Kopp, B., et al., *Analysis of ligand-stimulated CC chemokine receptor 5 (CCR5) phosphorylation in intact cells using phosphosite-specific antibodies.* J Biol Chem, 2003. **278**(4): p. 2190-8.

162. Venuti, A., et al., *The Abrogation of Phosphorylation Plays a Relevant Role in the CCR5 Signalosome Formation with Natural Antibodies to CCR5.* Viruses, 2017. **10**(1).

163. Skretas, G. and G. Georgiou, *Engineering G protein-coupled receptor expression in bacteria.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 14747-8.

164. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. **460**(7257): p. 894-898.

165. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.

166. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol Syst Biol, 2006. **2**: p. 2006 0008.

167. George, S., et al., *Generation of phospho-ubiquitin variants by orthogonal translation reveals codon skipping.* FEBS Lett, 2016. **590**(10): p. 1530-42.

168. Sprouffske, K. and A. Wagner, *Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves.* BMC Bioinformatics, 2016. **17**: p. 172.

169. R Core Development Team, *A language and environment for statistical computing*. 2019, R Foundation for Statistical Computing.

170. RStudio Team, *RStudio: Integrated Development for R*. 2020, RStudio.

171. Galili, T., et al., *heatmaply: an R package for creating interactive cluster heatmaps for online publishing.* Bioinformatics, 2018. **34**(9): p. 1600-1602.

172. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. **2**(8): p. 1896-906.

173. Perez-Riverol, Y., et al., *The PRIDE database and related tools and resources in 2019: improving support for quantification data.* Nucleic Acids Res, 2019. **47**(D1): p. D442-D450.

174. Shen, H., et al., *The Human Knockout Gene CLYBL Connects Itaconate to Vitamin B12.* Cell, 2017. **171**(4): p. 771-782 e11.