MINI-REVIEW

Leveraging nature's biomolecular designs in next-generation protein sequencing reagent development

Jennifer Tullman[1] · John P. Marino[1] · Zvi Kelman[1,2]

Abstract

Next-generation approaches for protein sequencing are now emerging that could have the potential to revolutionize the field in proteomics. One such sequencing method involves fluorescence-based imaging of immobilized peptides in which the N-terminal amino acid of a polypeptide is readout sequentially by a series of fluorescently labeled biomolecules. When selectively bound to a specific N-terminal amino acid, the NAAB (*N*-terminal amino acid binder) affinity reagent identifies the amino acid through its associated fluorescence tag. A key technical challenge in implementing this fluoro-sequencing approach is the need to develop NAAB affinity reagents with the high affinity and selectivity for specific N-terminal amino acids required for this biotechnology application. One approach to develop such a NAAB affinity reagent is to leverage naturally occurring biomolecules that bind amino acids and/or peptides. Here, we describe several candidate biomolecules that could be considered for this purpose and discuss the potential for developability of each.

Key points
• *Next-generation sequencing methods are emerging that could revolutionize proteomics.*
• *Sequential readout of N-terminal amino acids by fluorescent-tagged affinity reagents.*
• *Native peptide/amino acid binders can be engineered into affinity reagents.*
• *Protein size and structure contribute to feasibility of reagent developability.*

Keywords Affinity reagents · Aptamer · Fluoro-sequencing · NAAB · N-terminal amino acid binder · Protein sequencing

Introduction

Next-generation protein sequencing technologies are now emerging that have the potential to be as disruptive to the field of proteomics as next-generation DNA and RNA sequencing were to genomics and transcriptomics, respectively (Restrepo-Perez et al. 2018). Although RNA transcript levels within a cell or tissue can provide some information about protein abundance, the information derived from this measurement is limited since there is no general correlation between transcript abundance and protein level in a cell (Macaulay et al. 2017). Moreover, protein levels within a cell vary widely with up to 10 orders of magnitude differences between low and high abundance species (Ho et al. 2018; Ponomarenko et al. 2016; Schmidt et al. 2016). To identify low-abundance proteins, current proteomics methods largely rely on high-resolution mass spectrometry of purified biomarkers and/or detection using antibodies against known protein targets (Liu et al. 2019). In principle, these methods can accurately detect protein species down to attomolar concentrations in complex proteomic mixtures (Tsedilin et al. 2015). While highly effective, proteomic methods using detection with antibodies or identification by mass spectrometry rely on *a priori* knowledge of the proteins present in the sample. Such approaches therefore cannot easily identify novel or unanticipated protein species. Moreover, all proteomic techniques must contend with challenges of identifying protein products resulting from variation in gene translation as well as the numerous post-translational modifications (PTM) of proteins. New proteomic methods that enable improve non-targeted,

✉ Zvi Kelman
   zkelman@umd.edu

1   Institute for Bioscience and Biotechnology Research (IBBR), National Institute of Standards and Technology (NIST) and the University of Maryland (UMD), 9600 Gudelsky Drive, Rockville, MD 20850, USA

2   Biomolecular Labeling Laboratory, NIST/IBBR, Rockville, MD, USA

identification of low abundance protein species, and more robust detection of PTMs therefore continue to be sought (Parker et al. 2010; Yu et al. 2009).

Currently, there are several different technologies being developed for next-generation protein sequencing that could address gaps in proteomic measurement technologies (Callahan et al. 2020; Robertson and Reiner 2018; Rodriques et al. 2019; Tang 2018). These technologies include nanopores, electron-tunneling spectroscopy, fluorescent imaging of protease-digested proteins, the multiplexing of high-sensitivity mass-spectroscopy, and fluorescent imaging of immobilized peptides. For the latter, readout based on selective chemical tagging of amino acids with fluorophores has been proposed (Swaminathan et al. 2018). Alternatively, a fluorescently tagged protein, nucleic acid, or other reagent that binds specifically to the N-terminal amino acid on a peptide could allow sequential readout of the sequence. This biomolecule, referred to as a NAAB (N-terminal amino acid binder) affinity reagent (or binding reagent), is used to specifically bind the N-terminal amino acid of peptides generated after enzymatic digestion of a protein or proteomic sample (derived from an organism, cell, or tissue) (Fig. 1a). The peptides are immobilized on a surface via the C-terminal amino acid, leaving the N-terminus exposed to be recognized by the NAAB affinity reagent (Fig. 1b). The exposed N-terminal amino acid of each immobilized peptide is then identified via the binding of one of the NAAB affinity reagents. Following binding and fluorescent identification, the terminal amino acid is removed using Edman chemistry, exposing the next amino acid on the polypeptide chain for subsequent identification.

Biomolecules currently in development as NAAB affinity reagents are based on both protein and nucleic acid molecules. Here, we summarize the different candidate molecules under consideration for development as NAAB affinity reagents and the advantages and limitations of each. The amenability of each candidate to be engineered for new functions will also be discussed.

Required properties of a NAAB affinity reagent

In order to function as an NAAB affinity reagent in a protein sequencing apparatus, a candidate biomolecule needs to meet several requirements (Table 1). Of these requirements, one of the most important characteristics of a reagent is its ability to bind with high specificity. One would like the reagents to bind specific amino acids at the N-terminus of a peptide while discriminating against all other amino acids, including the same amino acid found at an internal position within a given peptide (Fig. 1b). In some cases, however, binding a class of amino acids (e.g., aromatic residues) may also be useful. Ideally, the reagents should be developed with only the terminal amino acid contributing to binding. However, for some
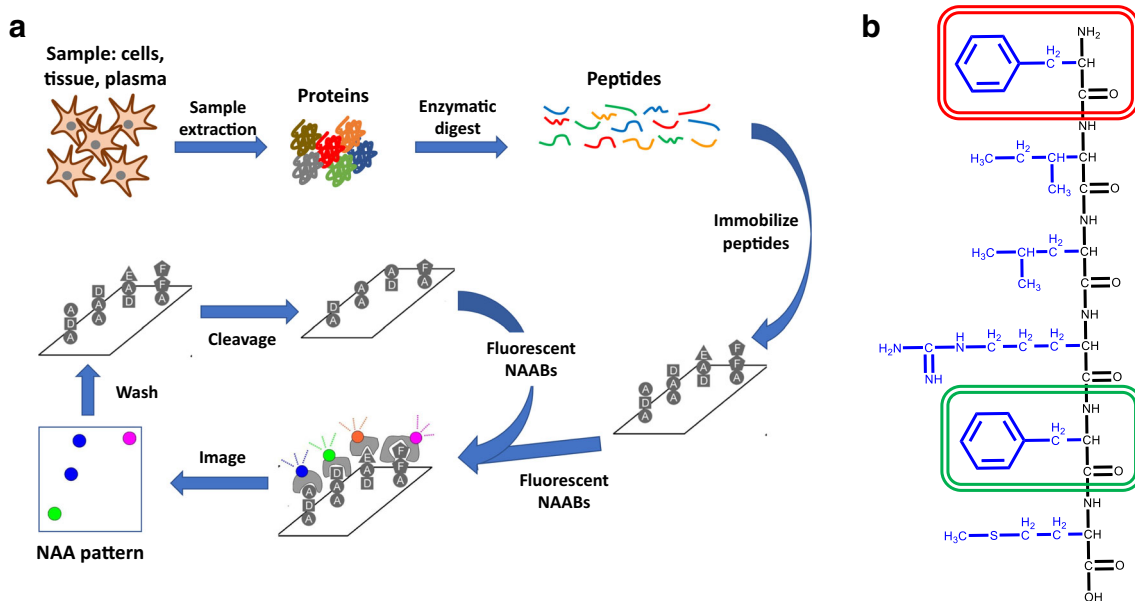

Fig. 1 Using NAABs in protein sequencing platform. a Schematic representation of the work flow of using NAABs for protein sequencing. The protein sample is extracted from the cell, tissue, etc., followed by enzymatic digestion with a specific protease. The peptides generated by the digestion are attached via the C-terminus onto a microscope slide. The fluorescently labeled NAABs are added to bind the N-terminal amino acid before an image is taken to identify the terminal residue. The NAABs are washed away form the peptides before the removal of the N-terminal residue via an Edman degradation. The process is repeated until the entire sequence of peptides is completed. b An example of the specificity required from a NAAB. In the peptide (Phe-Ile-Leu-Arg-Phe-Met), the NAAB needs to recognize the N-terminal Phe (red) but not the Phe in the middle of the peptide (green)

Table 1 Properties of candidates to be developed as affinity reagent

Affinity reagent	Size[1]	Oligomeric state	Target[2]	Specificity[3]
aaRS	*Various*	*Monomer* or **dimer**	Amino acid	*High*
PBP	*Small*	*Monomer*	Amino acid	*High*
Opp	**Large**	*Monomer*	Peptide	**Low**
Dpp	**Large**	*Monomer*	Peptide	**Low**
POT	**Large/very large**	*Monomer*	Peptide	**Low**
N-end rule				
ClpS	*Small*	*Monomer*	Peptide	*High*
UBR	**Very large**	*Monomer*	Peptide	*High*
Transferases	*Various*	**Various**	Peptide	*High*
Aminopeptidases	*Various*	*Monomer* or **multimer**	Peptide	*High*
Antibodies	**Very large**	*Monomer*	Peptide	*High*
Nanobodies	*Small*	*Monomer*	Peptide	*High*
RNA aptamer	*Various*	*Monomer*	Amino acid	*High*
DNA aptamer	*Various*	*Monomer*	Amino acid	*High*

Italicized are the desired properties of the affinity reagent and boldfaced are the undesirable properties

aaRS, aminoacyl-tRNA synthetases; *PBP*, periplasmic binding proteins; *Opp*, oligopeptide permease; *Dpp*, dipeptide permease; *POT*, proton-dependent oligopeptide transporter

[1] Small proteins are below 30 kDa and very large are above 100 kDa. When different size molecules or oligomeric states are in the group, they are marked in bold italics

[2] In general, molecules that recognize amino acids on peptide are preferred. However, in some cases, an affinity reagent can be developed from a molecule that recognizes free amino acids in solution

[3] High specificity denotes at least tenfold higher specificity in comparison to other targets

candidate reagent proteins, the residue at the penultimate positions of the target peptide (P2 and P3, the 2nd and 3rd amino acids from the N-terminus in the peptide) may influence the binding affinity. The molecular weight (size) and oligomeric state (complexity) of the NAAB affinity reagent should also be taken into consideration. The larger the protein, the more complicated it can be to engineer specific binding, and the greater the chance of non-specific binding to non-cognate peptides, the surface of the sequencing apparatus, and/or other NAAB affinity reagents. A large protein may also result in steric exclusion, preventing the reagent from binding a peptide that is attached to a surface. In addition, monomeric proteins are much preferred to limit the complexity of the system and make NAAB affinity reagent engineering simpler as many directed evolution approaches involve display of monomeric protein, and creating a DNA library of mutants can be simpler with smaller genes. In addition, if the reagent is an enzyme (e.g., peptidase), it may require mutations to knock out undesired enzymatic activity while maintaining binding affinity. The sequencing apparatus will also likely work in aqueous solution and thus, the reagents need to be soluble and compatible with other requirements of the sequencing workflow (e.g., pH and temperature ranges). Furthermore, it will be advantageous if the reagents that recognize different N-terminal residues can function under similar conditions such as buffer composition, temperature, and pH. This will allow their combination into "one pot" during the sequencing process. Some natural candidates for NAAB affinity reagents require cofactor for binding. This requirement for a cofactor may not be ideal in a "one pot" reaction as it may not be compatible with other binding reagents.

Protein-based NAAB affinity reagents

There are a number of protein families that recognize specific amino acids, either as free amino acids in solution or as a part of a polypeptide. These proteins provide potential leads for development of NAAB affinity reagents and could find use in their native form. In most cases, however, protein engineering of the native sequence will be required to improve their properties to meet the affinity and selectivity requirements for use in the next-generation sequencing application.

Aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases (aaRSs, also knowns as ARSs) are a family of enzymes that catalyze the attachment of the correct amino acid onto its cognate tRNA (Ibba and Soll 2000; Rajendran et al. 2018). In the first step of the reaction, the synthetase binds its cognate amino acid and, in an ATP-dependent manner, forms an aminoacyl-adenylate (enzyme-

amino acid-AMP complex), releasing a pyrophosphate in the process. In the next step, the adenylate-aaRS complex binds to the cognate tRNA, and the amino acid is transferred from the aa-AMP to the tRNA forming the "charged tRNA" which is used for protein synthesis.

aaRSs form a large and diverse family of proteins, each of which can correctly bind a specific amino acid corresponding to the specific tRNA (Rajendran et al. 2018). There are therefore at least 20 aaRSs in each organism that bind the 20 different canonical amino acids. Due to their crucial role in protein synthesis, the aaRS enzymes must have high fidelity in recognizing its cognate amino acid. Each mistake the enzyme makes will result in a mis-incorporated amino acid during protein synthesis. Thus, aaRSs might be considered good candidates to be developed as amino acid specific binding reagents. They exhibit the high specificity required for a NAAB affinity reagents. In addition, they are present in organisms from all three domains of life. The presence of the proteins in organisms from different environments may be beneficial for reagent development. For example, if a thermostable reagent is needed, the aaRS can be isolated from a thermophilic organism. Similarly, if a reagent that can tolerate low or high pH is required, the aaRS can be isolated from an acidophile or alkaliphile, respectively.

However, there are several aspects of aaRSs that make them less attractive as NAAB affinity reagents (Table 1). They evolved to recognize free amino acids with high specificity, but would need to be engineered to recognize an amino acid as the N-terminal residue of a peptide, and likely in the absence of the cognate tRNA. Engineering the aaRS to bind an amino acid in the context of residing at the N-terminus of peptide may affect the affinity and/or specificity of the enzyme. Another drawback is that some aaRS are large multidomain proteins, and many are active as homo- or hetero-multimers. On the other hand, several aaRSs have been engineered to bind non-canonical amino acids (ncAAs) (for example, see Crnkovic et al. 2019). This suggests that similar approaches may be used to engineer the proteins to bind terminal amino acids on peptides.

Periplasmic binding proteins

PBPs (periplasmic binding proteins) comprise a large number of diverse proteins found in bacteria and archaea, and are involved in the transport of small molecules and ions into the cytoplasm. PBPs play an essential role in providing essential nutrients such as vitamins, lipids, metal ions, peptides, and amino acids to the cell (Fehr et al. 2004; Sandhu et al. 2017). The proteins are monomeric and consist of two globular domains connected by a hinge region, with a ligand-binding site located in a cleft formed between the two domains. These proteins adopt two different conformations related to ligand binding: an open, ligand free structure, and a closed form

adopted upon ligand binding which involves a large movement of the hinge region between the two domains. The ligand-mediated conformational change is an attractive feature of PBPs for engineering and has been altered to develop proteins with new functions as biosensors, allosteric control elements, and more (for example, see Dwyer and Hellinga 2004; Ko et al. 2017; Ribeiro et al. 2019).

In several reported cases, it has been shown that PBPs form complexes with nanomolar affinities and that the proteins can be engineered for new functions (e.g., biosensor (Guntas et al. 2005; Ko et al. 2017; Tullman et al. 2016)). Moreover, several PBPs found in nature have evolved to bind free amino acids including Arg, His, Ile, Leu, Pro, and Val (for example, see Ausili et al. 2013; Chu et al. 2013; Fulyani et al. 2013; Magnusson et al. 2004). As such, PBPs represent another good set of candidates for development as NAAB affinity reagents. However, in nature, the PBPs evolved to bind free amino acids (Table 1). Similar to aaRSs, it is not clear if PBPs can be engineered to bind an amino acid at the N-terminal end of a peptide without losing specificity and/or affinity. The carboxyl end of the amino acid often contributes to ligand binding based on analysis of the crystal structures of apo PBPs as well as PBPs in complex with free amino acid ligands (for example, see PDB IDs 4G4P, 4KPT, 4KR5, 4KQP (Fulyani et al. 2013), 4PRS, 4PRH (Ruggiero et al. 2014), and 1WDN (Sun et al. 1998)). On the other hand, the wealth of structures available for PBPs and PBP-ligand complexes provides valuable information for guiding structure-based design. In addition, PBP proteins can be found in bacteria growing in different environments and thus, if needed, PBPs that can tolerate extreme conditions can also be isolated (for example, high temperature (Ausili et al. 2013)). Although the concerns with developing the PBPs as binding reagents are similar to those with aaRSs, an advantage of the PBPs over aaRS proteins is that this family of proteins is often active as monomers and is relatively small in size.

Peptide transporters

The proteins described above, aaRSs and PBPs, evolved to bind free amino acids in solution and not in the context of a polypeptide chain. Thus, they may not be amenable to be developed as NAAB affinity reagents where the amino acid to be recognized is located at a peptide amino terminus. Looking further into ideal protein families for NAAB affinity reagent development, one might first ask, which proteins naturally bind peptides? In this respect, there are three main systems involved in peptide transportation across the cell membrane. Although these proteins often bind peptides without regard to amino acid specificity, they may be good starting points for the engineering of binding reagents.

Oligopeptide permease systems

Oligopeptide transport systems from bacteria are part of the large family of ABC (ATP-binding cassette) transporters (Maqbool et al. 2015; Monnet 2003; Tam and Saier Jr. 1993; Tanaka et al. 2018). ABC transporters are multiple subunit complexes that include an extracellular protein(s) that capture the substrate, transmembrane protein(s), and membrane-associated proteins. One or more subunit contains ATPase activity, as the translocation of the cargo across the membrane is fueled by ATP hydrolysis.

In bacteria, one of the oligopeptide transport systems, Opp (oligopeptide permease), is composed of five subunits: OppABCDF (Ames 1986; Monnet 2003). The OppA protein captures the peptides from the external medium, two transmembrane proteins, OppB and OppC, form the pore needed for peptide translocation through the membrane, and two membrane-associated cytoplasmic ATPase proteins, and OppD and OppF, are members of the AAA+ (ATPases associated with diverse cellular activities) family of ATPases (Erzberger and Berger 2006; Ogura and Wilkinson 2001). The main difference between the Opp system in gram-negative and gram-positive bacteria is that the OppA in gram-negative bacteria is periplasmic, and in gram-positive, the protein is associated with the membrane via a NH_2-terminal lipomodification (Monnet 2003).

OppA proteins from several organisms have been studied, and they can bind peptides of various lengths from dipeptides to peptides several dozen amino acids in length (Detmers et al. 2000). OppA's ability to bind a large variety of peptides and its presence in bacteria and archaea (Gogliettino et al. 2010) suggest that OppA proteins could be good candidates for development as NAAB affinity reagents. However, several aspects of this protein make it less attractive. Although the protein binds different sizes of peptides, it shows very little amino acid specificity, preferring positively charged residues in positions P3 and P4 (Monnet 2003). Proteins that show some preference in amino acid binding may be a better starting point for protein engineering. In addition, the OppA proteins are fairly large (around 60 kDa) (Monnet 2003), making them less attractive as binding reagents and large for available platforms for protein engineering through directed evolution.

Dipeptide permease systems

In addition to the Opp systems, bacteria contain another ABC-type transporter with similar organization (Weinberg and Maier 2007). The Dpp (dipeptide permease) also contains five proteins (DppABCDF) and participates in di- and tri-peptide transport into the cell. Similar to OppA, the DppA protein is a periplasmic peptide binding protein responsible for capturing of di- and tri-nucleotides in the periplasmic space (Olson et al. 1991). Although di- or tri-peptides are too small to function in a sequencer apparatus, it is possible that the DppA protein can be engineered to bind longer peptides. However, as specificity is very important for a NAAB affinity reagent and DppA does not exhibit sequence specificity, it would also have to be engineered to meet this requirement. The need for intensive engineering makes DppA less likely to be an optimal candidate for development as a NAAB affinity reagent. However, only a small number of bacterial DppA proteins have been characterized. A DppA protein with amino acid specificity may yet be identified, which could allow for a better starting point for protein engineering.

Proton coupled peptide transporters

The POT (proton-dependent oligopeptide transporter) family of proteins (also called PTR (peptide transport) family) is found in all three domains of life and belongs to the MFS (major facilitator superfamily) family of transporters (Newstead 2015; Reddy et al. 2012). Although these are large proteins (50–110 kDa; 500–1000 amino acids), POT appears to be mainly involved in the intake of small peptides including di- and tri-peptides. The proteins are monomeric and contain 12–18 aa transmembrane domains that derive their energy of transport from the import of protons.

POT proteins evolved to bind peptides and thus may appear to be good candidates. However, they have several characteristics that may challenge their development as NAAB affinity reagents (Table 1). They bind only short, di- and tri-peptides, yet a robust peptide sequencer requires binding of longer peptides. In addition, the POTs bind peptides without sequence specificity; the protein sequencing application requires specific binding to the N-terminal amino acid. Furthermore, the POT proteins are large. Another major problem in the development of POT as binding reagents is that they are membrane-bound proteins. As discussed above, for the development of protein sequencers, soluble, well-folded proteins are required. Most membrane proteins are not well folded and/or functional when purified without the membrane, even when detergents are used. Therefore, although POT proteins have evolved to bind peptides, these proteins may be considered less-desirable candidates to be developed as NAAB affinity reagents.

The N-end rule

The N-end rule is a highly regulated protein degradation pathway and is conserved in all life forms (Sriram et al. 2011; Tasaki et al. 2012; Varshavsky 1996; Varshavsky 2019). This pathway determines the half-life of a protein based on the identity of its N-terminal residue. Therefore, the N-terminal residues can be classified into two broad groups: stabilizing or destabilizing. Proteins bearing an N-terminal destabilizing residue (N-degron) are recognized by a specific N-end rule recognition component (termed an N-recognin) and then delivered to a protease for degradation. For

example, in *Escherichia coli*, while positively charged and some aliphatic and aromatic residues on the N-terminus of a protein may have half-lives of few minutes, other amino acids at the N-terminus may result in proteins that are stable for several hours.

In eukarya and archaea, proteins destined for intracellular proteolysis are tagged for degradation by post-translational modification with a small protein: ubiquitin (Ub) in eukarya (Komander and Rape 2012; Swatek and Komander 2016), and small archaeal modifier proteins (SAMP) in archaea (Maupin-Furlow 2013). In both cases, the modifying proteins are covalently attached to Lys on the substrate, and the modified substrate is degraded by the proteasome. Most bacteria do not use a similar system. To date, the only exception is *Mycobacterium tuberculosis*, in which a prokaryotic ubiquitin-like protein (Pup) was identified and shown to be functionally analogous to Ub and SAMPs by providing the signal for degradation by the proteasome (Bode and Darwin 2014).

Many proteins participating in the N-end rule of bacteria and eukarya are good candidates to be developed as NAAB affinity reagents. Several of those are described below. A comprehensive list of candidate enzymes can be found in Varshavsky (2019).

ClpS

Most bacteria do not utilize protein modification as a signal for protein degradation via the N-end rule. The bacterial pathway contains two components, the ClpAP protease (Olivares et al. 2018) and the adaptor protein ClpS (Dougan et al. 2002). The ClpS protein binds an N-degron and delivers the protein to ClpAP for degradation. ClpS was shown to specifically recognize N-terminal Phe, Trp, Tyr, and Leu amino acids and this specificity makes ClpS a particularly good candidate to be developed as a binding reagent (Tullman et al. 2019) (Table 1). The protein displays two of the properties of a good binding reagent, the ability to bind peptides, and some specificity for specific N-terminal amino acids. However, one of the drawbacks of using ClpS is the effect of the amino acid in the P2 and P3 positions on binding affinity and selectivity. It was shown that different residues at the P2 and P3 positions affect ClpS binding to the N-terminal amino acid (Stein et al. 2016). Engineering of ClpS can reduce the effect of these penultimate amino acids and enhance its thermostability (Tullman et al. 2020).

UBR

In eukarya, the protein degradation by the ubiquitin system controls the levels of many intracellular proteins. A substrate of the Ub system is conjugated to Ub through the action of three enzymes: E1, E2, and E3. In those systems, the N-degrons are recognize by the UBR box domain of the E3 ubiquitin ligase, called N-recognins. In yeast, the N-recognin protein is expressed by the *ubr1* gene. The UBR1 contains

three types of substrate-binding sites. One site binds basic N-terminal residues (Arg, Lys, and His) on the N-terminus of proteins or peptides (Choi et al. 2010; Tasaki et al. 2009). The second binding site recognizes bulky hydrophobic residues (Leu, Ile, Phe, Tyr, and Trp) and the third binding site recognizes internal degron and not N-terminal ones. It was found that UBR box can also bind methylated peptides (Munoz-Escobar et al. 2017). The ability of the UBRs to bind proteins and peptides with some specificity for specific amino acid makes them potential candidates for development as NAAB affinity reagents. However, one of the drawbacks is their large size, being about 200 kDa. This size will limit the ability to engineer the protein for enhance specificity. On the other hand, there are two different sites that are involved in binding to hydrophobic and basic peptides. Thus, it may be possible to separate the two regions and generate smaller protein domains that are more amenable to protein engineering techniques. It is also possible that such smaller domains, similar in size to ClpS, may be sufficient for peptide binding (Lupas and Koretke 2003).

Transferases

Several transferases were also shown to participate in the N-end rule process. The L/F transferase (leucyl/phenylalanyl-tRNA-protein transferase) was shown to catalyze the transfer of Leu and Phe destabilizing residues from aminoacyl-tRNAs to the amino terminus of acceptor proteins containing an N-terminal Arg, Lys, and Met, thus tagging them for degradation (Ninnis et al. 2009; Watanabe et al. 2007). Another example of a transferase that can be developed as a NAAB affinity reagent is Ate1 proteins. In eukarya, the Ate1 gene encodes for an arginyl-transferases which transfer Arg to N-terminal Asp, Glu, and cysteine residues (Hu et al. 2005). The N-terminal Arg is recognized by the ubiquitin ligases leading to protein degradation. BPT is another example. This protein is an aminoacyl-transferase identified in *Vibrio vulnificus*. Although its sequence is similar to the eukaryotic arginyl-transferases, it conjugates Leu to Asp and Glu (Graciet et al. 2006). Another aminoacyl-transferase, termed ATEL1, was identified in *Plasmodium falciparum*. This protein shows sequence similarities to the bacterial L/F transferase but with similar specificity to the Ate1 protein (Graciet et al. 2006).

These peptidyl transferases could be good candidates for NAAB affinity reagent if the enzymatic activity can be eliminated via protein engineering. Additionally, it may be possible to eliminate the need for a tRNA cofactor but retain binding to the target peptide (Abramochkin and Shrader 1995).

Aminopeptidases

Aminopeptidases are a group of exopeptidases that catalyze the cleavage of the N-terminal amino acid from proteins or

peptides (Ferroa et al. 2014; Gonzales and Robert-Baudouy 1996; Sanderink et al. 1988). They constitute a large group of proteins that are present in all three domains of life. Aminopeptidases can be found in the cytoplasm, embedded in the membranes, anchored to the cell membrane, or secreted out of the cell. Aminopeptidases play a role in several physiological processes. Some participate in the catabolism of exogenously supplied peptides, while others are required for protein turnover in vivo. In addition, some aminopeptidases are involved in specific functions. For example, the bacterial and archaeal methionine-aminopeptidases are responsible for the cleavage of N-terminal methionine from newly synthesized proteins.

Aminopeptidases vary in size, specificity, activity, and biophysical property (Ferroa et al. 2014; Gonzales and Robert-Baudouy 1996; Sanderink et al. 1988). While some are active as monomers, others are multimeric. There are also various cofactors needed for activities (for example, Zn-metalloenzyme, Ca-metalloenzyme, cysteine-enzyme, serine-enzyme). Another difference between the different aminopeptidases is specificity. While many aminopeptidases have broad specificity and degrade any N-terminal amino acid, some are amino acid–specific and can only catalyze the removal of a specific amino acid such as Ala, Pro, Gly, and Met (Gonzales and Robert-Baudouy 1996).

The fact that aminopeptidases can recognize a specific amino acid makes them good candidates for development as a binding reagent, as these enzymes have also already evolved to bind peptides. And, although the enzyme will cleave the terminal amino acid, it has been shown that, at least in some cases, mutating the active site residue can prevent cleavage of the bound peptide while not affecting enzyme specificity (Thompson et al. 2003).

While the ability of aminopeptidases to bind peptides with a specific N-terminal amino acid could make them good candidates for development as NAAB affinity reagents, other aspects of the protein need to be considered including oligomeric structure, size, and the cofactor required for binding. In addition, in some cases, the residues in the P2, P3, and P4 have been shown to influence protein binding to the terminal amino acid (Xiao et al. 2010). As already mentioned, any effect of the downstream amino acid on binding is not desirable for a NAAB affinity reagent. It is possible that protein engineering may enable the alteration of the enzymes and reduce the effect of the P2–P4 positions on peptide binding.

Antibodies and nanobodies

Antibodies and nanobodies (also known as single-domain antibodies) are known to bind their ligand with high affinity and specificity. Thus, one would assume that they may be good candidates to be developed as NAAB affinity reagents. However, antibodies are large, averaging 150 kDa, tetrameric

complex with multiple domains which could significantly limit use as NAAB affinity reagents for protein sequencing. Nanobodies, on the other hand, are small, about 15 kDa, and composed of only one domain. Nevertheless, the ligands recognized by antibodies and nanobodies are fairly large and usually composed of several amino acids and often are not consecutive on the polypeptide chain. There are, however, exceptions, for example, the anti-phosphotyrosine antibodies that recognize phosphorylated tyrosine within a polypeptide chain but not specifically at the N-terminus of a protein/peptide (for example, Cobaugh et al. 2008; Mandell 2003). Although the epitopes are usually composed of several amino acids, there are examples of shorter single amino acid epitopes (for example, an antibody that can recognize di-Gly peptide; Wagner et al. 2011; Yoshida et al. 2015). Nevertheless, due to their limitations (Table 1), antibodies and nanobodies are not likely suitable for development into NAAB affinity reagents.

Nucleic acid–based NAAB affinity reagents

RNA aptamer

Riboswitches are segments of bacterial mRNA usually found in the 5′-untranslated regions (5′-UTRs) that play a regulatory role in the translation of proteins involved in metabolism and transport of metabolites (Breaker 2018; Sherwood and Henkin 2016). The small metabolite binds specifically to the riboswitch segment of the mRNA and directly regulates protein expression in response to the metabolite cellular concentrations. In general, riboswitches consist of two parts, the aptamer domain that specifically binds the small metabolite and a second domain that is involved in regulating gene expression (Breaker 2018; Sherwood and Henkin 2016). Several mechanisms for riboswitch actions have been studied including the formation of termination signals when the level of the metabolite is high, thus resulting in an aborted transcription. Another mechanism is the formation of a RNA structure that masks the ribosome binding site on the mRNA, thus preventing translation when the metabolite levels are high. It was also found that in some cases the riboswitch also functions as a ribozyme that cleaves itself (i.e., the mRNA) when its cognate metabolite levels are high.

Several riboswitches have specifically evolved to bind amino acids. Those include the Gln riboswitch (Weinberg et al. 2010), Gly riboswitch (Mandal et al. 2004), and Lys riboswitch (Mandal et al. 2003). Although the naturally occurring riboswitches evolved to bind only an amino acid and not a terminal amino acid on a peptide, they are potential candidates for further engineering to make them useful reagents. In addition, only the aptamer part of the riboswitch is needed to make a binding reagent and thus may be more amenable to modifications without affecting affinity to the

specific amino acid. To date, naturally occurring riboswitches have been identified against only a few amino acids. However, studies have shown that aptamers that recognize other amino acids can be developed including those which recognize Val (Majerfeld and Yarus 1994), Ile (Legiewicz and Yarus 2005), Tyr (Mannironi et al. 2000), and other amino acids (McKeague and Derosa 2012). There are many selection protocols designed to isolate aptamers that bind specific targets from a large random sequence pool (Blind and Blank 2015). Based on the amino acid aptamers that have been studied, one would expect that with the right screening scheme aptamers that recognize other amino acids and/or peptides could be identified.

However, one of the major drawbacks of using RNA as a binding reagent is its stability. RNA is very sensitive to nucleases and other degradation processes (for example, exposure to basic pH). Thus, RNA may not be stable during the sequencing process and thus may not be a good candidate for binding reagent. The incorporation of modified nucleotides (such as 2′-O-methyl rNTPs), however, may be used to improve aptamer stability.

Other possible aptamers

DNA could also be developed as NAAB affinity reagents. Several DNA aptamers have been developed to bind small molecule with high affinity and specificity (McKeague and Derosa 2012) and thus may also be developed to bind specific amino acids. Although DNA is more stable and easier to manipulate, its folding is not as versatile as RNA and thus, it may not be possible to design a DNA molecule to bind with the required specificity and affinity to N-terminal amino acids.

One can also consider using non-natural, nucleotide-based, NAAB affinity reagents such as PNA (peptide nucleic acid) and LNA (locked nucleic acid) (Jepsen et al. 2004; Siddiquee et al. 2015).

Concluding remarks

Although, currently, there is no working prototype of a next-generation protein sequencer, the field is moving forward at a rapid pace, and peptide fingerprinting is already a reality (Swaminathan et al. 2018; van Ginkel et al. 2018). For fluoro-sequencing based on sequential identification of the N-terminal amino acid of a polypeptide, one of the main challenges is the development of NAAB affinity reagents that bind N-terminal amino acids of a peptide or protein with high specificity and affinity. Here, we have described the properties required for NAAB affinity reagents and suggested some native systems that could form the basis for candidates to be developed as reagents. However, new candidates with the required properties are still to be discovered (for example,

see the glycine-specific N-end rule pathway (Timms et al. 2019)). A combination of naturally evolving biomolecules, together with engineering approaches to improve their utility as NAAB affinity reagents, could provide the necessary tools to address this critical need to advance this approach to next-generation protein sequencing.

Compliance with ethical standards

References

Abramochkin G, Shrader TE (1995) The leucyl/phenylalanyl-tRNA-protein transferase. Overexpression and characterization of substrate recognition, domain structure, and secondary structure. J Biol Chem 270(35):20621–20628. https://doi.org/10.1074/jbc.270.35.20621

Ames GF (1986) Bacterial periplasmic transport systems: structure, mechanism, and evolution. Annu Rev Biochem 55:397–425. https://doi.org/10.1146/annurev.bi.55.070186.002145

Ausili A, Staiano M, Dattelbaum J, Varriale A, Capo A, D'Auria S (2013) Periplasmic binding proteins in thermophiles: characterization and potential application of an arginine-binding protein from *Thermotoga maritima*: a brief thermo-story. Life (Basel) 3(1):149–160. https://doi.org/10.3390/life3010149

Blind M, Blank M (2015) Aptamer selection technology and recent advances. Mol Ther Nucleic Acids 4:e223. https://doi.org/10.1038/mtna.2014.74

Bode NJ, Darwin KH (2014) The Pup-proteasome system of mycobacteria. Microbiol Spectr 2(5). https://doi.org/10.1128/microbiolspec.MGM2-0008-2013

Breaker RR (2018) Riboswitches and translation control. Cold Spring Harb Perspect Biol 10(11). https://doi.org/10.1101/cshperspect.a032797

Callahan N, Tullman J, Kelman Z, Marino J (2020) Strategies for development of a next-generation protein sequencing platform. Trends Biochem Sci 45(1):76–89. https://doi.org/10.1016/j.tibs.2019.09.005

Choi WS, Jeong BC, Joo YJ, Lee MR, Kim J, Eck MJ, Song HK (2010) Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nat Struct Mol Biol 17(10):1175–1181. https://doi.org/10.1038/nsmb.1907

Chu BC, DeWolf T, Vogel HJ (2013) Role of the two structural domains from the periplasmic *Escherichia coli* histidine-binding protein HisJ. J Biol Chem 288(44):31409–31422. https://doi.org/10.1074/jbc.M113.490441

Cobaugh CW, Almagro JC, Pogson M, Iverson B, Georgiou G (2008) Synthetic antibody libraries focused towards peptide ligands. J Mol Biol 378(3):622–633. https://doi.org/10.1016/j.jmb.2008.02.037

Crnkovic A, Vargas-Rodriguez O, Soll D (2019) Plasticity and constraints of tRNA aminoacylation define directed evolution of aminoacyl-tRNA synthetases. Int J Mol Sci 20(9). https://doi.org/10.3390/ijms20092294

Detmers FJ, Lanfermeijer FC, Abele R, Jack RW, Tampe R, Konings WN, Poolman B (2000) Combinatorial peptide libraries reveal the ligand-binding mechanism of the oligopeptide receptor OppA of Lactococcus lactis. Proc Natl Acad Sci U S A 97(23):12487–12492. https://doi.org/10.1073/pnas.220308797

Dougan DA, Reid BG, Horwich AL, Bukau B (2002) ClpS, a substrate modulator of the ClpAP machine. Mol Cell 9(3):673–683. https://doi.org/10.1016/s1097-2765(02)00485-9

Dwyer MA, Hellinga HW (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. Curr Opin Struct Biol 14(4):495–504. https://doi.org/10.1016/j.sbi.2004.07.004

Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct 35:93–114

Fehr M, Ehrhardt DW, Lalonde S, Frommer WB (2004) Minimally invasive dynamic imaging of ions and metabolites in living cells. Curr Opin Plant Biol 7(3):345–351. https://doi.org/10.1016/j.pbi.2004.03.015

Ferroa ES, Riolib V, Castro LM, Fricker LD (2014) Intracellular peptides: from discovery to function. EuPA Open Proteomics 3:143–151abstr

Fulyani F, Schuurman-Wolters GK, Zagar AV, Guskov A, Slotboom DJ, Poolman B (2013) Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria. Structure 21(10):1879–1888. https://doi.org/10.1016/j.str.2013.07.020

Gogliettino M, Balestrieri M, Pocsfalvi G, Fiume I, Natale L, Rossi M, Palmieri G (2010) A highly selective oligopeptide binding protein from the archaeon Sulfolobus solfataricus. J Bacteriol 192(12):3123–3131. https://doi.org/10.1128/JB.01414-09

Gonzales T, Robert-Baudouy J (1996) Bacterial aminopeptidases: properties and functions. FEMS Microbiol Rev 18(4):319–344. https://doi.org/10.1111/j.1574-6976.1996.tb00247.x

Graciet E, Hu RG, Piatkov K, Rhee JH, Schwarz EM, Varshavsky A (2006) Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proc Natl Acad Sci U S A 103(9):3078–3083. https://doi.org/10.1073/pnas.0511224103

Guntas G, Mansell TJ, Kim JR, Ostermeier M (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc Natl Acad Sci U S A 102(32):11224–11229. https://doi.org/10.1073/pnas.0502673102

Ho B, Baryshnikova A, Brown GW (2018) Unification of protein abundance datasets yields a quantitative Saccharomyces cerevisiae proteome. Cell Syst 6(2):192–205 e3. https://doi.org/10.1016/j.cels.2017.12.004

Hu RG, Sheng J, Qi X, Xu Z, Takahashi TT, Varshavsky A (2005) The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437(7061):981–986. https://doi.org/10.1038/nature04027

Ibba M, Soll D (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem 69:617–650. https://doi.org/10.1146/annurev.biochem.69.1.617

Jepsen JS, Sorensen MD, Wengel J (2004) Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology. Oligonucleotides 14(2):130–146. https://doi.org/10.1089/1545457041526317

Ko W, Kim S, Lee HS (2017) Engineering a periplasmic binding protein for amino acid sensors with improved binding properties. Org Biomol Chem 15(41):8761–8769. https://doi.org/10.1039/c7ob02165h

Komander D, Rape M (2012) The ubiquitin code. Annu Rev Biochem 81:203–229. https://doi.org/10.1146/annurev-biochem-060310-170328

Legiewicz M, Yarus M (2005) A more complex isoleucine aptamer with a cognate triplet. J Biol Chem 280(20):19815–19822. https://doi.org/10.1074/jbc.M502329200

Liu Y, Chen X, Zhang Y, Liu J (2019) Advancing single-cell proteomics and metabolomics with microfluidic technologies. Analyst 144(3):846–858. https://doi.org/10.1039/c8an01503a

Lupas AN, Koretke KK (2003) Bioinformatic analysis of ClpS, a protein module involved in prokaryotic and eukaryotic protein degradation. J Struct Biol 141(1):77–83. https://doi.org/10.1016/s1047-8477(02)00582-8

Macaulay IC, Ponting CP, Voet T (2017) Single-cell multiomics: multiple measurements from single cells. Trends Genet 33(2):155–168. https://doi.org/10.1016/j.tig.2016.12.003

Magnusson U, Salopek-Sondi B, Luck LA, Mowbray SL (2004) X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity. J Biol Chem 279(10):8747–8752. https://doi.org/10.1074/jbc.M311890200

Majerfeld I, Yarus M (1994) An RNA pocket for an aliphatic hydrophobe. Nat Struct Biol 1(5):287–292. https://doi.org/10.1038/nsb0594-287

Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell 113(5):577–586. https://doi.org/10.1016/s0092-8674(03)00391-x

Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science 306(5694):275–279. https://doi.org/10.1126/science.1100829

Mandell JW (2003) Phosphorylation state-specific antibodies: applications in investigative and diagnostic pathology. Am J Pathol 163(5):1687–1698. https://doi.org/10.1016/S0002-9440(10)63525-0

Mannironi C, Scerch C, Fruscoloni P, Tocchini-Valentini GP (2000) Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. RNA 6(4):520–527. https://doi.org/10.1017/s1355838200991763

Maqbool A, Horler RS, Muller A, Wilkinson AJ, Wilson KS, Thomas GH (2015) The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. Biochem Soc Trans 43(5):1011–1017. https://doi.org/10.1042/BST20150135

Maupin-Furlow JA (2013) Ubiquitin-like proteins and their roles in archaea. Trends Microbiol 21(1):31–38. https://doi.org/10.1016/j.tim.2012.09.006

McKeague M, Derosa MC (2012) Challenges and opportunities for small molecule aptamer development. J Nucleic Acids 2012:748913–748920. https://doi.org/10.1155/2012/748913

Monnet V (2003) Bacterial oligopeptide-binding proteins. Cell Mol Life Sci 60(10):2100–2114. https://doi.org/10.1007/s00018-003-3054-3

Munoz-Escobar J, Matta-Camacho E, Cho C, Kozlov G, Gehring K (2017) Bound waters mediate binding of diverse substrates to a ubiquitin ligase. Structure 25(5):719–729 e3. https://doi.org/10.1016/j.str.2017.03.004

Newstead S (2015) Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. Biochim Biophys Acta 1850(3):488–499. https://doi.org/10.1016/j.bbagen.2014.05.011

Ninnis RL, Spall SK, Talbo GH, Truscott KN, Dougan DA (2009) Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. EMBO J 28(12):1732–1744. https://doi.org/10.1038/emboj.2009.134

Ogura T, Wilkinson AJ (2001) AAA+ superfamily ATPases: common structure–diverse function. Genes Cells 6(7):575–597

Olivares AO, Baker TA, Sauer RT (2018) Mechanical protein unfolding and degradation. Annu Rev Physiol 80:413–429. https://doi.org/10.1146/annurev-physiol-021317-121303

Olson ER, Dunyak DS, Jurss LM, Poorman RA (1991) Identification and characterization of dppA, an *Escherichia coli* gene encoding a periplasmic dipeptide transport protein. J Bacteriol 173(1):234–244. https://doi.org/10.1128/jb.173.1.234-244.1991

Parker CE, Mocanu V, Mocanu M, Dicheva N, Warren MR (2010) Mass spectrometry for post-translational modifications. In: Alzate O (ed) Neuroproteomics. Frontiers in Neuroscience, Boca Raton, pp 93–108

Ponomarenko EA, Poverennaya EV, Ilgisonis EV, Pyatnitskiy MA, Kopylov AT, Zgoda VG, Lisitsa AV, Archakov AI (2016) The size of the human proteome: the width and depth. Int J Anal Chem 2016:7436849–7436846. https://doi.org/10.1155/2016/7436849

Rajendran V, Kalita P, Shukla H, Kumar A, Tripathi T (2018) Aminoacyl-tRNA synthetases: structure, function, and drug discovery. Int J Biol Macromol 111:400–414. https://doi.org/10.1016/j.ijbiomac.2017.12.157

Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH Jr (2012) The major facilitator superfamily (MFS) revisited. FEBS J 279(11):2022–2035. https://doi.org/10.1111/j.1742-4658.2012.08588.x

Restrepo-Perez L, Joo C, Dekker C (2018) Paving the way to single-molecule protein sequencing. Nat Nanotechnol 13(9):786–796. https://doi.org/10.1038/s41565-018-0236-6

Ribeiro LF, Amarelle V, Ribeiro LFC, Guazzaroni ME (2019) Converting a periplasmic binding protein into a synthetic biosensing switch through domain insertion. Biomed Res Int 2019:4798793–4798715. https://doi.org/10.1155/2019/4798793

Robertson JWF, Reiner JE (2018) The utility of nanopore technology for protein and peptide sensing. Proteomics 18(18):e1800026. https://doi.org/10.1002/pmic.201800026

Rodriques SG, Marblestone AH, Boyden ES (2019) A theoretical analysis of single molecule protein sequencing via weak binding spectra. PLoS One 14(3):e0212868. https://doi.org/10.1371/journal.pone.0212868

Ruggiero A, Dattelbaum JD, Staiano M, Berisio R, D'Auria S, Vitagliano L (2014) A loose domain swapping organization confers a remarkable stability to the dimeric structure of the arginine binding protein from *Thermotoga maritima*. PLoS One 9(5):e96560. https://doi.org/10.1371/journal.pone.0096560

Sanderink GJ, Artur Y, Siest G (1988) Human aminopeptidases: a review of the literature. J Clin Chem Clin Biochem 26(12):795–807. https://doi.org/10.1515/cclm.1988.26.12.795

Sandhu P, Kumari M, Naini K, Akhter Y (2017) Genome scale identification, structural analysis, and classification of periplasmic binding proteins from *Mycobacterium tuberculosis*. Curr Genet 63(3):553–576. https://doi.org/10.1007/s00294-016-0664-5

Schmidt A, Kochanowski K, Vedelaar S, Ahrne E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M (2016) The quantitative and condition-dependent *Escherichia coli* proteome. Nat Biotechnol 34(1):104–110. https://doi.org/10.1038/nbt.3418

Sherwood AV, Henkin TM (2016) Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. Annu Rev Microbiol 70:361–374. https://doi.org/10.1146/annurev-micro-091014-104306

Siddiquee S, Rovina K, Azriah A (2015) A review of peptide nucleic acid. Adv Tech Biol Med 3(2):1000131

Sriram SM, Kim BY, Kwon YT (2011) The N-end rule pathway: emerging functions and molecular principles of substrate recognition. Nat Rev Mol Cell Biol 12(11):735–747. https://doi.org/10.1038/nrm3217

Stein BJ, Grant RA, Sauer RT, Baker TA (2016) Structural basis of an N-Degron adaptor with more stringent specificity. Structure 24(2):232–242. https://doi.org/10.1016/j.str.2015.12.008

Sun YJ, Rose J, Wang BC, Hsiao CD (1998) The structure of glutamine-binding protein complexed with glutamine at 1.94 A resolution: comparisons with other amino acid binding proteins. J Mol Biol 278(1):219–229. https://doi.org/10.1006/jmbi.1998.1675

Swaminathan J, Boulgakov AA, Hernandez ET, Bardo AM, Bachman JL, Marotta J, Johnson AM, Anslyn EV, Marcotte EM (2018) Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. Nat Biotechnol 36:1076–1082. https://doi.org/10.1038/nbt.4278

Swatek KN, Komander D (2016) Ubiquitin modifications. Cell Res 26(4):399–422. https://doi.org/10.1038/cr.2016.39

Tam R, Saier MH Jr (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol Rev 57(2):320–346

Tanaka KJ, Song S, Mason K, Pinkett HW (2018) Selective substrate uptake: the role of ATP-binding cassette (ABC) importers in pathogenesis. Biochim Biophys Acta Biomembr 1860(4):868–877. https://doi.org/10.1016/j.bbamem.2017.08.011

Tang L (2018) Next-generation peptide sequencing. Nat Methods 15(12):997. https://doi.org/10.1038/s41592-018-0240-7

Tasaki T, Zakrzewska A, Dudgeon DD, Jiang Y, Lazo JS, Kwon YT (2009) The substrate recognition domains of the N-end rule pathway. J Biol Chem 284(3):1884–18895. https://doi.org/10.1074/jbc.M803641200

Tasaki T, Sriram SM, Park KS, Kwon YT (2012) The N-end rule pathway. Annu Rev Biochem 81:261–289. https://doi.org/10.1146/annurev-biochem-051710-093308

Thompson MW, Govindaswami M, Hersh LB (2003) Mutation of active site residues of the puromycin-sensitive aminopeptidase: conversion of the enzyme into a catalytically inactive binding protein. Arch Biochem Biophys 413(2):236–242. https://doi.org/10.1016/s0003-9861(03)00123-1

Timms RT, Zhang Z, Rhee DY, Harper JW, Koren I, Elledge SJ (2019) A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. Science 365(6448):eaaw4912. https://doi.org/10.1126/science.aaw4912

Tsedilin AM, Fakhrutdinov AN, Eremin DB, Zalesskiy SS, Chizhov AO, Kolotyrkina NG, Ananikov VP (2015) How sensitive and accurate are routine NMR and MS measurements? Mendeleev Commun 25:454–456. https://doi.org/10.1016/j.mencom.2015.11.019

Tullman J, Nicholes N, Dumont MR, Ribeiro LF, Ostermeier M (2016) Enzymatic protein switches built from paralogous input domains. Biotechnol Bioeng 113(4):852–858. https://doi.org/10.1002/bit.25852

Tullman J, Callahan N, Ellington B, Kelman Z, Marino JP (2019) Engineering ClpS for selective and enhanced N-terminal amino acid binding. Appl Microbiol Biotechnol 103(6):2621–2633. https://doi.org/10.1007/s00253-019-09624-2

Tullman J, Christensen M, Kelman Z, Marino JP (2020) A ClpS-based N-terminal amino acid binding reagent with improved thermostability and selectivity. Biochem Eng J 154:107438

van Ginkel J, Filius M, Szczepaniak M, Tulinski P, Meyer AS, Joo C (2018) Single-molecule peptide fingerprinting. Proc Natl Acad Sci U S A 115(13):3338–3343. https://doi.org/10.1073/pnas.1707207115

Varshavsky A (1996) The N-end rule: functions, mysteries, uses. Proc Natl Acad Sci U S A 93(22):12142–12149. https://doi.org/10.1073/pnas.93.22.12142

Varshavsky A (2019) N-degron and C-degron pathways of protein degradation. Proc Natl Acad Sci U S A 116(2):358–366. https://doi.org/10.1073/pnas.1816596116

Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C (2011) A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics 10(10):M111 013284. https://doi.org/10.1074/mcp.M111.013284

Watanabe K, Toh Y, Suto K, Shimizu Y, Oka N, Wada T, Tomita K (2007) Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. Nature 449(7164):867–871. https://doi.org/10.1038/nature06167

Weinberg MV, Maier RJ (2007) Peptide transport in *Helicobacter pylori:* roles of dpp and opp systems and evidence for additional peptide transporters. J Bacteriol 189(9):3392–3402. https://doi.org/10.1128/JB.01636-06

Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol 11(3):R31. https://doi.org/10.1186/gb-2010-11-3-r31

Xiao Q, Zhang F, Nacev BA, Liu JO, Pei D (2010) Protein N-terminal processing: substrate specificity of *Escherichia coli* and human methionine aminopeptidases. Biochemistry 49(26):5588–5599. https://doi.org/10.1021/bi1005464

Yoshida Y, Saeki Y, Murakami A, Kawawaki J, Tsuchiya H, Yoshihara H, Shindo M, Tanaka K (2015) A comprehensive method for detecting ubiquitinated substrates using TR-TUBE. Proc Natl Acad Sci U S A 112(15):4630–4635. https://doi.org/10.1073/pnas.1422313112

Yu KH, Barry CG, Austin D, Busch CM, Sangar V, Rustgi AK, Blair IA (2009) Stable isotope dilution multidimensional liquid chromatography-tandem mass spectrometry for pancreatic cancer serum biomarker discovery. J Proteome Res 8(3):1565–1576. https://doi.org/10.1021/pr800904z

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