Characterizing Protein Kinase Substrate Specificity Using the Proteomic Peptide Library (ProPeL) Approach

Joshua M. Lubner,[1,4] Jeremy L. Balsbaugh,[2] George M. Church,[3] Michael F. Chou,[3,4] and Daniel Schwartz[1,4]

[1]University of Connecticut, Department of Physiology and Neurobiology, Storrs, Connecticut
[2]University of Connecticut, Proteomics & Metabolomics Facility, Center for Open Research Resources & Equipment, Storrs, Connecticut
[3]Harvard Medical School, Department of Genetics, Boston, Massachusetts and Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts
[4]Corresponding authors: *joshua.lubner@uconn.edu, mchou@genetics.med.harvard.edu, daniel.schwartz@uconn.edu*

Characterizing protein kinase substrate specificity motifs represents a powerful step in elucidating kinase-signaling cascades. The protocol described here uses a bacterial system to evaluate kinase specificity motifs *in vivo*, without the need for radioactive ATP. The human kinase of interest is cloned into a heterologous bacterial expression vector and allowed to phosphorylate *E. coli* proteins *in vivo*, consistent with its endogenous substrate preferences. The cells are lysed, and the bacterial proteins are digested into peptides and phosphoenriched using bulk TiO_2. The pooled phosphopeptides are identified by tandem mass spectrometry, and bioinformatically analyzed using the pLogo visualization tool. The ProPeL approach allows for detailed characterization of wildtype kinase specificity motifs, identification of specificity drift due to kinase mutations, and evaluation of kinase residue structure-function relationships. © 2018 by John Wiley & Sons, Inc.

Keywords: kinase specificity • mass spectrometry • protein kinase • phosphorylation motif • proteomic peptide library • ProPeL • pLogo

How to cite this article:
Lubner, J. M., Balsbaugh, J. L., Church, G. M., Chou, M. F., & Schwartz, D. (2018). Characterizing protein kinase substrate specificity using the proteomic peptide library (ProPeL) approach. *Current Protocols in Chemical Biology*,*10*, e38. doi: 10.1002/cpch.38

INTRODUCTION

Protein kinases are enzymes that catalyze the covalent addition of phosphate to specific amino acids within protein substrates as post-translational modifications. Such alter can often alter the biological function of the target protein, so understanding the relationship between kinases and their substrates can provide important biological insights. However, traditional co-immunoprecipitation methods do not work to identify these transient interactions, so alternative approaches including the one described here have been developed.

Kinases discriminate their substrates, in part, by recognizing short linear patterns of amino acids or "motifs" that surround the phosphoacceptor residue (Pinna & Ruzzene, 1996; Ubersax and Ferrell Jr, 2007), and the identification of these motifs has proven to be a

Current Protocols in Chemical Biology e38, Volume 10
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpch.38
© 2018 John Wiley & Sons, Inc.

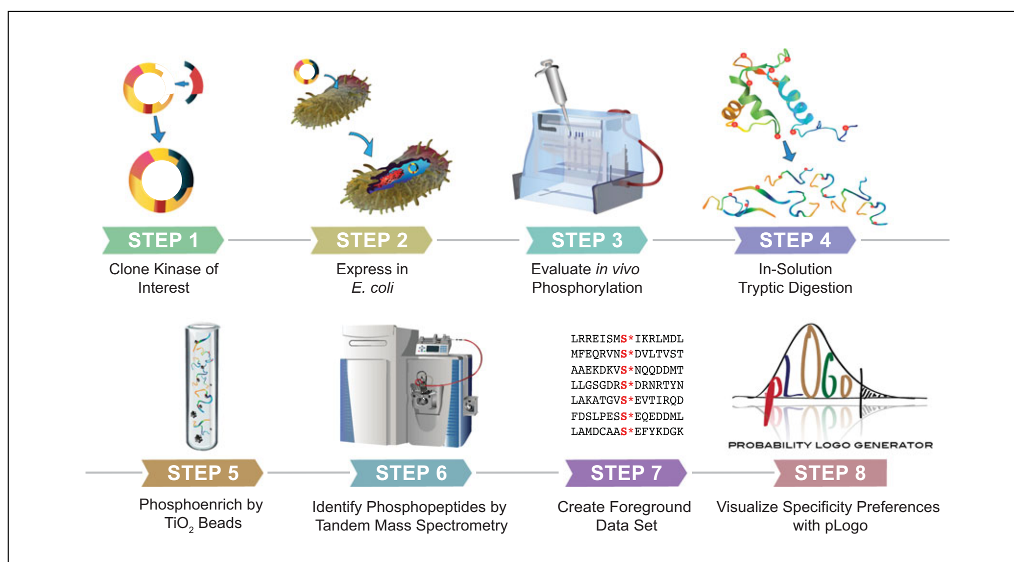

Figure 1 Schematic overview of the experimental ProPeL workflow. A kinase of interest is cloned and expressed in *E. coli*. Resulting bacterial phosphorylation is evaluated by SDS-PAGE with Pro-Q Diamond and Coomassie staining. Lysate is digested, phosphoenriched and identified by tandem mass spectrometry. Data sets are visualized with pLogo (O'Shea et al., 2013).

powerful tool for substrate hypothesis generation (Miller et al., 2008; Obenauer, Cantley, & Yaffe, 2003). This unit presents a non-radioactive, bacterial approach for querying protein kinase substrate specificity *in vivo* termed ProPeL (for Proteomic Peptide Library, Chou et al., 2012). In this method, a human kinase is expressed in *E. coli* cells (which have a very low background level of endogenous phosphorylation). The bacterial proteome functions as a substrate library for the human kinase to phosphorylate *in vivo*, consistent with its distinct specificity. In this way, the *E. coli* acts as a living mini reaction vessel, facilitating thousands of simultaneous *in vivo* phosphorylation events and generating thousands of kinase-specific phosphorylation sites that are isolated and identified by liquid chromatography tandem mass spectrometry (LC-MS/MS). Using our laboratory's suite of computational tools, we can extract and visualize kinase specificity motifs, and make high-confidence predictions of downstream targets. ProPeL can also be used to evaluate the influence of disease-associated mutations on kinase substrate specificity (Lubner et al., 2017).

The Basic Protocol describes the overall ProPeL workflow, which is represented in Figure 1. The major steps include expression of the kinase of interest and *in vivo* phosphorylation of bacterial proteins, tryptic digestion, phosphopeptide enrichment, phosphopeptide identification by LC-MS/MS, and computational analysis. Prior to carrying out ProPeL, the kinase of interest must be cloned into an appropriate bacterial expression vector (Strategic Planning), and *in vivo* activity may need to be optimized (Troubleshooting). In the event that the kinase cannot be expressed in an active form in *E. coli*, it is possible to perform an *in vitro* version of ProPeL using recombinant (or endogenously purified) kinase (Alternate Protocol).

STRATEGIC PLANNING

At the start of a new ProPeL project, it is important to design the correct kinase insert, and choose an appropriate bacterial expression vector. A successful ProPeL experiment necessarily requires the expression of a soluble, constitutively active protein kinase. This may require expressing a truncation that omits inhibitory sequences (such as the C-terminal PKC inhibitory tail) or mimics caspase cleavage (as is required for full activation of MST3). In other instances, it is as simple as expressing the naked catalytic subunit, as

is the case for PKA. Plasmid design is also an important variable that can enhance kinase solubility (see Critical Parameters). Although the background serine/threonine/tyrosine phosphorylation in *E. coli* is only around 0.9% (Hansen et al., 2013; Macek et al., 2008; Soares, Spät, Krug, & Macek, 2013; Potel et al., 2018), it is still advisable to create a kinase-dead mutant as a negative control. This is most easily achieved by mutating the catalytic aspartate to an asparagine (within the HRD motif) or the invariant lysine to an alanine (VAIK motif), as these residues are essential for catalysis (Gibbs & Zoller, 1991; Hanks, Quinn, & Hunter, 1988). The best indicator of a successful ProPeL result is the demonstration of strong *in vivo* phosphorylation of bacterial proteins. Therefore, it is critical to optimize expression conditions prior to mass spectrometry sample preparation (see Critical Parameters and Troubleshooting).

While we have found success with several phylogenetically distant kinases using the standard *in vivo* ProPeL approach, there are nevertheless instances of kinases that are challenging for the system. Kinases with highly complex activation requirements (such as involvement in multiple activation cascades or requirements for large protein scaffolding structures), cytoplasmic conditions that are unsustainable for *E. coli* growth, or kinases that are toxic to *E. coli* through their activity would be poor targets for ProPeL. Similarly, a kinase that is part of a cascade, such as the MAPK kinases, will be unsuitable for *in vivo* ProPeL. While it is possible to recapitulate activating cascades by co-expressing kinases in *E. coli* (Khokhlatchev et al., 1997), the greater the complexity of the cascade, the more difficult it is to determine which individual phosphorylation sites should be attributed to each individual kinase. However, those kinases are suitable candidates for the *in vitro* ProPeL approach, provided they can be successfully purified in the active state (see Alternate Protocol).

E. COLI KINASE EXPRESSION AND MASS SPECTROMETRY SAMPLE PREPARATION

BASIC PROTOCOL

This approach expresses an active protein kinase in *E. coli*, facilitating the *in vivo* phosphorylation of bacterial proteins. Following expression, the cells are harvested, lysed, and evaluated for kinase activity. Bacterial proteins are tryptically digested, and enriched for phosphopeptides using TiO_2. The resulting samples are ready for sequence identification by LC-MS/MS.

Materials

Appropriate *E. coli* cell strain (see Critical Parameters)
Bacterial expression vector with appropriate kinase insert
LB plates and liquid broth, and appropriate antibiotic (see recipes)
Isopropyl β-D-1-thiogalactopyranoside (IPTG; Promega, cat. no. V3951 or V3955)
Lysis buffer (see recipe)
BCA Assay Kit (Thermo Fisher Scientific, cat. no. PI23225)
SDS-PAGE Gel, Laemmli loading buffer, and running buffers (see recipes)
PeppermintStick ladder (Fisher Scientific, cat. no. P27167)
All Blue Protein Standards (Bio-Rad, cat. no. 1610373), *optional*
Fix solution (see recipe)
Water (double distilled or Ultrapure)
Pro-Q Diamond Phosphoprotein Gel Stain (Fisher Scientific, cat. no. P33300)
Pro-Q Diamond Phosphoprotein Destain Solution (see recipe)
GelCode Blue (Thermo Fisher Scientific, cat. no. PI24592)
Chloroform, HPLC grade (Fisher Scientific, cat. no. C607SK-4)
Methanol, LC/MS grade (Fisher Scientific, cat. no. A456-500)
Dithiothreitol (DTT; Fisher Scientific, cat. no. BP172)
Iodoacetamide, mass spectrometry grade (Sigma Aldrich, cat. no. I1149)

Tris·Cl, pH 8.2 (Sigma Aldrich, cat. no. T6066)
Calcium chloride (Fisher Scientific, cat. no. AC349615000)
Trypsin, sequencing grade modified (Promega, cat. no. V5111 or V5117)
Trifluoracetic acid, LC/MS grade (Fisher Scientific, cat. no. A116-50)
Desalting wash solution A (see recipe)
Desalting wash solution B (see recipe)
Desalting elution solution (see recipe)
Acetonitrile, LC/MS grade (Fisher Scientific, cat. no. A955-1)
Liquid nitrogen
TiO_2 binding solution (see recipe)
TiO_2 elution solution B (see recipe)
Desalting wash solution C (see recipe)
Titansphere TiO_2 5 μm beads (GL Sciences, cat. no. 1400B500)

Sterile pipette tips
Shaking bacterial culture incubator
500-ml Erlenmeyer flasks
Refrigerated centrifuge
Probe sonicator
5-ml disposable sterile syringe with Luer Lock (Fisher Scientific cat. no. 14-829-45)
0.22-μm sterile syringe filters (Fisher Scientific cat. no. SLGL0250S)
Room temperature 15-ml conical tube shaker
Electrophoresis chamber
Gel imager
Vortex mixer
15-ml conical tubes
tC18 SEP-Pak cartridges (Waters, cat. no. WAT054925)
1.5-ml microcentrifuge tubes
SpeedVac
Empore SPE Disks C18 (Sigma-Aldrich, cat. no. 66883-U)
Kel-F hub (KF), point style 3, gauge 16 needle (Hamilton Company, cat. no. 90516)
Plunger assembly N, RN, LT, LTN for model 1702 (Hamilton Company, cat. no. 1122-01)

NOTE: The authors believe that overexpression of a kinase-dead mutant is a better negative control than either an un-induced culture, or induction with an empty vector. The negative control kinase will more closely mimic the cellular stress of heterologous protein overexpression, and does not pose the contamination risk that may be encountered as a result of leaky expression of an active kinase.

Kinase expression and in vivo phosphorylation of bacterial proteins

This protocol assumes that the appropriate kinase-coding sequence has been cloned into a bacterial expression vector, and transformed into an appropriate *E. coli* cell strain. The following steps are for standard protein expression using an IPTG-inducible vector. Optimal protein expression conditions need to be determined empirically, and the expression steps should be adjusted accordingly.

1. Using a sterile pipette tip, streak a fresh LB agar plate (+ appropriate antibiotic) from a bacterial glycerol stock. Incubate plate upside down overnight at 37°C.
2. Inoculate a well-isolated colony in 5 ml LB medium (+ appropriate antibiotic) and grow overnight at 37°C and 250 rpm overnight.

3. Inoculate 100 ml LB medium (+ appropriate antibiotic) with 2 ml of overnight culture in a 500-ml Erlenmeyer flask.

4. Grow bacteria to an OD_{600} of 0.4-0.6 (mid-log) and induce with 0.5 mM IPTG. Incubate for 3 to 24 hr at 37°C and 250 rpm.

   This is an optimization point for kinase expression and phosphorylation. See Critical Parameters and Troubleshooting.

5. *Optional*: Collect a 1-ml aliquot in a separate tube and store up to 6 months at 4°C for later analysis by Pro-Q Diamond.

6. Pellet the cells by centrifuging for 15 min at 6000 × *g*, 4°C.

   Pellet may be stored up to 1 year at −80°C, but it is best to proceed promptly.

Lysis and evaluating the success of in vivo phosphorylation

When preparing the sample for LC-MS/MS, use all LC/MS-grade solvents and Eppendorf brand microcentrifuge tubes for sample preparation. If evaluating an aliquot during the kinase expression optimization phase, ACS-grade reagents are acceptable. ACS-grade solvents may be used for all SDS-PAGE steps.

7. Prepare lysis buffer, add at 5 ml/g of wet pellet, and resuspend by pipet mixing.

8. Lyse cells by sonication, using 15-sec pulses on 15% power, until solution is no longer opaque.

   To prevent cells from over-heating, keep the tubes on ice (between and during sonications) with at least 1 min rest between pulses. The solution will be colored, but should be clear.

9. Centrifuge the solution for 30 min at 20,000 × *g*, 4°C. Save the clarified supernatant and discard the pelleted cellular debris. If necessary, repeat centrifugation to further clarify.

10. Filter the lysate with a disposable syringe and 0.22-µm filter attachment to further remove cellular debris.

11. Quantify samples by BCA assay (or by NanoDrop using the protein A280 measurement if evaluating an aliquot during the optimization phase for kinase expression).

    Note that a NanoDrop A280 measurement is less accurate for quantifying protein concentration, and tends to overestimate protein concentration in crude cell lysate by a factor of 3 to 4× relative to a BCA assay. Accordingly, additional sample should be loaded when using NanoDrop readings. Using a NanoDrop is acceptable for optimization and gel evaluation, but when preparing a sample for mass spectrometry a BCA assay is critical for accurate protein quantification.

12. For each sample, separate 25 µg (or 75 µg if using NanoDrop A280 measurement) by SDS-PAGE, with 2 µl PeppermintStick Phosphoprotein ladder (and 5 µl All Blue Protein Standards, *optional*).

Stain the gel

Analyze with Pro-Q Diamond stain as described below, according to manufacturer's instructions. All incubations should be carried out on a rocker at room temperature.

13. Immerse the gel in 100 ml fix solution and incubate for 30 min. Discard the fix solution and add100 ml fresh fix solution. Incubate for at least 30 min.

    This is a pause point, as gel can be left in fix solution overnight.

14. Discard the fix solution and wash with 100 ml ultrapure water. Incubate for 10 min, discard, and repeat twice for a total of three water washes.

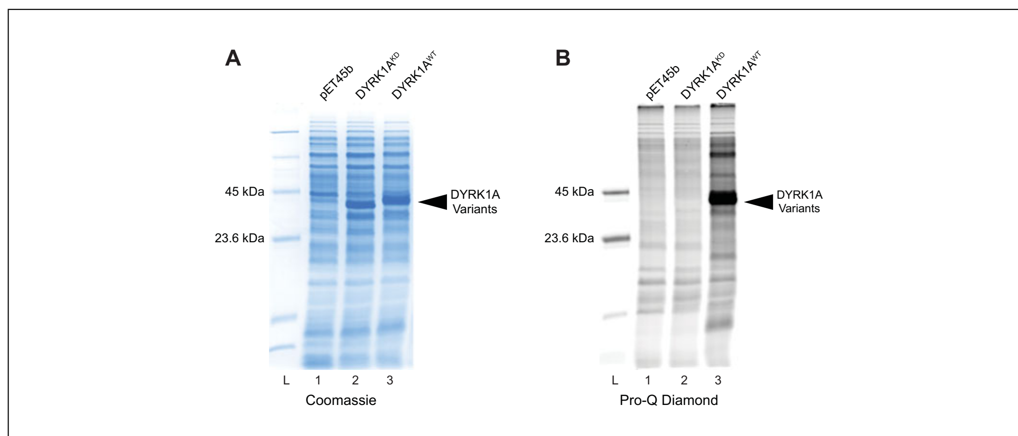

Figure 2 For the attached replacement figure, please scale the image to the maximum size to aid in clarity. Also, note that this is a new image, and is therefore no longer a modification of the previous publication Lubner et al. 2016. Example of Desired Kinase Expression and Activity. These gels illustrate the desired level of kinase expression and *in vivo* activity with the DYRK1A^{WT} kinase. (**A**) SDS-PAGE with Coomassie staining with robust expression of both (2) DYRK1A^{KD} and (3) DYRK1A^{WT} at the expected molecular weight of 43 kDa, relative to (1) empty vector pET45b. (**B**) Pro-Q Diamond staining reveals robust autophosphorylation and efficient phosphorylation of bacterial substrates over a wide molecular weight range for (3) DYRK1A^{WT} relative to (1) empty vector pET45b or (2) DYRK1A^{KD} negative controls.

All subsequent incubations must be done in the dark, as Pro-Q Diamond is light sensitive.

15. Add 60 ml Pro-Q Diamond stain and incubate 90 min.
16. Discard the stain and add 90 ml Pro-Q Diamond destain solution, incubating for 30 min. Discard destain and repeat twice more for a total of three destain washes. Rinse with 100 ml ultrapure water for 5 min, discard, and repeat the water wash once.
17. Visualize on an appropriate imager (Typhoon, ChemiDoc etc.) using the following wavelengths: Ex: 555 nm, Em: 580 nm. Adjust the signal such that only the two phosphoprotein bands (23 kDa and 40 kDa) on the PeppermintStick ladder are clearly visible (18 kDa band may be faintly visible).

To control for loading differences (which can change the level of background signal), it is important to perform a total protein stain. We use GelCode Blue according to manufacturer's instructions, but other stains (such as Coomassie staining) are acceptable.

18. Add 20 ml GelCode Blue and incubate on a rocker at room temperature for at least an hour. Preferably, leave the gel overnight in GelCode Blue for clearest signal.
19. Destain using ultrapure water. For best results, change water several times until background signal has been completely removed. Gel may be left in water overnight.
20. Image using the Coomassie setting (and white light conversion screen).

If autophosphorylation of the kinase of interest is evident and/or there is a marked increase in phosphorylation of proteins throughout the gel relative to the negative control (see Troubleshooting and Figure 2), proceed with the remaining steps for LC-MS/MS sample preparation. Otherwise, optimize kinase expression/activity (see Critical Parameters and Troubleshooting).

Protein reduction, alkylation and tryptic digestion

The following protocol is for the preparation of sample from 10 mg whole cell protein lysate (as quantified by BCA assay). Sample volumes can be scaled as needed. We have observed excellent results starting with as low as 1 mg crude lysate, but this is contingent

on the activity of the target kinase within the *E. coli*. It should be noted that the presence of lipids may cause an overestimation of protein concentration. Therefore, we recommend delipidating an excess of sample (e.g., 20 mg) by methanol/chloroform extraction, and then quantifying protein concentration by BCA.

Methanol/chloroform extract excess crude protein lysate as follows:

21. Add 4× sample volume methanol and briefly vortex 1 to 2 sec. Add 1× sample volume choloroform and briefly vortex 1 to 2 sec. Add 3× sample volume water and briefly vortex 1 to 2 sec.

22. Centrifuge the sample for 10 min at 14,000 × *g*, 4°C.

    There will be a lower chloroform (containing lipids) layer, a middle protein disc, and a top, aqueous layer. If separation is insufficient, increase centrifugation time, but DO NOT increase centrifuge speed. Remove top aqueous layer without disturbing protein disc.

23. Add 4× sample volume methanol and briefly vortex 1 to 2 sec. Centrifuge for 10 min at 14,000 × *g*, 4°C. Remove as much methanol as possible and air dry (approximately 5 to 10 min).

    Over-drying the pellet will make it very difficult to resuspend. For best results, proceed immediately to the next step.

24. Resuspend the pellet in sufficient lysis buffer, quantify by BCA, and adjust with additional lysis buffer to obtain 1 ml sample at a final protein concentration of 10 mg/ml. Pipet mixing can be aided by gently vortexing the sample and using heat for short durations (do not exceed 50°C). Sample can also be allowed to resolubilize overnight at 4°C for best results.

    Heating the sample to 50°C (step 24) or 56°C (Step 25) is not thought to be sufficient to lead to loss of phosphoester (i.e., phosphoserine, phosphothreonine, phosphotyrosine) phosphate groups. While other phosphoamino acids are significantly less stable (such as phosphoramidates like phosphohistidine), phosphoesters are very stable under a variety of conditions.

25. Add DTT from a fresh 0.5 M stock (10.1 µl of 0.5 M DTT) to a final concentration of 5 mM DTT. Incubate for 25 min at 56°C.

    During this step, the protein is unfolded by heat and DTT denaturation. Avoid temperatures above 60°C, which can cause urea-based carbamylation of lysines and protein N-termini.

26. Allow the mixture to cool to room temperature, and add iodoacetamide from a fresh 0.5 M stock to a final concentration of 14 mM iodoacetamide (29.1 µl of 0.5 M iodoacetamide). Incubate for 30 min at room temperature in the dark.

    During this step, the exposed free cysteine residues are alkylated to prevent disulfide bond formation. Iodoacetamide is light sensitive. Store the 0.5 M iodoacetamide stock solution in the dark, and carry out the alkylation and quenching steps in the dark.

27. Quench alkylation by adding an additional 5 mM DTT from a 0.5 M stock (10.5 µl of 0.5 M DTT). Incubate 15 min at room temperature in the dark.

28. Transfer the sample into a 15-ml conical tube, and dilute the sample 1:5 by adding 4.2 ml of 25 mM Tris·Cl (pH 8.2), to reduce urea concentration from 8 M to 1.6 M.

29. Add $CaCl_2$ from a 0.1 M stock to a final concentration of 1 mM (53 µl of 0.1 M $CaCl_2$).

30. Add 100 µg trypsin (for an enzyme:substrate ratio of 1:100), and incubate for 16 hr at 37°C.

31. Allow the digest to cool to room temperature, and stop digestion by acidification with 25 µl trifluoracetic acid to 0.4% (v/v). Verify that pH <2.5, otherwise, add additional trifluoracetic acid. Centrifuge for 10 min at 2500 × *g*, room temperature, and transfer the supernatant into a fresh 15-ml conical tube.

    Peptides may be stored for a short period of time at 4°C, but it is best to proceed to desalting immediately.

Peptide desalting

Peptides must be desalted prior to TiO_2 phosphoenrichment. When using tC18 SEP-Pak cartridges, each cartridge capacity is 5%, so up to 10 mg can be purified on a single 200 mg cartridge. Other capacities can be used for different sample volumes, with the wash steps below scaled proportionally by bed volume. Sample processing can be accomplished by gravity flow, or the use of a vacuum manifold.

32. Prepare the following solutions: Desalting wash solution A, desalting wash solution B, and desalting elution solution.
33. Condition the cartridge using 6 ml of acetonitrile followed by 2 ml of desalting elution solution.
34. Equilibrate with 6 ml of desalting wash solution A.
35. Load the sample from step 31.
36. Wash/desalt with 6 ml of desalting wash solution A.
37. Wash with 500 µl desalting wash solution B.
38. Elute into a clean 1.5-ml tube with 1 ml desalting elution solution.
39. Measure protein concentration by NanoDrop A280 (or other protein estimation method) and determine total quantity of protein in sample. Transfer 4 mg of peptides into a clean 1.5-ml tube.
40. Snap-freeze with liquid nitrogen and SpeedVac until the sample is a white (sometimes yellowish) pellet.

    Avoid over-drying pellet, as this will make it very difficult to resuspend. Samples can be stored up to 1 year at −20°C if TiO_2 phosphoenrichment cannot be performed immediately, but for best results proceed immediately to the next step.

Combined TiO_2 phosphoenrichment and terminal desalting

Phosphopeptides will be isolated by bulk TiO_2 enrichment. Combining the TiO_2 and terminal desalting steps helps to prevent loss and/or carryover of beads. Commercial desalting tips may be purchased (Thermo Fisher Scientific cat. no. PI89873); however, we routinely prepare StageTips in-house using five C18 plugs packed gently into a 200-µl tip, according to Rappsilber, Mann, & Ishihama, 2007. StageTips should be prepared during the TiO_2 incubation.

41. Prepare the following solutions: TiO_2 binding solution, TiO_2 elution solution B, desalting wash solution C, and desalting elution solution.
42. Add 1 ml TiO_2 binding solution to dried down peptides and allow to resolubilize for >30 min. Transfer into 15-ml conical tube and add an additional 2.6 ml TiO_2 binding solution (final volume 3.6 ml TiO_2 binding solution).

    Due to the lactic acid, the TiO_2 binding solution can result in gloves becoming sticky. Gloves exposed to TiO_2 binding solution may stick to tubes.

43. Condition enough TiO_2 beads to allow for a 1:1 ratio of beads to peptides based on post-SEP-Pak NanoDrop reading (e.g., 4 mg beads for every 4 mg peptides).

44. Condition the beads by washing in 50× bead volume of TiO_2 binding solution (50 µl TiO_2 binding solution per 1 mg beads) and centrifuge for 30 sec at 600 × *g*, room temperature. Remove the supernatant and repeat TiO_2 binding solution conditioning step.

45. Resuspend the beads in the appropriate volume of TiO_2 binding solution to obtain a bead concentration of 10 µg/µl (e.g., resuspend 4 mg of beads in 400 µl of binding solution).

46. Add 400 µl TiO_2 beads to 4 mg of resolubilized peptides from step 42. Final peptide concentration is 1 mg/ml.

    The TiO_2 beads settle rapidly. To avoid adding an incorrect volume of beads, resuspend the TiO_2 bead slurry by pipet mixing immediately before dispensing to each peptide sample.

47. To bind phosphopeptides, incubate in a conical tube shaker at maximum speed for 1 hr at room temperature.

48. Pellet the beads by centrifugation for 30 sec at 600 × *g*, room temperature, and remove the supernatant. Be careful not to remove beads. Binding buffer supernatant and all subsequent washes may be saved as "non-phosphopeptides" for analysis, if desired.

49. Wash the beads with 1 ml TiO_2 binding solution.

50. Pellet the beads and remove the supernatant. Repeat the wash two more times with fresh TiO_2 binding solution for a total of three washes.

51. Resuspend the beads in 200 µl TiO_2 binding solution.

    Perform all subsequent steps by centrifuging at 2000 × g, room temperature, for the minimum time required to pass the liquid through the StageTip (~30 sec/50 µl). In order to make sure StageTips do not over-dry, it may be necessary to centrifuge some StageTips longer than others.

52. Condition StageTips with 50 µl methanol.

53. Pre-clear StageTips by washing with 50 µl desalting elution solution, and then equilibrate by washing twice with 50 µl TiO_2 binding solution.

54. Load the TiO_2 bead slurry onto the top of the StageTip and centrifuge, saving flow-through with any residual "non-phosphopeptides," if desired.

55. Wash the combined StageTips/TiO_2 column twice with 150 µl TiO_2 binding solution. This can be added to the flow-through, if desired.

56. Equilibrate the combined StageTips/TiO_2 column with 100 µl desalting wash solution C.

57. Elute phosphopeptides with 150 µl TiO_2 elution solution. Repeat once.

    At this stage, phosphopeptides will be retained on the C18 disc.

58. Wash with 100 µl desalting wash buffer C.

59. Elute phosphopeptides off the disc with 100 µl desalting elution solution.

60. Dry the eluent in a SpeedVac, and store up to 1 year at −20°C until analysis by LC-MS/MS (see Support Protocol 1 for mass spectrometry acquisition and analysis procedures).

Phosphopeptide list creation and filtering

Prior to motif visualization, mass spectrometry data must be filtered for appropriate sequences and converted into a set of phosphorylation-centered 15mers. This is accomplished by removing undesirable peptide matches within the Phospho(STY) modification-specific file, and bioinformatically determining the *in vivo* 15-residue context centered around the phosphoacceptor.

61. Within the Phospho(STY) modification-specific file, remove any peptides that match to a reverse database, or are identified as a contaminant.

62. If additionally searching against the human proteome, remove any autophosphorylation sites on the kinase of interest.

    While these sites are often of biological interest, autophosphorylation sites (particularly those that occur in cis) frequently do not conform to consensus motifs and therefore should be removed prior to motif analysis.

63. Filter phosphopeptides to only retain high-confidence sites with a localization probability of ≥0.9. This is most easily accomplished by a text find-and-replace where "(1)" and "(0.9" are replaced with "*".

    Note that this localization probability value is based on the software detailed in Support Protocol 1. If a different search algorithm is used (or parameters are changed), this value may need to be determined empirically.

    Each site must be converted to a modification-centered 15mer; therefore, any phosphorylation site that is too close to a protein terminus to be extended to a centered 15mer is discarded. This is the same procedure as in motif-x analyses (Chou & Schwartz, 2011; Schwartz & Gygi, 2005). We have created a Web tool, PeptidExtender, which accomplishes this task, and is freely accessible at https://schwartzlab.uconn.edu/pepextend.

64. Paste the peptide sequences into the input box in the top left corner of the PeptidExtender Web page. This will cause "modification markers" to populate. Select "*" as the modification marker, select "right of modified residue" for position, and enter a target sequence width of 15. Select the "*E. coli*" proteome as the extension database. Click "extend peptides!" to create a list of unique phosphorylation-centered 15mers.

    PeptidExtender automatically filters out non-selected potential markers (e.g., non-amino acid characters such as numbers and brackets), deletes redundant sequences, and removes any sequence that fails to generate a full 15mer. The output from PeptidExtender is correctly formatted to be directly pasted into pLogo, although additional negative-control subtraction is necessary prior to motif visualization.

    It is critical to remove endogenous E. coli phosphorylation sites from the foreground data set. We have curated a master negative control list (Lubner et al., 2017), which was generated by pooling phosphopeptides previously identified in negative control experiments (Chou et al., 2012), previously identified endogenous E. coli phosphorylation sites (Macek et al., 2008; Soares et al., 2013), sites identified in Hansen et al., 2013, and Potel et al., 2018, and phosphorylation sites identified in empty vector and kinase dead negative control experiments. This list is available as Supplementary Table S1. Any additional sites identified in endogenous E. coli and negative control experiments should be added. Phosphorylation sites on this master negative control list must be removed from each target kinase data set to generate a final list of kinase-specific phosphorylation sites. We typically make use of the webtool Venny (Oliveros, J.C., 2007), which can be freely accessed at https://bioinfogp.cnb.csic.es/tools/venny/. After control subtraction, phosphorylation site lists from all runs can be merged within each kinase variant

(redundant sites will be automatically removed by our software). These final data sets can be used for motif visualization by pLogo (see Support Protocol 2) or motif analysis by motif-x (Chou & Schwartz, 2011; Schwartz & Gygi, 2005).

IN VITRO KINASE REACTION

ALTERNATE PROTOCOL

As mentioned above, there are several challenges associated with heterologous kinase expression in *E. coli*. However, many of the positive features of ProPeL are retained in an *in vitro* version of the protocol. Conceptually, the issues of kinase expression and/or activation are solved through purification of recombinant or endogenous kinase from an alternative source, and activation with the required co-factors. By adding recombinant kinase to *E. coli* lysate in a traditional *in vitro* kinase reaction, the target kinase is still able to phosphorylate bacterial proteins, which can be isolated as in the standard ProPeL workflow. This format still allows the high signal to noise ratio and direct link afforded by the low endogenous *E. coli* serine/threonine/tyrosine phosphorylation, and reactions still take place with full-length protein substrates. Identification by LC-MS/MS is the same, allowing high throughput identification, and the preservation of intra-motif correlations due to the proteomic context of each phosphopeptide.

There are a few different options for producing recombinant kinase. The target kinase can still be expressed in *E. coli*, and then affinity purified. This would be most useful in the case of re-constituting a kinase cascade, where the individual kinases can be purified separately by using distinct affinity tags. Alternatively, other host systems such as yeast, insect, or human cells may be employed to produce recombinant kinase. Finally, the use of cell-free protein systems allow for production of post-translationally modified kinases without the need for cell culture (Oza et al., 2015). Optimal kinase reaction conditions need to be determined empirically.

Additional Materials *(also see Basic Protocol)*

Untransformed or empty vector *E. coli* bacterial stock
Recombinant Kinase (purchase or purify in-house)
Appropriate Kinase Reaction Buffer (determine empirically)
Adenosine 5′-triphosphate (ATP) (Fisher Scientific cat. no. BP413-25)

Preparation of bacterial substrate library

For the *in vitro* version of ProPeL, bacterial proteins still function as a substrate library. Using either an untransformed bacterial stock or an empty vector stock, cells are grown and harvested similarly to the Basic Protocol, but lysed in kinase reaction buffer.

1. Using a sterile pipette tip, streak a fresh LB agar plate (+ appropriate antibiotic) from a bacterial glycerol stock. Incubate the plate upside down overnight at 37°C.

2. Inoculate a well-isolated colony in 5 ml LB medium (+ appropriate antibiotic) and grow overnight at 37°C and 250 rpm.

3. Inoculate 100 ml LB medium (+ appropriate antibiotic) with 2 ml of overnight culture in a 500-ml Erlenmeyer flask.

4. *Optional*: Grow bacteria to an OD_{600} of 0.4–0.6 (mid-log) and induce with 0.5 mM IPTG.

5. Incubate for 3 to 24 hr at 37°C and 250 rpm.

6. Pellet the cells by centrifugation for 15 min at 6000 × *g*, 4°C.

7. Pellet may be stored up to 1 year at −80°C, but it is best to proceed promptly.

When preparing sample for LC-MS/MS, use all LC/MS grade solvents and Eppendorf brand microcentrifuge tubes for sample preparation. If evaluating an aliquot during the kinase expression optimization phase, ACS grade reagents are acceptable. ACS grade solvents may be used for all SDS-PAGE steps.

8. Prepare kinase reaction buffer, add 5 ml of kinase reaction buffer per gram of wet pellet, and resuspend by pipet mixing.
9. Lyse the cells by sonication, using 15-sec pulses on 15% power, until solution is no longer opaque.

   To prevent cells over-heating, tubes should be kept on ice (in between and during sonication) with at least 1 min rest between pulses. Solution will be discolored, but should be clear.

10. Clarify the solution by centrifugation for 30 min at 20,000 × *g*, 4°C. Save the clarified supernatant and discard the pelleted cellular debris. If necessary, repeat centrifugation to further clarify.
11. Filter the lysate with a disposable syringe and 0.22-μm filter attachment to further remove cellular debris.
12. Quantify the samples by BCA assay.

In vitro kinase reaction and evaluation of kinase activity

In this protocol, recombinant kinase is incubated with bacterial lysate in an *in vitro* kinase reaction. Optimal reaction conditions should be determined empirically using small-scale reactions before proceeding to the full 10 mg reaction. After this section, sample preparation for LC-MS/MS is resumed as in the Basic Protocol.

13. Transfer 10 mg bacterial lysate into a 1.5-ml tube.
14. Add appropriate volume recombinant kinase and 1× kinase reaction buffer, and incubate 3 hr at 30°C.

    This is an optimization point. Appropriate buffer conditions, kinase:substrate ratio, incubation duration and temperature may need to be determined empirically.

15. For each sample, separate 25 μg (or 75 μg if using NanoDrop A280 measurement) by SDS-PAGE, with 2 μl PeppermintStick Phosphoprotein ladder (and 5 μl All Blue Protein Standards, *optional*).
16. Analyze with Pro-Q Diamond stain, according to manufacturer's instructions. All incubations should be carried out on a rocker at room temperature.
17. Immerse the gel in 100 ml fix solution and incubate for 30 min. Discard the fix solution and add fresh 100 ml fix solution. Incubate for at least 30 min.

    This is a pause point, as gel can be left in fix solution overnight.

18. Discard the fix solution and wash with 100 ml ultrapure water. Incubate for 10 min, discard, and repeat twice for a total of three water washes.

    All subsequent incubations must be done in the dark, as Pro-Q Diamond is light sensitive.

19. Add 60 ml Pro-Q Diamond stain and incubate for 90 min.
20. Discard stain and add 90 ml Pro-Q Diamond destain solution, incubating for 30 min. Discard destain and repeat twice more for a total of three destain washes. Rinse with 100 ml ultrapure water for 5 min, discard, and repeat water wash once.

21. Visualize on an appropriate imager (Typhoon, ChemiDoc etc.) using the following wavelengths: Ex: 555 nm, Em: 580 nm. Adjust the signal such that only the two phosphoprotein bands (23.k kDa and 40 kDa) on the PeppermintStick ladder are clearly visible (18 kDa band may be faintly visible).

    To control for loading differences (which can change the level of background signal), it is important to perform a total protein stain. We use GelCode Blue according to manufacturer's instructions, but other stains (such as Coomassie staining) are acceptable.

22. Add 20 ml GelCode Blue and incubate on a rocker at room temperature for at least an hour. Preferably, leave gel in GelCode Blue overnight for clearest signal.
23. Destain using ultrapure water. For best results, change water several times until background signal has been completely removed. Gel may be left in water overnight.
24. Image using the Coomassie setting (and white light conversion screen).

    If phosphorylation is acceptable, proceed with the remaining steps for LC-MS/MS sample preparation in the Basic Protocol, beginning with Protein Reduction, Alkylation and Tryptic Digestion. Otherwise, optimize kinase activity before proceeding.

PHOSPHOPEPTIDE IDENTIFICATION BY LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY

SUPPORT PROTOCOL 1

The following protocol represents our current instrumentation, and is provided as a reference.

Materials

Q Exactive HF Orbitrap mass spectrometer (Thermo Scientific)
Ultimate 3000 RSLC (Thermo Scientific)
250-mm nanoEase M/Z peptide BEH C18 column
130 Å, 1.7 µm particle size, 75 µm i.d (Waters, cat. no. 186008795)
10-µm silica PicoTip emitter (New Objective, cat. no. FS360-20-10-N-20-C12)

Liquid chromatography and mass spectrometry analysis

Dried and enriched phosphopeptides are resuspended in 40 µl of 0.1% formic acid in water. Peptide identification is achieved using electrospray ionization (ESI) and nanoLC-MS/MS on a Q Exactive HF Orbitrap mass spectrometer (QE-HF) coupled to an Ultimate 3000 RSLC operated in nanoflow mode. A 250-mm nanoEase M/Z peptide BEH C18 column is fitted to a 10-µm silica PicoTip emitter to permit ESI directly into the QE-HF inlet. For all samples, 1 µl is loaded and subject to a 150 min, 300 nl/min linear reversed-phase gradient (Solvent A: 0.1% formic acid in water, Solvent B: 0.1% formic acid in acetonitrile) as follows: initial 4% solvent B hold for 10 min, increase to 30% solvent B over 90 min, increase to 90% solvent B over 20 min, 90% solvent B hold for 10 min, then decrease to 4% solvent B over 2 min, followed by a re-equilibration period for 18 min. Column temperature is set to 50°C for the entire gradient. The QE-HF is operated in positive ion mode with a spray voltage of 1.5 kV. The capillary temperature is set to 250°C and all source gas flows are turned off. A Top 15 data-dependent (dd) MS/MS method is used that implements the following parameters for full MS scans: 1 microscan, 60,000 resolution at 200 *m*/z, 1e6 AGC target, 60 msec max ion time, and 300 to 1800 *m*/*z* mass range. MS/MS scans are acquired with the following parameters: 15,000 resolution at 20 *m/z*, 1e5 AGC target, 40 msec max IT time, 2.0 *m*/*z* isolation window, 0 *m*/*z* isolation offset, 200 to 2000 *m*/*z* mass range, 27 normalized collision energy, peptide match "preferred," exclude isotopes "on," a 30 sec dynamic exclusion window, and charge state exclusion set to exclude +1 and >+8 ions. All spectral data are collected in profile mode. A QC analysis of tryptic BSA peptides

(*optional*) is analyzed between each ProPeL injection to minimize sample carryover and gauge instrument performance stability. Similar instrument parameters are used but a shorter, 60 min gradient is performed in place of the 120 min gradient previously described.

Phosphopeptide identification and site localization

Raw files are searched with the Andromeda search engine and MaxQuant (Cox & Mann, 2008) against the UniProt *E. coli* strain B/BL21-DE3 proteome database (Proteome ID UP000002032). A reversed *E. coli* protein sequence database is automatically generated by the software and searched concurrently. All searches include the following parameters: 4.5 and 20 ppm mass tolerances for precursor and fragment ions, respectively, trypsin enzyme specificity, up to 2 missed cleavages, fixed carbamidomethyl C modification, variable phosphorylation of serine/threonine/tyrosine, oxidation of methionine, and acetylation of protein N-termini. Minimum peptide length is set to 5 and the contaminant database is included. All results are filtered at a 1% false discovery rate at the peptide spectrum match, protein and site levels. All other parameters are kept at MaxQuant default settings (version 1.6.0.1 at time of publication).

SUPPORT PROTOCOL 2

MOTIF VISUALIZATION WITH PLOGO

pLogos (O'Shea et al., 2013) depict residues proportional to the log-odds of their binomial probabilities with respect to a given background. In a pLogo, the most statistically significant residues appear closest to the *x*-axis, with residues above the *x*-axis indicating overrepresentation and those below the *x*-axis indicating underrepresentation. Given the existence of one or more different residues at a given substrate position, it is possible to compute conditional probabilities of all remaining amino acids and positions to determine significant positions *given* specific residues at specific positions. We refer to this as "fixing" a given residue at a given position, which allows for the exploration of correlated or uncorrelated residues across positions in the kinase specificity motif. Fixed positions within the pLogo (e.g., the central position) are depicted on a grey background, and red horizontal lines denote the $p = 0.05$ significance threshold (after Bonferroni correction). pLogos can be scaled for clarity. For each pLogo, the foreground data is the list of phosphorylation-centered 15mers, with negative control sites removed. The *E. coli* background data set is generated by pLogo through alignment of all unique phosphoacceptor-centered 15mers in the *E. coli* proteome. Below are basic instructions for generating pLogos.

1. Access the pLogo Web site (and register for an account if desired): https://plogo.uconn.edu
2. Paste desired foreground data set of aligned 15mers into the box on the left of the page.
3. Select "Protein" and then "e. coli k12" from the available backgrounds on the right of the page.
4. *Optional*: If logged into a personal account, the user may add a job name.
5. Click the "generate pLogo" button in the center of the page.
6. Residues can be fixed or unfixed by clicking on them; however, residues that do not achieve statistical significance may not be fixed. Alternatively, users may fix significant residues by checking the corresponding box in the "statistics" tab to the left of the pLogo.

7. The zoom can be changed by clicking the “customize” tab, and either hitting the ± buttons, or entering a value. Clicking “renormalize” will rescale the pLogo back to its default size.

   For additional functionality and explanations, see:

   O’Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, & Schwartz D. (2013). pLogo: A probabilistic approach to visualizing sequence motifs. Nat Methods 10, 1211–1212.

REAGENTS AND SOLUTIONS

Desalting elution solution (50 ml)

Combine the following into a 50-ml Pyrex medium bottle:

24.75 ml water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
25 ml acetonitrile, LC/MS grade (Fisher Scientific, cat. no. A955-1)
250 µl acetic acid, LC/MS grade (Fisher Scientific, cat. no. A11350)
Mix thoroughly
Store up to 1 year at room temperature

Desalting wash solution A (100 ml)

Combine the following in a 100-ml Pyrex medium bottle:

99.9 ml water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
100 µl trifluoracetic acid, LC/MS grade (Fisher Scientific, cat. no. A116-50)
Mix thoroughly
Store up to 1 year at room temperature

Desalting wash solution B (50 ml)

Combine the following in a 50-ml Pyrex medium bottle:

49.75 ml water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
250 µl acetic acid, LC/MS grade (Fisher Scientific, cat. no. A11350)
Mix thoroughly
Store up to 1 year at room temperature

Desalting wash solution C (50 ml)

Combine the following in a 50-ml Pyrex medium bottle:

49.5 ml water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
500 µl formic acid, LC/MS grade (Fisher Scientific, cat. no. A117-50)
Mix thoroughly
Store up to 1 year at room temperature

Fix solution (1 liter)

Combine the following in a 1-liter Pyrex medium bottle:

700 ml water (double distilled or Ultrapure)
500 ml methanol, ACS grade (Fisher Scientific, cat. no. A412-4)
100 ml acetic acid, glacial ACS grade (Fisher Scientific, cat. no. A38SI2-12)
Mix thoroughly
Store up to 1 year at room temperature

Laemmli loading buffer, 6× (10 ml)

Combine the following in a 15-ml conical tube:

1.2 g sodium dodecyl sulfate (SDS; Fisher Scientific, cat. no. BP166-500)

6 mg bromphenol blue (Fisher Scientific, cat. no. 403160050)
4.7 ml glycerol (Promega, cat. no. H5433)
1.2 ml Tris·Cl (pH 6.8) (Sigma Aldrich, cat. no. T6066)
2.1 ml water (double distilled or Ultrapure)
Mix thoroughly
Solution may be heated to 37°C to aid solubilization (which may take a period of several hours). Once fully dissolved, add 0.93 g dithiothreitol (DTT) and mix thoroughly. Divide into 0.5 ml aliquots and store up to 1 year at −20°C.

LB (lysogeny broth) liquid medium (1 liter)

Combine the following in a 1-liter Pyrex medium bottle:

10 g tryptone (Sigma-Aldrich, cat. no. 61044·1KG)
5 g yeast Extract (Sigma-Aldrich, cat. no. 09182-1KG-F)
10 g NaCl (Fisher Scientific, cat. no. S671-500)
Water (double distilled or Ultrapure)
Add water to ∼950 ml and mix thoroughly Adjust pH to 7.5 with NaOH
Add water to bring to 1-liter final volume
Split between two Pyrex medium bottles (<600 ml per bottle) and autoclave on liquid cycle
Once cooled, combine bottles and store up to 6 months at 4°C
Discard medium if it appears cloudy.
If adding antibiotic, wait until liquid media has cooled to 55°C. Adding antibiotic while the liquid medium is still hot may cause the antibiotic to degrade.

LB (lysogeny broth) solid plates (500 ml)

5 g tryptone (Sigma-Aldrich, cat. no. 61044-1KG)
2.5 g yeast extract (Sigma-Aldrich, cat. no. 09182-1KG-F)
5 g NaCl (Fisher Scientific, cat. no. S671-500)
7.5 g agar (Fisher Scientific, cat. no. BP1423-500)
Water (double distilled or Ultrapure)
Add dry reagents in a 1-liter Pyrex medium bottle, and add water to final volume 500 ml and mix thoroughly. It is not necessary to adjust the pH. Autoclave on liquid cycle, and then cool autoclaved medium in a water bath set to 55°C. Once cooled, add appropriate antibiotic if desired, swirl gently or rotating end-over-end to mix. Pour medium into ∼0.5-in. depth into sterile petri dishes. Prop the lid (to minimize dust) and allow the plates to cool in sterile area. Once cool, store the plates in plastic sleeve up to 3 months at 4°C. Discard any plates with visible growth.
If adding antibiotic, wait until liquid medium has cooled to 55°C. Adding antibiotic while the liquid medium is still hot may cause the antibiotic to degrade.

Lysis buffer (10 ml)

500 µl of 1 M Tris·Cl (pH 8.2) (Sigma Aldrich, cat. no. T6066)
750 µl of 1 M NaCl (Fisher Scientific, cat. no. S671-500)
4.80 g urea (Fisher Scientific, cat. no. BP169-500)
100 µl of 100× Halt Protease Inhibitor Cocktail, EDTA-free (Thermo Fisher Scientific, cat. no. PI87785)
100 µl of 100× Halt Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific, cat. no. PI78420)
100 µl of 100 mM phenylmethylsulfonyl fluoride (MP Biomedicals, cat. no. 195381)
Water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
Add Tris·Cl, NaCl, and urea into a clean 15-ml conical tube. Add water to approximately 9.5 ml and mix thoroughly. Add protease and phosphatase

inhibitors, and PMSF immediately prior to use. Add water to bring to 10 ml final volume, and store at room temperature.

Make buffer fresh daily.

Pro-Q Diamond destain solution (1 liter)

Combine the following in a 1-liter Pyrex medium bottle:
750 ml water (double distilled or Ultrapure)
200 ml acetonitrile, ACS grade (Fisher Scientific, cat. no. A21-4)
100 ml of 1 M sodium acetate (pH 4.0) (Fisher Scientific, cat. no. S210-500)
Mix thoroughly
Store up to 1 year at room temperature

SDS-PAGE gel

12% separating gel
3.4 ml water (double distilled or Ultrapure)
4 ml Bis/Acrylamide (37.5:1) (Bio-Rad, cat. no. 1610158)
2 ml of 1.87 M Tris·Cl (pH 8.9) (Sigma Aldrich, cat. no. T6066)
100 µl of 10% SDS Solution (Fisher Scientific, cat. no. BP166-500)
5 µl TEMED (Fisher Scientific, cat. no. BP150·20)
0.5 ml ammonium persulfate (15 mg/ml solution) (Fisher Scientific, cat. no. BP179-100)

3.5% Stacking Gel
0.826 ml water (double distilled or Ultrapure)
0.232 ml Bis/Acrylamide (Bio-Rad, cat. no. 1610158)
0.4 ml of 0.312 M Tris·Cl (pH 6.7) (Sigma Aldrich, cat. no. T6066)
20 µl of 10% SDS Solution (Fisher Scientific, cat. no. BP166-500)
1 µl TEMED (Fisher Scientific, cat. no. BP150-20)
0.533 ml ammonium persulfate (15 mg/ml solution) (Fisher Scientific, cat. no. BP179-100)

Assemble gel casting stand (Bio-Rad, cat. no. 1658050) and plates (glass short plates, Bio-Rad, cat. no. 1653308; 1.5-mn glass spacer plate, Bio-Rad, cat. no. 1653312). Mix *separating gel* reagents gently (avoid introducing bubbles), and pour separating gel until the level reaches approximately 1-in. below the top of the cassette. Layer isobutanol (use a 1:1 ratio of isobutanol:water, and use only the top, less dense layer) on top of the separating gel to prevent bubbles and level the gel. Allow the gel to polymerize for ~30 min. Pour off isobutanol and rinse with water. Mix *stacking gel* reagents gently (avoid introducing bubbles), and pour stacking gel to the top of the cassette. Insert an appropriate comb (e.g., 10-well; Bio-Rad, cat. no. 1653365 or 15-well comb; Bio-Rad, cat. no. 1653366) and allow the gel to polymerize for approximately 30 min. Gels can be wrapped in plastic wrap and stored up to several days at 4°C.

SDS-PAGE running buffer, 10×

30 g Tris base (Sigma Aldrich, cat. no. T6066)
144 g glycine (Fisher Scientific, cat. no. BP381-1)
10 g SDS (Fisher Scientific, cat. no. BP166-500)
Water (double distilled or Ultrapure)
Add reagents and water to a final volume of 1 liter, and mix thoroughly. It is not necessary to adjust the pH (which should be pH 8.3). Store up to 1 year at room temperature.

TiO_2 binding solution (100 ml)

Combine the following in a 100-ml Pyrex medium bottle:

continued

33.44 ml water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
50 ml acetonitrile, LC/MS grade (Fisher Scientific, cat. no. A955-1)
16.56 ml lactic acid (Sigma Aldrich, cat. no. 69785-1L)
Mix thoroughly
Store up to 1 year at room temperature
The lactic acid makes the solution very sticky. Gloves exposed to TiO_2 binding solution may stick to tubes.

TiO_2 elution solution (50 ml)

0.34 g KH_2PO_4 (potassium phosphate monobasic; Fisher Scientific, cat. no. BP362-500)
Water, HPLC or LC/MS grade (Fisher Scientific, cat. no. W6-1)
Add dry reagents in a 50-ml Pyrex medium bottle. Add water to ~45 ml and mix thoroughly. Adjust pH to 10 with NaOH. Add water to bring to 50 ml final volume, mix thoroughly, and store up to 1 year at room temperature.

COMMENTARY

Background Information

There are several existing strategies for determining kinase specificity motifs. The simplest approach is to align experimentally determined phosphorylation sites for a given kinase on the phosphoacceptor, and extract motifs bioinformatically. However, due to the complex, overlapping nature of kinase signaling cascades, it has been difficult to unambiguously pair identified phosphorylation sites with their upstream kinase. As a result of this bottleneck, several strategies have been developed for determining protein kinase substrate specificity motifs. Early success came from the use of an oriented peptide library, in a strategy pioneered by Cantley and colleagues (Songyang et al., 1994). This method utilized a synthetic peptide library of ≥2.5 billion distinct sequences, which all contained a single phosphoacceptor in the center of the peptide flanked by 4 variable positions upstream and downstream, and linker residues at the termini. The kinase of interest was incubated with the library and ^{32}P-ATP in an *in vitro* kinase reaction. Phosphopeptides were then separated by ferric iminodiacetic acid, and sequenced by Edman degradation. Observed residue frequencies in each position were compared to their respective abundance (obtained by sequencing the peptide library) to identify important determinants. While successful, this approach has several drawbacks. Notably, identification by Edman degradation is laborious and time-consuming, and technical limitations prevent querying tryptophan or cysteine (which interfere with sequencing due to oxidation), or additional phosphoacceptors (which would render the phosphorylation site ambiguous). In order to eliminate these challenges, the technique was refined by Turk and colleagues to a matrix format. Peptide libraries are arranged in a grid, with a single fixed residue/position for each well (representing all the possible amino acids in each position), while the other positions are randomized. After incubation with the kinase and ^{32}P-ATP, the reactions are spotted onto a membrane, and phosphorylation preferences are read by measuring radioactive incorporation at different fixed residue positions (Hutti et al., 2004). While these methods provide a wealth of data, the combinatorial libraries are prohibitively expensive for most labs. Additionally, they require the use of radioactivity and large amounts of recombinant kinase. Importantly, these approaches use peptide substrates (which are less physiologically relevant than protein substrates), and there is no ability to discern correlation between motif positions.

The next milestone came with the utilization of depleted cell lysates, which could function as a proteomic library. Crucially, a proteomic library (unlike a random library) is amenable to sequencing by tandem mass spectrometry. This approach was first demonstrated by Huang and colleagues (Huang, Tsai, Chen, Wu, & Chen, 2007), wherein rat uterus homogenate was fractionated by Strong Anion Exchange (SAX), treated with phosphatases (to remove endogenous phosphorylation), and heated to inactivate endogenous kinases and all phosphatases. The resulting fractions were incubated with recombinant kinase and *cold* ATP to allow the kinase of interest to phosphorylate the depleted cellular protein fractions (in the absence of ^{32}P-ATP). The reaction mixture was digested with trypsin and phosphoenriched by IMAC,

followed by identification by tandem mass spectrometry. Other versions of this approach inactivated endogenous kinases by tryptic digestion prior to SCX fractionation and dephosphorylation (Kettenbach et al., 2012) or by the addition of the irreversible ATP-competitive analog 5′-4-fluorosulphonylbenzoyladenosine (FSBA) (Knight et al., 2012). While the use of a proteomic library (and therefore the ability to identify sites by tandem mass spectrometry) is a significant advantage, these methods still require large amounts of recombinant kinase, run the risk of contamination in the event of incomplete dephosphorylation or residual endogenous kinase activity, and in some cases require that specificity preferences be queried by peptide (rather than protein) substrates.

The ProPeL approach described in this article has several important benefits and advantages over existing strategies. As the phosphorylation reactions occur *in vivo* in the *E. coli* cytoplasm, the living host produces both the kinase and substrate proteins, and regulates the environment (pH, ionic and co-factor concentrations, etc.). This obviates the need to purify catalytically active kinase, and means that the target kinase interacts with substrates under conditions that are more physiologically relevant than an *in vitro* reaction. The substrates themselves are full-length *E. coli* proteins, which offer two distinct advantages. First, the kinase reaction occurs with phosphoacceptors within protein substrates that are able to adopt a physiologically appropriate folded structure, rather than peptides that may not fully recapitulate the environment surrounding the phosphoacceptor. Second, a proteomic background allows unambiguous sequence and site identification using tandem mass spectrometry. This not only allows for high-throughput sequencing, but also puts each phosphorylation site in sequence context, which allows for intra-motif correlation between positions. As direct phosphorylation of bacterial substrates by the kinase of interest is measured, there is no need for radioactive material (typically ^{32}P-ATP), making ProPeL safer than traditional approaches. The actual reaction and sample preparation are also significantly cheaper than combinatorial peptide library approaches, although access to a mass spectrometer is required.

Critical Parameters

The success of a ProPeL experiment crucially depends on the ability to express soluble, active kinase to facilitate the *in vivo* phosphorylation of bacterial substrates. While the majority of eukaryotic proteins expressed in *E. coli* were easily purified in their correctly folded state (Braun et al., 2002), there are several factors that can affect heterologous protein expression (Dyson, Shadbolt, Vincent, Perera, & McCafferty, 2004). It is reasonable to expect that expressing different foreign kinases within *E. coli* will lead to differential exogenous phosphorylation, necessitating different optimal expression conditions to maximize *in vivo* phosphorylation of bacterial substrates by each kinase. There are several variables that can be adjusted to improve *in vivo* expression and activity.

Choosing the correct E. coli cell strain

In our hands, the single variable that has impacted target kinase expression the most has been selecting the correct cell strain. *E. coli* expression of the 61 tRNAs is different from eukaryotic species, which can hinder heterologous protein expression for transcripts that are rich in codons that are under-utilized in *E. coli* (Kane, 1995). This can be ameliorated by exhaustive codon optimization of the target kinase coding sequence, or by the usage of so-called codon-optimized strains. For example, the Rosetta2 strains supply tRNAs for the codons AUA, AGG, AGA, CUA, CCC, GGA, and CGG. Over-expression of these tRNAs can improve expression of the target kinase.

Alternatively, expression of the target kinase may be cytotoxic, preventing any cells that are competent to express the protein from surviving. A commonly used strategy is to switch to a resistant host strain, such as the C41(DE3) and C43(DE3) “Walker” strains, which exhibit elevated ability to express membrane-bound and toxic proteins (Miroux & Walker, 1996). These strains were later characterized as having mutations in the *lac*UV5 promoter (which controls expression of T7 RNA polymerase) and therefore exhibited more gradual target protein expression (Wagner et al., 2008).

Typically, we start by expressing a new plasmid in both C41(DE3) and Rosetta2 cells. If the target kinase expresses in the C41(DE3) strain, it implies that codon usage is not a concern, and typically we observe more robust expression in C41(DE3) strains than in Rosetta2. However, when codon utilization appears to be an issue, expression is often *only* successful in Rosetta2, with no expression in C41(DE3). We have also begun to use C43(DE3) cells with some success for kinases that failed to express in C41(DE3).

Protein induction and growth conditions

Protein expression is commonly achieved using the inducible T7 promoter pET system, whereby induction is initiated by addition of an allolactose analog, usually isopropyl β-D-1-thiogalactopyranoside (IPTG) (Donovan, Robinson, & Glick, 1996). Controlling the concentration of IPTG used and the temperature and duration of induction is critical for optimizing target kinase expression. Typical IPTG concentrations range from 0.1 to 1 mM, and while increasing concentrations can lead to faster protein production, this has the danger of overwhelming the cells and causing cytotoxicity. Similarly, while a longer induction may allow more time for kinase expression, this can be detrimental if the kinase hinders growth.

While IPTG induction is very successful in many instances, induction may overwhelm the cell, inhibiting further protein expression and in some cases killing plasmid-bearing cells. As an alternative to IPTG induction, protein expression can be accomplished using the autoinduction system, which uses a complex media containing both glucose and lactose. The *E. coli* preferentially metabolizes the glucose, growing to high density while target protein expression is suppressed by glucose. As the glucose is depleted the cells switch to lactose, producing allolactose, and inducing target protein expression in robust, high density cultures (Studier, 2005).

Enhancing protein solubility

ProPeL relies upon the ability of the target kinase to phosphorylate bacterial proteins *in vivo*. Therefore, it is critical that the kinase is not only expressed, but is also soluble during expression. The major challenge is to prevent the over-expressed kinase from being sequestered in inclusion bodies–insoluble aggregates of mis-folded proteins (Schein, 1989). Inclusion bodies form when exposed, hydrophobic stretches of insoluble, mis-folded, or partially folded proteins stick together via intermolecular β-sheet structures (Fink, 1998). The dynamics of inclusion body formation are such that there are typically a small number of inclusion bodies seeded by an incorrectly folded protein intermediate. Although various studies have suggested refolding strategies both *in vivo* (Zhao et al., 2012) and *in vitro* (Santos et al., 2012), the consensus is that it is critical to prevent inclusion body formation from ever occurring to maximize yield (Fink, 1998; Schein, 1989).

Within the context of ProPeL, there are several strategies that can be attempted for increasing solubility. At the start of a new ProPeL project, it is important to design the correct insert. Reducing the size of the target kinase by only expressing the catalytic domain (if appropriate) will help increase solubility, as lower molecular weight proteins tend to exhibit superior soluble expression (Dyson et al., 2004). The addition of fusion-protein tags such as 6XHis, GST, and MBP may also increase solubility, although there are no universal rules for how a tag may help or hinder solubility, and the inclusion of a tag may interfere with kinase function (Guerrero, Ciragan, & Iwaï, 2015; Rosano & Ceccarelli, 2014). Lowering the temperature during induction has also been demonstrated to improve protein solubility (Baldwin, 1986). Finally, co-expressing the target protein with molecular chaperones such as the GroELS or DnaK system is a common strategy to aide protein folding (Nishihara, Kanemori, Yanagi, & Yura, 2000; Marco et al., 2007). Unfortunately, to date this step has provided the most significant challenge for ProPeL, with limited success in the cases of protein kinases that appear insoluble.

Troubleshooting

In troubleshooting *in vivo* activity, it is advantageous to first identify whether the issue is protein expression, protein solubility, or kinase activity. The first question that must be answered is a relatively simple one: Is the target kinase expressed? Kinase expression is easily evaluated by lysing an aliquot in the standard denaturing lysis buffer and evaluating expression by SDS-PAGE and Coomassie staining. An example of the desired level of kinase expression is provided in Figure 2. It is also possible to evaluate expression by western blotting, but this may only be necessary if the expected kinase molecular weight is the same as highly expressed endogenous *E. coli* proteins. If the kinase is not sufficiently expressed as to be easily detected by Coomassie staining, then it is likely not expressed sufficiently to provide adequate *in vivo* activity. It is also worth noting that expression patterns for endogenous *E. coli* proteins change with expression conditions, so it is important to run negative controls (i.e., a kinase-dead mutant, see Strategic Planning) for each expression condition to generate an accurate background for comparison. If there is no detectable expression, this may indicate one of several problems. Strategies for overcoming codon bias, cytotoxicity, and solubility have

been discussed above in the Critical Parameters section. Considerations for kinase activation requirements have been discussed in the Strategic Planning section, and should be used to inform the choice between the Basic Protocol and Alternate Protocol.

Anticipated Results

In our experience, the number of unique phosphorylation sites discovered for a particular kinase of interest most strongly correlates with the level of expression and *in vivo* activity. With the instrumentation setup described in this protocol, we have observed that the total number of unique phosphorylation sites identified in a single run can range from 200 to 500 sites for one kinase to over 1500 sites for a different kinase. Technical replicates significantly increase detected phosphopeptides (Ham et al., 2008), although in our experience 2 to 3 technical replicates are typically sufficient.

Kinase motifs can at times be highly specific, where precise residue positions are favored, presumably due to the unique steric considerations, angles, and electrostatic properties associated with the individual residues. These motifs are frequently represented as consensus sequences, which offer the idealized version of the motif. A classic example is the PKA motif, which today we can summarize as the consensus sequence [R/K][R/K]x[S/T]Φ, where [R/K] indicates a preference for either arginine or lysine, Φ represents a hydrophobic amino acid, and "x" indicates no preference. While the PKA motif is well-defined, specificity motifs can also be more general, such as simple recognition of charge or hydrophobicity, and with positional flexibility. It is important to consider motif results holistically, allowing for both specific and general preferences when interpreting the pLogo for a given kinase. It is also informative to evaluate phosphoacceptor motifs separately–often there can be subtle differences between preferences surrounding serine compared with threonine phosphoacceptors, even for the same kinase.

Time Considerations

The Basic Protocol takes ∼1 week to complete on average. Starting with an overnight bacterial starter culture, expression typically takes up to a day. Cell harvesting, lysis, quantification and SDS-PAGE analysis occurs on day 2, and typically takes a full day. The most efficient workflow is to perform a methanol/chloroform extraction at the end of the day, and allow the protein disc to resolubilize overnight. The tryptic digestion should not proceed longer than 16 hr, so we typically begin reduction and alkylation steps around 2 pm on day 3. Tryptic digestion occurs overnight. SEP-Pak peptide desalting, lyophilization, and TiO_2 enrichment can be accomplished in a single day, but we often store lyophilized peptides (pre-TiO_2) at −20°C overnight. Analysis by LC-MS/MS takes approximately 4 hr for each run including the BSA QC standard analysis. MaxQuant and Andromeda processing of the raw spectra typically takes approximately 0.5–3 hr, depending on the number of simultaneous searches and number of threads available on the server, while data filtering and pLogo analysis can be completed in under an hour. There are several pause points that are noted throughout the protocol. We have successfully stored cell pellets at −80°C for several months, and have stored lyophilized peptides (pre-TiO_2) at −20°C for several months. Cell lysate can be stored at 4°C for several months.

Acknowledgements

This work was supported in whole or in part by grants awarded to G.M.C. from the Department of Energy (DE-FG02-02ER63445), and to D.S. from the University of Connecticut Research Foundation, the University of Connecticut Office of the Vice President for Research, and the National Institute of Neurological Disorders and Stroke (1R21NS096516).

Literature Cited

Baldwin, R. L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proceedings of the National Academy of Sciences*, *83*, 8069–8072. doi: 10.1073/pnas.83.21.8069.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., & LaBaer, J. (2002). Proteome-Scale Purification of Human Proteins from Bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 2654–2659. doi: 10.1073/pnas.042684199.

Chou, M. F., Prisic, S., Lubner, J. M., Church, G. M., Husson, R. N., & Schwartz, D. (2012). Using Bacteria to Determine Protein Kinase Specificity and Predict Target Substrates. *PLoS One*, *7*, e52747. doi: 10.1371/journal.pone.0052747.

Chou, M. F., & Schwartz, D. (2011). Biological sequence motif discovery using motif-x. *Current Protocols in Bioinformatics/Editoral Board, Andreas D. Baxevanis ... [et Al.]*, Chapter 13:UNIT *13*, 15–24. doi: 10.1002/0471250953.bi1315s35.

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates,

individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*, 1367–1372. doi: 10.1038/nbt.1511.

Donovan, R. S., Robinson, C. W., & Glick, B. R. (1996). Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *Journal of Industrial Microbiology*, *16*, 145–154. doi: 10.1007/BF01569997.

Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., & McCafferty, J. (2004). Production of soluble mammalian proteins in Escherichia coli: Identification of protein features that correlate with successful expression. *BMC Biotechnology*, *4*, 32. doi: 10.1186/1472-6750-4-32.

Fink, A. L. (1998). Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Folding and Design*, *3*, R9–R23. doi: 10.1016/S1359-0278(98)00002-9.

Gibbs, C. S., & Zoller, M. J. (1991). Rational scanning mutagenesis of a protein kinase identifies functional regions involved in catalysis and substrate interactions. *Journal of Biological Chemistry*, *266*, 8923–8931.

Guerrero, F., Ciragan, A., & Iwaï, H. (2015). Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expression and Purification*, *116*, 42–49. doi: 10.1016/j.pep.2015.08.019.

Ham, B. M., Yang, F., Jayachandran, H., Jaitly, N., Monroe, M. E., Gritsenko, M. A., . . . Smith, R. D. (2008). The Influence of Sample Preparation and Replicate Analyses on HeLa Cell Phosphoproteome Coverage. *Journal of Proteome Research*, *7*, 2215–2221. doi: 10.1021/pr700575m.

Hanks, S. K., Quinn, A. M., & Hunter, T. (1988). The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science*, *241*, 42–52. doi: 10.1126/science.3291115.

Hansen, A.-M., Chaerkady, R., Sharma, J., Díaz-Mejía, J. J., Tyagi, N., Renuse, S., . . . Pandey, A. (2013). The *Escherichia coli* Phosphotyrosine Proteome Relates to Core Pathways and Virulence. *PLOS Pathogens*, *9*, e1003403. doi: 10.1371/journal.ppat.1003403.

Huang, S.-Y., Tsai, M.-L., Chen, G.-Y., Wu, C.-J., & Chen, S.-H. (2007). A Systematic MS-Based Approach for Identifying in vitro Substrates of PKA and PKG in Rat Uteri. *Journal of Proteome Research*, *6*, 2674–2684. doi: 10.1021/pr070134c.

Hutti, J. E., Jarrell, E. T., Chang, J. D., Abbott, D. W., Storz, P., Toker, A., . . . Turk, B. E. (2004). A rapid method for determining protein kinase phosphorylation specificity. *Nature Methods*, *1*, 27–29. doi: 10.1038/nmeth708.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current Opinion in Biotechnology*, *6*, 494–500. doi: 10.1016/0958-1669(95)80082-4.

Kettenbach, A. N., Wang, T., Faherty, B. K., Madden, D. R., Knapp, S., Bailey-Kellogg, C., & Gerber, S. A. (2012). Rapid Determination of Multiple Linear Kinase Substrate Motifs by Mass Spectrometry. *Chemistry & Biology*, *19*, 608–618. doi: 10.1016/j.chembiol.2012.04.011.

Khokhlatchev, A., Xu, S., English, J., Wu, P., Schaefer, E., & Cobb, M. H. (1997). Reconstitution of Mitogen-activated Protein Kinase Phosphorylation Cascades in Bacteria EFFICIENT SYNTHESIS OF ACTIVE PROTEIN KINASES. *Journal of Biological Chemistry*, *272*, 11057–11062. doi: 10.1074/jbc.272.17.11057.

Knight, J. D., Tian, R., Lee, R. E., Wang, F., Beauvais, A., Zou, H., . . . Kothary, R. (2012). A novel whole-cell lysate kinase assay identifies substrates of the p38 MAPK in differentiating myoblasts. *Skeletal Muscle*, *2*, 5. doi: 10.1186/2044-5040-2-5.

Lubner, J. M., Dodge-Kafka, K. L., Carlson, C. R., Church, G. M., Chou, M. F., & Schwartz, D. (2017). Cushing's syndrome mutant PKAL205R exhibits altered substrate specificity. *FEBS Letters*, *591*, 459–467. doi: 10.1002/1873-3468.12562.

Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., & Mann, M. (2008). Phosphoproteome Analysis of *E. coli* Reveals Evolutionary Conservation of Bacterial Ser/Thr/Tyr Phosphorylation. *Molecular & Cellular Proteomics*, *7*, 299–307. doi: 10.1074/mcp.M700311-MCP200.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., & Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC Biotechnology*, *7*, 32. doi: 10.1186/1472-6750-7-32.

Miller, M. L., Jensen, L. J., Diella, F., Jørgensen, C., Tinti, M., Li, L., . . . Linding, R. (2008). Linear Motif Atlas for Phosphorylation-Dependent Signaling. *Science Signaling*, *1*, ra2–ra2. doi: 10.1126/scisignal.1159433.

Miroux, B., & Walker, J. E. (1996). Overproduction of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*, *260*, 289–298. doi: 10.1006/jmbi.1996.0399.

Nishihara, K., Kanemori, M., Yanagi, H., & Yura, T. (2000). Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in Escherichia coli. *Applied and Environmental Microbiology*, *66*, 884–889. doi: 10.1128/AEM.66.3.884-889.2000.

Obenauer, J. C., Cantley, L. C., & Yaffe, M. B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research*, *31*, 3635–3641. doi: 10.1093/nar/gkg584.

O'Shea, J. P., Chou, M. F., Quader, S. A., Ryan, J. K., Church, G. M., & Schwartz, D. (2013). pLogo: A probabilistic approach to visualizing sequence motifs. *Nature Methods*, *10*, 1211–1212. doi: 10.1038/nmeth.2646.

Oza, J. P., Aerni, H. R., Pirman, N. L., Barber, K. W., ter Haar, C. M., Rogulina, S., … Jewett, M. C. (2015). Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nature Communications*, *6*, 8168. doi: 10.1038/ncomms9168.

Pinna, L. A., & Ruzzene, M. (1996). How do protein kinases recognize their substrates? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1314*, 191–225. doi: 10.1016/S0167-4889(96)00083-3.

Potel, C., M., Lin, M.-H., Heck, A. J. R., & Lemeer, S. (2018). Widespread bacterial protein histidine phosphorylation revealed by mass spectrometry-based proteomics. *Nature Methods*, *15*(3), 187–190.

Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, *2*, 1896–1906. doi: 10.1038/nprot.2007.261.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*, Available at: Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029002/ [Accessed June 13, 2016]. doi: 10.3389/fmicb.2014.00172.

Santos, C. A., Beloti, L. L., Toledo, M. A. S., Crucello, A., Favaro, M. T. P., Mendes, J. S., … Souza, A. P. (2012). A novel protein refolding protocol for the solubilization and purification of recombinant peptidoglycan-associated lipoprotein from Xylella fastidiosa overexpressed in Escherichia coli. *Protein Expression and Purification*, *82*, 284–289. doi: 10.1016/j.pep.2012.01.010.

Schein, C. H. (1989). Production of Soluble Recombinant Proteins in Bacteria. *Nature Biotechnology*, *7*, 1141–1149. doi: 10.1038/nbt1189-1141.

Schwartz, D., & Gygi, S. P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nature Biotechnology*, *23*, 1391–1398. doi: 10.1038/nbt1146.

Soares, N. C., Spät, P., Krug, K., & Macek, B. (2013). Global Dynamics of the Escherichia coli Proteome and Phosphoproteome During Growth in Minimal Medium. *Journal of Proteome Research*, *12*, 2611–2621. doi: 10.1021/pr3011843.

Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., & Cantley, L. C. (1994). Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Current Biology*, *4*, 973–982. doi: 10.1016/S0960-9822(00)00221-9.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*, 207–234. doi: 10.1016/j.pep.2005.01.016.

Ubersax, J. A., & Ferrell, Jr, J. E. (2007). Mechanisms of specificity in protein phosphorylation. *Nature Reviews Molecular Cell Biology*, *8*, 530–541. doi: 10.1038/nrm2203.

Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., … de Gier, J. W. (2008). Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences*, *105*, 14371–14376. doi: 10.1073/pnas.0804090105.

Zhao, Y., He, W., Liu, W.-F., Liu, C.-C., Feng, L.-K., Sun, L., … Hang, H.-Y. (2012). Two Distinct States of Escherichia coli Cells That Overexpress Recombinant Heterogeneous β-Galactosidase. *Journal of Biological Chemistry*, *287*, 9259–9268. doi: 10.1074/jbc.M111.327668.

Key References

Chou et al., 2012. See above.
Original proof-of-concept publication for ProPeL.

Lubner et al., 2017. See above.
Publication using the current ProPeL workflow.

Internet Resources

https://schwartzlab.uconn.edu/pepextend
PeptidExtender Web tool, used in this protocol to map tryptic phosphopeptides to the reference E. coli proteome and create modification-centered 15mers.

https://plogo.uconn.edu/
Motif visualization tool.

https://motif-x.med.harvard.edu
Motif discovery tool.