Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-019-09624-2

BIOTECHNOLOGICALLY RELEVANT ENZYMES AND PROTEINS

Engineering ClpS for selective and enhanced N-terminal amino acid binding

Jennifer Tullman[1] · Nicholas Callahan[1] · Benjamin Ellington[1] · Zvi Kelman[1,2] · John P. Marino[1]

Received: 30 August 2018 / Revised: 16 December 2018 / Accepted: 17 December 2018

Abstract

One of the central challenges in the development of single-molecule protein sequencing technologies is achieving high-fidelity sequential recognition and detection of specific amino acids that comprise the peptide sequence. An approach towards achieving this goal is to leverage naturally occurring proteins that function through recognition of amino (N)-terminal amino acids (NAAs). One such protein, the N-end rule pathway adaptor protein ClpS, natively recognizes NAAs on a peptide chain. The native ClpS protein has a high specificity albeit modest affinity for the amino acid Phe at the N-terminus but also recognizes the residues Trp, Tyr, and Leu at the N-terminal position. Here, we employed directed evolution methods to select for ClpS variants with enhanced affinity and selectivity for two NAAs (Phe and Trp). Using this approach, we identified two promising variants of the *Agrobacterium tumefaciens* ClpS protein with native residues 34–36 ProArgGlu mutated to ProMetSer and CysProSer. In vitro surface binding assays indicate that the ProMetSer variant has enhanced affinity for Phe at the N-terminus with sevenfold tighter binding relative to wild-type ClpS, and that the CysProSer variant binds selectively to Trp over Phe at the N-terminus while having a greater affinity for both Trp and Phe. Taken together, this work demonstrates the utility of engineering ClpS to make it more effective for potential use in peptide sequencing applications.

Keywords Peptide sequencing · Protein engineering · Biosensor · N-end rule pathway

Introduction

Methods for high-throughput detection and quantification of single or low-abundant proteins in mixtures would satisfy an unmet need that spans the fields of proteomics, synthetic biology, and precision medicine (Tyers et al. 2003; Faulkner et al. 2015). In 2014, a preliminary draft of the human proteome contained 86% of the known proteins based on predicted open reading frames (ORFs) from the genomic data available at this time (Kim et al. 2014). However, due to discrepancies in genomic ORF annotation in which short ORFs or genes with internal initiation sites are often mis-annotated, this number is somewhat of a moving target. As of 2017, ~18% of proteins were still considered "missing" as per Human Proteome Project metrics (Baker et al. 2017). While next-generation DNA sequencing technologies have enabled reductions in cost and time for nucleotide sequencing over the past decade that dramatically advanced genomic research (Goodwin et al. 2016), the field of proteomics has seen more steady, but modest advances towards throughput and completeness in proteome analysis. Similar to the effect that disruptive technologies have had on DNA sequencing, progress in proteomic research could be propelled forward dramatically with analogously disruptive advances in technologies for peptide and protein sequencing. In this respect, methods developed for oligonucleotide sequencing could provide a theoretical framework for next-generation protein sequencing technologies. However, there are additional technical challenges when it comes to directly interrogating amino acid residues as

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00253-019-09624-2) contains supplementary material, which is available to authorized users.

✉ Jennifer Tullman
tullmanj@ibbr.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, IBBR, Rockville, MD, USA

Published online: 24 January 2019

opposed to nucleotides. First, there are 20 possible amino acids per position in the peptide chain in comparison to only 4 nucleotides in an oligonucleotide. Second, post-translational modifications of amino acids and N-terminal acetylation will ultimately need to be accounted for, just as in DNA sequencing epigenetic features such as methylation can also be detected with bisulfite sequencing. Finally, the unique chemical moieties present on each amino acid side chain present different and perhaps more difficult challenges when it comes to distinguishing them than is the case for the different nitrogenous bases of DNA.

A potential natural resource for an amino acid recognition reagent are proteins that are used in cells to detect or interact specifically with a particular amino acid or post-translationally modified amino acid. For example, there are 20 tRNA synthetases, which are each highly specific for 1 of the 20 amino acids (Perona and Hadd 2012). Additionally, many proteases are able to uniquely identify a target amino acid and cleave before or after that residue on a peptide chain. However, it is important that any new method for detecting amino acids, in the context of sequencing, be coupled with the ability to identify the location of that amino acid on the peptide chain. Therefore, an amino acid-binding protein that is selective for an amino acid at a structurally unique chain position, such as the N-terminus of a peptide, is an appealing candidate for development into a sequencing reagent. This type of a reagent would allow the positional information of the amino acid to be directly coupled to the binding event used to detect that amino acid.

The N-end rule pathway functions to degrade proteins in cells as part of a regulated process for maintaining protein homeostasis. In general, the bacterial system consists of adaptor and chaperone proteins that specifically recognize the proteins to be degraded and can unfold and deliver them to a protease core for destruction, or in organisms such as yeast, which contain a ubiquitin pathway, the targeted protein may be ubiquitylated on a lysine residue and then degraded (Bachmair et al. 1986). In eukaryotes, the pathway depends on different N-terminal amino acids (NAAs) than the subset utilized in the bacterial N-end rule pathway, and the different nomenclature and proteins involved are well described in the literature (Varshavsky 1996). In bacteria, for example, the Clp protease that performs the degradation interacts with different chaperones divided into class I (ClpA, ClpC, ClpD, or ClpE) or class II (ClpX, ClpY) among different systems from proteobacteria, actinobacteria, or cyanobacteria. This pairing leads to a convenient nomenclature to describe the protease core complex (e.g., ClpAP or ClpXP). An adaptor protein in bacteria, ClpS, interacts with some class I chaperones and is functionally an N-recognin, as it specifically recognizes the substrate to be degraded through an N-degron tag (Dougan et al. 2002; Erbse et al. 2006; Román-Hernández et al. 2011; Nishimura and van Wijk 2015).

ClpS provides a promising scaffold from which to develop an amino acid-binding reagent for protein sequencing. Importantly, it is a "gatekeeper" of the specificity of this protein degradation machinery in many organisms. ClpS and homologous domains such as one of the UBR-box (E3 ligase) N-recognin domains (Lupas and Koretke 2003) exist in bacteria, yeast, mammals, and plants (Graciet and Wellmer 2010; Nishimura et al. 2013) and target different N-degron tags with either type 1 (Arg, Lys, or His) or type 2 (Tyr, Phe, Trp, or Leu) destabilizing residues. The N-degron tags that ClpS proteins from bacteria recognize are sequences containing the type 2N-terminal primary destabilizing residues (Wang et al. 2008b). Moreover, cyanobacteria contain two different ClpS protein paralogues (Tryggvesson et al. 2015), which interact with different types of Clp protease cores, and exhibit different specificity. Similarly, α-proteobacteria contain two different ClpS proteins, presumably to achieve enhanced control over which N-degron containing substrates are targeted for degradation by fine-tuning the expression levels of the ClpS adaptors (Stein et al. 2016). In addition to the wide natural variation in substrate recognition sequences, ClpS can be engineered for new specificity. For example, it was evolved for use in a "post-translational proofreading" system to detect non-standard amino acids such as *p*-acetyl-phenylalanine (Kunjapur et al. 2018). Finally, the ClpS protein family has the advantages of being relatively small (MW ~13 kDa) and having no native proteolytic or enzymatic activity that would be unnecessary and potentially complicating in the context of a protein sequencing application.

Based on previous studies, we selected the *Plasmodium falciparum* ClpS protein and the ClpS2 of *Agrobacterium tumefaciens* as the starting scaffolds for our protein engineering efforts. The *P. falciparum* ClpS protein has enhanced affinity for the NAA Phe (K_D = 600 nM) (Ahyoung et al. 2016) compared to other ClpS proteins that have been studied to date (Schuenemann et al. 2009), although these studies experiments did not directly compare the affinities using the same technique or target peptide. In this respect, the second residue from the N-terminus has been shown to also play a role in binding affinity (Stein et al. 2016). The *P. falciparum* ClpS protein also has the unique ability to recognize isoleucine (Tan et al. 2016) unlike the other homologs that have been reported (Wang et al. 2008a). The *A. tumefaciens* ClpS2 protein has the drawback of lower overall affinities, but the advantage of higher specificity for Phe when compared to Tyr, Trp, and Leu. Therefore, both of these proteins were chosen as starting points for a directed evolution approach with the goal of selecting mutants with higher affinity and selectivity for target NAAs to demonstrate the utility of the ClpS family of proteins for the development of NAA binding reagents (NAABs).

Materials and methods

Plasmid construction

The wild-type *P. falciparum* ClpS gene (Accession XM_001349904) was a gift from Dr. Egea (UCLA). It was cloned into the pET15b vector (catalog # 69661, Millipore Sigma, Burlington, MA, USA) with an additional His_6-tag at the N-terminus of the protein, followed by a TEV protease cleavage site, using the *Nco*I and *Xho*I restriction sites. The *A. tumefaciens* ClpS gene (Accession NP_355189.1) was amplified from the genomic DNA, strain GV3101 from the C58 chromosomal background, which was a gift from Dr. Shunyuan Xiao (IBBR), and cloned into the pET15b vector in the same manner. Both genes were cloned into the pCTCON2 vector for yeast display (catalog # 41843, Addgene, Watertown, MA, USA) by amplifying the gene by polymerase chain reaction (PCR) with primers to add the *Nhe*I and *Bam*HI sites, and then ligating this to the vector such that the gene is located to the C-terminal end of the aga2-encoding gene, separated by a factor XA cleavage site, an HA-tag and a $[GGGGS]_{x3}$ spacer. There is also a C-terminal myc-tag before the stop codon. A vector was also created that contains no gene insert, but rather a *Sac*II site and a *Spe*I site, so that the vector can be linearized at this site for homologous recombination.

Random mutagenesis library creation

To create the random mutagenesis libraries, error-prone PCR was used. The gene encoding the ClpS protein of interest was amplified using the HA-tag for (CCATACGACGTTCC AGACTAC) and T7 (TAATACGACTCACTATAGGG) primers in a reaction containing 0.2 mM dATP, 1 mM dCTP, 0.2 mM dGTP, 1 mM dTTP, 10 mM $MgCl_2$, and 0.5 mM $MnCl_2$, 1× Taq reaction buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl) without $MgCl_2$, and Taq DNA polymerase. The PCR product was used for homologous recombination, as described below. The libraries each contained at least 1 million naïve members and greater than 50% of the colonies contained at least one mutation from the 10 colonies sequenced.

Saccharomyces cerevisiae (yeast) transformation

EBY100 strain *S. cerevisiae* (ATCC® MYA4941™, Manassas, VA, USA) were transformed with pCTCON2 plasmids containing the wild-type ClpS genes using the Frozen EX Yeast Transformation II kit (Zymo Research, Irvine, CA, USA) and subsequently grown on selective media as the pCTCON2 plasmid harbors the ability to synthesize tryptophan. Synthetic dextrose media supplemented with casamino acids lacking tryptophan (SD-CAA) and containing 100 μg/mL ampicillin was used to grow all the yeast used in this study. Protein surface expression was induced by resuspending the cells in synthetic galactose media supplemented with casamino acids lacking tryptophan (SG-CAA).

Site-saturation library creation/homologous recombination in yeast

EBY100 *S. cerevisiae* cells were grown overnight to an OD_{600} of 3 in yeast extract- peptone-dextrose (YPD) media at 30 °C. This was used to inoculate a 100 mL culture of YPD to OD_{600} 0.3. After 5 h, when cells had grown to OD_{600} 1.0, the cells were transferred to 50-mL conical tubes and centrifuged at 3000*g* for 3 min at 4 °C. The cell pellet was washed twice with 50-mL ice-cold sterile water and then washed once with 50-mL ice-cold electroporation buffer (1 M sorbitol/1 mM $CaCl_2$). The cells were conditioned for electroporation by resuspending the cell pellet in 20 mL 0.1 M LiAc, 10 mM DTT and shaking for 30 min at 30°. The cells were centrifuged as above and washed with 25 mL per tube of electroporation buffer before being resuspended in 200 μL of electroporation buffer to reach a final volume of about 1 mL. Cells were kept on ice until electroporation.

For electroporation, 400 μL of competent cells prepared as above were incubated with the vector and insert, in a 1:3 ratio, and kept on ice for 5 min. The vector used was the pCTCON2 plasmid described above containing the *clpS* gene of interest, and digested within the *clpS* gene with the restriction enzyme *Ale*I. The insert used was the error prone PCR library obtained as described above or the NNK primer (TAAGCTCT ACAAGGTCATGCTGCTGAATGACGACTATA CGNNKNNKNNKTTTGTCACCGGTGTGCTGAA GGCCGTCTTTCGCATGAGCG) for the site-saturation library. The cells were then transferred to a 0.2 cm electroporation cuvette and electroporated on the pre-set yeast settings (1.5 kV, 25 μF, Bio-Rad GenePulser Xcell, Hercules, CA, USA). The cells were transferred to a tube containing 4 mL of YPD media (Catalog # Y1375, Sigma, St. Louis, MO, USA) and 4 mL of 1 M sorbitol and incubated at 30 °C for 1 h, 225 rpm. The cells were then centrifuged and resuspended in SD-CAA media and dilutions were plated to calculate library size, and the rest was grown in a flask containing 250 mL of SD-CAA media and passaged once before selections or sorting.

Library selection

The magnetic-activated cell sorting (MACS) and fluorescence-activated cell sorting (FACS) were performed using slightly modified protocols from the Pacific Northwest National Lab Yeast Display ScFV Antibody Library User's Manual (Feldhaus et al. 2003) and the Methods in Molecular Biology Flow Cytometry Protocols (Hawley and Hawley

2004). Briefly, the yeast displaying a library of mutant ClpS proteins were grown in SD-CAA media overnight at 30 °C until the OD_{600nm} was approximately 4.0. The yeast were used to inoculate a fresh culture at an OD_{600nm} of 1.0 in a mixture of 80% SG-CAA/20% SD-CAA and incubated for 24 h at 20 °C. Approximately 10^9 yeast were washed and resuspended in 1 mL of Dulbecco's phosphate buffered saline containing 0.5% bovine serum albumin (PBS/BSA) (DPBS, Catalog # 14190-136, Thermo Fisher, Waltham, MA, USA; BSA, Catalog # A7906, Sigma, St. Louis, MO, USA) containing 10-μM biotinylated peptide at room temperature for 1 h. The yeast were pelleted by centrifugation at 3000*g* for 2 min. The supernatant was decanted to remove excess peptide and the pelleted yeast resuspended with 100 μL of streptavidin coated (Dynabeads M-280 streptavidin, 10 mg/mL, Catalog # 11205D, ThermoFisher, Waltham, MA, USA) or anti-biotin-coated magnetic beads (Anti-Biotin MicroBeads, Catalog # 130-090-485, Miltenyi Biotech, Bergisch Gladbach, Germany) and flowed over a MACs column (Catalog # 130-042-401, Miltenyi Biotech, LS column, Bergisch Gladbach, Germany).

After two rounds of MACs selection, the library was sorted by flow cytometry. The cells were induced to express surface-displayed protein as described above and then incubated with biotinylated peptide in different concentrations, streptavidin, R-phycoerythrin (PE), and anti-myc AF647 overnight at room temperature. A typical reaction contained 100 μL of cells (containing approximately 10^6 cells), 10 μL of peptide at a concentration between 10 nM and 10 μM, and 25 μL of a master mix containing 2 μL of the anti-myc antibody (c-myc Antibody (9E10) Alexa Fluor®647-200 μg/mL, catalog # sc-40 AF647, Santa Cruz Biotechnology, Dallas, TX, USA), 4 μL of the SAPE (streptavidin, R-phycoerythrin conjugate) - 1 mg/mL (catalog # S866, ThermoFisher, Waltham, MA, USA) and 19 μL of PBS/BSA for each sample.

The cells were then washed with PBS/BSA and sorted using a FACS Aria cytometer (BD Biosciences, San Jose, CA, USA) and collected in 1 mL of SD-CAA media. The amount of cells that bound the peptide improved with each round and were sorted with sequentially lower concentrations of peptide to increase the stringency of the selection. In general, 8 to 16 colonies were sequenced from the sorted libraries after the fourth and fifth rounds of selection.

Peptides

All peptides are named by indicating the first two residues, with the full sequence available in Table 1. All of the peptides used in this study had one of two sequences for the C-terminal end of the peptide, either XDEDLE or XGVECK, where the N-terminal amino acid is varied on a particular peptide scaffold. The X_1G_2 peptides also contained a biotin linked via the lysine side chain on the C-terminal residue. X_1G_2 peptides were ordered from Biomatik (Cambridge, Ontario, Canada) and received in 1 mg/vial lyophilized form. Peptides were resuspended in 1× DPBS and diluted to the appropriate concentration into the experiment buffer. The X_1D_2 peptides were synthesized in-house on a 20-μmol scale on a Gyros Protein Technology Tribute Peptide Synthesizer with amino acid reagents obtained from Novabiochem (Millipore Sigma, Burlington, MA, USA) and Gyros Protein Technologies (Uppsala, Sweden) and a Novatag (Millipore Sigma, Burlington, MA, USA) Biotin Resin. The synthesis results in a peptide with an ethylene diamine spacer and then the biotin moiety. The peptide was then cleaved from the resin using 3 mL of trifluoroacetic acid (TFA), phenol, water, and triisopropylsilane (TIPS) in an 88:5:5:2 ratio. The peptide was subsequently rinsed with ice-cold ether, pelleted by centrifugation at 4500*g* for 10 min at 4 °C, and decanted, three times. It was then dried under nitrogen overnight at room temperature and subsequently lyophilized and stored at − 20 °C until resuspension in the assay buffer.

Table 1 Peptide and protein nomenclature

Name	Sequence
X_1D_2	XDEDLE-biotin
X_1F_2	XFDEDLE-biotin
X_1G_2	XGVECK-biotin
Wild-type (WT)	*A. tumefaciens* ClpS2 [Pro_{34} Arg_{35} Glu_{36}]
Variant 1 (V1)	*A. tumefaciens* ClpS2 [Pro_{34} Met_{35} Ser_{36}]
Variant 2 (V2)	*A. tumefaciens* ClpS2 [Cys_{34} Pro_{35} Ser_{36}]
Variant 3 (V3)	*A. tumefaciens* ClpS2 [Cys_{34} Ser_{35} Trp_{36}]

Protein purification

The wild-type and mutant ClpS proteins were expressed in BL21 *Escherichia coli* cells (C2527I, New England Biolabs, Ipswich, MA, USA). Expression of wild-type and ProMetSer mutant proteins was induced with 0.5 mM IPTG when the OD_{600} nm reached 1.0 and incubated for 6 h at 37 °C. The cells expressing the CysProSer and CysSerTrp mutants were removed from the 37 °C incubator and cooled to 15 °C when the OD_{600} nm reached 0.5, then induced with 0.5 mM IPTG when the OD_{600} nm reached 1.0 and grown for 16 h at 15 °C. The cells were harvested by centrifugation at 5000*g* for 20 min, and the cell pellets frozen for future use.

Frozen cell pellets were resuspended in lysis buffer (100 mM Tris-HCl, 300 mM NaCl, 25 mM imidazole, pH 8.0), and sonicated using a Fisher Scientific (Waltham, MA, USA) FB505 sonicator (500 W) with a Cl334 probe at 20% amplitude for 4 s on, 20 s off, for 90 min, which results in 15-min total sonication time. The lysate was centrifuged at 20,000*g* for 40 min and then incubated for 1 h with chelating

fast-flow sepharose resin (catalog # 17-0575-02, GE Healthcare, Chicago, IL, USA) coated with nickel and pre-equilibrated in lysis buffer. The mixture was centrifuged at 1000*g* for 10 min and then the supernatant removed, and the resin resuspended in 5-mL lysis buffer and used to form a column. The column was then washed with 10 column volumes (CVs) of lysis buffer, and then 5 CVs of wash buffer (100 mM Tris-HCl, 300 mM NaCl, 75 mM imidazole, pH 8.0), before eluting with 5 CVs of elution buffer (100 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was then dialyzed using 10 kDa molecular weight cutoff (MWCO) dialysis tubing into 50 mM Tris-HCl, 300 mM NaCl, 1 mM DTT, 5% glycerol, pH 8.0. Each dialysis was performed for > 12 h, for a total of three times. The protein was removed from dialysis tubing, centrifuged 40 min at 20,000*g*, and the supernatant concentration measured by Bradford assay against a BSA standard curve according to manufacturer instructions (catalog # 500-0201, Bio-Rad, Hercules, CA, USA). The protein was then loaded onto a S200 26/60 size exclusion chromatography column pre-equilibrated in 2CV of 50 mM Tris-HCl, 300 mM NaCl, 5% glycerol, 1 mM DTT, pH 8.0. Five-milliliter fractions were collected and tracked at 280 nm; peaks were compared with the Bio-Rad (Hercules, CA, USA) gel filtration standard and further analyzed by SDS-PAGE (Supplemental Fig. 1). Fractions were combined, concentrated using an Amicon Ultra centrifugation unit with a 10 kDa MWCO (Sigma, St. Louis, MO, USA), centrifuged for 40 min at 20,000*g*, and measured by the Bradford assay as above.

The thermal stability of the variants was assessed using the Tycho NT.6 (NanoTemper Technologies, Munich, Germany). Each variant was loaded in 3× PBS at approximately 1 mg/mL concentration in capillary tubes and the intrinsic protein fluorescence recorded at 330 nm and 350 nm while heating the sample over a 35–95 °C at a rate of 30 °C per minute.

Yeast/peptide pull-down assay

A colorimetric pull-down assay was used to screen the FACS selected library variants with different peptide substrates in a high-throughput manner and determine the optimal candidates for in vitro characterization. Yeast displaying the library variants of interest were grown to saturation in SD-CAA media and transferred to SG-CAA media for surface expression at 20 °C for 24 h. The cultures were pelleted at 3000*g* for 2 min and washed with PBS, 0.5 mg/mL BSA, 0.1% Tween20 ("ELISA buffer") and diluted to $OD_{600nm} = 2.5$. A UV-transparent 96-well flat-bottomed polystyrene plate was pre-blocked with 1 mg/mL BSA and washed with ELISA buffer. Ten microliters of diluted cells were added to the wells and mixed with 90 μL of 1 μg/mL biotinylated peptide substrate in ELISA buffer. The mixture plate was incubated in a bench-top orbital shaker for 1 h, then centrifuged at 3000*g* to pellet the yeast cells bound to peptide and wash the excess peptide away by washing three times with 100 μL of buffer. The cells were resuspended in 100 μL of buffer with 1 μg/mL of streptavidin-HRP (horseradish peroxidase) (Thermofisher/Life Technologies, Waltham, MA, USA). Incubation and wash steps were repeated. Cells were resuspended in a final volume of 50 μL buffer and 50 μL of 1× tetramethylbenzidine (TMB, VWR, Radnor, PA, USA) was added to the wells. Reactions were incubated for 20 min at room temperature for oxidation of the TMB by horseradish peroxidase (HRP), which produces a blue coloration. Reactions were then quenched by adding 100 μL of 1 M HCl, which produces a yellow color that can be measured by the absorbance at 450 nm. The intensity is a function of the amount of streptavidin-HRP pulled down by interaction with the yeast cells, facilitated by biotinylated peptide binding to surface-expressed ClpS.

Each sample was measured in triplicate within the same row of the 96 well plate, allowing for four ClpS/substrate pairs per row, with the standard deviation of Abs_{450nm} taken as the error. Fluid was added to rows using a 12-channel hand pipette. In designing our assays, we stressed the difficulty of comparing row to row, because of the variance that could be introduced by human error into TMB addition and quenching. When testing a single library variant against a panel of peptides, we kept wells 1 through 3 empty as a check for artifacts, added only peptide to wells 4 through 6, added peptide and uninduced yeast to wells 7 through 9, and added peptide and induced yeast to wells 10 through 12. As the background in lanes 1 through 9 showed no increase in background absorbance from the peptide or media, the data displayed here is the average and standard deviation of lanes 10 through 12, except for the data labeled "no yeast," which is lanes 4 through 6.

Surface plasmon resonance

The surface plasmon resonance (SPR) data was collected on a Biacore (Uppsala, Sweden) T100 instrument by loading a streptavidin chip with the target biotinylated peptides in each of the four channels and then flowing the protein over the chip at each concentration for 180 s before washing with 50 mM Tris-HCl, 300 mM NaCl, 1 mM DTT, 5% glycerol, and pH 8.0 buffer for 180 s. The SPR assays for all variants were performed under the above optimized buffer conditions to obtain the greatest binding activity, as the activity and stability of ClpS is pH, and NaCl concentration dependent (data not shown). Additionally, the wild-type and mutant proteins purify as dimers that can be disrupted by the addition of DTT, indicating they are likely disulfide-linked. This has been shown in the literature (Colombo et al. 2018), with different homologs of ClpS having different propensities for dimerization. The four channels were used such that the negative control peptide (A_1G_2) was in channel 1, F_1G_2 in channel 2, W_1G_2 in channel 3, and Y_1G_2 in channel 4. The negative

control peptide channel was used as a baseline for subtraction, which accommodates both any non-specific binding and refractive index change. A steady state response at each concentration was plotted and fit to calculate the K_D, while the resonance units vs. time was used to calculate a k_{off} for each mutant protein with each peptide.

Results

Directed evolution of ClpS

Our approach was to cast a wide net via random mutagenesis and screening in order to select for proteins with higher affinity binding to the target NAA. In subsequent steps, we would then hone in on specific residues determined from these initial screens with a targeted mutagenesis step. Therefore, we began by applying an error-prone PCR approach to broadly and randomly mutagenize both the *P. falciparum clpS* gene and the *A. tumefaciens clpS2* gene. We then performed homologous recombination into a yeast display vector such that the gene of interest was fused to the C-terminal end of the *aga2* gene for display. The displayed proteins also contained a C-terminal myc-tag that could be used for detection as shown in Fig. 1a, where an Alexafluor647 labeled antibody could bind to the C-terminal myc-tag, while successful binding of the ClpS protein variant to a biotinylated peptide with the appropriate N-terminal amino acid was monitored by detection with a fluorescently labeled streptavidin. Using FACS, we assessed the affinity of each mutant for the N-terminal amino acid on the target peptide based on which quadrant the fluorescent signal fell into as shown in Fig. 1b. If the AlexaFluor647 signal is high, the full-length protein is accessible on the yeast surface for the antibody to bind and the cell will appear in Q1 or Q2. If there is also a high fluorescence signal along the *x*-axis (streptavidin-PE), then the peptide is presumably bound to the protein variant, and the signal will appear in Q2. As shown in Fig. 1c, a naïve library will contain a mixture of yeast cells, some of which do not express full-length variants and fall into Q3. Some of the variants will be expressed but not bind the peptide, as in Q1, and a small percentage of the library will contain variants which bind the peptide of interest as in Q2. For a given library, the top 5% to 10% of cells, those in the second quadrant which express mutant ClpS proteins that bind the peptide of interest (in this case F_1D_2), were sorted and grown in selective media. In the second (Fig. 1d) and third (Fig. 1e) rounds, the cells that were selected from the first or second rounds, respectively, were sorted again until the majority of the library members had an affinity for the peptide of interest. As illustrated, the percentage of cells displaying a

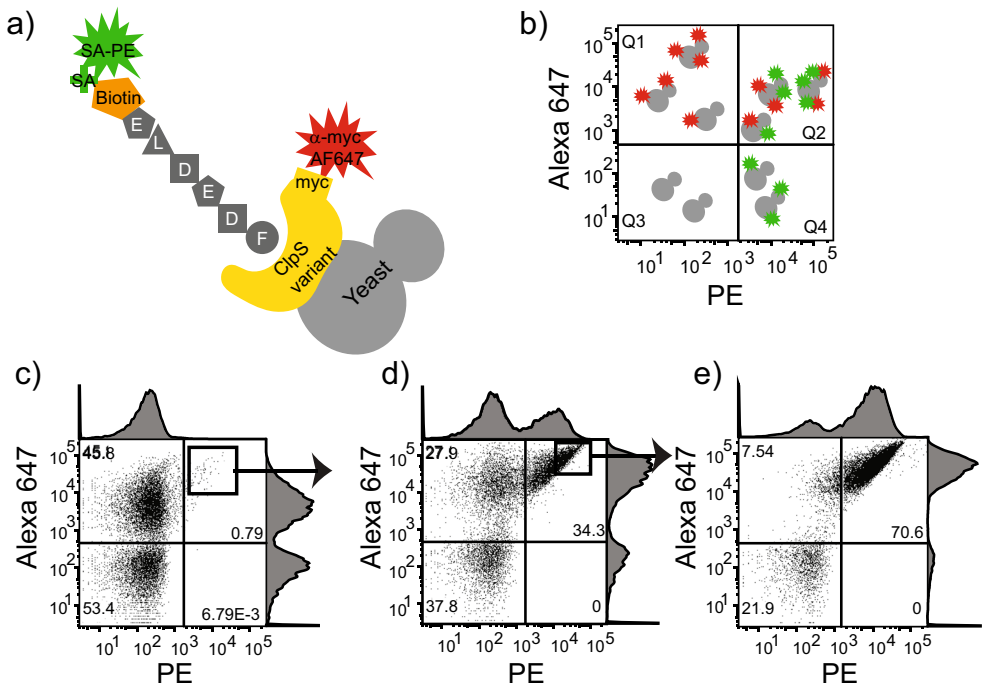

Fig. 1 **a** Cartoon depicting the fluorescent labeling scheme used for detection of peptide binding during flow cytometry. The myc tag is detected with an AlexaFluor 647 labeled anti-myc antibody (shown in red). The peptide is detected using streptavidin-PE (shown in green) which binds the biotinylated C-terminus of each peptide. **b** The expected flow cytometry results depicting yeast that display a non-binding protein in quadrant 1 (Q1), yeast that bind the peptide in Q2, yeast that do not display the protein in Q3, and yeast that exhibit non-specific binding to the peptide in Q4. **c** Representative flow cytometry plots displaying the increased PE fluorescence seen in each round after three rounds of selection of a library against a Phe peptide. The square and arrow in Q2 illustrate those cells that are carried on to the next round of sorting after outgrowth

protein that bound the peptide did increase with each round of selection. One representative library selection is shown in Fig. 1 where three rounds of selection were performed for this library. This process was repeated multiple times after the creation of each new mutant library. The sequence and corresponding nomenclature for the target peptides used in selections and screens is given in Table 1, where *X* stands for any amino acid. Additionally, Supplemental Table S1 gives a list of the libraries and selections performed by this method. Sequencing analysis of the individual clones from the final round of selection of the first error-prone PCR library for each protein indicated that there was indeed a hotspot of substituted residues corresponding to residues 122–124 in the *P. falciparum* protein, and the homologous residues, 34–36 in the *A. tumefaciens* protein. The mutated residues are highlighted in yellow in Fig. 2 and the hotspot residues are boxed in blue. The figure was generated using the ESPript 3.0 web utility, (Robert and Gouet 2014). These same residues were mutated in many of the constructs that showed altered and improved affinity for the various targeted peptides. The full sequencing results are shown in Supplemental Figs. S2 and S3. These residues are in a flexible loop near the opening of the peptide binding site on ClpS based on an available crystal structure from *P. falciparum* (PDB: 4O2X) (Ahyoung et al. 2016) and the structure of *A. tumefaciens* bound to phenylalaninamide (PDB: 4YJX) (Stein et al. 2016). Based on this observation, we created a second library of the *A. tumefaciens* ClpS2 in which these three residues were mutated to all 20 amino acids (a theoretical library size of 8000 constructs). This new library was further screened against either an F_1 peptide or a W_1 peptide, and after four rounds of selection, the hits were again sequenced and characterized. The variants that were chosen for characterization are given in Table 1 with the three residues that were mutated from the wild-type sequence listed. The full sequencing results of all variants selected from different rounds of sorting are listed in Supplemental Table S2. Some proteins behaved well in the yeast displayed context, but were not amenable to *E. coli* expression and purification, while others were easily purified from *E. coli* but had high background in the yeast-display format, which we have traced to non-specific binding to the streptavidin-HRP (data not shown). Therefore, we chose to characterize two of the variants (V2 and V3) in detail using the pull-down assay and two variants (V1 and V2) after expression and purification of the proteins from *E. coli* in SPR experiments.

Characterization of yeast-displayed ClpS mutant proteins

We designed a pull-down assay to screen the FACS-selected hits against many different NAA-containing peptides in high-throughput. As shown in Fig. 3a, yeast displaying a particular mutant ClpS protein were incubated with a panel of C-terminally biotinylated peptides containing different NAAs. Samples were then centrifuged, washed, and incubated with streptavidin-HRP and TMB substrate, followed by acid addition to quench the reaction. This was used to identify the yeast displaying proteins with differing specificity and/or affinity towards those peptides as compared to the wild-type protein warranting further characterization. Two protein variants from the FACS selection were characterized with a full panel of 20 peptides to assess whether the affinity for other NAA peptides was inadvertently changed. V2 had an improved affinity for the W_1G_2 peptide compared to the wild-type, as determined via flow cytometry of the yeast displaying these proteins (Supplemental Fig. S4) and via the pull-down assay described above (Fig. 2b and Supplemental Fig. S5a). Conversely, V3

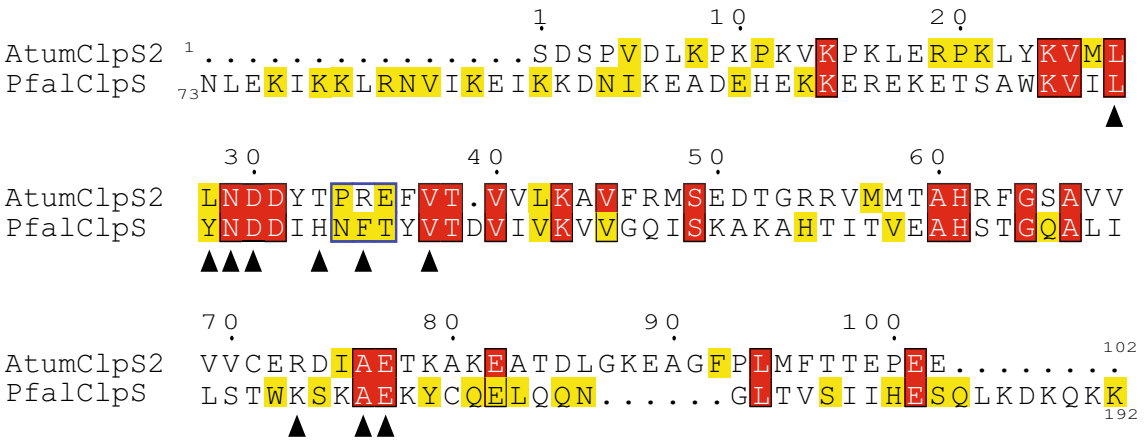

Fig. 2 Alignment of the *A. tumefaciens* ClpS2 (Accession NP_355189.1) and *P. falciparum* ClpS (Accession XM_001349904) protein sequences (21.43% identity). Identical positions are highlighted in red. The black triangles indicate proposed substrate contacts, based on the crystal structure of *A. tumefaciens* ClpS2 bound to L-phenylalaninamide (Stein et al. 2016). The residues highlighted in yellow are those that were found to be mutated in constructs selected from the initial error-prone libraries for increased Phe binding. The blue box around residues 34P, 35R, and 36E of the *A. tumefaciens* ClpS2 is to highlight the residues that were mutated to all 20 amino acids in the second library

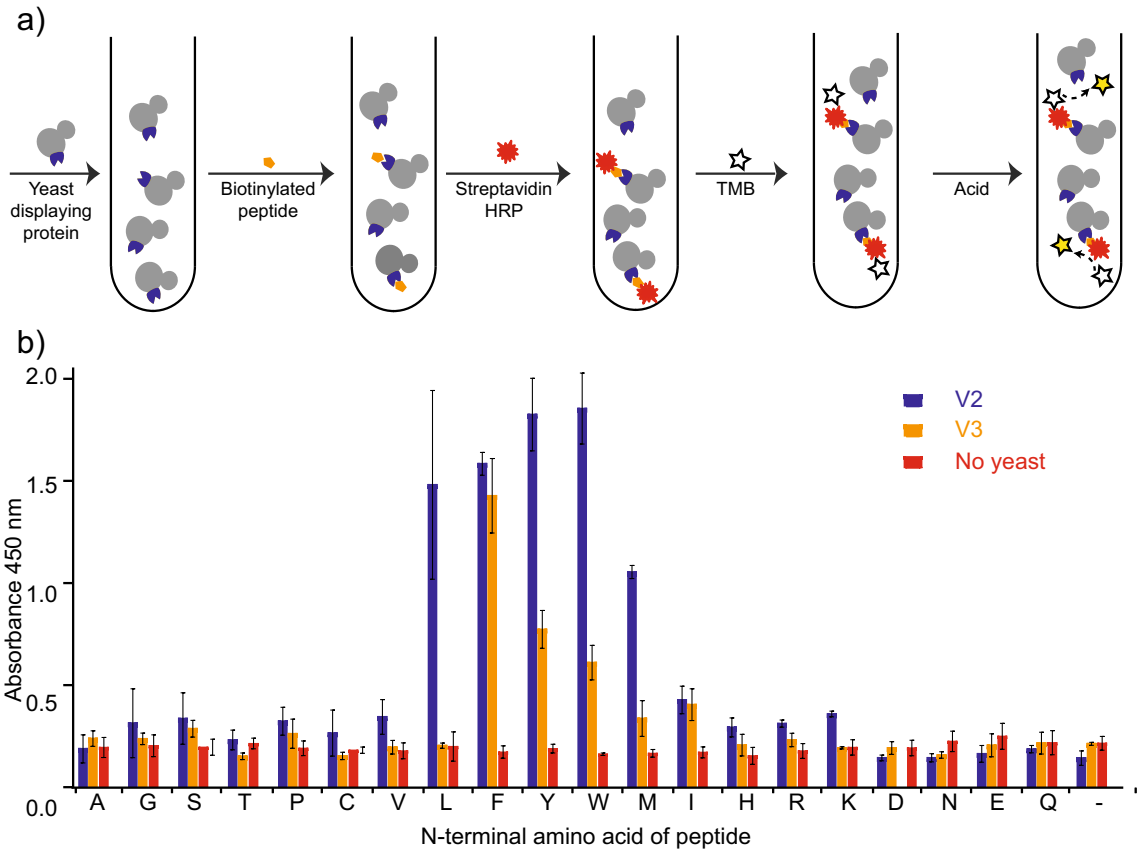

Fig. 3 **a** The pull-down assay workflow: incubation of the yeast with the peptide of interest, centrifugation, and washing to remove unbound peptide, labeling with the streptavidin-HRP, followed by incubation with TMB substrate and acid quenching which results in a yellow color change that could then be quantified in each well of a 96-well plate. **b** The assay described in part a was used with peptides containing the N-terminal residue labeled on the *x*-axis for the wild-type (black) and two variants (blue and yellow) of the ClpS2 protein. The no yeast control is shown in red. In this assay, the absorbance at 450 nm correlates with the number of yeast bound to the peptide. The error bars represent three replicate measurements of one biological sample

displayed increased affinity for the F_1G_2 peptide and higher selectivity towards the F_1G_2 peptide as compared to other NAAs. Neither mutant protein had any undesirable "off-target" affinity towards any NAA that was not already recognized by the wild-type protein, with one exception. V2 has an increased affinity towards Leu, which the wild-type *A. tumefaciens* ClpS protein does not bind. However, an affinity for Leu is exhibited by other ClpS homologs (Stein et al. 2016). Therefore, V2 appears to have regained some of the canonical binding activity of the ClpS protein family with respect to leucine. In fact, the affinity for every other residue besides the canonical (Phe, Trp, Tyr, Leu) binding residues was either the same or reduced compared to the wild-type (wild-type data shown in Supplemental Fig. S5a). Our pull-down assay also indicated, for the first time, that the wild-type ClpS2 protein has some binding activity towards Met, His, Arg, and Lys at the high concentrations used in the assay. The His and Arg binding was also confirmed via flow cytometry and this data is displayed in Supplemental Fig. S5b. Some differences between the observed binding for a particular residue is likely due to the different concentrations and format used for the assay and whether that was above or below the K_D for binding to that residue. Finally, the pull-down assay (Supplemental Fig. S5a) show the wild-type protein does bind Ile, which disagrees with the flow cytometry assay (Supplemental Fig. S5b) and previous reports (Stein et al. 2016) in which the wild-type protein was reported to show weak binding with an I_1R_2 and an I_1K_2 peptide when ClpS is in a non-yeast displayed context. From our pull-down assay, we also confirm the lack of binding to Val and Ala reported in the literature, and add the remaining amino acids to the non-binding category.

Characterization of the ClpS mutant protein binding kinetics

For peptide sequencing via some recently proposed methods, as described in the "Introduction" section, the reagents need to bind peptides affixed to a surface. To that end, we characterized two *A. tumefaciens* ClpS2 protein variants (V2, described above, and V1) in vitro in this context. Both variants were expressed and purified from *E. coli*; however, the third variant

could not be purified in a high enough concentration to perform the subsequent assays. V2 was not as thermally stable as the wild-type and V1 (Supplemental Fig. S6 and Supplemental Table S3); however, enough protein was obtained to assess its binding properties in vitro at temperatures below its melting temperature. The binding properties of each variant were analyzed using surface plasmon resonance (SPR). Since the SPR chip has four channels, we chose to use three peptides ending in the residues that are the most commonly recognized by the wild-type ClpS proteins (Phe, Trp, Tyr) and a negative control (Ala). The summary of the SPR data is displayed in Table 2, where the steady state values of the K_D are calculated from the experiments performed using concentrations ranging from 0 to 50 μM of protein for each variant. The association curves for V2 with each peptide ending in Phe, Trp, or Tyr, are given in Fig. 4a. The dissociation curves are shown in Fig. 4b, and the steady state plot calculated from the plateau of the association curve for each concentration is given in Fig. 4c. The corresponding data for V1 and the wild-type protein are given in the Supplemental Figs. S7 and S8, respectively. Significantly, a 6.6-fold decrease in the K_D for W_1G_2 peptide was observed for V2, compared to the wild-type protein, as determined by SPR with the peptide attached to the surface and the protein in solution. A 3.5-fold improved K_D was observed for V2 with the F_1G_2 peptide as well. This data supports the trend observed from the pull-down assay that also showed increased binding affinity, compared to wild-type, to Phe, Trp, and Tyr. Unfortunately, V1 exhibited non-specific binding in the pull-down assay and could not be compared to the SPR data. Via the SPR assay, V1 has little change in its affinity for tryptophan; however, it is improved 7.2-fold in affinity for the F_1G_2 peptide and 5.5-fold for Y_1G_2. Thus, the ClpS2 variants have increased affinity for either Phe or Trp, but due to the overall increases in affinity compared to the wild-type proteins it is difficult to discern the change in selectivity of the mutant proteins. Therefore, the binding affinity for each variant with respect to one peptide was divided by the affinity for each of the other peptides to give a relative specificity factor displayed in the three lower panels of Table 2. If the specificity factor is less than 1, the protein has higher affinity for a different NAA than the one tested in that panel and vice versa if the specificity factor is greater than 1. Presenting the binding data in this way clearly shows, as expected, that the wild-type protein has a modest specificity for N-terminal Phe. Although V2 is improved in K_D for both peptides with N-terminal Phe and Trp, V2 is more specific for Phe over Trp than the wild-type protein. Additionally, V1 is more specific for N-terminal Trp than for Phe. It is worth noting that residue 35 (mutated from Arg to Met in V1) is homologous to the *E. coli* ClpS1 residue M40, which is considered to be a gatekeeper residue that when mutated to Ala allows non-canonical NAAs such as Val to fit in the binding pocket (Wang et al. 2008b; Schuenemann et al. 2009). It is evolutionarily conserved as a Met in bacterial ClpS1 proteins, but is conserved in eukaryotic ClpS1 proteins as an Arg residue, and is sometimes found as Glu or Phe in other ClpS1-like and ClpS2 proteins (Nishimura et al. 2013). In the *Arabidopsis thaliana* ClpS1 protein, the Arg residue at

Table 2 Affinity and selectivity of ClpS variants for different NAAs

K_D (μM)			
Variant	Phe	Trp	Tyr
WT	13.0 ± 0.6	18.4 ± 0.4	63.0 ± 4.0
V1	1.8 ± 0.4	13.1 ± 1.3	11.6 ± 1.2
V2	3.7 ± 0.3	2.8 ± 0.3	36.1 ± 4.4
Selectivity for Phe			
Variant	Phe/Phe	Phe/Trp	Phe/Tyr
WT	1.0	0.71	0.21
V1	1.0	0.14	0.16
V2	1.0	1.32	0.10
Selectivity for Trp			
Variant	Trp/Phe	Trp/Trp	Trp/Tyr
WT	1.42	1.0	0.29
V1	7.28	1.0	1.13
V2	0.76	1.0	0.08
Selectivity for Tyr			
Variant	Tyr/Phe	Tyr/Trp	Tyr/Tyr
WT	4.86	3.43	1.0
V1	6.43	0.88	1.0
V2	9.82	12.91	1.0

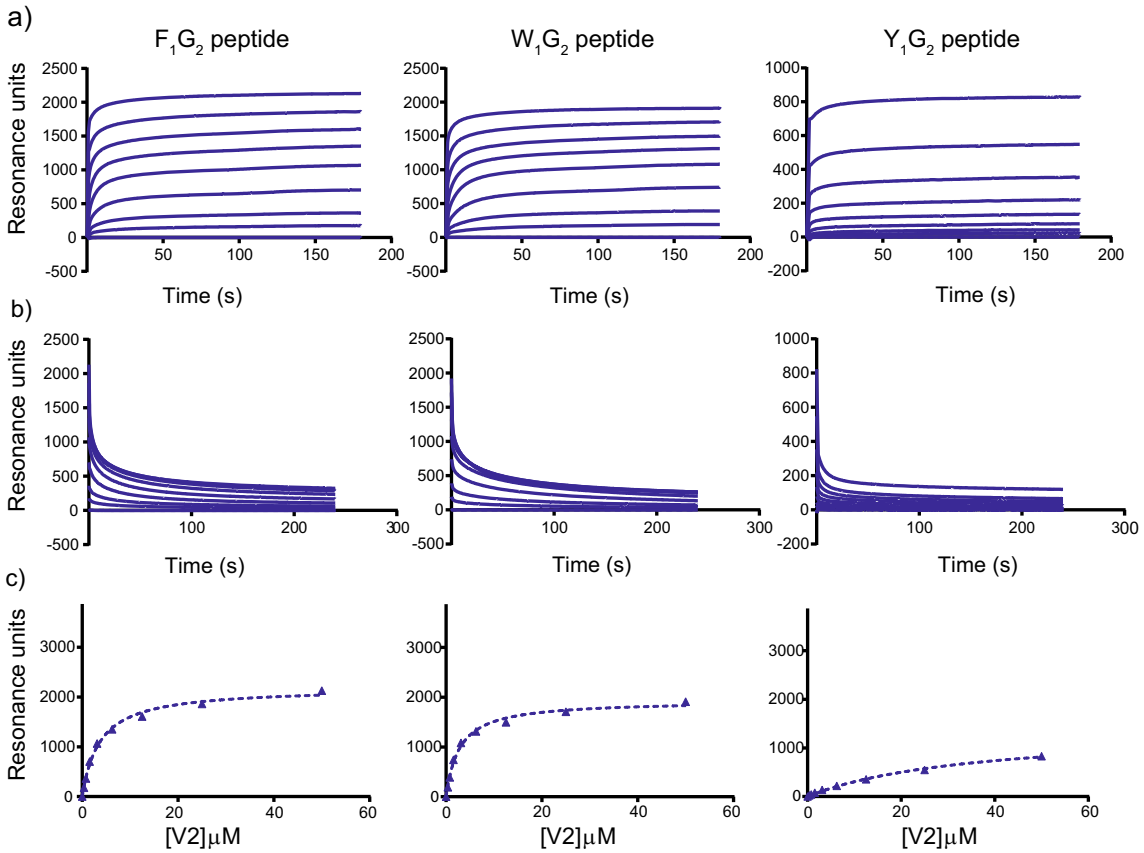

Fig. 4 **a** The association curves from an SPR experiment performed with the peptide of interest attached via biotin to a streptavidin chip, and V2 in solution at concentrations 0, 0.39, 0.78, 1.56, 3.13, 6.25, 12.5, 25, and 50 μM. **b** The dissociation curves for the same SPR experiment. **c** The steady-state response from the SPR association curves are plotted versus the concentration of V2 for three different peptides. The dashed lines indicate the fits used to calculate the stead-state K_D

this position was replaced with Met in a "back-to-bacterial consensus" mutant and the canonical bacterial binding affinity was recapitulated (Colombo et al. 2018). Thus, the V1 binding pocket is more similar to the bacterial ClpS1 proteins than its parent *A. tumefaciens* ClpS2 protein in this respect.

Finally, as the SPR experiment gives not only the steady-state K_D but also the on and off rates, it was determined that the main contribution to the improved K_D resulted from a slower off rate for the mutant proteins. It is noticeable from the SPR sensorgrams shown in Fig. 4b that the slope of the k_{off} curve is less steep for the peptide with NAA Phe than for the peptide with NAA Tyr, thus leading to the improved K_D. Note the difference in scale for the Tyr dissociation curve as compared to the Phe and Trp curves. Similarly, the dissociation rate of V1 is significantly slower than that for the wild-type protein (see Supplemental Fig. S7 and Supplemental Fig. S8). From the dissociation curves, the k_{off} was calculated for each protein with each NAA bearing peptide and the data is displayed in Table 3. The dissociation rate for variant 1 has slowed 9.4-fold over the wild-type protein for Phe; while, the variant 2 dissociation rates have slowed 27.9 and 51.3-fold over wild-type for Phe and Trp, respectively.

Table 3 Dissociation rate $k_{off}(s^{-1})$ for different NAAs

Variant	Phe	Trp	Tyr
WT	0.93	1.32	1.16
V1	0.1	0.76	0.5
V2	0.03	0.03	0.05

Discussion

In addition to advancing the field of proteomics, rapid protein/peptide sequencing would facilitate protein engineering by eliminating the currently required need to maintain a genotype-phenotype linkage, as is fulfilled by using yeast or phage display. Similarly, in combinatorial peptide drug library screening, the ability to sequence peptides would eliminate the need for barcoding strategies that are currently used. Currently, there is no single molecule protein sequencing technology available; however, many technologies are being proposed and repurposed to meet this challenge. For example, FRET pairs are being used to tag residues and detect amino

acids in a technique called "single-molecule peptide fingerprinting" that can be applied when the peptide is translocated through the Clp protease molecular machinery attached to a surface (Yao et al. 2015; van Ginkel et al. 2018). In an alternative approach, the ionic current fluctuation pattern within nanopores can be detected to directly discriminate between different groups of amino acids (Asandei et al. 2017). Engineered nanopores can also be used as a potential way to cleave a single amino acid, and, combined with mass spectrometry, identify that amino acid (Bush et al. 2017). Additionally, it has been proposed that partial sequencing, or fluorosequencing can be accomplished by immobilizing a peptide to a surface such that single-molecule fluorescence can be measured (Swaminathan et al. 2015). In this measurement modality, the fluorescence detection can be accomplished by either chemically modifying the amino acids based on their unique side chain chemistry, or by using a fluorescently labeled amino acid binding reagent. The approach of chemical modification has been successfully employed to detect cysteine and lysine residues (Swaminathan et al. 2019). The specificity and affinity requirements of a potential NAAB is dependent on the limits of detection of the technology being employed; however, some groups are exploring the theoretical limitations to which lower affinity NAABs can still potentially be useful reagents (Rodriques et al. 2018).

This work shows that NAABs that are enhanced in specificity and affinity for some NAAs can indeed be engineered from ClpS proteins, which natively recognize some NAAs, using directed evolution and yeast-based FACS screening methods. The initial directed evolution design targeted all of the ClpS residues in a random fashion, rather than relying on structural insights into the exact location of the binding pocket. However, many of the residues that were ultimately found to be mutated were indeed in or near the binding pocket and recapitulate properties that could perhaps have been rationally designed from critical assessment of the crystal structures available in the literature. The strategy of targeting the residues that frequently occurred after the first set of selections with a more in depth, focused library testing each of the 20 residues at these positions resulted in more hits during subsequent rounds of selection. However, there is still the possibility of evolving these proteins further, for enhanced selectivity and affinity for the Trp or Phe NAAs, or for other amino acid targets such as Tyr and Leu, by combining the mutations found in this work with those that may be found among the different homologs in the literature with known differences in specificity. Additionally, in a peptide sequencing context, ideally, the second amino acid must not significantly affect the binding of the NAAB or false positives could occur. The finalized NAAB reagents will ultimately need to be characterized against a panel of peptides to ensure that the neighboring sequence does not affect the sequencing fidelity, as has been done for the some of the wild-type ClpS proteins (Stein et al. 2016). Although there is a significant amount of characterization of this family of proteins in the literature, it has previously been studied in ways which focus on the properties of ClpS that contribute to the selection of substrates for degradation within a cell. As with any engineering effort, one must strive to understand as much as possible about a system in order to engineer it to perform a new or different task. From the standpoint of using ClpS homologs as potential tools or binding reagents, the criteria and attributes by which protein engineers assess this protein are beginning to evolve. For instance, assessing the ability of ClpS to bind non-standard amino acids that it would not normally encounter in nature contributed to the use of the protein to detect non-standard amino acid incorporation (Kunjapur et al. 2018). In our work, the finding that the ClpS mutants with higher NAA target affinities are driven primarily by reduced rates of dissociation (k_{off}) is encouraging from a binding reagent standpoint. It poises these mutant proteins as potentially more suitable binding reagents for NAA detection in the development of fluorescence-based next-generation sequencing technologies as the slowed rate of dissociation theoretically allow for a longer residence time for fluorescence-based imaging of the NAAB.

Although the yeast display system used in this study is robust and rapid, it does have the drawback that mutant proteins were not directly selected for reduced dissociation rates, but rather for improved overall affinity. Furthermore, the binding reaction occurs in solution, rather than with the peptide affixed to a surface, as would likely be the case in a peptide sequencing setting. However, as the SPR-based experiments were performed with the peptide affixed to a surface, we have shown that ClpS can clearly bind the peptides in a surface-attached configuration. Theoretically, additional detection reagents could be selected in a similar directed evolution manner as the proof-of-principle reagents described herein, and/or by rational design using these initial successful mutations as a guide. The yeast display system should be broadly applicable for the development of additional NAABs in combination with further protein engineering to ensure efficient activity in vitro in the surface adhered peptide context, fluorescent labeling, and stability to the length of incubation times, temperatures, and buffer conditions that may be necessary for the sequencing detection.

Acknowledgements The authors thank Courtnay Brand (IBBR/UMD) for synthesizing all the peptides used in this study; Kenneth Class of the UMD MPRI Flow Cytometry Core Facility; and student interns Mildred Devereux, Zaman Nasser, Eleanor Kelman, Caleb Weddington, Morgan O'Brien, Makenzie Christensen, and Kunal Dharmadhikari.

Certain commercial equipment, instruments, or materials are identified to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Funding Jennifer Tullman was supported by a National Institute of Standards and Technology (NIST) National Research Council (NRC) postdoctoral fellowship. This work was supported by a NIST Innovation in Measurement Science Award.

Compliance with ethical standards

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest A provisional patent application has been submitted based on some of the reagents developed in this work.

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

References

Ahyoung AP, Koehl A, Vizcarra CL, Cascio D, Egea PF (2016) Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen *Plasmodium falciparum*. Protein Sci 25:689–701. https://doi.org/10.1002/pro.2868

Asandei A, Rossini AE, Chinappi M, Park Y, Luchian T (2017) Protein nanopore-based discrimination between selected neutral amino acids from polypeptides. Langmuir 33:14451–14459. https://doi.org/10.1021/acs.langmuir.7b03163

Bachmair A, Finley D, Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234:179–186. https://doi.org/10.1126/science.3018930

Baker MS, Ahn SB, Mohamedali A, Islam MT, Cantor D, Verhaert PD, Fanayan S, Sharma S, Nice EC, Connor M, Ranganathan S (2017) Accelerating the search for the missing proteins in the human proteome. Nat Commun 8:1–13. https://doi.org/10.1038/ncomms14271

Bush J, Maulbetsch W, Lepoitevin M, Wiener B, Mihovilovic Skanata M, Moon W, Pruitt C, Stein D (2017) The nanopore mass spectrometer. Rev Sci Instrum 88:113307. https://doi.org/10.1063/1.4986043

Colombo CV, Rosano GL, Mogk A, Ceccarelli EA (2018) A gatekeeper residue of ClpS1 from *Arabidopsis thaliana* chloroplasts determines its affinity towards substrates of the bacterial N-end rule. Plant Cell Physiol 59:624–636. https://doi.org/10.1093/pcp/pcy016

Dougan DA, Reid BG, Horwich AL, Bukau B (2002) ClpS, a substrate modulator of the ClpAP machine. Mol Cell 9:673–683. https://doi.org/10.1016/S1097-2765(02)00485-9

Erbse A, Schmidt R, Bornemann T, Schneider-Mergener J, Mogk A, Zahn R, Dougan DA, Bukau B (2006) ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. Nature 439:753–756. https://doi.org/10.1038/nature04412

Faulkner S, Dun MD, Hondermarck H (2015) Proteogenomics: emergence and promise. Cell Mol Life Sci 72:953–957. https://doi.org/10.1007/s00018-015-1837-y

Feldhaus JM, Siegel RW, Feldhaus JM, Wittrup KD (2003) Yeast display scFv antibody library user's manual. Pacific Northwest National Laboratory, Richland

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. https://doi.org/10.1038/nrg.2016.49

Graciet E, Wellmer F (2010) The plant N-end rule pathway: structure and functions. Trends Plant Sci 15:447–453. https://doi.org/10.1016/j.tplants.2010.04.011

Hawley TS, Hawley RG (2004) Flow cytometry protocols. Humana Press, Totowa, NJ

Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LDN, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TSK, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A (2014) A draft map of the human proteome. Nature 509:575–581. https://doi.org/10.1038/nature13302

Kunjapur AM, Stork DA, Kuru E, Vargas-Rodriguez O, Landon M, Söll D, Church GM (2018) Engineering posttranslational proofreading to discriminate nonstandard amino acids. Proc Natl Acad Sci U S A 115:619–624. https://doi.org/10.1073/pnas.1715137115

Lupas AN, Koretke KK (2003) Bioinformatic analysis of ClpS, a protein module involved in prokaryotic and eukaryotic protein degradation. J Struct Biol 141:77–83. https://doi.org/10.1016/S1047-8477(02)00582-8

Nishimura K, van Wijk KJ (2015) Organization, function and substrates of the essential Clp protease system in plastids. Biochim Biophys Acta 1847:915–930. https://doi.org/10.1016/j.bbabio.2014.11.012

Nishimura K, Asakura Y, Friso G, Kim J, Oh S-H, Rutschow H, Ponnala L, van Wijk KJ (2013) ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in *Arabidopsis*. Plant Cell 25:2276–2301. https://doi.org/10.1105/tpc.113.112557

Perona JJ, Hadd A (2012) Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. Biochemistry 51:8705–8729. https://doi.org/10.1021/bi301180x

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–W324. https://doi.org/10.1093/nar/gku316

Rodriques S, Marblestone A, Boyden E (2018) A theoretical analysis of single molecule protein sequencing via weak binding spectra. bioRxiv:1–25. https://doi.org/10.1101/352310

Román-Hernández G, Hou JY, Grant RA, Sauer RT, Baker TA (2011) The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. Mol Cell 43:217–228. https://doi.org/10.1016/j.molcel.2011.06.009

Schuenemann VJ, Kralik SM, Albrecht R, Spall SK, Truscott KN, Dougan DA, Zeth K (2009) Structural basis of N-end rule substrate recognition in *Escherichia coli* by the ClpAP adaptor protein ClpS. EMBO Rep 10:508–514. https://doi.org/10.1038/embor.2009.62

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

Swaminathan J, Boulgakov AA, Marcotte EM (2015) A theoretical justification for single molecule peptide sequencing. PLoS Comput Biol 11:1–26. https://doi.org/10.1371/journal.pcbi.1004080

Swaminathan J, Boulgakov AA, Hernandez ET, Bardo AM, Bachman JL, Marotta J, Johnson AM, Anslyn EV, Marcotte ME (2019, 2018) Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. Nat Biotechnol. https://doi.org/10.1038/nbt.4278

Tan JL, Ward L, Truscott KN, Dougan DA (2016) The N-end rule adaptor protein ClpS from *Plasmodium falciparum* exhibits broad substrate specificity. FEBS Lett 590:3397–3406. https://doi.org/10.1002/1873-3468.12382

Tryggvesson A, Ståhlberg FM, Töpel M, Tanabe N, Mogk A, Clarke AK (2015) Characterization of ClpS2, an essential adaptor protein for

the cyanobacterium *Synechococcus elongatus*. FEBS Lett 589: 4039–4046. https://doi.org/10.1016/j.febslet.2015.11.026

Tyers M, Tyers M, Mann M, Mann M (2003) From genomics to proteomics. Nature 422:193–197. https://doi.org/10.1038/nature01510

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

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

Wang KH, Oakes ESC, Sauer RT, Baker TA (2008a) Tuning the strength of a bacterial N-end rule degradation signal. J Biol Chem 283: 24600–24607. https://doi.org/10.1074/jbc.M802213200

Wang KH, Roman-Hernandez G, Grant RA, Sauer RT, Baker TA (2008b) The molecular basis of N-end rule recognition. Mol Cell 32:406–414. https://doi.org/10.1016/j.molcel.2008.08.032

Yao Y, Docter M, van Ginkel J, de Ridder D, Joo C (2015) Single-molecule protein sequencing through fingerprinting: computational assessment. Phys Biol 12:55003. https://doi.org/10.1088/1478-3975/12/5/055003