Published in final edited form as:

Nat Chem Biol. 2018 February 14; 14(3): 188–192. doi:10.1038/nchembio.2572.

The ABCs of PTMs

Karl W. Barber and Jesse Rinehart

Department of Cellular & Molecular Physiology and the Systems Biology Institute, Yale University, West Haven, Connecticut, USA

Abstract

Post-translational modifications (PTMs) are ubiquitous in all forms of life and often modulate critical protein functions. Recent chemical and biological advances have finally enabled scientists to precisely modify proteins at physiologically-relevant positions ushering in a new era of protein studies.

The grand promise of synthetic biology is that scientists will one day be able to make organisms from scratch to perform user-defined functions. Fueled by the what-I-cannot-create-I-do-not-understand Feynman mentality, the basic building blocks of biology have been chopped up and reassembled to design full genomes, novel genetic circuits and *de novo* proteins, demonstrating our rapidly expanding abilities to tame biological molecules to do our bidding^{1–3}. This frame of mind is not unique to bioengineers, as laboratories around the world routinely synthesize (“create”) recombinant proteins to answer (“understand”) fundamental biological questions. However, with current technologies, we are trying to write words with an incomplete alphabet.

Virtually all proteins in eukaryotes undergo PTMs that can be transformative in protein folding and function. Without expanding our ability to make modified proteins, we are constrained to a limited understanding the consequences of these important regulations. Despite decades of advances in protein synthesis, we usually operate with a toolkit limited to the canonical 20 amino acids, unable to easily insert or append modifications at desired, physiologically-relevant positions. Enzymes responsible for phosphorylation, acetylation, glycosylation and other processes have been successfully deployed to drive modifications of proteins of interest (Fig. 1a), but there are several shortcomings to these methods. First, the enzyme responsible for a site-specific protein modification of interest is often unknown or difficult to synthesize. Second, many of these enzymes will recognize and modify multiple sites within the same target protein, making analysis of the role of modification at a particular position more challenging. Third, especially regarding complex processes such as glycosylation and ubiquitylation, PTMs are sequential and complex, occurring in a step-wise manner and requiring various enzymes to properly decorate the protein surface. Another common technique to investigate PTMs is to substitute an amino acid “mimetic” at the position of interest, such as acidic residues for phosphorylation or glutamine for

Competing financial interests

The authors declare no competing financial interests.

acetyllysine. This technique, however, does not fully capture the size, charge distribution or dynamic nature of a bona fide PTM, and most natural PTMs have no suitable mimetic.

Fortunately, we have found the missing letters in our protein alphabet: recent techniques have enabled the synthesis of proteins containing amino acid modifications at the positions of interest, providing new layers of control and intricacy in the proteins we construct in the laboratory. In this commentary, we will discuss the “spelling rules” in the new age of PTM studies: our abilities to identify protein modifications, advances in chemical and synthetic biology to study PTMs and expand our recombinant repertoire, and how this lays the foundation for the next generation of protein modification studies. This commentary will only address a handful of human-relevant PTMs, but we hope the ideas presented will encourage new experimentation addressing a broader scope of modifications.

I before E

The first rule in PTM studies is “identification” before “experimentation”. Most interest in protein modifications spawns from physiological relevance. PTMs can have potent effects on protein folding and function, and are frequently associated with both healthy and disease states (Box 1). Before generating a protein containing PTMs, we must first identify positions within the protein that are modified in the cell. One of the most commonly employed techniques is antibody-based detection. However, antibodies raised against a single epitope can only be used to interrogate the status of a modification (present or absent) and cannot be used to identify novel modified sites by immunoblot. Antibody generation to investigate PTMs also requires chemical synthesis of a modified peptide antigen and time-consuming animal immunization, and the degree of specificity of the resulting antibody varies greatly.

Box 1

Common PTMs and their physiological relevance

Although hundreds of classes of PTMs have been identified, phosphorylation is one of the most studied because of its prevalence in eukaryotic signaling, the availability of reagents for detection and enrichment, and by historical virtue of being one of the first identified PTMs. In eukaryotes, kinases are responsible for the phosphorylation of serine, threonine and tyrosine residues, covalently adding a phosphoryl group with a -2 charge at physiological pH that can greatly impact protein folding, function, and stability. Dysregulation of kinase activity or protein phosphorylation site mutations can contribute to cancer initiation or progression, and hyperphosphorylation of certain proteins is associated with certain disease states, such as the tau protein in Alzheimer’s disease.

Glycosylation, another frequently studied PTM, is the attachment of a glycan to a protein, most commonly on an asparagine residue. N-linked glycosylation is initiated in the endoplasmic reticulum and the glycan is further processed along the secretory pathway. This results in complex and often heterogeneous glycoprotein products, and the final products are starkly different between organisms. N-linked glycosylation, like many other PTMs, can greatly impact protein structure and function, and mutations within many oligosaccharide processing enzymes can result in human disease, such as developmental

delays. Importantly, many human antibodies are glycosylated within the Fc region, impacting the engendered effector function and alluding to the interest in synthesizing properly glycosylated therapeutic antibodies.

Other PTMs, such as acetylation and methylation, are also quite prevalent. These modifications have been most extensively studied in the context of histones, where they contribute to the architectural remodeling of chromatin. Reader proteins or domains (such as bromodomains for acetyllysine) are able to selectively bind to these histone modifications and recruit accessory proteins for signal transduction. Although not confined solely to epigenetic regulation, acetylation and methylation are both known to play important roles in transcriptional activation and repression.

The most comprehensive and definitive PTM mapping techniques rely on identification by enzymatic digest followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Complex protein samples can be rapidly studied using LC-MS/MS and modifications can be easily identified in the resulting spectra due to characteristic mass differences between modified and unmodified peptides. Searching for common modifications in protein samples can both confirm PTMs situated at previously identified positions and discover new sites of modification. LC-MS/MS conveniently offers a means to monitor the presence of thousands of modifications, and differential abundances across experimental conditions may provide mechanistic insight into the cellular conditions and machinery responsible for PTMs.

High-throughput PTM discovery still has limitations for a variety of reasons. For example, PTMs often have low stoichiometry (i.e. perhaps only several molecules of a particular protein are modified at a given time). This means that, in a complex trypsin-digested protein sample, peptides containing the modification of interest may be in the vast minority and

easily missed. Labile modifications may be unstable during sample preparation and modified peptides may be difficult to detect due to inefficient ionization. To combat this unfavorable distribution, modification enrichment techniques have been developed. Many modifications can be pulled out of complex mixtures using immobilized antibodies that recognize a specific PTM, such as ubiquitylation, acetylation and phosphorylation. However, antibody-based enrichment can often incur bias driven by biochemical properties of the peptides and sequence context of the PTM. Hydrazine chemistry and lectin affinity immobilization are commonly used to enrich for glycoproteins, but these methods preferentially isolate certain types of oligosaccharides. Phosphorylation can also be robustly enriched using several other techniques, such as immobilized metal affinity chromatography or using titanium dioxide, contributing to the identification of hundreds of thousands of putative phosphorylation sites occurring in natural proteomes⁴.

Although PTMs like glycosylation and phosphorylation are very common in eukaryotes, they are likely relatively overrepresented in scientific literature compared to other PTMs because of the availability of enrichment techniques targeted to these modifications. The future of PTM mapping will rely heavily on improved enrichment techniques for diverse classes of modifications. New mass spectrometers with greater speed and sensitivity along with advanced method development to maintain modifications and ionize modified peptides will also aid in PTM discovery. Challenges still remain in identifying multiple modifications within the same protein molecule, as bottom-up approaches make it impossible to discern whether two separate modified peptides originated from the same original molecule within a diverse population of proteins. Another outstanding problem is that currently unknown types of PTMs are difficult to identify using mass spectrometry since only known modifications can be identified using standard mass spectra search algorithms. In fact, a large proportion of peptides observed by mass spectrometry often remain unidentified; peptides modified in unpredicted ways certainly contribute to some of these unassigned spectra. New PTMs will likely be identified using low-throughput biochemical techniques and possibly with chemical tagging strategies that couple PTM enzyme activity with substrate discovery. Once new classes of modifications are discovered, we will probably see evidence of these PTMs in LC-MS/MS datasets both new and old.

Except after C

“Chemical” and synthetic approaches have opened the floodgates of PTM-related experimentation. PTM localization studies certainly aid in identifying sites of modification that will be of interest to investigators, but if the enzyme responsible for the modification is unknown or difficult to prepare, generating modified protein may be difficult. Chemical biology has come to the rescue, offering a wide host of methods to investigate various PTMs. Using chemical approaches to synthesize modified proteins permits incorporation of modifications at desired positions that do not rely on endogenous cellular signaling. In this way, PTMs incorporated using chemical biology can lead to discovery of biological function.

One of the most common methods to create modified proteins is using native chemical ligation (Fig. 1b). In one popular version of this scheme known as expressed protein

ligation⁵, a protein is isolated from a host organism containing a self-excising peptide that will yield a free C-terminal cysteine. In parallel, a peptide containing the PTM of interest and a thioester moiety is made via solid phase synthesis. Reaction of the two components results in a full-length, modified protein. However, this technique is generally only useful to generate proteins with modifications close to a terminus and, in most methods, requires the presence or addition of a cysteine residue. This technique can also cause problems with protein structure since the ligated peptide sequence is not present during initial protein folding. On the positive side, the resulting protein is scarless (i.e. there is a peptide bond at the site of the ligation), and some methods that do not require the presence of a cysteine have been developed.

Certain amino acids within a protein can provide a reactive handle for further semisynthetic approaches to the generation of modified proteins (Fig. 1c). For example, specific protein glycoforms can be made by substituting asparagine for cysteine at desired positions and reacting the protein with a glycosyl iodoacetamide, yielding a very similar product to N-glycosylation⁶. This type of reaction can also provide much more homogeneous product than protein isolated directly from eukaryotic cells. One problem with this particular approach is that any cysteine residues not participating in disulfide bridges may be subject to conjugation of the modification. Additionally, although the resulting molecule will strongly resemble a protein with the desired PTM, sulfur atoms remain, and thus these chemical groups only serve as close substitutes for true PTMs.

The last decade has brought about tremendous advances in PTM synthesis from a completely different angle: genetic code expansion of living organisms for ribosomal incorporation of modified amino acids. To accomplish this, an orthogonal translation system (OTS) is developed to aminoacylate a tRNA with a noncanonical amino acid (ncAA), which is then added site specifically to a growing peptide chain (Fig. 1d)^{7,8}. This technique most often utilizes the rare UAG stop codon and has been used in bacteria, yeast, mammalian cells, and other model organisms. Natural PTMs such as phosphoserine (our laboratory's specialty) can now be co-translationally inserted into proteins, obviating the need for post-translational manipulation and allowing a high level of control over the location of modification(s)⁹. These techniques can also be applied to designer ncAAs used as a reactive handle for PTM addition, such as the site-specific insertion of a click-reactive residue for ubiquitin conjugation¹⁰, an approach that could certainly be extended to other PTMs. One problem with OTSs is protein purity; natural amino acids may still be incorporated (albeit at low frequencies) at the targeted site due to competition from near-cognate tRNAs (i.e. anticodon mismatch) or promiscuous endogenous aminoacyl tRNA synthetases mischarging tRNA^{UAG}, although the majority of OTSs exhibit very low cross-reactivity with natural amino acids. In our own hands, we have seen that ncAA incorporation level is position-dependent, and some recombinant modified proteins are much more difficult to synthesize than others. Moreover, bulky or charged ncAA side chains may not be amenable to ribosomal synthesis; further engineering of the ribosome may be helpful to improve the accommodation of certain PTM classes. Difficulties may also arise for modified amino acids that cannot be distinguished from their unmodified counterparts even after attempts in synthetase evolution. Regardless, given the general success of OTS methodology, more effort to expand the genetically encoded PTM landscape is warranted.

Another problem with OTS platforms is competition with the translational machinery at the ribosome, since all 64 sense and stop codons are already naturally assigned. The most common codon used for ncAA incorporation, UAG, is naturally a stop codon that triggers translational termination by release factor 1. Although most OTSs have been successfully developed and deployed in competition with release factor 1, UAG decoding remained ambiguous in all contexts. An elegant solution to this problem was recently developed wherein all 321 instances of the TAG stop codon in *E. coli* were replaced with TAA, resulting in the world's first genomically recoded organism¹¹. This genome editing maneuver allowed the deletion release factor 1 and permitted the subsequent reassignment of UAG to other ncAAs using an OTS, therefore turning a stop codon into a sense codon. Recoded *E. coli* paired with OTSs provide perhaps the simplest conduit for chemical biologists to enter the PTM sphere since the technology is firmly rooted in universal techniques of recombinant protein expression in *E. coli* that are already set up in most labs. The stage is set for developing a broader array of OTSs for physiologically relevant PTMs with an eye towards multi-site incorporation of different PTMs. The future of PTM chemical biology in this arena will require more recoding to reassign more codons as well as engineering new sets of OTS elements that can decipher various codons that are not cross-reactive^{12,13}.

Or when sounded as A

The best news for the chemical biology community is that much of the legwork has already been done: strategies to synthesize modified proteins have been established, and they are primed and ready for "applications". PTMs in general can act as important structural elements and greatly impact enzymatic function (Fig. 2a). For instance, phosphorylation within an activation loop can act as an on/off switch for protein kinases, and phosphoserine incorporation as an ncAA can yield constitutively active kinase preparations^{9,14}. Nonstandard amino acid incorporation can also be used to prepare recombinant modified proteins for crystallization. For example, a crystal structure of acetylated human cyclophilin A synthesized heterologously in *E. coli* has provided insight into the inhibitory nature of site-specific acetylation within this protein as well as this PTM's role in HIV infection¹⁵. While chemical ligation/semisynthetic methods have been used for decades for functional studies of diverse modified protein modifications from ubiquitylation to farnesylation, OTS-based techniques have not been widely used beyond initial proof-of-principle studies. In the near future, we will likely see a surge in use of genetically encoded PTMs to understand the biological function of naturally modified proteins.

Genetically encoded PTM detector elements have also recently been developed, such as methyllysine biosensors that can assess the methylation status of histones using a bioluminescent complementation system¹⁶. Our improved abilities to synthesize proteins containing PTMs hint at the concomitant development of improved biosensors for PTMs, which could find utility in basic biology studies or for diagnostic purposes. Pure preparations of modified proteins can serve as scaffolds for protein or aptamer binding, and OTS platforms will enable directed evolution experiments to develop new PTM-binding proteins (Fig. 2b). Dispatching these new biosensors will then allow for the facile detection of PTMs of interest in their native context in cell culture or in model organisms.

Lastly, the doors have swung wide open for drug discovery and other therapeutic assays using synthetically modified proteins as either direct targets or as substrates of targeted enzymes (Fig. 2c). Many disease-relevant cellular signaling states are associated with increased modification status of proteins. For example, deregulation of ubiquitylation systems is associated with many types of cancers and neurodegenerative diseases. Although once considered an undruggable system, there has been renewed interest recently in identifying small molecules to modulate protein ubiquitylation and degradation rates¹⁷. Chemical synthesis and synthetic biology approaches have been used to generate ubiquitylated proteins or modified ubiquitin^{18,19}, which can serve as important reagents for inhibitor screens. We believe that many classes of PTMs will be relevant for these types of studies, as aberrant signaling involving modified proteins is implicated or causative in many diseases. Additionally, the generation of protein crystal structures containing modifications should also greatly facilitate structure-based drug design targeting enzymes whose activity is regulated by PTMs.

As in neighbor and weigh

The “way” forward holds much promise for PTM studies that are supported by complete biosynthesis of modified proteins. Advancing our ability to make recombinant human proteins with physiologically relevant modifications for *in vitro* studies is often approached from the angle of humanizing non-human cells. This is already being actively addressed in terms of introducing new ncAAs to the genetic code of *E. coli* and yeast, allowing for the synthesis of proteins containing human-relevant PTMs. Progress has been made toward more extensively recoded organisms with multiple codons available for ncAA incorporation (Fig. 3a), and the development of organisms utilizing genomic nonstandard base pairs or orthogonal ribosomes that decode quadruplet codons provide entirely new codons to encode ncAAs^{7,13,20} although these organisms raise questions of how many ncAAs incorporated in the same protein will prove interesting or relevant. Still, differences in protein folding and compartmentalization between human cells and other organisms will continue to impede the production of many functional recombinant human proteins.

One of the most daunting challenges in PTM studies is the investigation of their function in native contexts, either in mammalian cell culture or in animals, in order to study the effects of PTMs on protein stability, localization, and intermolecular interactions. Genetic code expansion via genomic recoding in human cells or lab animals would require *de novo* genome synthesis and is purely speculative, yet the synthetic biology community is already making preparations²¹. Another large issue with current PTM studies is that true modifications are often targeted for removal by endogenous enzymes. In humans, PTMs exist in a state of flux, constantly being added and removed from proteins. This means that OTS-installed PTMs can be recognized and eliminated by enzymes in mammalian systems and other organisms. For this reason, some ncAA sidechains consisting of nonhydrolyzable analog versions of PTMs have been genetically encoded^{22,23}. Future strategies to improve modified protein synthesis in native contexts may rely on genome engineering to create new orthogonal synthesis compartments that sequester modified proteins from catabolic enzymes, or new cell lines that exhibit less hydrolytic activity (e.g. phosphatases), although the host biology may be too perturbed to be representative of wildtype systems. A system to

generate less immunogenic protein glycoforms in yeast was recently described, in which the endogenous glycosylation system was modified to yield less complex modifications, a principle that could certainly be applied to other PTMs and perhaps in mammalian cells²⁴. Another promising avenue is the use of photocaged ncAAs, which can be incorporated into proteins and activated with pulses of light. These may provide incredible spatiotemporal control over PTMs to observe their effects in native systems, and will hopefully be expanded to various types of modifications and further functionalized in mammalian systems.

Thus far, PTM investigations have been largely limited to individual sites and a handful of proteins. One recent forward-thinking study employed chemical ligation strategies to create libraries of modified histones to systematically evaluate the effects of various PTMs on nucleosome architecture²⁵. Future studies will certainly examine the effects of combinatorial modifications and systems-level questions about the implications of PTMs. The behavior of modifications can vary greatly depending on sequence context (i.e. other surrounding modifications), and proteins in complex mixtures can exhibit differences in functionality due to the number of molecules present or higher-order intermolecular interactions. The decreased cost of DNA synthesis will allow for the construction of large-scale gene libraries either directly encoding modified proteins via genetic code expansion or providing reactive side chains for conjugation of interesting moieties (Fig. 3b). We should soon be able to construct entire proteomes that contain diverse modifications at varying positions, and to determine how PTM perturbations effect complex signaling environments.

Finally, another unexplored avenue is the use of synthetic PTMs for experiments in evolutionary biology and molecular evolution. By using gene libraries encoding modified amino acids at various locations within the same protein, positional scanning or directed evolution experiments can explore new protein fitness landscapes to evaluate why and how proteins have evolved modifiability at precise positions (Fig. 3c). Alternatively, the introduction of PTMs into synthetic gene libraries can be used to evolve improved or novel functionality by virtue of more complex molecules, although these experiments are certainly not restricted to naturally occurring PTMs. Biotechnology companies optimizing heterologous natural product synthesis pathways in yeast or *E. coli* should consider including modified amino acids when they randomize targeted regions of enzymes to improve functionality. Depending on the capabilities of the host organism, incorporation of PTMs at new positions could also be used to develop either biorthogonal circuits or new circuits that interact with endogenous proteins.

With new tools in hand, scientists now have access to a vast array of chemical and biological techniques to install PTMs at positions of interest within recombinant proteins. The accessibility of these methods makes the identification and the synthesis of these modified proteins more tractable than ever. With these new letters added to our biological alphabet, we will finally be able to build fully functional humanized proteins and study their roles in healthy systems and pathogenesis.

Acknowledgments

K.W.B. acknowledges support from the NSF (DGE-1122492) and J.R. is supported by the NIH (R01GM125951; R01GM117230; P01DK017433; U54CA209992).

References

1. Hutchison CA, et al. *Science*. 2016; 351
2. Nielsen AAK, et al. *Science*. 2016; 352
3. Boyken SE, et al. *Science*. 2016; 352:680–687. [PubMed: 27151862]
4. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. *Nature Methods*. 2007; 4:231–237. [PubMed: 17293869]
5. Muir TW, Sondhi D, Cole PA. *Proceedings of the National Academy of Sciences*. 1998; 95:6705–6710.
6. Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL. *Chemistry & Biology*. 2001; 8:133–145. [PubMed: 11251288]
7. Chin JW. *Nature*. 2017; 550:53–60. [PubMed: 28980641]
8. Liu CC, Schultz PG. *Annual Review of Biochemistry*. 2010; 79:413–444.
9. Park HS, et al. *Science*. 2011; 333:1151–1154. [PubMed: 21868676]
10. Rösner D, Schneider T, Schneider D, Scheffner M, Marx A. *Nature Protocols*. 2015; 10:1594–1611. [PubMed: 26401915]
11. Lajoie MJ, et al. *Science*. 2013; 342:357–360. [PubMed: 24136966]
12. Italia JS, et al. *Nature Chemical Biology*. 2017; 13:446–450. [PubMed: 28192410]
13. Lajoie MJ, Söll D, Church GM. *Journal of Molecular Biology*. 2016; 428:1004–1021. [PubMed: 26348789]
14. Pirman NL, et al. *Nature Communications*. 2015; 6:8130.
15. Lammers M, Neumann H, Chin JW, James LC. *Nature Chemical Biology*. 2010; 6:331–337. [PubMed: 20364129]
16. Sekar TV, Foygel K, Gelovani JG, Paulmurugan R. *Analytical Chemistry*. 2015; 87:892–899. [PubMed: 25506787]
17. Huang X, Dixit VM. *Cell Research*. 2016; 26:484–498. [PubMed: 27002218]
18. Zhang M, et al. *Nature Methods*. 2017; 14:729–736. [PubMed: 28553966]
19. Ordureau A, et al. *Proceedings of the National Academy of Sciences*. 2015; 112:6637–6642.
20. Zhang Y, et al. *Nature*. 2017; 551:644–647. [PubMed: 29189780]
21. Boeke JD, et al. *Science*. 2016; 353:126–127. [PubMed: 27256881]
22. Rogerson DT, et al. *Nature Chemical Biology*. 2015; 11:496–503. [PubMed: 26030730]
23. Luo X, et al. *Nature Chemical Biology*. 2017; 13:845–849. [PubMed: 28604693]
24. Meuris L, et al. *Nature Biotechnology*. 2014; 32:485–489.
25. Dann GP, et al. *Nature*. 2017; 548:607–611. [PubMed: 28767641]

Figure 1.

Methods to synthesize recombinant proteins containing PTMs. **(a)** *In vitro* reaction using an enzyme capable of catalyzing the transfer of a PTM of interest to a substrate. The enzyme responsible for the PTM must be known and it must be able to be isolated in its active form, which is often difficult to achieve via heterologous expression. **(b)** Native chemical ligation. This technique allows generation of a protein containing a desired PTM, but normally the modification must occur near a protein terminus. **(c)** Conjugation using a reactive sidechain. This technique can be used to generate a protein with a nearly-identical modified sidechain by placing a cysteine at a desired position and performing an *in vitro* reaction. However, proteins with multiple accessible cysteines will be prone to off-target reactions. **(d)** Genetically encoded ribosomal incorporation using an orthogonal translation system to insert a pre-modified amino acid at position of interest. This system relies on the development of an orthogonal translation system to insert a noncanonical amino acid in response to the UAG codon, which may not be possible for certain large PTMs that cannot be accommodated by the translational machinery.

Figure 2.

Basic applications of modified recombinant proteins. **(a)** Assay to assess the effect of a PTM on enzyme activity. Kinetic comparisons of unmodified and various modified forms of the enzyme can be performed to elucidate the effect of PTMs at specific positions. **(b)** Identification of PTM-dependent interactors, or development of new PTM biosensors. Modified proteins can be used as scaffolds for identifying or evolving new PTM-dependent binding interactions. **(c)** Drug screens to identify inhibitors of PTM-activated proteins. Modified proteins are important reagents to identify candidate small molecule inhibitors that specifically target only modified forms of a protein of interest. Modified proteins are often implicated in disease and can serve as excellent therapeutic targets.

Figure 3.

Future of PTM studies. **(a)** Extensively recoded organisms with more unassigned codons will allow for multi-PTM insertion. By reducing the redundancy of the genetic code and creating more unassigned codons, this creates the opportunity to dramatically rewrite the genetic code such that numerous codons are given entirely new meanings, allowing for the study of the effects of multiple classes of PTMs that exist simultaneously within the same protein molecule. **(b)** Gene pools to allow synthesis of modified protein libraries to assess behavior within complex populations. PTM-omics studies will be made possible by allowing the incorporation of PTMs in large-scale recombinant human protein collections. High-throughput screens will determine which library members are most functionalized by the PTM of interest. **(c)** Synthetic gene libraries used to identify PTM sites that affect protein function, or multiple ncAA incorporation in organisms with multiple available codons to perform combinatorial studies. These types of gene libraries will allow us to further explore the relationship between protein structure and function.