The Locational Impact of Site-Specific PEGylation: Streamlined Screening with Cell-free Protein Expression and Coarse-grain Simulation.

Kristen M. Wilding, Addison K. Smith, Joshua W. Wilkerson, Derek B. Bush, Thomas A. Knotts IV, Bradley C. Bundy

Department of Chemical Engineering, Brigham Young University, Provo, Utah

ABSTRACT: Although polyethylene glycol (PEG) is commonly used to improve protein stability and therapeutic efficacy, the optimal location for attaching PEG onto proteins is not well understood. Here, we present a cell-free protein synthesis-based screening platform which facilitates site-specific PEGylation and efficient evaluation of PEG attachment efficiency, thermal stability, and activity for different variants of PEGylated T4 lysozyme, including a di-PEGylated variant. We also report developing a computationally-efficient coarse-grain simulation model as a potential tool to narrow experimental screening candidates. We use this simulation method as a novel tool to evaluate the locational impact of PEGylation. Using this screen, we also evaluated the predictive impact of PEGylation site solvent accessibility, conjugation site structure, PEG size, and double PEGylation. Our findings indicate that PEGylation efficiency, protein stability, and protein activity varied considerably with PEGylation site, variations which were not well predicted by common PEGylation guidelines. Overall our results suggest current guidelines are insufficiently predictive, highlighting the need for experimental and simulation screening systems such as the one presented here.

Keywords: cell-free protein synthesis, unnatural amino acid, site-specific, PEG, PEGylation, coarse-grain simulation

Since its invention in the 1970's, PEGylation has proved to be a valuable tool for pharmaceutical applications.[10-13] Several PEGylated therapeutics are already available for clinical use including the top 10 selling pharmaceutical Neulasta.[11, 14, 15] PEGylated therapeutics are reported to have improved pharmacokinetics and reduced immunogenicity, due to slower renal filtration and increased resistance to degradation and aggregation.[2, 6, 9, 16, 17] Enzyme biocatalysts could similarly benefit from PEGylation through improved stability and greater hydrodynamic radius, leading to improved recoverability and retention in matrices.[14, 18] However, commercially available PEGylated proteins to date are non-specifically PEGylated, targeting multiple natural residues such as lysine or cysteine,[1, 19] or are produced by targeting of naturally occurring, uniquely reactive sites, such as the N- or C- terminus,[15, 20] disulfide bonds,[25] or at less prevalent natural amino acids such as cysteine which have been mutated into the protein.[26, 27] Because these techniques limit the sites available for targeting, the tethering locations on the protein may be in suboptimal locations such that conjugation would hinder protein stability or important protein-protein interactions,[20] both of which can dramatically reduce the protein's activity. In some

proteins, application of these techniques may also require extensive mutagenesis in order to prevent undesired PEGylation where the targeted moiety occurs naturally in the protein.

Site-specific insertion of unnatural amino acids (uAA) via stop codon suppression offers the ability to conjugate a protein at potentially any site with minimal mutation.[30-32] The wide variety of uAA with useful side-chain chemistries adds a flexibility to uAA-based conjugation which makes it a powerful tool for producing optimally PEGylated proteins. However, the optimal site for PEG conjugation is not well understood and activities of different analogs can vary greatly.[31, 33] In order to mitigate the costs associated with development of optimized PEGylated proteins, improved guidelines are necessary to inform conjugate design.

E. coli-based cell-free protein synthesis (CFPS) offers an ideal platform for rapidly and economically screening various sites for uAA incorporation[34, 35] and PEGylation, providing a promising tool for both developing guidelines to inform PEGylated protein design and identifying optimal PEGylated proteins from a pool. The flexibility of the CFPS has enabled cell-free synthesis of a wide variety of challenging proteins, including cytotoxic proteins,[36] disulfide-bonded therapeutics,[37] virus-like particles,[38-40] proteins requiring chaperones,[41] and antibodies.[30] Although the *E. coli*-based cell-free system is currently unable to replicate mammalian glycosylation patterns, recent advances in engineering glycosylation into the system and advances in uAA incorporation are promising developments towards this end. Here we present a cell-free based approach that allows rapid, scalable assessment of PEGylated proteins, enabling optimization of PEGylated proteins in a more time- and resource-efficient manner.

As a proof-of-concept study, we apply our system to evaluate existing guidelines for PEGylated protein design, specifically (1) PEG size and number as a factor in PEGylation efficiency,[10] protein stability,[1-5, 9] and protein activity,[1, 6-8] (2) PEGylation site solvent accessibility as a predictor of efficient PEGylation,[21-24] and (3) flexible loops as preferred conjugation sites to minimize stability and activity loss.[13, 28, 29] Finally, to further enhance the screening process, we also develop a coarse-grain molecular simulation to inform candidate site selection and demonstrate the utility of the simulation for coarse PEGylation site assessment.

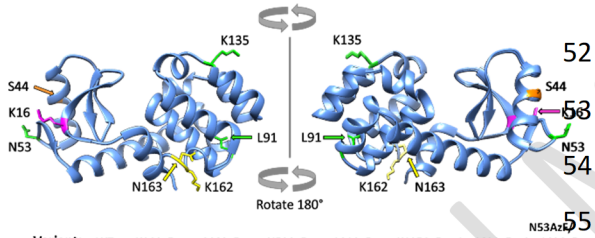
1 RESULTS AND DISCUSSION

1.1 CFPS-Based Screening System

In this study we first seek to demonstrate the utility of a cell-free expression system to quickly assess the impact of PEG size and conjugation site on the properties of a target protein. Using CFPS as opposed to *in vivo* methods reduced resource costs and accelerates protein preparation from weeks to days by (1) eliminating cell culturing of each protein variant, enabling protein synthesis in ~8 hours, (2) allowing microliter scale protein synthesis reactions for efficient reagent use, and (3) simplifying protein purification. Additionally, in less than 5 hours and with less than 2 ug of each protein sample, we are able to screen PEGylation efficiency and thermal stability in parallel with protein activity. We selected T4 lysozyme (T4 Lyz, EC 3.2.1.17) as a model enzyme for this study, and at the conclusion of our study, we evaluate our results against several generic PEGylation guidelines.

A ASA-View Relative Solvent Accessible Surface Area

K16	S44	N53	L91	K135	Ins163*
0.588	0.597	0.94	0.03	0.71	0.77



B

Variant:	WT	K16AzF	S44AzF	N53AzF	L91AzF	K135AzF	Ins163AzF	N53AzF/ Ins163AzF
pAzF	-	- +	- +	- +	- +	- +	- +	-
T4 Lysozyme								
Yields (ug/mL)	650 ± 90	160 ± 40	240 ± 60	490 ± 170	500 ± 90	450 ± 40	500 ± 170	280 ± 90

Figure 1 - T4 Lysozyme variants and yields. A: ASA-View relative solvent accessible surface areas, normalized to the average surface area of each amino acid in 30 random configurations of a Gly-X-Gly tripeptide (*SASA reported for Ins163 insertion site is an average of the ASA-View values for the SASA of K162 and N163), and Chimera image showing locations of these residues in T4 Lyz (showing PDB ID code 2LZM). Residues depicted in green are part of unstructured loops, while structured sites are represented in pink or orange (pink for beta sheet, orange for alpha helix). The residues surrounding the insertions site are shown in yellow, and are also part of an unstructured loop; B: CFPS Yields and Autoradiogram. Full-length T4 Lyz CFPS yields and sample autoradiogram. Data shown are averages and standard deviations of n = 2-5 separate reactions for all variants.

1.1.1 CFPS of Screening Pool

Six sites were selected for uAA p-azido-L-phenylalanine (AzF) incorporation, spanning a range of solvent accessible surface areas (SASA) and structures. The selected sites are as follows: substitutions at sites K16, S44, N53, L91, and K135, and an amber stop insertion between residues K162 and N163 (Ins163). Locations of the selected sites along with their corresponding relative SASA values, as scored by ASA-View,[42] are reported in Figure 1A. Plasmids were constructed to incorporate the amber stop codon, TAG, at each of these sites, as detailed in the Materials and Methods.

We also constructed a double amber suppression variant, as di-PEGylation may further stabilize proteins compared to mono-PEGylation. [9, 43] Di-PEGylation may also provide additional advantages for therapeutic proteins, such as reduced vacuole

formation in the kidneys.[44] We constructed a double amber suppression variant with sites N53 and Ins163 (N53/Ins163), because they have high solvent accessibility and are located on opposite sides of the protein.

Using an *E. coli*-based cell-free system, AzF was incorporated as reported previously,[45] resulting in the production of the following lysozyme variants: K16AzF, S44AzF, N53AzF, K135AzF, Ins163AzF, and N53AzF/Ins163AzF. As our lab and others have observed previously,[31, 34, 46] yields of full-length lysozyme varied depending on the location of uAA incorporation. Average yields fell between 160-500 ug/mL (Figure 1B) – exceeding 25% of WT yield in all cases – and very low levels of full-length T4 Lyz expression were observed in the absence of AzF. The yield of the double-AzF-modified protein was 0.28 mg/mL, significantly exceeding previously reported yields for double AzF-incorporation via amber suppression.[8, 23, 29] The ability to achieve higher yields in hours, even for double amber suppression, is a primary advantage of the CFPS expression system as a basis for a conjugate screening platform. Full-length AzF-incorporated variants were then rapidly purified using spin columns, as detailed in the Materials and Methods.

Using the cell-free system enabled sufficient protein yields for our screen from ≤ 400 uL reactions in just 8+ hours with minimal optimization. As such, CFPS reduces time and resource costs of the screen by eliminating the need for cell culturing of each construct. This system also simplifies purification by eliminating the lysis and clarification steps between expression and purification, allowing direct addition of the expression reaction to a spin column for rapid, small-scale purifications.

1.1.2 PEGylation Efficiency Screening

Purified T4 Lyz variants were PEGylated with 5kDa and 20kDa PEG using a strain-promoted azide-alkyne cycloaddition (SPAAC) reaction. The SPAAC reaction, illustrated in Figure 2A, is advantageous because it can be done rapidly and at physiological conditions without additional components such as protecting groups or copper catalysts. [45, 47, 48] SPAAC reactions were performed using 5 uM purified T4 Lyz variant and 20 or 50 equivalents, respectively, of 20kDa or 5kDa PEG with a dibenzocyclooctyne (DBCO) terminal group providing the strained alkyne DBCO-mPEG. The DBCO-mPEG equivalents were doubled for PEGylation of N53AzF/Ins163AzF. These conditions provided high click efficiency at most of the chosen sites.

Average PEGylation efficiency is shown in Figure 2B as calculated by SDS-PAGE electrophoresis, autoradiography, and densitometry (Supplemental Figure S1). While the results reported here are from autoradiography of the protein gels in order to insure that PEG did not interfere with staining, densitometry

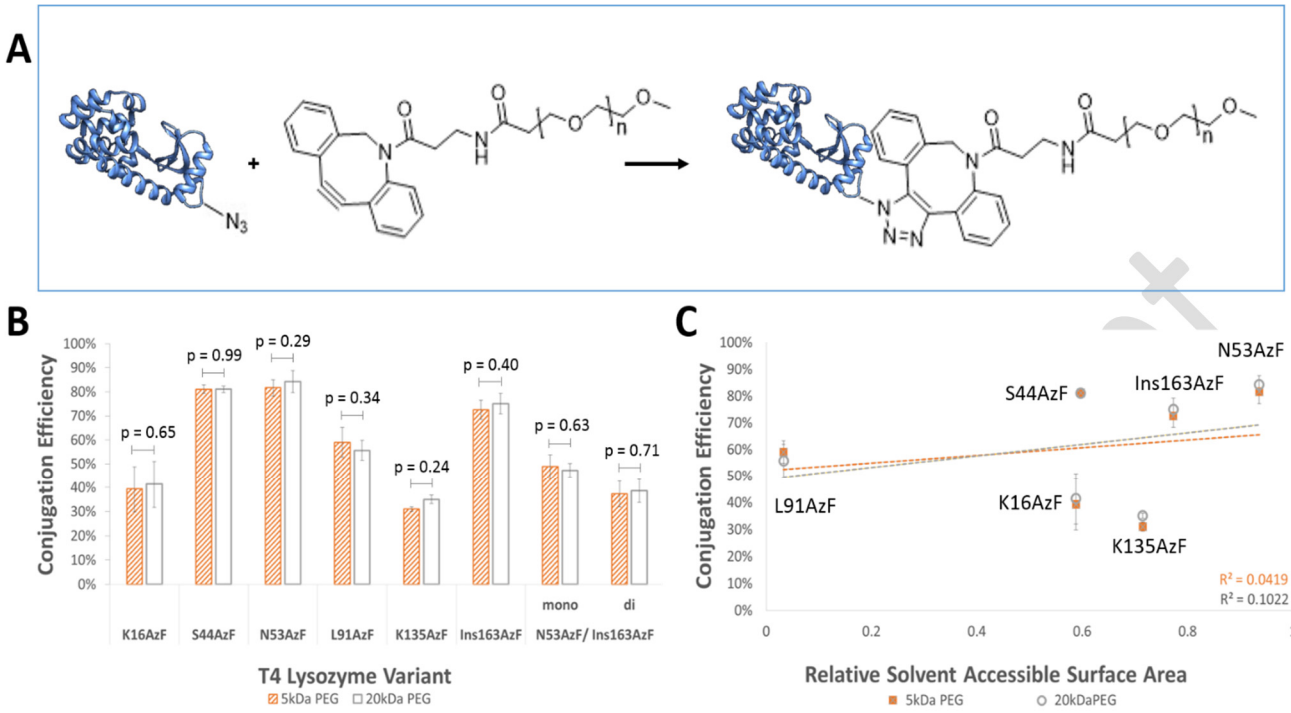


Figure 2 – SPAAC reaction scheme & efficiency. A: SPAAC Reaction between DBCO-mPEG and AzF-substituted T4 Lysozyme; B: SPAAC Efficiency of T4 lysozyme variants (n = 2+, error bars represent one standard deviation); C: SPAAC Efficiency vs Relative SASA, note that relative SASA is not linearly correlated with conjugation efficiency in this study (n = 2+, error bars represent one standard deviation. When not visible, error bars are hidden under marker).

can also be done directly using the protein gels with very similar results, due to the high purity of the samples. As alluded to previously, the conjugation sites selected in this study were chosen in part to evaluate PEG size and conjugation site accessibility, which are common guidelines for the design of PEGylated proteins with high conjugation efficiency. [10, 21-24] The results of our screen revealed no significant impact of PEG size on click efficiency with the 5kDa and 20kDa unbranched-PEG molecules employed in our study, even for the di-PEGylated N53AzF/Ins163AzF (Figure 2B). This result suggests that the accessibility of the site to the larger DBCO group plays a stronger role in conjugation efficiency than the 15kDa increase in size of the flexible PEG chain. Indeed, the impact of PEG size on PEGylation efficiency may be influenced by the conjugation mechanism, and this possible relationship should be considered when optimizing PEGylation location. We also observed that while the fraction of unmodified N53AzF/Ins163AzF following SPAAC reaction is on par with that of N53AzF and Ins163AzF, the efficiency of dual PEGylation is lower than would be expected based on the conjugation efficiencies of each site individually (expected: 62%, 63%). These results suggest that steric hindrance from the first PEG chain may inhibit conjugation with the second PEG chain, despite the relatively large distance between the sites, and is an important factor when

considering the advantages of multi-PEGylation verses mono-PEGylation with longer PEG. This phenomenon is expected based on observations for dual-PEGylation using natural amino acids.[44]

Using our screen, we also found that higher SASA does not necessarily correlate with higher conjugation efficiency (Figure 2C). Using Chimera,[49] we also evaluated the local surface hydrophobicity of these sites, a trait which has been suggested to improve SPAAC efficiencies via interaction with the hydrophobic DBCO group.[24, 50] Again, we found no clear correlation (Supplementary Figure S2), suggesting that neither SASA or surface hydrophobicity are sufficient predictors of high-efficiency conjugation sites.

The discrepancies between common design guidelines and our results as described above further demonstrate the need to rapidly screen multiple PEGylation sites in parallel. Importantly, the small-scale expression reactions and low concentration conjugation reactions described above mitigate screening costs by reducing the amount of product needed.

1.1.3 Stability Screening

A primary motivation for PEGylating proteins is to improve both thermal stability and protease resistance.[1, 9, 51] Hence, stability evaluation is a key step in the conjugate screening process. Stability against protease degradation has previously been shown to correlate with protein conformational stability.[9] Thus, in our screening approach we use changes in protein melting temperature to characterize the stability of the screened proteins and corresponding conjugates. Thermal shift assays provide a higher-throughput, more cost-effective method for assessing protein melting temperature when compared to more traditional methods such as differential scanning calorimetry by allowing rapid, accurate T_m characterization with small samples and low-cost reagents. Here, we use a protein thermal shift assay to evaluate stability of our screening pool using only 1-1.5 ug protein from each unpurified SPAAC reaction (3 replicates with ~0.3-0.5 ug of protein per replicate) in under 5 hours at less than $0.30 per screened protein or conjugate (~ $ 0.09 per replicate). The Protein Thermal Shift assay requires three orders of magnitude less protein at an order of magnitude lower protein concentration than traditional differential scanning calorimetry.[52]

Using the Protein Thermal Shift assay, control T_m values for WT, WT+PEG5kDa, and WT+PEG20kDa were determined to be 61.0°C ± 0.10, 61.1°C ± 0.81, and 60.7°C ± 0.39, respectively, at a pH of 7.4, which agrees well with values previously reported for WT T4 lysozyme in literature.[53-57] The change in T_m due to incorporation of AzF or PEGylation at the incorporated AzF residue(s) relative to the corresponding control WT T_m is shown as ΔT_m in Figure 3A ($\Delta T_m = T_{m,raw} - T_{m,control}$, where $T_{m,control}$ is

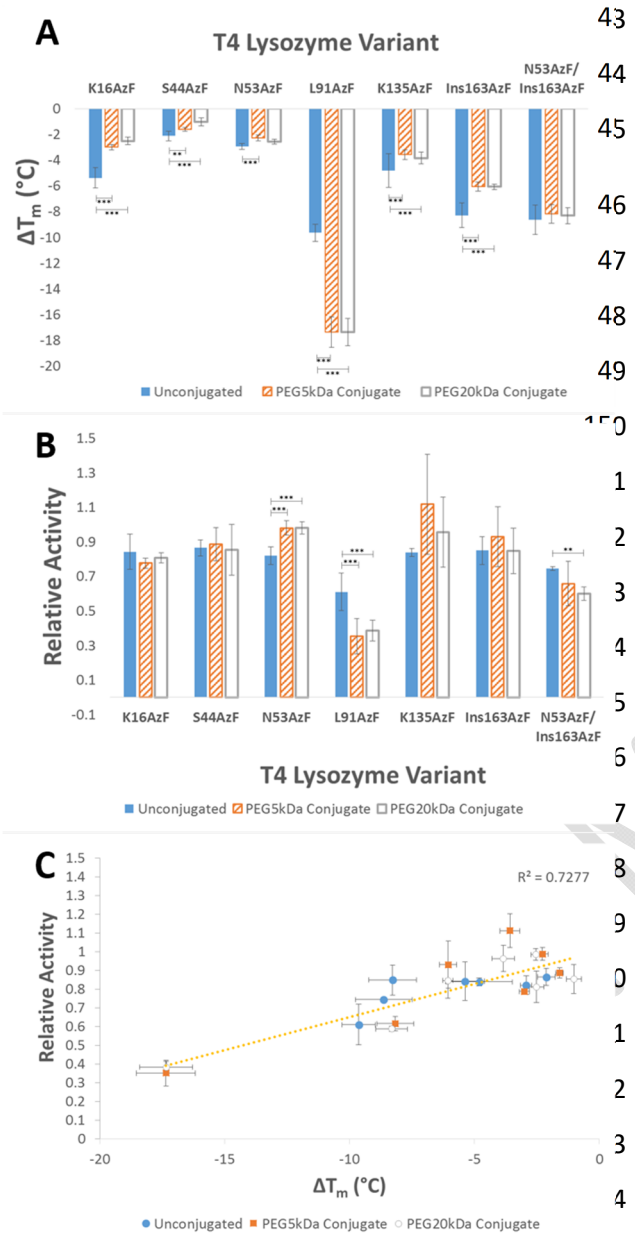


Figure 3 – T4 Lyz stability and activity. For A and B stars indicate significant differences, where *** = $p < 0.001$ and ** = $p < 0.01$. A: Thermal stability of T4 Lyz variants and conjugates. ΔT_m as determined from the Protein Thermal Shift Assay shown with one standard deviation, where n = 3-18 from 2-6 separate experiments; B: Relative Activity of T4 Lyz variants and conjugates. Data shown as averages and standard deviations of n = 3-11 from 3 separate experiments; C: Relative Activity vs ΔT_m. Data from A and B plotted against each other. Note the positive linear trend, which indicates that stability may potentially be used as a screen to eliminate less active conjugates before activity testing.

the T_m for WT, WT+PEG5kDa, or WT+PEG20kDa for the unconjugated, PEG5kDa conjugates, and PEG20kDa conjugates, respectively).

For all seven lysozyme variants, the magnitude of ΔT_m associated with AzF incorporation was greater than the magnitude of the change in ΔT_m as a result of PEGylation (Figure 3A, blue bars vs the difference between blue bars and either orange striped or white bars. This change in ΔT_m between the conjugated and unconjugated T4 Lyz variants will be referred to as $\Delta T_{m,conj}$, where $\Delta T_{m,conj} = \Delta T_{m,PEGylated} - \Delta T_{m,unconjugated}$). The observed impact on stability of even a single amino acid change demonstrates the challenge of engineering conjugation sites into proteins, as even minimal mutation can cause significant impact on protein stability. This further highlights the utility of the presented CFPS-based conjugate screening approach, as cell-free systems have been used to corporate a variety of uAAs,[45] providing an opportunity to minimize the impact on stability through optimization of the choice of uAA.

PEGylation had the greatest stabilizing impact on K16AzF and Ins163AzF, but modestly improved the stability at all mono-AzF T4 Lyz variants with the exception of L91AzF. Slight improvements in stability following double PEGylation of N53AzF/Ins163AzF were not significant, indicating that in this case stabilizing benefits did not compound

as previously reported.[9] The varying impact of PEGylation on T_m for the different T4 Lyz variants was also not well predicted by SASA (Figure S3B).

In all cases, increasing PEG size from 5kDa to 20kDa did not affect the protein stability (Figure 3A), which agrees with many previous reports.[3-5] Studies which have found increased conjugate thermal stability with increasing PEG size have focused on smaller PEG oligomers,[9, 58] which could explain the difference to the present study (see Supplemental Information for a more detailed discussion).

Our data also indicates that unstructured sites are not necessarily preferable to structured sites, either for uAA substitution or for PEGylation. Indeed, the site which was most tolerant of the AzF substitution was S44, which is located inside an alpha helix. By contrast, the least tolerant site, L91, is in an unstructured loop just outside an alpha helix. Similarly, the K16AzF variant was most stabilized by PEGylation, despite its location in a beta sheet. In contrast, PEGylation at an alpha helix (S44AzF) or at less structured sites was less stabilizing.

The dramatic destabilization from PEGylation at site L91 is interesting, given that this site was determined to be the optimal site for lysozyme immobilization.[28, 59] These results indicate that the effects on protein stability from conjugation to a polymer differ significantly from the effects due to conjugation to a surface. A more detailed discussion of these differences is contained in the Supplemental Materials.

Overall, our results highlight the need for screening approaches such as the one presented herein due to the lack of predictive factors for PEGylation-based stabilization. The CFPS-based screening approach provides a method for rapid, cost-effective screening of the effect of PEGylation at a variety of sites on protein thermal stability.

1.1.4 Activity Screening

Retained protein activity is another essential metric in the design of PEGylated protein. Fortunately, a reduction in specific activity following PEGylation can be compensated for, in many cases, by a corresponding increase in stability.[10] However, minimizing the negative effects of PEGylation on activity would require less PEGylated protein in its final application, reducing costs and negative side effects of therapeutics. As PEG size and number have been previously reported to impact the activity of PEGylated proteins, we applied our screen to evaluate these effects on T4 Lyz activity. [1, 6-8] The activity of each T4 Lyz variant and its PEG5kDa and PEG20kDa conjugates was determined using the Enzchek Lysozyme Assay (ThermoFisher Scientific) as reported in Figure 3B.

The activities of unconjugated K16AzF, S44AzF, N53AzF, K135AzF, and Ins163AzF were not statistically different ($p > 0.05$). However, all are greater than the activities of L91AzF and Ins163AzF. These results suggest that, like stability, activity of AzF substituted variants are not well predicted by PEGylation site secondary structure. The reduction in the activity of the unconjugated N53AzF/Ins163AzF was less than the sum of the reductions from the two individual AzF incorporations (N53AzF and Ins163AzF), suggesting that the activity losses in multi-uAA incorporations should not be considered additive. L91AzF had the lowest activity of any unconjugated T4 Lyz variant, including the double suppression variant, at only 61% WT activity.

Of the six mono-PEGylated lysozyme variants tested, only PEGylation of N53AzF clearly increased the activity. While unexpected, improved activity following PEGylation is not without precedent.[2, 60] PEGylation at both the structured sites K16AzF and S44AzF and at the unstructured sites K135AzF and Ins163AzF did not significantly change the activity compared to the unconjugated variants. This suggests that the effect of PEGylation on activity is not well predicting by conjugation site secondary structure.

In stark contrast to the other 5 mono-PEGylated variants, PEGylation of L91AzF resulted in an additional loss of activity, with conjugates retaining only about 60% of the activity of unconjugated L91AzF. This decrease in activity correlates with the significant decrease in stability seen as a result of PEGylation of L91AzF. There may be a general correlation between high activity retention and SASA, given the reduction in activity following PEGylation of L91AzF, minimal impact of PEGylation on activity for K16AzF and S44AzF, and improvement in activity following PEGylation of N53AzF. However because only PEGylation of L91AzF and N53AzF significantly changed the activity, the utility of the trend in predicting sites with the highest activity retention is uncertain (Figure S4). Still, SASA could be a useful guideline in narrowing potential PEGylation sites to exclude sites with very low SASA.

In addition, activity of diPEGylated N53AzF/Ins163AzF decreased with 20kDa PEG. As neither PEGylation of N53AzF nor Ins163AzF with 20kDa PEG had a negative effect on activity, and because PEGylation of the double ambers suppression variant had no significant effect on the stability of the variant, this decrease in activity can reasonably be attributed to reduced accessibility of the active site to the macromolecular substrate (Micrococcus lysodeikticus cell walls) due to steric hindrance from the two large PEG chains.

In all cases, there was no significant difference between the activities of the 5kDa PEG conjugates vs the 20kDa PEG conjugates. PEG size has previously been reported to be negatively correlated with in vitro

activity,[1, 6, 7, 61] but other studies have also shown that in some cases activity is independent of PEG size.[62, 63] For example, another study in which IFN was PEGylated with 5kDa, 10kDa, and 20kDa PEG at a disulfide bridge and the activity impact was found to be size independent.[63]

Literature also reports that point mutations in a protein have minimal, localized impact on the structure,[56] so, given their distance from the active site, it is less likely that any of these mutations impact activity by deforming the active site. Thus, it is reasonable that any changes in activity would be related to changes in the dynamics of the protein. In analyzing the data in this study, there indeed appears to be a correlation between T4 Lyz analog stability and activity, as shown in Figure 3C. This relationship between stability and enzymatic activity could be used as a screening tool to eliminate variants which are likely less active based on their stability. This would reduce costs of PEGylation screening, as activity assays are more expensive than the stability assay. It is important to note that more data is necessary to confirm the existence of this trend for proteins beyond T4 Lyz, however, the presented screen could facilitate verification of this trend for other proteins. Verification of this trend would be especially important for proteins with more than 2 folding states, as destabilization of certain domains may have a smaller impact on activity than others. However, our study indicates that stability could be a useful metric by which to eliminate less active candidates in a large screen without directly testing the activity of all candidates.

1.2 Evaluating Common PEGylation Guidelines with Experimental Screen Results

In the decades since the introduction of PEGylation, several recommendations have evolved towards the design of PEGylated proteins. A synopsis of key similarities and differences between the effects of PEGylation observed experimentally through the present screen and some common PEGylation guidelines in literature is summarized in Table 1.

Our observations regarding PEG size support the claim that, in general, large PEGs are more advantageous than small PEGs for conjugate optimization, as we found no significant negative impacts of PEG size on the stability or activity of T4 Lyz which might offset the improved pharmacokinetics generally reported for conjugates with larger PEG chains.[44, 64-66] In our evaluation of *in vitro* assays of a double site-specifically PEGylated conjugate, we found no motivation for double site-specific PEGylation due to the lack of additional stabilizing benefit and the negative effects on activity. However, as it is well documented that increased PEG weight improves pharmacokinetics of a therapeutic, these results suggest that there may be an optimal trade-off between specific activity and half-life which could potentially be achieved through double site-specific PEGylation. Finally, we evaluated trends in click efficiency, activity, and stability

relating to the uAA incorporation/PEGylation site, finding that current design metrics were insufficiently predictive of the effects of PEGylation observed through our screen.

Overall, the experimental screen enabled the identification of S44, followed closely by N53, as the best sites for PEGylation of the sites screened due to their high PEGylation efficiency, high thermal stability, and high retention of activity. It should be noted that PEGylation had the greatest stabilizing effect on K16AzF, K135AzF, and Ins163AzF, however these sites were less tolerant of the AzF incorporation and had lower conjugation efficiencies. The locational effects of site-specific PEGylation were not well predicted by the common design guidelines evaluated in the screen. While some guidelines, such as conjugation at sites with high SASA, may have merit in some aspects of PEGylation engineering, their predictive capacity is incomplete. For example, K135 may be correctly predicted by SASA to have high activity retention, however PEGylated K135AzF would be more costly to produce commercially than similarly active S44AzF due to its low PEGylation efficiency. Hence, the screening approach presented is an attractive tool to facilitate the efficient screening of PEGylation sites in order to optimize the multiple facets of PEGylated protein design.

1.3 Enhancing CFPS-Based Screen with Coarse-grain Simulation

Considering the necessity of an experimental screen, an *in silico* tool to narrow candidate PEGylation sites would be useful in further reducing the costs of PEGylated protein design. Protein stability is a driving factor in protein PEGylation and is highly dependent on PEGylation site, therefore we sought to integrate a rapid molecular dynamics simulation as part of the screening process. Specifically, the potential of coarse grain simulation was evaluated as a pre-screen tool to inform selection of candidate conjugation sites, thereby reducing the size of the screening pool. Coarse-grain molecular dynamics simulations have previously been used to correctly identify an optimal site for T4 Lyz immobilization.[28, 59] The method employs replica exchange molecular dynamics to calculate the heat capacity (C_v) of the molecule as a function of temperature. Peaks in the C_v curve occur during structural transitions (unfolding/folding), and changes in the thermal stability can be obtained by comparing the temperatures at which the peaks occur in the C_v profiles of WT T4 Lyz and PEGylated T4 Lyz. The computational efficiency of coarse-grain, molecular simulation make it especially appealing as a screening tool because many simulations can be done quickly and in parallel.

We modified the coarse-grain simulation to include a conjugated PEG chain at a specified site. The methods used are those that have been outlined previously[59, 67-69] with a few modifications to include PEG, as detailed in the Materials and Methods. While the mechanism behind PEG-based stabilization is not well understood, there are two primary theories: PEG stabilizes proteins through 1) direct interaction with the protein surface[58, 70, 71] or 2) entropic interactions with the solvent.[9, 70] In this study we limited the simulation to evaluate only the entropic effect. The entropy-only approach has been successfully applied to immobilized proteins,[67, 72] and it has been previously reported that the stabilizing effect of PEG is primarily entropic.[9] The polymer-protein interactions were thus purely repulsive. The tethering of PEG to the protein was done with a simple harmonic restraint without changing the tethering residue to AzF. Currently AzF cannot be included in the model due to insufficient experimental data for paramterization of AzF in the coarse-grain interactions. Therefore, predicted changes in melting temperature correspond to the change in melting temperature solely due to PEGylation ($\Delta T_{m,conj}$). The three-dimensional structure of T4 Lyz, needed for the model, was obtained from PDB ID 2LZM and is expected to be a reasonable representation of the structures of AzF-substituted lysozyme as any structural perturbations resulting from such substitutions have been previously reported to be minimal and highly localized.[56] However, we excluded the variants involving an AzF insertion (Ins163AzF and N53AzF/Ins163AzF) from the simulation due to concerns that the WT structure may by a less accurate representation of the structure for these variants.

We compared the simulation-predicted $\Delta T_{m,conj}$ to the experimentally determined values in our screen, with the results shown in Figure 4. The predictions from the entropy-based PEGylated T4 Lyz simulations agree well with the results from experiment. While the simulation did not predict the exact change in melting temperature due to PEGylation, it accurately predicted the relative change in stability after PEGylation. For both the PEG5kDa and PEG20kDa, the simulation accurately predicted the most stabilizing site for conjugation to be residue 16 and the least stabilizing site to be residue 91. The simulation also accurately

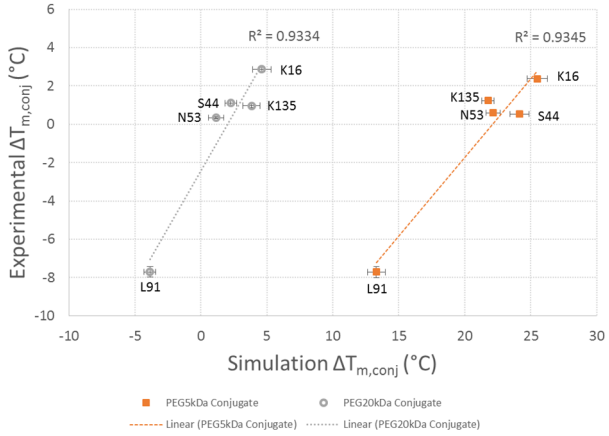


Figure 4 – Comparison of coarse-grain simulation and experiment. $\Delta T_{m,conj}$ for 20kDa PEG conjugates (gray) and 5kDa PEG conjugates (orange). Averages and standard errors of $\Delta T_{m,conj}$ calculated from simulation (n = 10+) and protein thermal shift assay (n = 3-18,); When error bars for experimental are not visible, they are hidden under the marker.

predicted the remaining 3 sites to be slightly less stabilizing than conjugation at residue 16 but significantly more stabilizing than conjugation at site 91. As such, the entropy-only simulation could be a powerful tool for informing design of a screening pool (Example videos of simulation results provided in the supplementary information). Adapting the simulation to include enthalpic effects of PEG-protein interactions may further improve the predictive capacity of the simulation. Indeed, previous results have suggested that the enthalpic interaction effects of PEGylation can be destabilizing in some cases,[9] which could account for the over-prediction of $\Delta T_{m,conj}$ by the entropy-only model. Interestingly, the simulation seems to provide better quantitative estimates of $\Delta T_{m,conj}$ for the 20kDa conjugates. We hypothesize that this is due to a more dominant impact of entropy with the larger polymer. Hence, incorporation of PEG-protein interaction parameters may also allow the model to predict the relative impact of PEGylation with different lengths of PEG.

Overall, our use of coarse grain simulations indicates that valuable insight can be gained about comparative effects of PEGylation from a simple, entropy-based model. This model could be incorporated as part of the screening process to create a hybrid simulation- and experimental-based screening method in which all of the sites on a target protein which are available to conjugation are first screened with the coarse grain model. The simulations in this study ran to completion in under 11 hours, potentially enabling simultaneous evaluation of all potential conjugation sites on a protein in less than 24 hours. Sites which are predicted by the simulation to be least stabilizing could then be eliminated from the subsequent experimental screen in order to conserve resources by reducing the number of candidate conjugation sites. Further improvements in the model, including parameterization of AzF and PEG-protein interactions, may enhance the capacity of the model for quantitative ΔT_m prediction.

2 CONCLUSIONS

We have presented a screening system based on cell-free protein synthesis in order to rapidly assess PEGylated protein variants with microliter-scale reactions. Our system allows high yield of uAA-modified proteins, including doubly modified proteins. Recent work demonstrating cell-free expression of a variety of challenging proteins illustrates the adaptability of the CFPS expression system, providing guidelines by which the presented system could be tailored to screen a variety of PEGylated proteins.[30, 36-41] Coarse-grain simulations shows promise as a tool for rapid *in silico* screening of potential conjugation sites to further expedite the screening process by narrowing the experimental screening pool, conserving lab resources. We demonstrate the system's utility in evaluating the effects of PEGylation at various sites on T4 lysozyme and

show that common guidelines were insufficient for predicting these effects. Our results and analysis highlight shortcomings in the current understanding and guidelines governing PEGylated protein design, and demonstrate the potential of a CFPS screening system to efficiently evaluate protein PEGylation sites in the absence of specific design heuristics.

Table 1 – Comparison to design recommendations in literature.

		PEGylation Recommendations in Literature	Observations with CFPS Screen
PEG Size		- Larger PEGs offer improved pharmacokinetics and stability relative to smaller PEGs[a] - ΔT_m independent of PEG size[b] - *In vitro* activity is inversely proportional to attached PEG mass[c]	- No significant difference in stability or activity of 20kDa conjugates vs 5kDa conjugates
PEG Number		- Higher modification number correlates with lower activity[d] - PEGylation at two individually stabilizing sites results in further stability improvement, though effects are not always additive[e] - Single large PEG is preferable to multiple small PEGs[f]	- Double PEGylation with 20kDa PEG decreased activity slightly - Double PEGylation at two individually stabilizing sites did not significantly improve stability
Conjugation Site	**Efficiency**	- High SASA (Residues with ASA-View score > 0.4) improves conjugation efficiency[g] - Buried residues in hydrophobic pockets may conjugate with high efficiency in SPAAC reactions[h]	- SASA was not predictive of conjugation efficiency - No clear correlation to surface hydrophilicity/hydrophobicity
	Stability/ Activity	- Conjugation at unstructured loops may minimize the strain on protein structure[i]	- General trend of increased activity retention after conjugation at higher SASA may help to narrow experimental screens by eliminating sites with very low SASA - No clear correlation between stabilizing potential and PEGylation site secondary structure or SASA. -

References: a[1, 2], b [3-5], c[1, 6-8], d[6], e[9], f[1], g[21-24], h[24], i[13, 28, 29]

3 METHODS

3.1 Extract Preparation

Extracts were prepared using an *Escherichia coli* BL21*(DE3) pEVOL-AzF strain, a kind gift from from Peter Shultz[73, 74]. The extract was prepared in a manner similar to that which has been described previously[45, 75-79] with a few modifications. Cells were grown at 37°C and 280 rpm in sequential growths. Growths were started in 5 mL of 2xYT media, incubated overnight, and moved into 100 mL 2xYT. The 100 mL growth was then incubated until an O.D. of 2.0 and then added to 900 mL L 2xYT media in a 2.5 L Tunair baffled shake flask (IBI Scientific, Peosta, IA) for a final volume of 1 L. When the 1 L growth reached an O.D. between 0.5 and 0.7, cells were induced with both 1 mL of 1 M isopropyl-1-thio-b-D-galactopyranoside (IPTG) and 0.20 g arabinose. Cells were then monitored and harvested in mid-log phase (An O.D. of ~2.0 in this work) at 8,000 rpm for 30 minutes. After washing in Buffer A, cells were re-suspended in Buffer A at a ratio of 1 mL per gram and lysed in three passes through an Avestin Emulsiflex B-15 cell disruptor (Ottawa, Canada) at 21,000 psi. Lysate was centrifuged at 12,000 rcf for 10 minutes, following which supernatant was removed and incubated at 37°C for 30 minutes. Extract was then flash frozen and stored at -80°C until use.

3.2 Plasmid Preparation

A cysteine-free T4 Lyz variant was obtained from Addgene (Cambridge, MA) and cloned into the pY71 plasmid, and a C-terminal strep-tag was added for purification purposes, as described previously.[28] Six variants – K16Amber, S44Amber, N53Amber, L91Amber, K135Amber, and Ins163Amber – were created using the Quikchange II mutagenesis protocol (Agilent Technologies, Santa Clara, CA) [28]. A seventh variant with two Amber codons was created also using the Quikchange II mutagenesis protocol to insert an additional amber stop between K162 and N163 on the N53Amber variant. This variant is hereafter referred to as N53/Ins163AzF. Plasmids were purified for use in cell-free protein synthesis using a Qiagen Plasmid Maxi Kit (Valencia, CA).

3.3 Cell-Free Protein Synthesis

Cell-free protein synthesis was performed using the standard PANOxSP system, with a few modifications[45]. The reaction mixture was as follows, with components obtained from Sigma-Aldrich unless otherwise specified: 25% v/v *E. coli* pEVOL pAzF extract,25% v/v 19-amino acid PANOxSP mixture (Glutamate was added separately as a salt with Mg in order to optimize the Mg content of the reaction), 18 mM $Mg(Glu)_2$,

12 nM plasmid purified with Qiagen Plasmid Maxi Kit, 5 uM C_{14} leucine (PerkinElmer), 3 mM AzF (Chem Impex International, Wood Dale, IL), and the remaining volume distilled deionized water. For synthesis of the N53AzF/Ins163AzF AzF variant, AzF was added at 6 mM in order to improve yields. The reactions were assembled under a safe-light and incubated in darkness in order to preserve the azide group, which decays upon exposure to UV or near-UV light[45, 74, 80]. Reactions were performed in 15 mL falcon tubes (GeneMate) at 300 – 400 uL volumes and incubated overnight (~15 hours in this work) at 30°C. Negative control reactions were performed in 50 uL volumes in 2 mL microcentrifuge tubes (GeneMate) for the same time at 30°C. Purified synthetase was not added as yields were sufficiently high with only the synthetase provided in the prepared pEVOL extract. Total protein synthesis yields were determined using liquid scintillation as discussed previously[45], using 5% trichloroacetic acid for protein precipitation. Synthesis of full-length protein was verified by running 3 uL of the CFPS reaction on a NuPAGE 10% Bis-Tris Gel (Invitrogen, Carlsbad, CA). The gels were run according to manufacturer's instructions, at 200V for 35 minutes using MES buffer. After running, gels were stained with SimplyBlue SafeStain (Invitrogen), dried, and autoradiograms were performed using Kodak MR Autoradiogram Films with 2 day exposure time. Yields of full-length variants were determined from densitometry using ImageJ software[81] to compare the relative band intensities to that of WT lysozyme and scale the yields accordingly to the WT yields.

3.4 T4 Lysozyme Purification

T4 Lyz was purified using Strep-Tactin Spin Columns (IBA Life Sciences, Gottingen, Germany) according to manufacturers' specifications, with the following variations. To improve recovery, CFPS samples were run through the spin columns three times and columns were then washed 5 times with the provided Buffer W. T4 Lyz was then eluted according to the procedure specified for high concentration. Liquid scintillation was used to determine the concentration of the purified product. Using the total protein yields calculated from the scintillation of the CFPS reaction, the CPM/mg/mL was determined for each sample, which was then used to calculate the concentration of T4 Lyz variant in the purified samples.

3.5 PEGylation Reactions

Conjugation reactions were performed using strain-promoted azide-alkyne cycloaddition, or SPAAC. The number of PEG equivalents was optimized in order to obtain maximal conjugation with the minimal allowable PEG. The optimal PEG equivalents was determined to be 20 for 20kDa PEG and 50 for 5kDa PEG. Reactions were assembled similarly to protocols which have been described previously,[82, 83] with the following specifications. In a PCR tube, 5 uM lysozyme was combined with 20 or 50 equivalents of 20kDa

or 5kDa DBCO-mPEG (Click Chemistry Tools, Scottsdale, AZ), respectively, in PBS buffer. PEG equivalents were doubled for the double amber suppression variant in order to obtain higher conjugation efficiency. The reactions were incubated at 37°C and 280 rpm for 18 hours. Although complete conjugation with SPAAC reactions has been reported with much shorter reaction times[30, 83], an 18 hour reaction time was used in order to maximize conjugation at low-efficiency sites and thereby obtain more accurate stability and activity data for these lysozyme analogs. Control reactions were also assembled with WT T4 lysozyme, mimicking the conditions of the SPAAC reactions for both PEG sizes and unconjugated controls. These reactions provided a baseline to account for the effects of incubation time and unconjugated PEG time on the stability and activity of the lysozyme.

3.6 PEGylation Efficiency Analysis

To determine efficiency of each PEGylation reaction, a sample of each reaction was run on a NuPAGE 10% Bis-Tris Gel at 200V for 35 minutes using MES buffer. The gels were stained with SimplyBlue SafeStain (Invitrogen), dried, and used to produce an autoradiogram. Autoradiograms were done using Kodak MR Autoradiogram Film with a 2 day exposure. Using ImageJ, the relative intensities of the PEG-shifted bands and the un-shifted bands were calculated. The click efficiency was determined as the intensity of the PEG-shifted band divided by the sum of the intensity of the un-shifted and shifted band(s) in the gel lane. PEGylation efficiencies were calculated from the autoradiogram in order to eliminate error from possible interference of PEG with staining.

3.7 Stability Analysis

3.7.1 Stability Assay

Stability was analyzed by examining the shifts in the protein melting temperature, T_m. Melting temperatures were determined using the Protein Thermal Shift Assay (Thermo Fisher Scientific, Carlsbad, CA) and the corresponding Protein Thermal Shift Software, version 1.3. This assay uses a hydrophobic dye which fluoresces when it binds to the hydrophobic regions of the protein as they are exposed during melting. The protein is combined with the dye and gradually stepped through increasing temperatures in a real-time PCR machine, which monitors the change in florescence. The florescence curve is then used to determine the melting temperature. Melting temperatures were calculated using the derivative method in order to skewing of a Boltzmann fit by any remaining peak for the unconjugated protein. The derivative method identifies the T_m by computing a second-derivative to identify the inflection point of the fluorescence curve.

The melt reactions were assembled according to the manufacturer's instructions, with the following specificaitons: 5 uL Protein Thermal Shift Buffer, 7.5 uL PBS buffer, 5 uL click reaction, 2.5 uL Diluted Protein Thermal Shift Dye (8x) diluted with PBS buffer. Reactions were done in triplicate, assembled in a 96-well, semi-skirted Framestar Fast Plate (Midsci, St. Louis, MO) and covered with a MicroAmp Optical Adhesive Film (Applied Biosystems,Thermo Fisher Scientific). The assay was set up using StepOne Software v2.3, and run in a StepOnePlus™ Real-Time PCR System (Applied Biosystems). In order to obtain maximal resolution of the T_m, the assay was run using a standard ramp speed from a starting temperature of 25°C to 99°C. Analysis was done using the derivative method. Results are reported as a change from the average T_m of the corresponding WT reaction (no PEG, 5kDa PEG, or 20kDa PEG), ΔT_m. For the simulation comparison, the experimental data is reported as a difference in T_m between a PEGylated T4 Lyz variant and the corresponding unconjugated variant, $\Delta T_{m,conj}$.

3.7.2 Simulation-predicted T_m

Coarse-grain simulations were performed using the Go-like model and replica-exchange algorithm which has been described previously.[59, 67-69] In this coarse-grain model, each residue is approximated as a single site, centered at the location of the C_α atom of the residue in the crystal structure. The model input files were created using the MMTSB Web Service (mmtsb.org/webservices/gomodel.html)[84, 85] based on the T4 Lyz structure from the Protein Data Bank (PDB ID: 2LZM). PEG monomers were approximated as single residues with a radius of 1.48 angstroms and a center-to-center distance of 3.7 angstroms, and the DBCO connecting group was approximated as a residue with a radius of 5.18 angstroms. These distances were obtained using a GaussView[86] model of the SPAAC product and attached PEG chain. Bond energies were also obtained from GaussView.[86] Other relevant parameters, including Lennard-Jones parameters, were chosen to resemble those of the peptide bonds within the protein. Additional details pertaining to the model form are included in the Supplementary Materials.

PEG chains were added to the protein model extending linearly away from the attached site. The PEG chain was then allowed to equilibrate to a more realistic conformation in NVE MD equilibrium simulations with constraints in place to hold the protein in the properly folded conformation while the PEG polymer equilibrated. Results of these equilibration simulations were then used as starting points for replica exchange simulations where an additional unconstrained equilibrium phase ensured system equilibration before production steps were recorded. Replica exchange simulations were done in the NVT ensemble using three Nose-Hoover thermostats, a time step of 3 fs, and a mass of 7.81338×10^{-22} kg*angstroms^2. Each simulation

contained 20 x 10^6 equilibration steps and 60 x 10^6 productions steps. Sixty-six boxes, with temperature steps of 1.5K between boxes in the 21K range surrounding the expected melting temperatures (345K – 366K for 5kDa conjguates, 327K – 348K for WT and 20kDa conjguates) and steps of 3K for the remaining range, were used for all simulations with a box size of 4000 x 4000 x 4000 angstroms. The replica exchange simulations were run 10 times for all cases except for K16 and S44 PEGylation models, which were run 20 times due to slightly more scatter in the data. The results for all runs were averaged and the standard error reported.

The average relative native contacts, representing the remaining interactions pertaining to the protein secondary and tertiary structure relative to the 100% folded states, was plotted versus temperature. This plot shows how the degree of protein folding decreases over temperature as the protein unfolds, and is analogous to the fluorescence versus temperature plot obtained from the experimental assay. To mimic the experimental analysis, the derivative of the relative native contacts was calculated numerically with respect to temperature using a central difference formula. The temperature at which the magnitude of this derivative was greatest was taken to be T_m, as it represents the point where the rate of change in the degree of protein folding is greatest, and corresponds with the point where the rate of increase in fluorescence is greatest. This point also corresponded with the maximum heat capacity value, which is expected as T_m is also often determined to be the temperature at which heat capacity is a maximum. The melting points calculated from simulation were then compared to the experimentally determined T_m.

3.8 Activity Assay

Activity of each lysozyme construct was determined via the EnzChek Lysozyme Assay Kit (Thermo Fisher Scientific). Reactions were performed according to the manufacturer's instructions, with the following specifications and modifications. EnzChek buffer was not used, as it has been reported that T4 Lyz is less active at high ionic strengths.[87] Unpurified PEGylated lysozyme from each SPAAC reaction was added to each reaction at a final lysozyme concentration of 0.06 uM. This lysozyme concentration was selected because it did not cause the activity assay to saturate and required less dilution than other suitable concentrations, which eliminated unnecessary error. In a black 96-well untreated polystyrene plate (VWR), 1.2 uL of unpurified SPAAC reaction was added to 48.8 uL ddH2O and mixed thoroughly by pipetting. The 1 mg/mL fluorogenic substrate was diluted 20x in ddH2O, and 50 uL was added to each well. The plate was then immediately placed into a plate reader (Synergy MX, BioTek) and incubated at 37°C. Fluorescence was monitored every 1.5 minutes for 2 hours (494nm/518nm). Activity was determined from the endpoint

fluorescence of each well and normalized to the endpoint activity of the corresponding WT reaction. Each sample was assayed in n=3+. Relative activity of each PEGylated lysozyme analog was calculated from the mixture activity data, assuming additive activities of the PEGylated and unPEGylated lysozyme. WT activity data for a set of assays was averaged, and then conjugate activity (a_c) was calculated for each replicate according to Equations 1, where x_c and x_{uc} is the fraction of conjugated and unconjugated T4 Lyz, respectively, a_{rxn} is the activity of the reaction mixture, and a_{uc} is the activity of the unconjugated variant. The activity for di-PEGylated N53AzF/Ins163AzF (a_{c2}) was calculated as shown in Equation 2, where the activity of mono-PEGylated N53AzF/Ins163AzF (a_{c1}) is calculated as shown in Equation 3, where $a_{N53AzF,c}$ and $a_{N53AzF,uc}$ is the activity of conjugated and unconjugated N53AzF, respectively, with analogous activities for Ins163AzF and N53AzF/Ins163AzF. Calculated activities were then averaged over all replicates and over all sets of assays.

$$a_c = 1/x_c * (a_{rxn} - x_{uc}a_{uc}) \quad (1)$$

$$a_{c2} = 1/x_{c2} * (a_{rxn} - a_{uc}x_{uc} - a_{c1}x_{c1}) \quad (2)$$

$$a_{c1} = 0.5\left(\frac{a_{N53AzF,c}}{a_{N53AzF,uc}} + \frac{a_{Ins163AzF,c}}{a_{Ins163AzF,uc}}\right) * a_{N53AzF/Ins163AzF,uc} \quad (3)$$

4 AUTHOR INFORMATION

Corresponding Author

*Email: bundy@byu.edu

ORCID

orcid.org/0000-0003-4438-183X

Author Contributions

K.M.W. performed experiments and coarse-grain simulations, analyzed data, helped with model development, and wrote the paper. A.K.S. adapted the coarse-grain model for the PEGylated lysozymes and helped with running the simulations and analyzing simulation data. T.A.K. developed the simulation codes and assisted in simulation results analysis and writing the paper. B.C.B. assisted in experimental analysis and writing the paper.

Notes

The authors declare no competing financial interests.

5 ACKNOWLEDGEMENTS

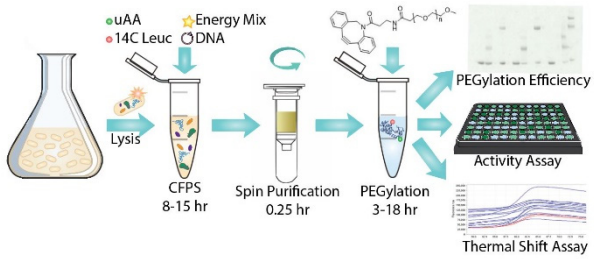
This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. #1247046, by the National Science Foundation CBET Division

CAREER Award (#1254148) and by the National Science Foundation DMR Divison Award (#1710574). The authors thank Dr. Peter Schultz (Scripps Research Institute) for his generous gift of the pEVOL-pAzF plasmid and Dr. Brent Nielsen (Brigham Young University) for his generous giving of expertise and access to real-time PCR equipment. The authors are also grateful to the Fulton Super Computing Laboratory at Brigham Young University for computational resources. The authors declare no conflict of interest.

6 Supporting Information

Figures S1-S9, additional coarse-grain simulation details, and heat capacity and native contacts results from coarse-grain simulation.

7 Abstract Graphic



REFERENCES

[1] Bailon, P., Palleroni, A., Schaffer, C. A., Spence, C. L., Fung, W.-J., Porter, J. E., Ehrlich, G. K., Pan, W., Xu, Z.-X., Modi, M. W., Farid, A., Berthold, W., and Graves, M. (2001) Rational Design of a Potent, Long-Lasting Form of Interferon: A 40 kDa Branched Polyethylene Glycol-Conjugated Interferon α-2a for the Treatment of Hepatitis C, *Bioconjugate Chemistry 12*, 195-202.

[2] Chiu, K., Agoubi, L. L., Lee, I., Limpar, M. T., Lowe, J. W., and Goh, S. L. (2010) Effects of Polymer Molecular Weight on the Size, Activity, and Stability of PEG-Functionalized Trypsin, *Biomacromolecules 11*, 3688-3692.

[3] Rodríguez-Martínez, J. A., Solá, R. J., Castillo, B., Cintrón-Colón, H. R., Rivera-Rivera, I., Barletta, G., and Griebenow, K. (2008) Stabilization of α-Chymotrypsin upon PEGylation Correlates with Reduced Structural Dynamics, *Biotechnology and bioengineering 101*, 1142-1149.

[4] Plesner, B., Fee, C. J., Westh, P., and Nielsen, A. D. (2011) Effects of PEG size on structure, function and stability of PEGylated BSA, *European Journal of Pharmaceutics and Biopharmaceutics 79*, 399-405.

[5] Plesner, B., Westh, P., and Nielsen, A. D. (2011) Biophysical characterisation of GlycoPEGylated recombinant human factor VIIa, *International Journal of Pharmaceutics 406*, 62-68.

[6] da Silva Freitas, D., and Abrahão-Neto, J. (2010) Biochemical and biophysical characterization of lysozyme modified by PEGylation, *International Journal of Pharmaceutics 392*, 111-117.

[7] Zang, Q., Tada, S., Uzawa, T., Kiga, D., Yamamura, M., and Ito, Y. (2015) Two site genetic incorporation of varying length polyethylene glycol into the backbone of one peptide, *Chemical Communications 51*, 14385-14388.

[8] Grace, M. J., Lee, S., Bradshaw, S., Chapman, J., Spond, J., Cox, S., DeLorenzo, M., Brassard, D., Wylie, D., Cannon-Carlson, S., Cullen, C., Indelicato, S., Voloch, M., and Bordens, R. (2005) Site of Pegylation and Polyethylene Glycol Molecule Size Attenuate Interferon-α Antiviral and Antiproliferative Activities through the JAK/STAT Signaling Pathway, *Journal of Biological Chemistry 280*, 6327-6336.

[9] Lawrence, P. B., Gavrilov, Y., Matthews, S. S., Langlois, M. I., Shental-Bechor, D., Greenblatt, H. M., Pandey, B. K., Smith, M. S., Paxman, R., Torgerson, C. D., Merrell, J. P., Ritz, C. C., Prigozhin, M. B., Levy, Y., and Price, J. L. (2014) Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic and Proteolytic Stability, *Journal of the American Chemical Society 136*, 17547-17560.

[10] Veronese, F. M., and Pasut, G. (2005) PEGylation, successful approach to drug delivery, *Drug Discovery Today 10*, 1451-1458.

[11] Pasut, G., and Veronese, F. M. (2012) State of the art in PEGylation: The great versatility achieved after forty years of research, *Journal of Controlled Release 161*, 461-472.

[12] Pasut, G., Sergi, M., and Veronese, F. M. (2008) Anti-cancer PEG-enzymes: 30 years old, but still a current approach, *Advanced Drug Delivery Reviews 60*, 69-78.

[13] Mei, B., Pan, C., Jiang, H., Tjandra, H., Strauss, J., Chen, Y., Liu, T., Zhang, X., Severs, J., Newgren, J., Chen, J., Gu, J.-M., Subramanyam, B., Fournel, M. A., Pierce, G. F., and Murphy, J. E. (2010) Rational design of a fully active, long-acting PEGylated factor VIII for hemophilia A treatment, *Blood 116*, 270.

[14] Pelegri-O'Day, E. M., Lin, E.-W., and Maynard, H. D. (2014) Therapeutic Protein–Polymer Conjugates: Advancing Beyond PEGylation, *Journal of the American Chemical Society 136*, 14323-14332.

[15] Maiser, B., Dismer, F., and Hubbuch, J. (2014) Optimization of Random PEGylation Reactions by Means of High Throughput

Screening, *Biotechnology and Bioengineering 111*, 104-114.
[16] Kontermann, R. E. (2011) Strategies for extended serum half-life of protein therapeutics, *Current Opinion in Biotechnology 22*, 868-876.
[17] Dhalluin, C., Ross, A., Leuthold, L.-A., Foser, S., Gsell, B., Müller, F., and Senn, H. (2005) Structural and Biophysical Characterization of the 40 kDa PEG−Interferon-α2a and Its Individual Positional Isomers, *Bioconjugate Chemistry 16*, 504-517.
[18] Illanes, A., Cauerhff, A., Wilson, L., and Castro, G. R. (2012) Recent trends in biocatalysis engineering, *Bioresource Technology 115*, 48-57.
[19] Schumacher, D., and Hackenberger, C. P. R. (2014) More than add-on: chemoselective reactions for the synthesis of functional peptides and proteins, *Current Opinion in Chemical Biology 22*, 62-69.
[20] Levine, P. M., Craven, T. W., Bonneau, R., and Kirshenbaum, K. (2014) Intrinsic bioconjugation for site-specific protein PEGylation at N-terminal serine, *Chemical Communications 50*, 6909-6912.
[21] Lim, S. I., Cho, J., and Kwon, I. (2015) Double clicking for site-specific coupling of multiple enzymes, *Chemical Communications 51*, 13607-13610.
[22] Qiu, H., Boudanova, E., Park, A., Bird, J. J., Honey, D. M., Zarazinski, C., Greene, B., Kingsbury, J. S., Boucher, S., Pollock, J., McPherson, J. M., and Pan, C. Q. (2013) Site-Specific PEGylation of Human Thyroid Stimulating Hormone to Prolong Duration of Action, *Bioconjugate Chemistry 24*, 408-418.
[23] Lim, S. I., Hahn, Y. S., and Kwon, I. (2015) Site-specific albumination of a therapeutic protein with multi-subunit to prolong activity in vivo, *Journal of Controlled Release 207*, 93-100.
[24] Arpino, J. A. J., Baldwin, A. J., McGarrity, A. R., Tippmann, E. M., and Jones, D. D. (2015) In-Frame Amber Stop Codon Replacement Mutagenesis for the Directed Evolution of Proteins Containing Non-Canonical Amino Acids: Identification of Residues Open to Bio-Orthogonal Modification, *PLOS ONE 10*, e0127504.
[25] Brocchini, S., Godwin, A., Balan, S., Choi, J.-w., Zloh, M., and Shaunak, S. (2008) Disulfide bridge based PEGylation of proteins, *Advanced Drug Delivery Reviews 60*, 3-12.
[26] Rosendahl, M. S., Doherty, D. H., Smith, D. J., Carlson, S. J., Chlipala, E. A., and Cox, G. N. (2005) A Long-Acting, Highly Potent Interferon α-2 Conjugate Created Using Site-Specific PEGylation, *Bioconjugate Chemistry 16*, 200-207.
[27] Doherty, D. H., Rosendahl, M. S., Smith, D. J., Hughes, J. M., Chlipala, E. A., and Cox, G. N. (2005) Site-Specific PEGylation of Engineered Cysteine Analogues of Recombinant Human Granulocyte-Macrophage Colony-Stimulating Factor, *Bioconjugate Chemistry 16*, 1291-1298.
[28] Wu, J. C. Y., Hutchings, C. H., Lindsay, M. J., Werner, C. J., and Bundy, B. C. (2015) Enhanced Enzyme Stability Through Site-Directed Covalent Immobilization, *Journal of Biotechnology 193*, 83-90.
[29] Moatsou, D., Li, J., Ranji, A., Pitto-Barry, A., Ntai, I., Jewett, M. C., and O'Reilly, R. K. (2015) Self-Assembly of Temperature-Responsive Protein–Polymer Bioconjugates, *Bioconjugate Chemistry 26*, 1890-1899.
[30] Zimmerman, E. S., Heibeck, T. H., Gill, A., Li, X., Murray, C. J., Madlansacay, M. R., Tran, C., Uter, N. T., Yin, G., Rivers, P. J., Yam, A. Y., Wang, W. D., Steiner, A. R., Bajad, S. U., Penta, K., Yang, W., Hallam, T. J., Thanos, C. D., and Sato, A. K. (2014) Production of Site-Specific Antibody–Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System, *Bioconjugate Chemistry 25*, 351-361.
[31] Cho, H., Daniel, T., Buechler, Y. J., Litzinger, D. C., Maio, Z., Putnam, A.-M. H., Kraynov, V. S., Sim, B.-C., Bussell, S., Javahishvili, T., Kaphle, S., Viramontes, G., Ong, M., Chu, S., Gc, B., Lieu, R., Knudsen, N., Castiglioni, P., Norman, T. C., Axelrod, D. W., Hoffman, A. R., Schultz, P. G., DiMarchi, R. D., and Kimmel, B. E. (2011) Optimized clinical

performance of growth hormone with an expanded genetic code, *Proceedings of the National Academy of Sciences 108*, 9060-9065.

[32] Tian, F., Lu, Y., Manibusan, A., Sellers, A., Tran, H., Sun, Y., Phuong, T., Barnett, R., Hehli, B., Song, F., DeGuzman, M. J., Ensari, S., Pinkstaff, J. K., Sullivan, L. M., Biroc, S. L., Cho, H., Schultz, P. G., DiJoseph, J., Dougher, M., Ma, D., Dushin, R., Leal, M., Tchistiakova, L., Feyfant, E., Gerber, H.-P., and Sapra, P. (2014) A general approach to site-specific antibody drug conjugates, *Proceedings of the National Academy of Sciences 111*, 1766-1771.

[33] Schinn, S.-M., Bradley, W., Groesbeck, A., Wu, J. C., Broadbent, A., and Bundy, B. C. (2017) Rapid in vitro screening for the location-dependent effects of unnatural amino acids on protein expression and activity, *Biotechnology and Bioengineering 114*, 2412-2417.

[34] Schinn, S.-M., Bradley, W., Groesbeck, A., Wu, J., Broadbent, A., and Bundy, B. Rapid In vitro Screening for the Location -Dependent Effects of Unnatural Amino Acids on Protein Expression and Activity, *Biotechnology and Bioengineering Accepted*.

[35] Salehi, A. S. M., Smith, M. T., Schinn, S.-M., Hunt, J. M., Muhlestein, C., Diray-Arce, J., Nielsen, B. L., and Bundy, B. C. Efficient tRNA degradation and quantification in Escherichia coli cell extract using RNase-coated magnetic beads: A key step toward codon emancipation, *Biotechnology Progress*, n/a-n/a.

[36] Salehi, A. S. M., Smith, M. T., Bennett, A. M., Williams, J. B., Pitt, W. G., and Bundy, B. C. (2016) Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system, *Biotechnology Journal 11*, 274-281.

[37] Yang, J. H., Kanter, G., Voloshin, A., Levy, R., and Swartz, J. R. (2004) Expression of active murine granulocyte-macrophage colony-stimulating factor in an Escherichia coli cell-free system, *Biotechnology Progress 20*, 1689-1696.

[38] Smith, M. T., Varner, C. T., Bush, D. B., and Bundy, B. C. (2012) The incorporation of the A2 protein to produce novel Qβ virus-like particles using cell-free protein synthesis, *Biotechnology progress 28*, 549-555.

[39] Bundy, B. C., and Swartz, J. R. (2011) Efficient disulfide bond formation in virus-like particles, *Journal of Biotechnology 154*, 230-239.

[40] Bundy, B. C., Franciszkowicz, M. J., and Swartz, J. R. (2008) Escherichia coli-based cell-free synthesis of virus-like particles, *Biotechnology and Bioengineering 100*, 28-37.

[41] Park, C.-G., Kim, T.-W., Oh, I.-S., Song, J. K., and Kim, D.-M. (2009) Expression of functional Candida antarctica lipase B in a cell-free protein synthesis system derived from Escherichia coli, *Biotechnology Progress 25*, 589-593.

[42] Ahmad, S., Gromiha, M., Fawareh, H., and Sarai, A. (2004) ASAView: Database and tool for solvent accessibility representation in proteins, *BMC Bioinformatics 5*, 51-51.

[43] Loong, B. K., and Knotts, T. A. (2014) Communication: Using multiple tethers to stabilize proteins on surfaces, *The Journal of Chemical Physics 141*, 051104.

[44] Xu, J., Bussiere, J., Yie, J., Sickmier, A., An, P., Belouski, E., Stanislaus, S., and Walker, K. W. (2013) Polyethylene Glycol Modified FGF21 Engineered to Maximize Potency and Minimize Vacuole Formation, *Bioconjugate Chemistry 24*, 915-925.

[45] Bundy, B. C., and Swartz, J. R. (2010) Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein−Protein Click Conjugation, *Bioconjugate Chemistry 21*, 255-263.

[46] Albayrak, C., and Swartz, J. R. (2013) Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation, *Nucleic Acids Research 41*, 5949-5963.

[47] Kolb, H. C., and Sharpless, K. B. (2003) The growing impact of click chemistry on drug

discovery, *Drug Discovery Today 8*, 1128-1137.
[48] Thirumurugan, P., Matosiuk, D., and Jozwiak, K. (2013) Click Chemistry for Drug Development and Diverse Chemical–Biology Applications, *Chemical Reviews 113*, 4905-4979.
[49] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—A visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.
[50] Reddington, S. C., Tippmann, E. M., and Dafydd Jones, D. (2012) Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry, *Chemical Communications 48*, 8419-8421.
[51] Lee, B. K., Kwon, J. S., Kim, H. J., Yamamoto, S., and Lee, E. K. (2007) Solid-Phase PEGylation of Recombinant Interferon α-2a for Site-Specific Modification: Process Performance, Characterization, and in Vitro Bioactivity, *Bioconjugate Chemistry 18*, 1728-1734.
[52] Johnson, C. M. (2013) Differential scanning calorimetry as a tool for protein folding and stability, *Archives of Biochemistry and Biophysics 531*, 100-109.
[53] Matsumura, M., Becktel, W. J., Levitt, M., and Matthews, B. W. (1989) Stabilization of phage T4 lysozyme by engineered disulfide bonds, *Proceedings of the National Academy of Sciences of the United States of America 86*, 6562-6566.
[54] Gray, T. M., and Matthews, B. W. (1987) Structural analysis of the temperature-sensitive mutant of bacteriophage T4 lysozyme, glycine 156----aspartic acid, *Journal of Biological Chemistry 262*, 16858-16864.
[55] Wetzel, R., Perry, L. J., Baase, W. A., and Becktel, W. J. (1988) Disulfide bonds and thermal stability in T4 lysozyme, *Proceedings of the National Academy of Sciences of the United States of America 85*, 401-405.
[56] Baase, W. A., Liu, L., Tronrud, D. E., and Matthews, B. W. (2010) Lessons from the lysozyme of phage T4, *Protein Science : A Publication of the Protein Society 19*, 631-641.
[57] Mooers, B. H. M., Baase, W. A., Wray, J. W., and Matthews, B. W. (2009) Contributions of all 20 amino acids at site 96 to the stability and structure of T4 lysozyme, *Protein Science : A Publication of the Protein Society 18*, 871-880.
[58] Pandey, B. K., Smith, M. S., Torgerson, C., Lawrence, P. B., Matthews, S. S., Watkins, E., Groves, M. L., Prigozhin, M. B., and Price, J. L. (2013) Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length, *Bioconjugate Chemistry 24*, 796-802.
[59] Wei, S., and Knotts, T. A. (2011) Effects of tethering a multistate folding protein to a surface, *The Journal of Chemical Physics 134*, 185101.
[60] Hartley, A. M., Zaki, A. J., McGarrity, A. R., Robert-Ansart, C., Moskalenko, A. V., Jones, G. F., Craciun, M. F., Russo, S., Elliott, M., Macdonald, J. E., and Jones, D. D. (2015) Functional modulation and directed assembly of an enzyme through designed non-natural post-translation modification, *Chemical Science 6*, 3712-3717.
[61] Harris, J. M., Martin, N. E., and Modi, M. (2001) Pegylation, *Clinical Pharmacokinetics 40*, 539-551.
[62] Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M., and Schultz, P. G. (2004) Site-specific PEGylation of proteins containing unnatural amino acids, *Bioorganic & Medicinal Chemistry Letters 14*, 5743-5745.
[63] Balan, S., Choi, J.-w., Godwin, A., Teo, I., Laborde, C. M., Heidelberger, S., Zloh, M., Shaunak, S., and Brocchini, S. (2007) Site-Specific PEGylation of Protein Disulfide Bonds Using a Three-Carbon Bridge, *Bioconjugate Chemistry 18*, 61-76.
[64] Kim, T. H., Swierczewska, M., Oh, Y., Kim, A., Jo, D. G., Park, J. H., Byun, Y., Sadegh-Nasseri, S., Pomper, M. G., Lee, K. C., and Lee, S.

(2013) Mix to Validate: A Facile, Reversible PEGylation for Fast Screening of Potential Therapeutic Proteins In Vivo, *Angewandte Chemie International Edition 52*, 6880-6884.

[65] Cong, Y., Pawlisz, E., Bryant, P., Balan, S., Laurine, E., Tommasi, R., Singh, R., Dubey, S., Peciak, K., Bird, M., Sivasankar, A., Swierkosz, J., Muroni, M., Heidelberger, S., Farys, M., Khayrzad, F., Edwards, J., Badescu, G., Hodgson, I., Heise, C., Somavarapu, S., Liddell, J., Powell, K., Zloh, M., Choi, J.-w., Godwin, A., and Brocchini, S. (2012) Site-Specific PEGylation at Histidine Tags, *Bioconjugate Chemistry 23*, 248-263.

[66] Nairn, N. W., Shanebeck, K. D., Wang, A., Graddis, T. J., VanBrunt, M. P., Thornton, K. C., and Grabstein, K. (2012) Development of Copper-Catalyzed Azide–Alkyne Cycloaddition for Increased in Vivo Efficacy of Interferon β-1b by Site-Specific PEGylation, *Bioconjugate Chemistry 23*, 2087-2097.

[67] Wei, S., and Knotts, T. A. (2010) Predicting stability of alpha-helical, orthogonal-bundle proteins on surfaces, *The Journal of Chemical Physics 133*, 115102.

[68] Bush, D. B., and Knotts, T. A. (2015) Communication: Antibody stability and behavior on surfaces, *The Journal of Chemical Physics 143*, 061101.

[69] Bush, D. B., and Knotts, T. A. (2017) Probing the effects of surface hydrophobicity and tether orientation on antibody-antigen binding, *The Journal of Chemical Physics 146*, 155103.

[70] Chao, S.-H., Matthews, S. S., Paxman, R., Aksimentiev, A., Gruebele, M., and Price, J. L. (2014) Two Structural Scenarios for Protein Stabilization by PEG, *The Journal of Physical Chemistry B 118*, 8388-8395.

[71] Yang, C., Lu, D., and Liu, Z. (2011) How PEGylation Enhances the Stability and Potency of Insulin: A Molecular Dynamics Simulation, *Biochemistry 50*, 2585-2593.

[72] Knotts, T. A., Rathore, N., and de Pablo, J. J. (2008) An Entropic Perspective of Protein Stability on Surfaces, *Biophysical Journal 94*, 4473-4483.

[73] Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia coli, *Journal of the American Chemical Society 124*, 9026-9027.

[74] Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli, *Journal of Molecular Biology 395*, 361-374.

[75] Smith, M. T., Hawes, A. K., Shrestha, P., Rainsdon, J. M., Wu, J. C., and Bundy, B. C. (2014) Alternative fermentation conditions for improved Escherichia coli-based cell-free protein synthesis for proteins requiring supplemental components for proper synthesis, *Process Biochemistry 49*, 217-222.

[76] Smith , M. T., Berkheimer , S. D., Werner , C. J., and Bundy, B. C. (2014) Lyophilized E***scherichia coli*** -based cell-free systems for robust, high-density, long-term storageLyophilized Escherichia coli-based cell-free systems for robust, high-density, long-term storage, *BioTechniques 56*, 8.

[77] Earl, C. C., Smith, M. T., Lease, R. A., and Bundy, B. C. (2017) Polyvinylsulfonic acid: A Low-cost RNase inhibitor for enhanced RNA preservation and cell-free protein translation, *Bioengineered*, 1-8.

[78] Salehi, S. A., Smith, M. T., Bennet, A. M., Williams, J. B., Pitt, W. G., and Bundy, B. C. (2015) Cell-free Protein Synthesis of a Cytotoxic Cancer Therapeutic: Onconase Production and a Just-add-water Cell-free System, *Journal of Biotechnology*.

[79] Smith, M. T., Bennett, A. M., Hunt, J. M., and Bundy, B. C. (2015) Creating a completely "cell-free" system for protein synthesis, *Biotechnology Progress 31*, 1716-1719.

[80] Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013) A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli, *Biochemistry 52*, 1828-1837.

[81] Abràmoff, M. D., and Ram, S. J. (2004) Image Processing with ImageJ, *Biophotonics International 11*, 7.
[82] Debets, M. F., van Berkel, S. S., Dommerholt, J., Dirks, A. J., Rutjes, F. P. J. T., and van Delft, F. L. (2011) Bioconjugation with Strained Alkenes and Alkynes, *Accounts of Chemical Research 44*, 805-815.
[83] Debets, M. F., van Berkel, S. S., Schoffelen, S., Rutjes, F. P. J. T., van Hest, J. C. M., and van Delft, F. L. (2010) Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition, *Chemical Communications 46*, 97-99.
[84] Karanicolas, J., and Brooks Iii, C. L. (2003) Improved Gō-like Models Demonstrate the Robustness of Protein Folding Mechanisms Towards Non-native Interactions, *Journal of Molecular Biology 334*, 309-325.
[85] Karanicolas, J., and Brooks, C. L. (2002) The origins of asymmetry in the folding transition states of protein L and protein G, *Protein Science 11*, 2351-2361.
[86] Dennington, R., Keith, T. A., and Millam, J. M. (2016) GaussView, Version 5, Semichem Inc., Shawnee Mission, KS.
[87] Jensen, H. B., and Kleppe, K. (1972) Effect of Ionic Strength, pH, Amines and Divalent Cations on the Lytic Activity of T4 Lysozyme, *European Journal of Biochemistry 28*, 116-122.