ACS Biomaterials SCIENCE & ENGINEERING

Article

pubs.acs.org/journal/abseba

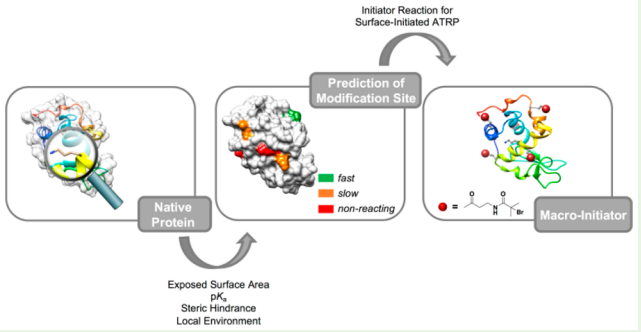
Tertiary Structure-Based Prediction of How ATRP Initiators React with Proteins

Sheiliza Carmali,[†] Hironobu Murata,[†] Erika Amemiya,[†] Krzysztof Matyjaszewski,[†] and Alan J. Russell*[,†,‡]

[†]Center for Polymer-Based Protein Engineering and [‡]Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, United States

Supporting Information

ABSTRACT: The growth of polymers from the surface of proteins via controlled radical polymerization depends on the attachment of small molecule initiators to amino acid residues. Our ability to control and harness the power of polymer-based protein engineering is reliant on the accuracy of prediction where and how fast atom transfer radical polymerization (ATRP) initiators will react with a protein surface. We performed a systematic characterization of the reaction between a bromine-functionalized *N*-hydroxysuccinimide amine-reactive ATRP initiator and the amino groups in lysozyme and chymotrypsin. The tertiary structures of the proteins were used to predict computationally α-amino group and lysine side-chain accessibility and analyze the chemical and structural environment of the amino groups. To predict reactivity from accessibility calculations, a probe radius that resembled the size of the initiator molecule was used. Experimental data showed that the rate of initiator−protein modification at each amine site was related to surface accessibility but not the pK_a of amino groups. Further refinements of the prediction of where the initiator modified the protein and in what sequence were achieved by considering the local environment of each amino group.



KEYWORDS: *protein, ATRP, initiator, reactivity, amines, modification, structure-based, prediction*

INTRODUCTION

Protein function is dependent on tertiary structure and a delicate balance of interactions between the protein and its environment through protein surface-solvent interactions. For enzymes, nature's catalysts, biological activity is completely dependent on whether substrates and products can access and depart from the active site of the protein. Because of the link between structure and function, extensive effort has been placed on mapping protein structure[1,2] and predicting protein−protein interactions.[3,4]

Polymer-based protein engineering (PBPE) by surface-initiated ATRP induces structural changes to the protein by generating protein−polymer conjugates with dense polymer shells that impact protein function.[5−13] Our understanding of how to grow polymer chains and analyze polymer−protein conjugates is increasing exponentially,[14,15] yet our ability to use the tertiary structure of a protein to predict the outcome of polymer−protein conjugate synthesis has remained limited. Significant hurdles also remain in understanding how to predict protein−polymer interactions. Protein−polymer conjugate properties are considerably influenced by the site of modification on the protein surface combined with the chemical and physical properties of the polymer.[16] Decoupling of these properties is critical in determining how to rationally tailor protein−polymer interactions. To achieve the full potential of polymer-enhanced biomacromolecular systems, we must learn to reproducibly tailor, control, and direct the reaction between polymers or initiators and proteins.

As we strive to control initiator−protein reactions, we have focused on determining the site and rate of reaction between amine-reactive ATRP initiators and protein amino groups[11,17−20] (Figure 1). Herein, we define the combination of site and rate as the specificity of the initiator−protein reaction. Over the last several decades, researchers, including ourselves, have suggested that side-chain pK_a,[18,20] exposed surface area (ESA),[11,17,19,21] polarity, and hydrophobicity[22] can influence the specificity of the side-chain−initiator/polymer interaction.

Amine modification in proteins targets mainly lysine residues due to their intrinsically high reactivity and the perception that their side-chain amino groups are accessible.[23] At neutral pH, the positively charged, polar side chain of lysine residues generally drives lysine-containing regions of the protein backbone to the surface during folding. Przybylski and co-workers[24] reported that upon acetylation of lysine residues in lysozyme, the relative reactivities for the modified lysines

Received: May 5, 2017
Accepted: July 15, 2017
Published: July 15, 2017

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

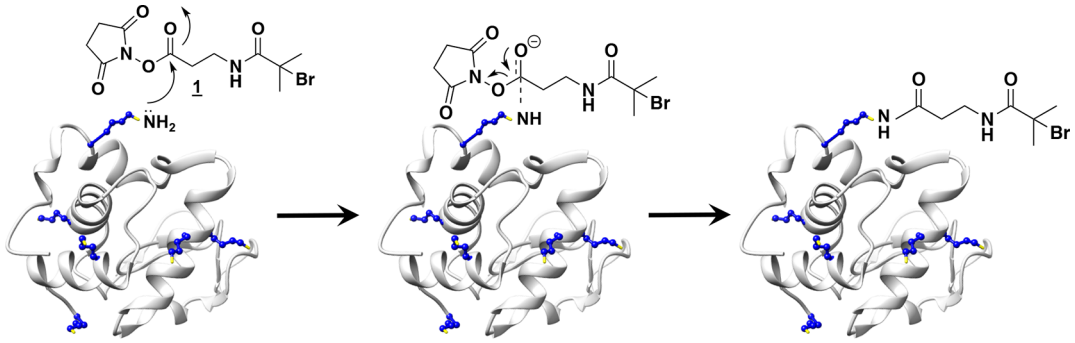


Figure 1. Modification of the amino groups in lysozyme with bromine-functionalized *N*-hydroxysuccinimide (NHS) ATRP initiator **1**. The amino group, functioning as a nucleophile, attacks the electron-deficient carbonyl group in **1** leading to the displacement of the NHS group and consequent covalent attachment of the initiator molecule to the protein.

correlated well with their exposed surface areas. Kodíček's group[25] also focused on the reactivity of amino acids with diethylpyrocarbonate (DEP) and found ESA to be an important factor in modification ability. However, in many proteins, when those lysines are coupled in tight electrostatic interactions the amine functionality can also become buried and inaccessible. Thus, determining the exposed surface areas for targeted amine groups on proteins can provide a good indication of their reactivity.

Historically, the ESA of a protein has been used to quantify hydrophobic burial and to further understand protein-folding.[26] More recently, amino acid exposed surface areas have been used for secondary structure prediction[27] and to help determine potential residues as active sites.[28−30]

ESA has generally been defined as the surface area of a region of a protein that is accessible to a water molecule with a radius of 1.4 Å.[26] However, larger spherical probe sizes have been used to predict antigenic areas on a protein surface.[31,32] Probes of varying radii have also been used to detect pockets on protein surfaces and potential ligand binding sites.[33]

In addition to ESA, pK_a has been suggested as an important predictor of where modification may preferentially occur on the surface of proteins.[34−36] This is often the case for the N-terminal amino group of a protein that will also react vigorously with amine-reactive moieties. The pK_a of the amine determines the relative amounts of nonreactive acid (NH_3^+) and reactive conjugate base (NH_2) at a given pH. The lower pK_a of the α-amino group from the N-terminus of a protein (7.7 ± 0.5[37]) can lead to higher reactivity when compared to the pK_a's of lysine ε-amino group that are typically between 9.3 and 10.5. This difference in pK_a has been exploited in site-specific approaches to modification at the N-terminal group.[38−42]

To accurately model lysine side-chain and N-terminal group accessibility to ATRP initiator molecules, we explored the use of a probe radius resembling the size of the initiator for ESA calculations. Two model proteins, lysozyme and chymotrypsin, were used. Chymotrypsin is a well characterized protease that participates in the digestion of proteins consumed as food in the small intestine.[43] Lysozyme is an antimicrobial enzyme and an important component of the innate immune system.[44] Both proteins have been widely used for protein conjugation,[6,7,45−47] have well characterized three-dimensional structures, are easily available, and are generally stable proteins. These features make them ideal workhorses through which to unravel the underlying principles that govern the tertiary structure dependence of amine reactivity. Enzymatic and chemical digestion studies of protein−initiator complexes followed by peptide mass mapping were used to identify initiator modification sites. Experimental results were correlated with tertiary structure-derived predictions of calculated ESA, pK_a, local charge, and environment for the targeted amino group.

EXPERIMENTAL SECTION

Materials. Hen egg-white lysozyme (Lyz) from *Gallus gallus*, α-chymotrypsin (CT) from bovine pancreas, 2-nitro-5-thiocyanatobenzoic acid (NTCB), glycine, and guanidine HCl were purchased from Sigma-Aldrich (St Louis, MO) and used without further purification. Bromine-functionalized *N*-hydroxysuccinimide ATRP initiator **1** synthesis was carried out as described previously.[7] Dialysis tubing (molecular weight cut off 15 kDa Spectra/Por, Spectrum Laboratories Inc., CA) for macroinitiator isolation, ZipTipsC_{18} microtips (Millipore, catalogue no. ZTC1 8M0 08), and In-Solution Tryptic Digestion and Guanidination Kits (catalogue no. 89895) were purchased from Thermo Fisher Scientific (Pittsburgh, PA).

Lysine and N-Terminus Surface Accessibility Analysis. The tertiary structures of Lyz from *Gallus gallus*[48] and Bovine CT were downloaded from the Protein Data Bank (PDB code: 1LYZ and 4CHA, respectively). The ionization state of the side-chain residues in both proteins was corrected to pH 8.0 using Avogadro software.[49] Absolute values for accessible surface areas (Å^2) of lysine side chains and the α-amino groups from the N-termini were calculated using Discovery Studio Visualizer version 4.5.0.15071 (Dassault Systèmes). Accessible surface area values were determined with a probe radius of 1.4 Å (equivalent to the radius of water molecule) and 4.2 Å (average radius of bromine-functionalized NHS initiator, as determined by molecular dynamics simulation (Supporting Information)). The accessible surface area values for lysine side chains and α-amino groups from the N-termini were used as absolute values obtained from the calculations.

Prediction of Lysine and pK_a. The pK_a of the α- and ε-amino groups was predicted using PROPKA (Supporting Information). Lysine residues are ionizable residues which can exchange protons with their environment. However, upon protein folding lysine residues (and other ionizable residues) are transferred from a solution-like environment to an environment which is determined by the tertiary structure of the protein. Additionally, in a folded protein, lysine residues will be closer to other ionizable groups in the protein, allowing interactions with permanent charges and dipoles. These structural and electrostatic effects can alter the pK_a value of the ε-amino group in the lysine side chain as well as the α-amino group from the N-terminus. The PROPKA method consists of an empirical approach to determining pK_a values by calculating the effect of the protein environment on the pK_a value of an amino acid side chain.[50] For the ε-amino group in lysine residues and the α-amino group from the N-terminus, the starting pK_a in the calculations are 10.5 and 8.0, respectively.

Reaction between ATRP Initiator 1 and Lyz. A family of lysozyme macroinitiators with varying degrees of modification was

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

prepared using bromine-functionalized ATRP initiator **1**. ATRP initiator **1** (1.6−24.6 mg, 0.005−0.07 mmol) and Lyz (50 mg, 0.004 mmol protein, 0.02 mmol NH_2 group in lysine residues and N-terminus) were dissolved in 10 mL of 0.1 M sodium phosphate buffer, pH 8.0. The solution was stirred at 4 °C for 2 h. After this time, the reaction was dialyzed against deionized water, using dialysis tubing with a molecular weight cutoff of 15 kDa, for 24 h at 4 °C, and then lyophilized.

Reaction between ATRP Initiator 1 and CT. A family of chymotrypsin macroinitiators with varying degrees of modification was prepared using bromine-functionalized ATRP initiator **1**. ATRP initiator **1** (0.7−29.4 mg, 0.002−0.09 mmol) and CT (50 mg, 0.002 mmol protein, 0.03 mmol NH_2 group in lysine residues and N-terminus) were dissolved in 10 mL of 0.1 M sodium phosphate buffer, pH 6.0, 7.0, or 8.0. The solution was stirred at 4 °C for 2 h. After this time, the reaction was dialyzed against deionized water using dialysis tubing with a molecular weight cutoff of 15 kDa, for 24 h at 4 °C, and then lyophilized.

Trypsin Digestion of Protein−Initiator Conjugates. Trypsin digests were used to generate peptide fragments from which protein−initiator attachment sites could be determined using MALDI mass spectrometry. Samples were digested according to the protocol described in the In-Solution Tryptic Digestion and Guanidination Kit. In brief, 20 μg of protein or protein−initiator complexes (10 μL of a 2 mg/mL protein solution in ultrapure water) were added to 15 μL of 50 mM ammonium bicarbonate with 1.5 μL of 100 mM dithiothreitol (DTT) in a 0.5 mL Eppendorf tube. The reaction was incubated for 5 min at 95 °C. Thiol alkylation was conducted by the addition of 3 μL of 100 mM iodoacetamide aqueous solution to the protein solution and incubation in the dark for 20 min at room temperature. After this time, 1 μL of 100 ng of trypsin was added to the protein solution, and the reaction was incubated at 37 °C for 3 h. An additional 1 μL of 100 ng of trypsin was subsequently added. The reaction was terminated after a total reaction time of 5 h by the addition of trifluoroacetic acid (TFA). Digested samples were purified using ZipTipC_{18} microtips and eluted with 2 μL of matrix solution (20 mg/mL sinapinic acid in 50% acetonitrile with 0.1% TFA) directly onto a MALDI-TOF plate for subsequent analysis. The molecular weight of the expected peptide fragments before and after digestion was predicted using PeptideCutter (ExPASy Bioinformatics Portal, Swiss Institute of Bioinformatics). A complete list of identified tryptic peptides can be found in Supporting Information.

2-Nitro-5-thiocyanobenzoic (NTCB) Digestion of Protein−Initiator Conjugates. All buffers and reagents were prepared fresh for the NTCB reactions. The following procedure was carried out for both lysozyme− and chymotrypsin−initiator modified complexes. Native protein or initiator-modified protein complexes (20 μg) were dissolved in 1 M glycine, 6 M guanidine−HCl at pH 10.0 (15 μL) and treated with 1.5 μL of 100 mM of DTT for 5 min at 95 °C. After this time, 5 μL for lysozyme samples or 24 μL of 22 mM NTCB (approximately 20-fold excess of the total cysteine content in protein) was added, and digestion was carried out at 37 °C for 18 h.[51] The reaction was stopped by the addition of 3 μL of TFA, and digested samples were purified using ZipTipC_{18} microtips and eluted with 2 μL of matrix solution (20 mg/mL sinapinic acid in 50% acetonitrile with 0.1% TFA) directly onto a MALDI-TOF plate for subsequent analysis. The molecular weight of the expected peptide fragments before and after digestion was predicted using PeptideCutter (ExPASy Bioinformatics Portal, Swiss Institute of Bioinformatics). The identified cyanylated peptides are listed in Supporting Information.

Determination of the Reaction Rate between ATRP Initiator 1 and Lyz. Excess ATRP initiator **1** (2.46 mg, 0.0073 mmol) was added to native lysozyme (5 mg, 0.00035 mmol protein, 0.002 mmol NH_2 group in lysine residues and N-terminus) and dissolved in 1 mL of 0.1 M sodium phosphate buffer, pH 8.0, and placed at 4 °C. At t = 0 s (prior to the addition of initiator **1**) and after every 30 s for a total time of 5 min, 20 μL aliquots were taken. The reaction was stopped by lowering the pH through the addition of 1 μL of TFA. Typically, *N*-hydroxysuccinimide activated substrates react efficiently with amine groups in aqueous solutions under pH conditions close to those found physiologically (pH ∼ 6−9).[52] Addition of TFA allowed the pH to drop and prevented significant aminolysis. Samples were diluted to 1 mg/mL with deionized water for subsequent trypsin digestion (as described above) and/or MALDI-TOF analysis. To determine the reaction rate of individual amine residues, 10 μL aliquots taken from the initiator reaction were added immediately to 15 μL of 50 mM ammonium bicarbonate with 1.5 μL of 100 mM DTT in a 0.5 mL Eppendorf tube. The reaction was incubated for 5 min at 95 °C, and subsequent steps for trypsin digestion were carried out as described previously. Plots of peak height ratio versus time were used to calculate an initial rate (s^{-1}).

Determination of the Reaction Rate between ATRP Initiator 1 and CT. Excess ATRP initiator **1** (2.94 mg, 0.0087 mmol) was added to native chymotrypsin (5 mg, 0.00019 mmol protein, 0.003 mmol NH_2 group in lysine residues and N-terminus) and dissolved in 1 mL of 0.1 M sodium phosphate buffer, pH 8.0, and placed at 4 °C. At t = 0 s (prior to the addition of initiator **1**) and after every 30 s for a total time of 5 min, 20 μL aliquots were taken. The reaction was stopped by the addition of 1 μL of TFA. Samples were diluted to 1 mg/mL with deionized water for subsequent trypsin digestion (as described above) and/or MALDI-TOF analysis. To determine the reaction rate of individual amine residues, 10 μL aliquots taken from the initiator reaction were added immediately to 15 μL of 50 mM ammonium bicarbonate with 1.5 μL of 100 mM DTT in a 0.5 mL Eppendorf tube. The reaction was incubated for 5 min at 95 °C, and subsequent steps for trypsin digestion were carried out as described previously. Plots of peak height ratio versus time were used to calculate an initial rate (s^{-1}).

MALDI-TOF Analysis. Matrix assisted laser desorption ionization-time-of-flight mass spectroscopy (MALDI-TOF MS) measurements were recorded using a PerSeptive Voyager STR MS with nitrogen laser (337 nm) and 20 kV accelerating voltage with a grid voltage of 90%. At least 300 laser shots covering the complete spot were accumulated for each spectrum.

For the determination of the molecular weight of synthesized macroinitiator complexes, sinapinic acid (20 mg/mL) in 50% acetonitrile with 0.1% trifluoroacetic acid was used as matrix. Protein solution (1.0 mg/mL) was mixed with an equal volume of matrix, and 2 μL of the resulting mixture was loaded onto a silver sterling plate target. Apomyoglobin, cytochrome *C*, and aldolase were used as calibration samples. The extent of modification was determined by subtracting the protein−initiator conjugates' m/z values from the protein lysozyme m/z and dividing by the molecular weight of the initiator (220.9 g/mol).

Relative aminolysis rates were determined by profile analysis of the digestion samples. While MALDI-TOF does not show a quantitative relationship between ion current and amount of analyte in the sample matrix, the relative signal intensities within the mass spectrum are constant. The relative amount of initiator modification on each lysine and α-amino group from the N-terminus was defined as the ratio of the peak height of the corresponding modified peptide fragment to that of a nonmodifiable peptide fragment. For lysozyme, the nonmodifiable fragment used was peptide ^{62}WWCNDGR with 997.66 m/z. For chymotrypsin, IQK^{82} with 689.11 m/z was used since K82 was found to not react with initiator **1**.

Molecular weights of peptide fragments obtained in protein digests were determined after the solutions were purified by use of ZipTipC_{18} microtips. Bradykinin fragment, angiotensin II (human), and insulin oxidized B chain (bovine) were used for calibration.

RESULTS AND DISCUSSION

Lysozyme. Lysozyme contains seven possible sites that could react with initiator **1**; one α-amino group of N-terminal (K1) and the six ε-amino groups of the lysine residues (K1, K13, K33, K96, K97, and K116).[53] Previous studies using X-ray crystallography have shown that the α- and ε-amino groups in lysozyme are in different structural and chemical environments, due to the presence of intra- and intermolecular interactions with neighboring residues and local charge variations.[54] We

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

hypothesized that the reactivity of the individual amino groups to initiator **1** would be tertiary structure-dependent.

Degree of Protein Modification by ATRP Initiator 1. We first exposed lysozyme to increasing amounts of bromine-functionalized initiator **1** in 0.1 M sodium phosphate buffer, pH 8, at 4 °C for 2 h. We aimed to understand whether initiator **1** would react equally with all accessible amino groups or if certain amino groups would react faster than others. Lysozyme macroinitiators (Lyz-Br) were purified by dialysis and analyzed by MALDI-TOF (Figure 2). The number of initiators in each

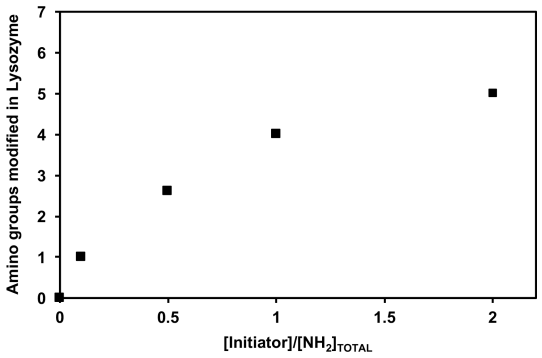


Figure 2. Dependence of initiator molecules reacting with the surface of lysozyme as a function of the initiator to the amino groups' stoichiometry.

macroinitiator complex was calculated by subtracting the Lyz-Br m/z values from native lysozyme m/z and dividing by the molecular weight of initiator **1** (Mw = 220.9 g/mol; Supporting Information). In the presence of excess initiator **1** (2 equiv with respect to the total number of amino groups in lysozyme), a maximum of five of the seven available amine groups were modified, indicating that two amine groups in lysozyme were nonreactive.

A decrease in the amino group-to-initiator ratio led to decreased density of surface modification. At lower stoichiometry (0.1 equiv), only one initiator molecule was coupled to the surface of lysozyme (Lyz-Br_1). This presented an opportunity to explore whether such minimally modified proteins were selectively modified. We were interested in whether the Lyz-Br_1 complex was consistently modified at the same amino group or if multiple single sites would generate a family of protein–initiator complexes.

Predicting Exposed Surface Areas of Amino Groups in Lysozyme. We first determined the exposed surface areas of all the amino groups in lysozyme. Absolute values for ESA have generally been calculated from a "rolling ball" algorithm which measures the surface area between the midpoint of the probe sphere and the surface of the protein.[55] Thus, the probe size will impact the observed ESA. We first used a spherical probe with the radii of a water molecule (1.4 Å) (Figure 3). The maximum ESA of lysine residues in an unfolded protein has been reported to be 211 $Å^2$ with the side chain having a maximum exposure of 114 $Å^2$.[23] It should be noted that in a folded protein the exposed surface area has been found to be approximately 45% lower.[23] Since there were no reported ESA data for the α-amino groups, we calculated absolute ESA from a reference tripeptide (Gly-Lys-Gly), which in the extended conformation was found to be 52 $Å^2$.

We next calculated the ESA using a probe size that approximated that for bromine–initiator **1**, which we had

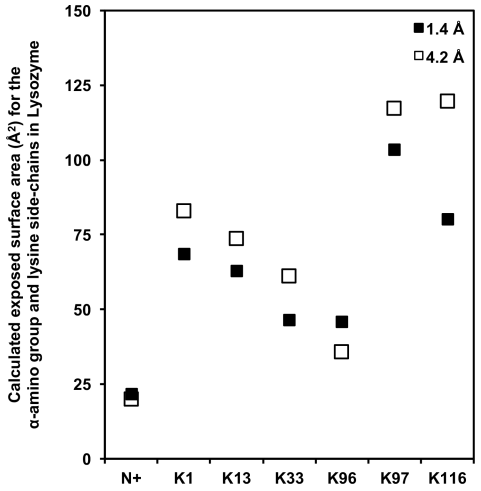


Figure 3. Calculated exposed surface area for the α-amino group and lysine side chains in native lysozyme. Computational probe radii of 1.4 Å, representing a water molecule (shaded squares), and 4.2 Å, representing the average radius of a bromine-functionalized NHS ATRP initiator **1** (nonshaded squares), were used.

determined previously from a short molecular dynamics simulation in water (4.2 Å). We anticipated that the reactivity of the α-amino group and lysine side chains with initiator **1** would be sharply dependent on appropriately calculated ESA. Interestingly, the lysine side chains in lysozyme had a larger ESA (Lys 96 having the lowest ESA of 46 $Å^2$) than the α-amino group (22 $Å^2$)[56] (Figure 3). Generally, it would be expected that the N-and C-termini, as charged residues, would be on the protein surface.[57] However, charge burial can create strong stabilizing electrostatic interactions in proteins.[58]

Comparison between ESA calculations using both probe sizes showed a more pronounced distinction between the exposed surface areas of the lysine side chains with the initiator probe (4.2 Å). Overall, both ESA calculations maintained the trend of accessibility among all amino groups with the exception of K116. An increased ESA for K116 was obtained with the initiator probe. This indicated that K116 would be more accessible to the initiator molecule than otherwise anticipated from ESA calculations with a water probe. This observation was consistent with previously reported molecular dynamics simulations of lysozyme that also showed K116 as the most accessible lysine residue.[59]

These ESA calculations can also be extended to other initiator molecules and even polymers. As with the initiator **1**, an adequate probe radius for the coupling reagent/molecule would be necessary. Molecular dynamics simulation studies or experimentally determined hydrodynamic size (in the case of polymers) may be explored to estimate the probe size to be used in the ESA calculations. An accurate probe size would allow for a better representation of the accessible interaction between amino groups on a protein and the coupling reagent/molecule.

In our study, ESA calculations reflected the use of a single initiator structure and the coupling reaction through NHS chemistry. This approach could be also expanded to other amine-reactive chemistries[60] (e.g., carbodiimides, carbonates, anhydrides, aldehydes, or isocyanates) with additional computational studies. The structure–reactivity dependence of the amino–initiator interactions could also be evaluated using molecular docking studies.[61,62] Moreover, the flexibility and

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086–2097

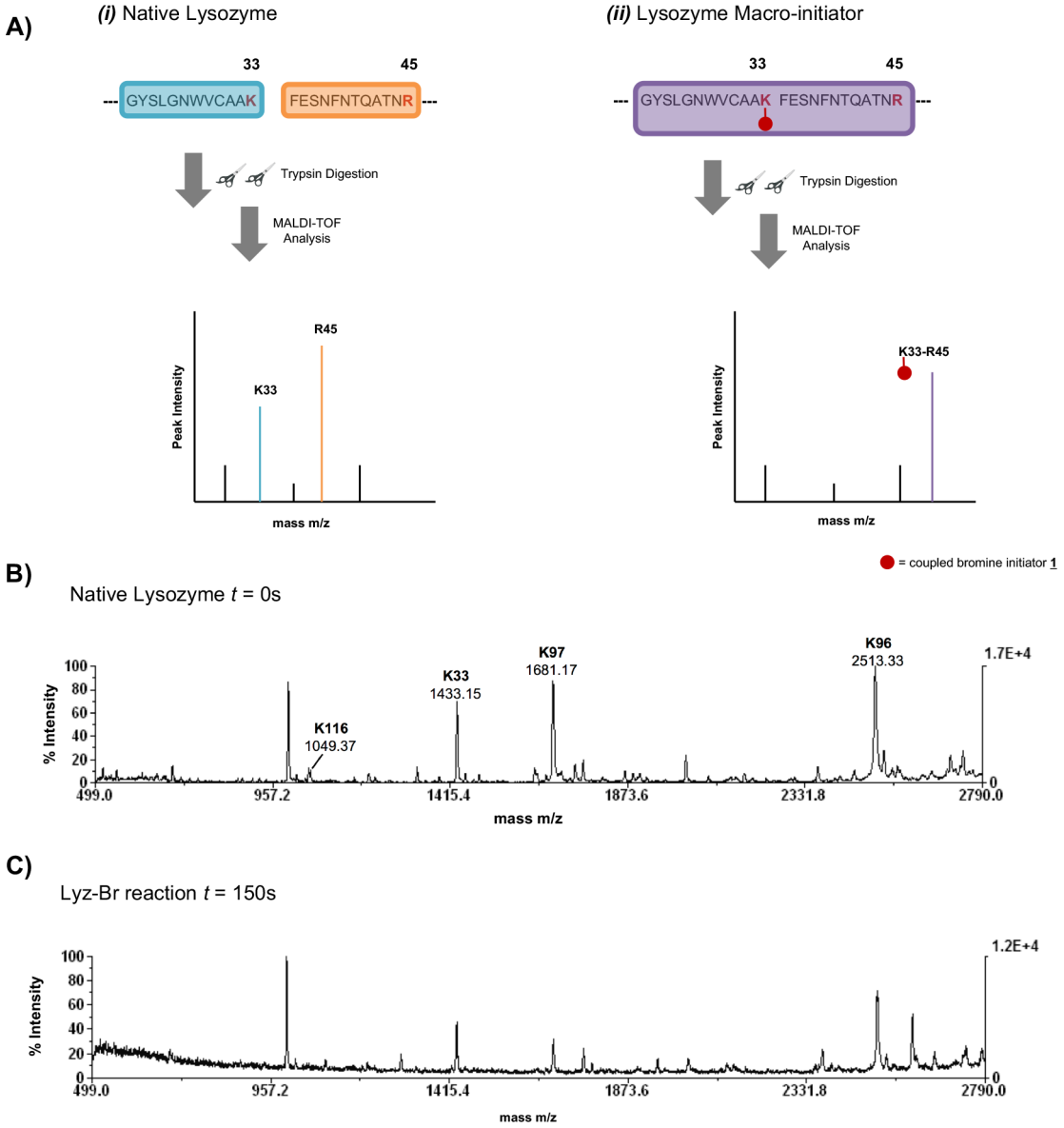


Figure 4. (A) Schematic representation of the trypsin digestion of lysozyme at K33. Trypsin cleaved the peptide backbone at K33 yielding characteristic peaks in MALDI-TOF for the peptide fragment containing K33 and the subsequent fragment with R45. When K33 is modified with bromine initiator **1**, trypsin is unable to cut, and the peaks for K33 and R45 do not appear in the mass spectra. Alternatively, a combined fragment is formed with a distinct mass. MALDI-TOF spectra acquired after trypsin digestion of (B) native lysozyme, prior to the addition of bromine initiator **1** and (C) lysozyme−initiator reaction after 150 s.

conformational changes that occur from the amino−initiator coupling reaction could be further understood using molecular dynamics simulations.

Experimental Determination of the Site of Lysozyme Modification by ATRP Initiator 1. To understand if ESA calculations with the initiator probe could be used to predict the site and sequence of lysozyme modification by initiators, we needed to identify where the ATRP initiators were covalently attached to the surface of lysozyme. Trypsin digestion studies were carried out on the various Lyz-Br macroinitiators prepared previously (Figure 2), and the resulting peptide fragments were analyzed by MALDI-TOF. Alternatively, other mass spectrometry techniques, such as LC-MS, could have been used for peptide fragment identification. However, MALDI-TOF was chosen due to its higher acquisition speed and ease of data analysis.

Trypsin hydrolyzes peptide bonds at the carboxy-terminal side of lysine and arginine residues.[63] We hypothesized that modification with initiator **1** would prevent trypsin hydrolysis at modified lysines.[11] This would translate to a decrease or disappearance of the expected peptide fragment peak in the mass spectra (Figure 4A).

As mentioned previously, lysozyme contains six lysine residues that may be cleaved by trypsin. Tryptic digestion of native lysozyme resulted in four peaks that led to the identification of K116 (1049.37 m/z), K33 (1433.15 m/z, $[M+NH_4]^+$), K97 (1681.17 m/z, $[M+NH_4]^{2+}$), and K96 (2513.33 m/z, $[M+NH_4]^+$) (Figure 4B). Comparison of the mass spectrum for fully modified lysozyme macroinitiator (Lyz-Br_5), after trypsin digestion, showed a significantly different fragmentation pattern. The peptide fragment peak corresponding to K116 disappeared in Lyz-Br_5, while fragment peaks for

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

K33 and K97 were significantly reduced in peak height. These results indicated that initiator modification had occurred at these three sites (Table 1). Interestingly, the peptide fragment for K96 was present for all reaction times with Lyz-Br_5, suggesting that K96 had not reacted with initiator 1. Although this could have resulted from reduced intrinsic reactivity, we believe that the proximity of K96 to K97 may have generated steric hindrance.

Table 1. Identification of the Site of Modification in Lysozyme−Initiator Complexes after Trypsin and NTCB Digestion

	peptide fragment peaks[a]				
	K13	K33	K96	K97	K116
native Lyz	+	+	+	+	+
Lyz-Br_1	+	+	+	+	−
Lyz-Br_3	+	−	+	−	−
Lyz-Br_5	−	−	+	−	−

[a]Peptide fragment peaks corresponding to K13, K33, K96, K97, and K116 identified by MALDI-TOF. Symbols in the table indicate the presence of a peak (+) or the disappearance/decrease of a peak (−) after digestion.

Chemical digestion using (2-nitro-5-thiocyanobenzoic acid) NTCB was also performed to confirm the site of modification in each sample (Supporting Information). NTCB reacts specifically by cyanylation with cysteine thiols. Under alkaline conditions, the N-terminal peptide bond of the cyanylated residue for native and modified lysozyme was cleaved, and the peptide fragments were mapped to the protein sequence by MALDI-TOF. The fragmentation pattern was compared for native and initiator-modified proteins. MALDI TOF analysis of the digests showed that K33, K97, and K116 had been modified on the surface of the lysozyme. This was in agreement with previous trypsin digestion studies. Additionally, NTCB cleavage also allowed for the identification of K13 as a modified site (Table 1).

Naturally, if the digested protein sample contained a mixture of modification sites, this technique would not identify them separately. It was therefore important to use stoichiometry to produce macroinitiators with single sites of modification (Lyz-Br_1), followed by samples with two or more initiators per protein molecule. Digestion studies with Lyz-Br_1 confirmed that initiator monosubstitution occurred only at K116 (Table 1). This result implied that upon reaction with initiator 1, modification occurred at a preferred amino site.

These data allowed us to demonstrate definitively that the reacting amino groups in lysozyme were K13, K33, K97, and K116, while K96 was found to be nonreacting. We next focused on determining the order in which the initiator molecules reacted with the amino groups on the surface of lysozyme.

Experimental Determination of the Sequence in Which Lysozyme Amino Groups React with ATRP Initiator 1. To determine the sequence of reaction for the α- and ε-amino groups, if there was one, we next reacted the lysozyme with excess initiator 1. At gradually increasing times, we enzymatically digested the resulting protein−initiator complexes and analyzed the resulting peptide fragments by MALDI-TOF. To gain insight into the reaction rates for each amino group, we inferred a rate from the ratio of the peak height of the corresponding modified peptide fragment to that of a nonmodifiable peptide fragment. Peak ratios on MALDI are the best estimate we have for rate but are distinctly imperfect since the efficiency of MALDI will be sample dependent. The rate was calculated from the slope of the initial decay in peak height. For lysozyme, peptide fragment ^{62}WWCNDGR (997.66 m/z) was chosen as the nonmodifiable lysine-free fragment.

Fragment ^{117}GTDVQAWIR, used as a surrogate to identify K116, was found to disappear almost immediately. The rate of modification for K116 (1.96 s^{-1}) indicated that it was the most reactive and available group in lysozyme to react with ATRP initiator 1. K33 and K97 were also found to react rapidly (0.70 and 0.6 s^{-1}, respectively) with initiator 1 but always after K116. In contrast, K96 showed a very slow decrease in relative peak height for peptide fragment NLCNIPCSALLS-SDITASVNCAK^{96} (0.277 s^{-1}). As noted previously, this fragment was still present after 2 h indicating that K96 was an extremely slow reacting residue. The proximity of K96 to K97 and the lower accessibility of K96 undoubtedly contributed to the essentially nonreacting behavior of the residue.

The combined data from the site and rate of modification were conclusive in that each amino group had distinct reactivity. Careful experiments allowed us to determine the sequence in which the initiator reacted with the protein surface.

Chymotrypsin. Lysozyme is a small protein from which the beginning of a cohesive hypothesis could emerge. We next selected a second protein, CT, which was larger and more complex, to determine if similar trends in reactivity were observable. CT has one α-amino group and 14 lysine side chains, with 15 total possible sites of modification with bromine initiator 1.

Degree of Chymotrypsin Modification by ATRP Initiator 1. Previously, we have shown that upon reaction with excess initiator 1, a maximum modification of 12 of the 15 available amine groups could be obtained.[6] We initially dissolved native chymotrypsin in phosphate buffer (0.1 M pH 8.0 at 4 °C). The enzyme was then reacted with increasing amounts of initiator 1. At the maximum stoichiometry used for this experiment, 11 amine residues were modified. Increasing the stoichiometric ratio between amino groups and initiator generated enzyme conjugated with an increased number of initiator sites (Figure 5A).

We also explored the impact of pH on the degree of initiator modification of CT (Figure 5B). Excess initiator 1 was reacted with native chymotrypsin in sodium phosphate buffer at three different pHs (0.1 M pH 6.0, 7.0, or 8.0 for 2 h at 4 °C). After dialysis, CT-Br complexes were analyzed by MALDI-TOF. Since at pH 6.0 (versus 7.0 and above) there would be more amino groups that were in the protonated state than the unprotonated state, it was not surprising that at pH 6.0 fewer amino groups reacted with initiator 1. Our core interest remained in determining the site and sequence in which the amino groups in CT reacted with initiator 1.

Predicting Exposed Surface Areas of Amino Groups in Chymotrypsin. The exposed surface areas for the amino groups in CT were calculated using probes with the radii of water or initiator molecules (Figure 6). Calculations using the water probe showed that ESA varied, in general from 100−150 $Å^2$. Given that the ESA's were all similar, then if CT behaved in a manner similar to lysozyme, we would have predicted that CT-Br_1 would consist of a mixture of proteins modified at multiple sites. In contrast, ESA calculations using the initiator probe were much more distinct. Residues K36, K79, K93, and

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

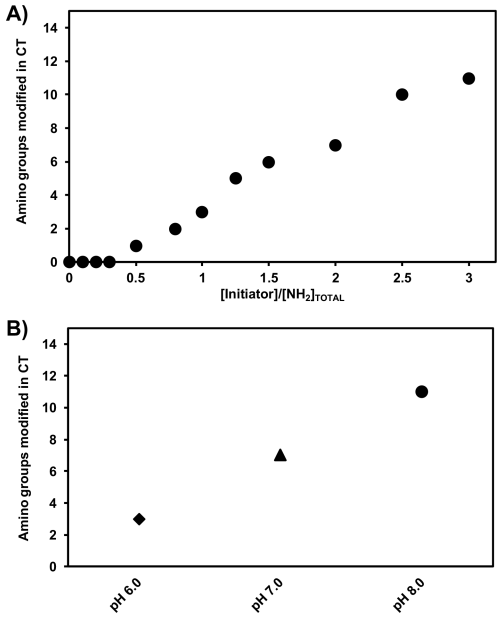


Figure 5. Dependence of initiator molecules reacting with the surface of chymotrypsin as a function of (A) initiator to amino group stoichiometry and (B) pH.

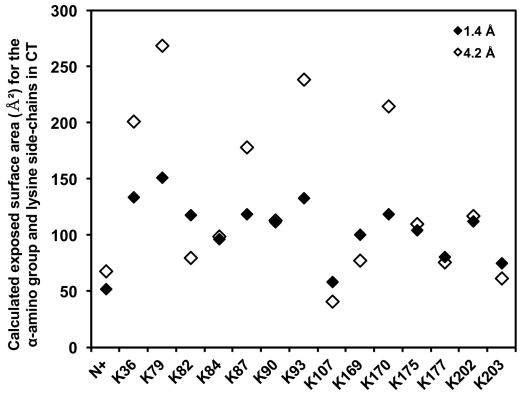


Figure 6. Calculated exposed surface area for the α-amino group and lysine side chains in native chymotrypsin. Computational probe radii of 1.4 Å, representing a water molecule (shaded squares), and 4.2 Å, representing the average radius of a bromine-functionalized NHS ATRP initiator **1** (nonshaded squares), were used.

K170 were found to have absolute ESA above 150 Å^2, which suggested increased reactivity with initiator at these amino sites in comparison to the other lysine residues available in CT.

Residues K107 and K203 have unusually low calculated exposed surface areas when determined with the initiator probe. We hypothesized that these two lysine residues would not react with the ATRP initiator. We next sought to determine experimentally which amino groups in CT were modified.

Experimental Determination of the Site of Chymotrypsin Modification by ATRP Initiator 1. As with lysozyme, digestion studies with native CT and CT-Br macroinitiators were conducted to determine the site of modification. Initially, the most reactive residues were examined by varying the stoichiometry and pH of the initiator−CT reaction. From our first exploratory initiator reactions, excess initiator (2.5 equiv) at pH 6.0 led to the coupling of three initiator molecules to the surface of chymotrypsin. As expected, for pH 7.0 a lower stoichiometry (1.0 equiv) could be used to covalently attach the same number of initiators to chymotrypsin. Similarly, at pH 8.0 a stoichiometry of 0.8 equiv gave three initiator molecules per chymotrypsin. We hypothesized that in each case the first three amino groups to react were the same. Comparison of the mass spectrum obtained for NTCB digested native chymotrypsin with all three CT-Br$_3$ conjugates showed that modification of chymotrypsin led to an identical fragmentation pattern. Two significant MALDI-TOF peaks were found to disappear in the CT-Br$_3$ conjugates. Peptide fragments ^{1}CGVPAIQPVLSGLSRIVNGEEAVPGSWPWQVSLQD^{36}KTGFHF (4,497.11 m/z) and fragment C^{202}K^{203}KNGAWTLVGIVSWGSST (2067.11 m/z) were not present in the spectra for CT-Br$_3$. These results suggested that the same three amine groups were the first to react with the ATRP initiator (N-terminal group, K36 and K202/203).

To further expand on which lysine residues were reacting and nonreacting, an additional enzymatic digestion was performed (Supporting Information). Fully modified chymotrypsin (CT-Br11) with 11 initiators coupled to the surface of chymotrypsin was digested with trypsin, and the resulting MALDI-TOF was compared to trypsin digested native chymotrypsin.

Again, we found that the N-terminal group, K36, and K202 reacted with the initiator. Peptide fragment DAMICAGASGVSSCMGDSGGPLVC^{202}K^{203}K was still present in CT-Br$_{11}$ after trypsin digestion (2662.00 m/z, [M + Na]$^+$), indicating that trypsin was able to cleave efficiently at K203. This led us to believe that the initiator reacted with K202 but not K203. Trypsin digestion studies, coupled with NTCB digestion, showed that the other reacting amino groups were K79, K90, K169, and K177, while the nonreacting lysines were K170, K82, K203, and K107 (Table 2).

Experimental Determination of the Sequence in which Chymotrypsin Amino Groups React with ATRP Initiator 1. To determine the sequence of reaction for the α- and ε-amino groups in CT, excess initiator **1** was added to native chymotrypsin, and reaction aliquots were removed every 30 s for trypsin digestion and subsequent MALDI-TOF analysis. As before, rates of reaction at each amino site were determined by MALDI profile analysis. When working with

Table 2. Identification of the Site of Modification in Chymotrypsin−Initiator Complexes after Trypsin and NTCB Digestion

	peptide fragment peaks[a]										
	N$^+$	K36	K79	K82	K90	K107	K169	K170	K177	K202	K203
native CT	+	+	+	+	+	+	+	+	+	+	+
CT-Br$_1$	−	+	+	+	+	+	+	+	+	+	+
CT-Br$_3$	−	−	+	+	+	+	+	+	+	−	+
CT-Br$_{11}$	−	−	−	+	−	+	−	+	−	−	+

[a]Peptide fragment peaks corresponding to the amino groups in chymotrypsin identified by MALDI-TOF. Symbols in the table indicate the presence of a peak (+) or the disappearance/decrease of a peak (−) after digestion.

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

lysozyme, we selected a nonmodifiable peak as our standard when calculating ratios. However, due to the number of lysine residues in CT, we were not able to clearly identify a peak corresponding to a peptide fragment without a lysine residue. Since we had previously identified the lysine residues that did not react with initiator **1**, we chose peptide fragment IQK[82] (689.11 m/z, $[2M+NH_4]^+$) with nonreacting K82 as the internal peptide standard. The peak corresponding to this fragment was found to be present in the MALDI spectra throughout the experiment

Quantification of the relative peak heights in MALDI-TOF trypsin digestion spectra from the CT-initiator reaction showed that the N-terminal group, K36, and K202 had the highest reactivities (0.82, 1.29, and 1.52 s^{-1}, respectively). As described above, these data were consistent with digestion data for the CT-Br_3 conjugate that showed that these amino groups were fast-reacting (Table 2).

It was interesting that the lower pK_a of the α-amino group from the N-terminus did not yield a fast-reacting site when using the ratio of MALDI peak heights method. However, for the CT-Br_1 conjugate, where only one initiator was modified onto the surface of CT, trypsin and NTCB digestion confirmed that the first amino group to react was the N-terminus. This discrepancy reminded us that the MALDI-TOF peak ratio method was a semiquantitative tool.

The estimated rates for the initiator−lysine reaction at K107 and K203 supported the view that these sites did not react (0.06 and 0.18 s^{-1}, respectively), while K79, K90, K169, and K177 (0.37, 0.36, 0.32, and 0.25 s^{-1}, respectively) were found to have midlevel rates. One obvious outlier in the analysis was K79, which had a low reaction rate but high exposed surface area. A closer look at the coordinate file for the X-ray structure of CT showed that K79 had a very high B-factor (30.68 Å^2). The B-factor is used to quantitate the uncertainty in precise atomic positions and can indicate disorder in the protein crystal.[64] This helped explain the unusually high exposed surface area and the lower reactivity of K79 and led us to conclude that care should be taken when predicting the reactivity of lysine side chains with high B-factors.

Fine-Tuning the Predicted Site and Rate of Initiator−Amine Reaction. With the compiled data set for the observed reactivities for the amino groups in lysozyme and chymotrypsin, we sought to elaborate a set of rules to predict initiator−protein specificity. To accurately predict the site and rate of modification, it was critical to identify the key factors that affected the reaction at a given amino site and understand why those factors were important. As described in more detail below, we identified five factors that influenced the initiator−protein specificity: (1) ESA of the amino group (determined with a probe size radius of the initiator molecule); (2) steric hindrance in the region of the amino group (especially between adjacent lysine groups); (3) intrinsic pK_a of the amino group; (4) secondary structure in the region of the amino group; and (5) local environment (charge and noncovalent interactions) of the amino group.

The first two factors, ESA and steric hindrance, directly affected the ability of the amino group to react with the initiator molecule. From the data obtained from lysozyme and chymotrypsin, we observed that a minimum ESA of 50 Å^2 was needed for an amino group to be modified. Nonreacting residues K96 (Lyz) and K107 (CT) have unusually low exposed surface areas (36 and 41 Å^2, respectively). Moreover, for two amino groups located in close proximity, both groups would not be modified even if their ESA was above 50 Å^2. In CT, two pairs of neighboring lysines were present; K169 and K170, and K202 and K203. Because of steric hindrance, only two of the four residues reacted with initiator **1**. Initially, our assumption was that a higher ESA would dictate which amino group would react preferentially. This was, in fact, the case for K202 which was modified with initiator **1**. However, for the pair K169/K170, digestion studies showed that the less exposed K169 was modified. To further understand this discrepancy, we predicted the respective pK_a values for the ε-amino groups in the lysine side chains using PROPKA.[50] Interestingly, both K170 and K203 had higher pK_a's (10.52 and 10.5, respectively) when compared to their neighboring lysine residues. Lysine side chain pK_as for K169 and K202 were predicted to be 10.41 and 10.27, respectively. Thus, we inferred that for amino groups that had a minimum ESA but were in close proximity, a slight difference in the protonation states could determine the site of modification.

To further validate the premise that the reactivity of neighboring residues was influenced by steric hindrance, a 20 ns molecular dynamics simulation of native chymotrypsin in water was performed (see Supporting Information and Figure S11). The proximity of the ε-amino groups in the neighboring residues K169 and K170 was compared to that of the pair K90 and K93 over the simulation time. For K169/K170, we observed that the distance between the respective ε-amino groups was as close as 5 Å at certain times. In contrast, the closest distance that K90 and K93 are separated in the amino acid sequence by two residues was 13 Å. Thus, we hypothesize that sufficient proximity between two neighboring residues would allow only one of them to be modified.

Having established which factors impacted the site of modification, we focused next on understanding those that influenced the rate of the reaction. ESA calculations, coupled with a complete structural and chemical analysis and digestion studies allowed us to classify the amino groups in Lyz and CT as fast-, slow-, and nonreacting families. The use of a large probe size, approximating the initiator size, generally predicted which amino group fell into which family.

To further understand any trends in reactivity, we divided the data set into α- and ε-amino groups. Experimental data for CT showed that the α-amino group was the most reactive residue in that protein. Given that this α-amino had an ESA of 68 Å^2 (probe radius 4.2 Å) which was above the minimum required ESA, we classified this amino group as a reacting residue.

The ESA, along with a lower pK_a when compared to the remaining ε-amino groups explained the high level of reactivity. Of course, with one data point we should resist the temptation to draw conclusions, but perhaps it was reasonable to surmise that if a given α-amino group was accessible to the initiator, then the amino group would be fast-reacting.

We next examined the ε-amino groups from the lysine side chains in Lyz and CT and considered the impact of secondary structure. We expected that secondary structure would impact the accessibility of the amino group to initiator **1**. In general, residues in the coil-turn and β-sheet conformation were more accessible than that in the helical fold.[23] We therefore thought that if ε-amino groups in the helical fold were less exposed to react with initiator **1**, then the rate of reaction at these amino sites would be slower. This was consistent with the experimental data we gathered. Residues K13, K33, and K97 (Lyz) in the helical fold were found to have a relatively slow reactivity. However, for residues in coil-turn or β-sheet

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

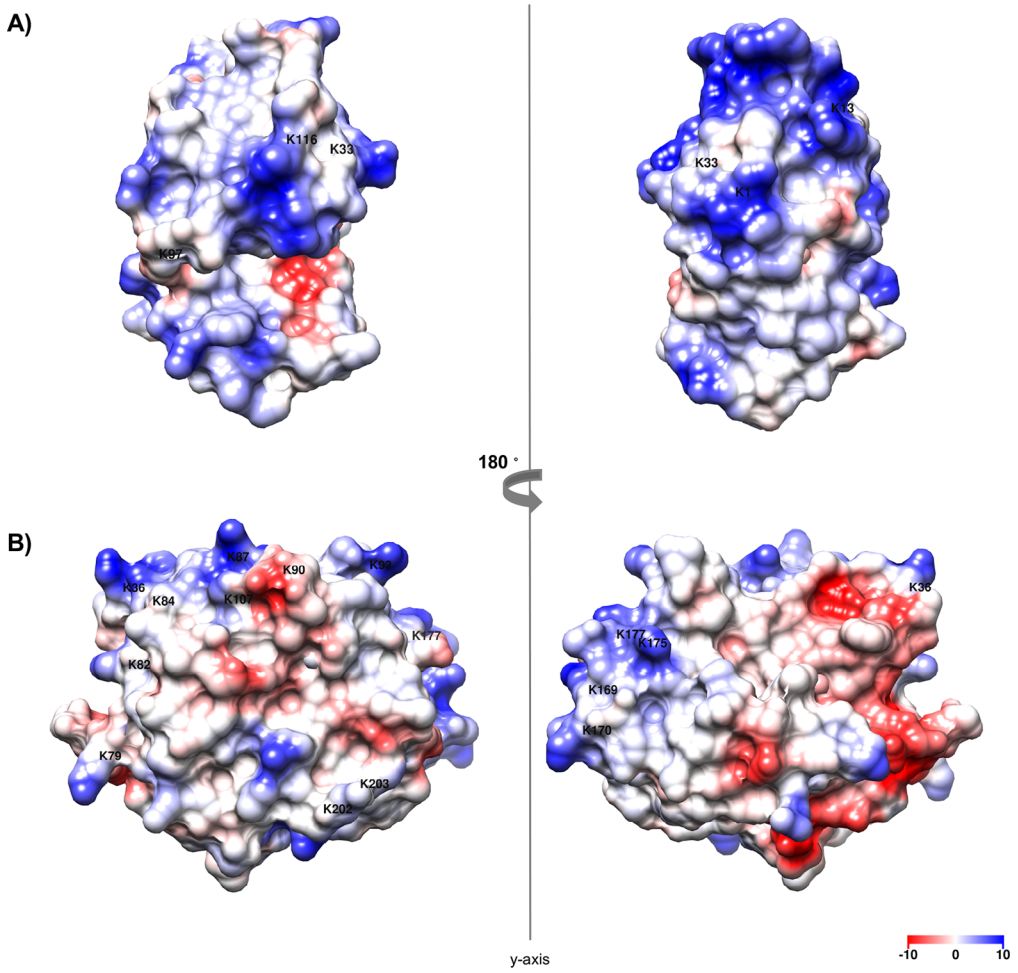


Figure 7. Electrostatic potential surfaces for (A) native lysozyme and (B) native chymotrypsin were calculated by Coulombic surface analysis using the USCF Chimera package. Electrostatic potentials were measured in kcal/(mol *e*), and negatively and positively charged regions are indicated in red and blue, respectively.

conformations, we observed a mixture of fast- and slow-reacting amino groups. For these, we next considered the local environment of the amino groups. Specifically, we analyzed the local charge surrounding the amino group in the protein and any noncovalent interactions. Lysozyme K116 was the fastest reacting amino group, followed by K97 and K33, with the slowest reacting group being K13. The proton on the ε-amino group in K116 participated as a proton donor with N106, which explained its increased relative nucleophilicity.

We next considered the reactivity of K33 since K13 showed a higher predicted exposed surface area, but K33 reacted faster. We hypothesized that the increased positive charge surrounding K13 likely led to reduced nucleophilicity (Figure 7A). Our finding of the higher reactivity of K116, K97, and K33 was not entirely unexpected. Previous research involving amine PEGylation of lysozyme had also reported a preferential modification at these three lysine residues.[24,65,66]

For CT, we observed that K36, which had a very high ESA (201.52 Å^2), reacted very rapidly. In contrast, the reactivity of K202 (116.84 Å^2) was unusually high, given that other lysines were more exposed (>200 Å^2). Local charge analysis of native chymotrypsin showed that in comparison to, for example, K79 (268.62 Å^2), K202 was in a less positively charged region of the protein. This could have explained the increased nucleophilicity of K202 when compared to that of other more exposed lysine residues (Figure 7B).

To further understand the importance of local charge, we compared K202 to K90, an amino group with similar ESA (116.8 and 113.2 Å^2, respectively) and pK_a values (10.27 and 10.13, respectively). Both residues were also found to be located in the β-sheet conformation in CT. While K202 reacted rapidly, rates determined by digestion studies showed that K90 reacted slower. Analysis of the local charge indicated that K90 was in a region of increased positive charge when compared to K202. This supported our assertion that local electrostatic and noncovalent interactions could determine amino group reactivity.

Clearly, all five factors that we assessed were found to impact the specificity of the initiator–protein reaction. These factors were then combined to generate a rules-based decision tree that could be used to predict the site and relative reactivity of the amino groups on a protein when modified with ATRP initiators (Figure 8). Importantly, between CT and Lyz we have 22 representative amino groups, each with a unique structure and environment, so we believe the data supports some useful conclusions. That said, these conclusions likely impact proteins of a similar size and shape as CT and Lyz, and further experiments will be needed to extend the generalizability of the structure–reactivity relationships that we measured. This

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086–2097

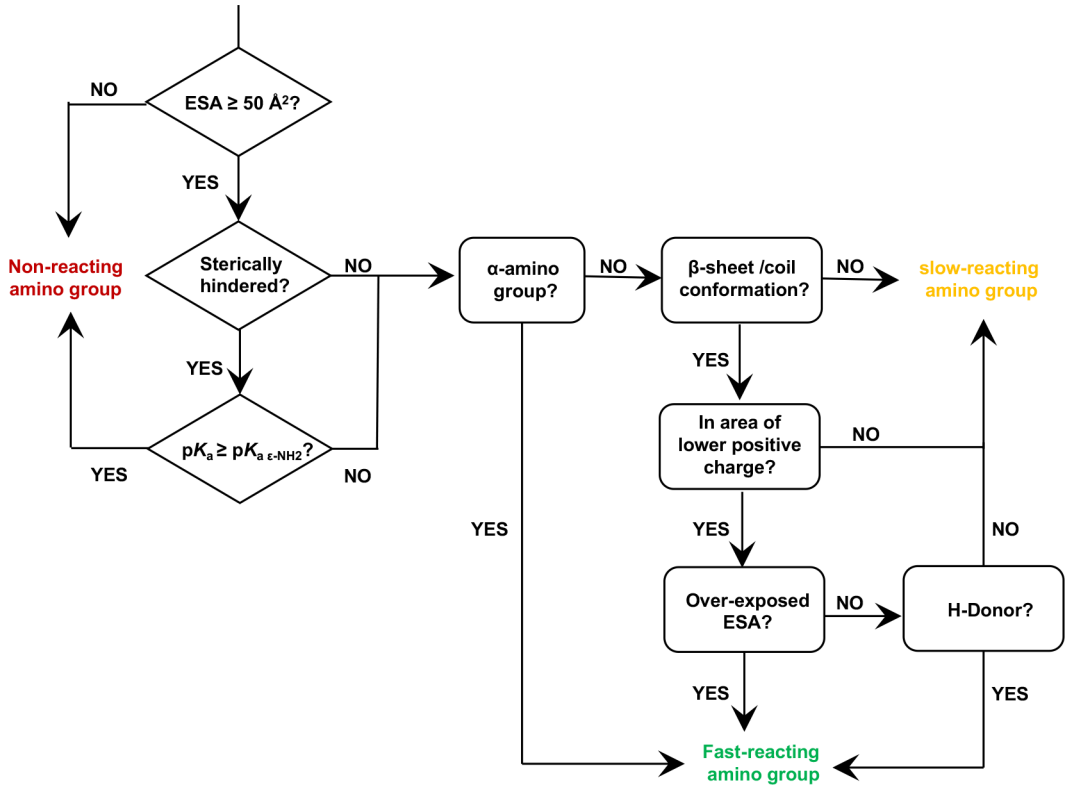


Figure 8. Rule-based flowchart to predict the rate of reaction of an amino group with ATRP initiator **1**.

decision tree was used for CT and Lyz to categorize amino groups as fast-, slow-, or nonreacting. We are now using the decision tree to predictably modify other proteins for polymer-based protein engineering.

CONCLUSIONS

Lysozyme and chymotrypsin were used as model proteins to determine initiator–protein specificity for protein modification with bromine functionalized ATRP initiator **1**. Protein–initiator complexes with varying extents of modification were prepared by changes in pH, time, and stoichiometry. Enzymatic and chemical digestion studies were used to identify the site and relative rates of modification at each amino site. A probe radius resembling ATRP initiator **1** was used for ESA calculations. Experimentally determined reactivities indicated that the exposed surface area was a major governing factor in initiator–protein reactions. Further understanding of these reactivities was achieved by a structural (secondary structure) and chemical (pK_a, charge and noncovalent interactions) analysis. The use of a probe size equivalent to the size of coupling reagent was important when identifying families of polymer-based protein engineering initiator reactivities and specificities. Exposed surface area calculations, accompanied by analysis of electrostatic and noncovalent intramolecular interactions, provided insights into the structure–reactivity relationships of amino groups.

The synthesis of the next generation of protein–polymer conjugates will depend on our ability to predict the effect of simple modification of the protein surface with controlled radical polymerization initiators. These data provide us with the synthetic tools to selectively modify and fine tune the grafting density and location of ATRP initiator molecules on the surface of proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsbiomaterials.7b00281.

Molecular dynamics simulation study for ATRP initiator; MALDI-TOF spectra for the characterization of protein-initiator conjugates; and data for the trypsin and NTCB digestion studies (PDF)

AUTHOR INFORMATION

Corresponding Author

*Fax: 412-268-5229. E-mail: alanrussell@cmu.edu.

ORCID

Krzysztof Matyjaszewski: 0000-0003-1960-3402
Alan J. Russell: 0000-0001-5101-4371

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the financial support provided by the Carnegie Mellon University Center for Polymer-based Protein Engineering.

ABBREVIATIONS

PBPE, polymer-based protein engineering; ATRP, atom transfer radical polymerization; NHS, *N*-hydroxysuccinimide; MALDI-TOF, matrix assisted laser desorption ionization-time-of-flight

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086–2097

REFERENCES

(1) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (23), 9615−20.

(2) Yee, A. A.; Savchenko, A.; Ignachenko, A.; Lukin, J.; Xu, X.; Skarina, T.; Evdokimova, E.; Liu, C. S.; Semesi, A.; Guido, V.; Edwards, A. M.; Arrowsmith, C. H. NMR and X-ray Crystallography, Complementary Tools in Structural Proteomics of Small Proteins. *J. Am. Chem. Soc.* **2005**, *127*, 16512−16517.

(3) Bock, J. R.; Gough, D. A. Predicting Protein-Protein Interactions from Primary Structure. *Bioinformatics* **2001**, *17* (5), 455−460.

(4) Shen, J. W.; Zhang, J.; Luo, X. M.; Zhu, W. L.; Yu, K. Q.; Chen, K. X.; Li, Y. X.; Jiang, H. L. Predicting Protein-Protein Interactions Based only on Sequences Information. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (11), 4337−4341.

(5) Campbell, A. S.; Murata, H.; Carmali, S.; Matyjaszewski, K.; Islam, M. F.; Russell, A. J. Polymer-based protein engineering grown ferrocene-containing redox polymers improve current generation in an enzymatic biofuel cell. *Biosens. Bioelectron.* **2016**, *86*, 446−453.

(6) Cummings, C.; Murata, H.; Koepsel, R.; Russell, A. J. Tailoring enzyme activity and stability using polymer-based protein engineering. *Biomaterials* **2013**, *34* (30), 7437−43.

(7) Cummings, C.; Murata, H.; Koepsel, R.; Russell, A. J. Dramatically Increased pH and Temperature Stability of Chymotrypsin using Dual Block Polymer-Based Protein Engineering. *Biomacromolecules* **2014**, *15* (3), 763−71.

(8) Cummings, C. S.; Murata, H.; Matyjaszewski, K.; Russell, A. J. Polymer-Based Protein Engineering Enables Molecular Dissolution of Chymotrypsin in Acetonitrile. *ACS Macro Lett.* **2016**, *5* (4), 493−497.

(9) Gao, W.; Liu, W.; Mackay, J. A.; Zalutsky, M. R.; Toone, E. J.; Chilkoti, A. In situ growth of a stoichiometric PEG-like conjugate at a protein's N-terminus with significantly improved pharmacokinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (36), 15231−6.

(10) Heredia, K. L.; Bontempo, D.; Ly, T.; Byers, J. T.; Halstenberg, S.; Maynard, H. D. In Situ Preparation of Protein-"Smart" Polymer Conjugates with Retention of Bioactivity. *J. Am. Chem. Soc.* **2005**, *127*, 16955−16960.

(11) Lele, B. S.; Murata, H.; Matyjaszewski, K.; Russell, A. J. Synthesis of Uniform Protein-Polymer Conjugates. *Biomacromolecules* **2005**, *6*, 3380−3387.

(12) Qi, Y.; Amiram, M.; Gao, W.; McCafferty, D. G.; Chilkoti, A. Sortase-catalyzed initiator attachment enables high yield growth of a stealth polymer from the C terminus of a protein. *Macromol. Rapid Commun.* **2013**, *34* (15), 1256−60.

(13) Nicolas, J.; San Miguel, V.; Mantovani, G.; Haddleton, D. M. Fluorescently tagged polymer bioconjugates from protein derived macroinitiators. *Chem. Commun.* **2006**, *45*, 4697−9.

(14) Averick, S.; Mehl, R. A.; Das, S. R.; Matyjaszewski, K. Well-defined biohybrids using reversible-deactivation radical polymerization procedures. *J. Controlled Release* **2015**, *205*, 45−57.

(15) Cobo, I.; Li, M.; Sumerlin, B. S.; Perrier, S. Smart hybrid materials by conjugation of responsive polymers to biomacromolecules. *Nat. Mater.* **2015**, *14* (2), 143−59.

(16) 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.; Price, J. L. Criteria for selecting PEGylation sites on proteins for higher thermodynamic and proteolytic stability. *J. Am. Chem. Soc.* **2014**, *136* (50), 17547−60.

(17) Berberich, J. A.; Yang, L. W.; Madura, J.; Bahar, I.; Russell, A. J. A stable three-enzyme creatinine biosensor. 1. Impact of structure, function and environment on PEGylated and immobilized sarcosine oxidase. *Acta Biomater.* **2005**, *1* (2), 173−81.

(18) Nordwald, E. M.; Kaar, J. L. Stabilization of Enzymes in Ionic Liquids via Modification of Enzyme Charge. *Biotechnol. Bioeng.* **2013**, *110* (9), 2352−2360.

(19) Toews, J.; Rogalski, J. C.; Kast, J. Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications. *Anal. Chim. Acta* **2010**, *676* (1−2), 60−7.

(20) Watkins, N. G.; Thorpe, S. R.; Baynes, J. W. Glycation of Amino Groups in Protein. *J. Biol. Chem.* **1985**, *260* (19), 10629−10636.

(21) Chen, X.; Muthoosamy, K.; Pfisterer, A.; Neumann, B.; Weil, T. Site-selective lysine modification of native proteins and peptides via kinetically controlled labeling. *Bioconjugate Chem.* **2012**, *23* (3), 500−8.

(22) Andersson, L. K.; Caspersson, M.; Baltzer, L. Control of Lysine Reactivity in Four-Helix Bundle Proteins by Site-Selective pKa Depression: Expanding the Versatiliy of Proteins by Postsynthetic Functionalisation. *Chem. - Eur. J.* **2002**, *8* (16), 3687−3697.

(23) Lins, L.; Thomas, A.; Brasseur, R. Analysis of Accessible Surface of Residues in Proteins. *Protein Sci.* **2003**, *12*, 1406−1417.

(24) Suckau, D.; Mak, M.; Przybylski, M. Protein Surface Topology-Probing by Selective Chemical Modification and Mass Spectrometric Peptide Mapping. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5630−5634.

(25) Hnízda, A.; Šantrůček, J.; Šanda, M.; Strohalm, M.; Kodíček, M. Reactivity of histidine and lysine side-chains with diethylpyrocarbonate — A method to identify surface exposed residues in proteins. *J. Biochem. Biophys. Methods* **2008**, *70* (6), 1091−1097.

(26) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.

(27) Heffernan, R.; Paliwal, K.; Lyons, J.; Dehzangi, A.; Sharma, A.; Wang, J.; Sattar, A.; Yang, Y.; Zhou, Y. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. *Sci. Rep.* **2015**, *5*, 11476.

(28) Ahmad, S.; Gromiha, M. M.; Sarai, A. Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information. *Bioinformatics* **2004**, *20* (4), 477−86.

(29) Glantz-Gashai, Y.; Meirson, T.; Samson, A. O. Normal Modes Expose Active Sites in Enzymes. *PLoS Comput. Biol.* **2016**, *12* (12), e1005293.

(30) Marsh, J. A.; Teichmann; Sarah, A. Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding. *Structure (Oxford, U. K.)* **2011**, *19* (6), 859−867.

(31) Novotny, J.; Handschumacher, M.; Haber, E.; Bruccoleri, R. E.; Carlson, W. B.; Fanning, D. W.; Smitht, J. A.; Rose, G. D. Antigenic Determinants in Proteins Coincide with Surface Regions Accessible to Large Probes (Antibody Domains). *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 226−230.

(32) Rubinstein, N. D.; Mayrose, I.; Pupko, T. A machine-learning approach for predicting B-cell epitopes. *Mol. Immunol.* **2009**, *46* (5), 840−7.

(33) Kawabata, T.; Go, N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins: Struct., Funct., Genet.* **2007**, *68* (2), 516−29.

(34) Larda, S. T.; Bokoch, M. P.; Evanics, F.; Prosser, R. S. Lysine methylation strategies for characterizing protein conformations by NMR. *J. Biomol. NMR* **2012**, *54* (2), 199−209.

(35) Larda, S. T.; Pichugin, D.; Prosser, R. S. Site-Specific Labeling of Protein Lysine Residues and N-Terminal Amino Groups with Indoles and Indole-Derivatives. *Bioconjugate Chem.* **2015**, *26* (12), 2376−83.

(36) Zhang, M.; Vogel, H. J. Determination of the Side Chain pKa Values of the Lysine Residues in Calmodulin. *J. Biol. Chem.* **1993**, *268* (30), 22420−22428.

(37) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* **2009**, *18* (1), 247−251.

(38) Kinstler, O.; Molineux, G.; Treuheit, M.; Ladd, D.; Gegg, C. Mono-N-terminal poly(ethylene glycol)−protein conjugates. *Adv. Drug Delivery Rev.* **2002**, *54*, 477−485.

(39) Hu, J.; Sebald, W. N-terminal specificity of PEGylation of human bone morphogenetic protein-2 at acidic pH. *Int. J. Pharm.* **2011**, *413* (1−2), 140−6.

(40) Chan, A. O.; Ho, C. M.; Chong, H. C.; Leung, Y. C.; Huang, J. S.; Wong, M. K.; Che, C. M. Modification of N-terminal alpha-amino

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097

groups of peptides and proteins using ketenes. *J. Am. Chem. Soc.* **2012**, *134* (5), 2589−98.
(41) Singudas, R.; Adusumalli, S. R.; Joshi, P. N.; Rai, V. A phthalimidation protocol that follows protein defined parameters. *Chem. Commun.* **2015**, *51*, 473−476.
(42) Baker, D. P.; Lin, E. Y.; Lin, K.; Pellegrini, M.; Petter, R. C.; Chen, L. L.; Arduini, R. M.; Brickelmaier, M.; Wen, D.; Hess, D. M.; Chen, L.; Grant, D.; Whitty, A.; Gill, A.; Lindner, D. J.; Pepinsky, R. B. N-Terminally PEGylated Human Interferon-â-1a with Improved Pharmacokinetic Properties and in Vivo Efficacy in a Melanoma Angiogenesis Model. *Bioconjugate Chem.* **2006**, *17*, 179−188.
(43) Blow, D. M. Structure and Mechanism of Chymotrypsin. *Acc. Chem. Res.* **1976**, *9* (4), 145−152.
(44) Johnson, L. N. The Structure and Function of Lysozyme. *Sci. Prog.* **1966**, *54* (215), 367−385.
(45) Gauthier, M. A.; Klok, H.-A. Polymer−Protein Conjugates: An Enzymatic A Activity Perspective. *Polym. Chem.* **2010**, *1* (9), 1352.
(46) Li, H.; Bapat, A. P.; Li, M.; Sumerlin, B. S. Protein Conjugation of Thermoresponsive Amine-Reactive Polymers Prepared by RAFT. *Polym. Chem.* **2011**, *2* (2), 323−327.
(47) Lucius, M.; Falatach, R.; McGlone, C.; Makaroff, K.; Danielson, A.; Williams, C.; Nix, J. C.; Konkolewicz, D.; Page, R. C.; Berberich, J. A. Investigating the Impact of Polymer Functional Groups on the Stability and Activity of Lysozyme-Polymer Conjugates. *Biomacromolecules* **2016**, *17* (3), 1123−34.
(48) Diamond, R. Real-space refinement of the structure of hen egg-white lysozyme. *J. Mol. Biol.* **1974**, *82* (3), 371−391.
(49) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization and Analysis Platform. *J. Cheminf.* **2012**, *4* (1), 17.
(50) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Genet.* **2005**, *61* (4), 704−21.
(51) Tang, H. Y.; Speicher, D. W. Identification of alternative products and optimization of 2-nitro-5-thiocyanatobenzoic acid cyanylation and cleavage at cysteine residues. *Anal. Biochem.* **2004**, *334* (1), 48−61.
(52) Lim, C. Y.; Owens, N. A.; Wampler, R. D.; Ying, Y.; Granger, J. H.; Porter, M. D.; Takahashi, M.; Shimazu, K. Succinimidyl Ester Surface Chemistry: Implications of the Competition between Aminolysis and Hydrolysis on Covalent Protein Immobilization. *Langmuir* **2014**, *30* (43), 12868−12878.
(53) Canfield, R. E. The Amino Acid Sequence of Egg White Lysozyme. *J. Biol. Chem.* **1963**, *288* (8), 2698−2707.
(54) Browne, W. J.; North, A. C. T.; Phillips, D. C.; Brew, K.; Vanaman, T. C.; Hill, R. L. A Possible Three-Dimensional Structure of Bovine α-Lactalbumin based on that of Hen's Egg-White Lysozyme. *J. Mol. Biol.* **1969**, *42*, 65−86.
(55) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. *J. Mol. Biol.* **1973**, *79*, 351−371.
(56) Spassov, V. Z.; Karshikov, A. D.; Atanasov, B. P. Electrostatic Interactions in Proteins. A Theoretical Analysis of Lysozyme Ionization. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1989**, *999*, 1−6.
(57) Jacob, E.; Unger, R. A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics* **2007**, *23* (2), e225−30.
(58) Kajander, T.; Kahn, P. C.; Passila, S. H.; Cohen, D. C.; Lehtiö, L.; Adolfsen, W.; Warwicker, J.; Schell, U.; Goldman, A. Buried Charged Surface in Proteins. *Structure* **2000**, *8* (11), 1203−1214.
(59) McGovern, R. E.; Snarr, B. D.; Lyons, J.; McFarlane, J.; Whiting, A. L.; Paci, I.; Hof, F.; Crowley, P. Structural study of a small molecule receptor bound to dimethyllysine in lysozyme. *Chemical science* **2015**, *6*, 442−449.
(60) The Chemistry of Reactive Groups. In *Bioconjugate Techniques*; Hermanson, G. T., Ed. Academic Press: San Diego, CA, 1996; pp 137−146.
(61) Gallina, A. M.; Bork, P.; Bordo, D. Structural analysis of protein-ligand interactions: the binding of endogenous compounds and of synthetic drugs. *J. Mol. Recognit.* **2014**, *27* (2), 65−72.
(62) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins: Struct., Funct., Genet.* **2006**, *65* (1), 15−26.
(63) Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3* (6), 608−14.
(64) Yuan, Z.; Bailey, T. L.; Teasdale, R. D. Prediction of protein B-factor profiles. *Proteins: Struct., Funct., Genet.* **2005**, *58* (4), 905−12.
(65) da Silva Freitas, D.; Abrahão-Neto, J. Biochemical and biophysical characterization of lysozyme modified by PEGylation. *Int. J. Pharm.* **2010**, *392* (1−2), 111−117.
(66) Lee, H.; Park, T. G. A Novel Method for Identifying PEGylation Sites of Protein Using Biotinylated PEG Derivatives. *J. Pharm. Sci.* **2003**, *92* (1), 97−103.

DOI: 10.1021/acsbiomaterials.7b00281
ACS Biomater. Sci. Eng. 2017, 3, 2086−2097