Determining protein structure by tyrosine bioconjugation

Mahta Moinpour[1], Natalie K. Barker[2], Lindsay E. Guzman[1], John C. Jewett[1], Paul R. Langlais[2], Jacob C. Schwartz[1]*

[1] Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA

[2] Department of Medicine, Division of Endocrinology, University of Arizona College of Medicine, Tucson, AZ 85721, USA

* Corresponding author: jcschwartz@email.arizona.edu

KEYWORDS: tyrosine, triazolinediones, protein folding, low complexity, conjugation

ABSTRACT

Exploration of protein structure by its solvent accessible surfaces has been widely exploited in structural biology. Amino acids most commonly targeted for covalent modification of the native folded protein are lysine and cysteine. Here we leveraged an ene-type chemistry targeting tyrosine residues to discriminate those solvent exposed from those buried. We find that 4-phenyl-3H-1,2,4-triazole-3,5(4H)-dione (PTAD) can conjugate the phenolic group of tyrosine in a manner heavily influenced by the orientation of the residue with respect to the protein surface. We developed a strategy to investigate protein structure by analyzing PTAD conjugations with free tyrosine, peptides, and proteins. We found this conjugation-based approach robust, sensitive to shifts in protein structure, and adaptable to a wide range of analytic technologies, including fluorescence, chromatography, or mass spectrometry. These studies show how established tyrosine-specific bioconjugation chemistry can expand the toolkit for applications in structural biology.

INTRODUCTION

Protein structure is driven by the interactions of the 20 amino acids with solvent and other amino acids[1]. This basic property has been exploited for decades in the developing methods that can discriminate protein structures and conformational states. Experimental approaches that analyze or manipulate solvent accessibility in proteins employ recombinant protein engineering, solvent exchanges, substrate interaction kinetics, and enzyme modifications[2]. Covalent modifications of proteins have also been used to distinguish solvent exposed residues from buried. Strong covalent bonds can offer the advantage of preserving a footprint of the protein's native structure during analyses with technologies that may require denaturing or degrading the protein. The most commonly employed chemistries that offer site-selectivity are targeting lysine and cysteine residues[3]. Lysine residues are charged and most often found on protein surfaces. Lysine is also abundant, making up an average of 5.9% of protein amino acid sequences. Cysteine comprises 1.9% of protein amino acid content, readily forms covalent disulfide bonds, and rarely found on protein surfaces.

By contrast, tyrosine comprises 3.2% of proteins, making it a balanced and sufficiently abundant target in native protein structures to potentially provide structural information without a need for recombinant engineering[4]. The amphipathic nature of its phenolic ring places tyrosine near the boundary to be categorized as hydrophobic or polar. Consequently, tyrosine residues are well-distributed in proteins between surfaces or buried in the hydrophobic core[4b, 5]. They are also enriched and frequently provide the strongest interactions at protein interfaces, also called hot spots, that bind small molecules, nucleic acids, or protein partners[5-6].

Options to target tyrosine for chemical conjugation have recently expanded through the development of strategies that employ electrophiles, such as aryl diazonium ions, Mannich reactions, and triazoledione compounds[7]. The Barbas group first established the usefulness and selectivity of an ene-like reaction between tyrosine and 4-phenyl-3H-1,2,4-triazole-3,5(4H)-dione (PTAD)[8]. This reaction was described as fast and selective, compatible with buffers suitable for proteins, and could proceed with a functionalized PTAD to be targeted by click chemistry[8a, 9]. More recent reports have exploited PTADs to conjugate nucleic acids[9d], fluorophores[8-9], glycans[9c], and to crosslink protein hydrogels[9b]. Studies to date have largely focused on pursuing single-site specificity[9a]. Leveraging PTAD to map the relative tyrosine exposure on the surfaces of a protein has not been pursued.

If made a practical target for structure-based investigations, tyrosine residues open a range of new possibilities for focused studies. Among these are studies of low complexity proteins, which recently have received greatly increased attention[10]. Certain low complexity domains have the ability to drive formation of non-membrane bound cellular organelles, also known as granular bodies, through a process referred to as phase separation[11]. Among the most studied proteins with phase separation properties are those with tyrosine-rich domains of repeating GYG or SYS motifs and are also mostly devoid of lysine or cysteine[11e, 12]. Tyrosine-rich, low complexity proteins are typically intrinsically disordered or lacking in rigid secondary structure elements, but are often implicated in disorder-to-order transitions through protein-protein binding[13]. Protein disorder renders the most popular NMR and X-ray crystallography methods incapable of providing high-resolution structure data and elevates the potential for a new method of structure analysis to reveal a wealth of otherwise unattainable information.

Here we characterize the products of PTAD conjugations to proteins and peptides. Approaches that modify PTAD reactivity toward proteins were evaluated, as was the utility of PTAD to distinguish relative solvent exposure of tyrosine residues across a protein surface. We reasoned that PTAD reactivity might discriminate the simple duality between buried residues and those at the surface. Our results are indicative of a more complex relationship between tyrosine conjugation and local structure, which may increase the potential for this approach to offer useful insights when more traditional structural biology techniques are limited.

METHODS

Materials. PTAD conjugation reagents were purchased commercially from Sigma-Aldrich (Germany) and used without further purification: 4-Phenyl-3H-1,2,4-triazole-3,5(4H)-dione (PTAD, cat. # 42579), 4-(4-(2-Azidoethoxy)phenyl)-1,2,4-triazolidine-3,5-dione, N3-Ph-Ur for e-Y-CLICK (PTAD-N_3, cat. # T511552), 1,3-Dibromo-5,5-dimethylhydantoin (DBH, cat. # 157902), DBCO-Cy3 (cat. # 77366), DBCO-Cy5 (cat. # 777374), and tyrosine (cat. # T3754). Tris-HCl was purchased from Goldbio (cat. # T-400-5). Urea was purchased from Invitrogen (Carlsbad, CA). Peptides and proteins were commercially available from Sigma-Aldrich: angiotensin II (cat. # A9525), peptide mixture (cat. # H2016), and myoglobin (cat. # M1822). Bovine serum albumin was purchased from VWR (cat. # 97062-508).

Tyrosine and peptide conjugation. Tyrosine stocks were dissolved in 1 M HCl and diluted into 1:1 water and acetonitrile for conjugation reactions. The PTAD-N_3 pre-cursor was oxidized by briefly incubating with equimolar DBH until a color change to cranberry red was observed. These

were combined, vortexed, and incubated at room temperature for 1 hour. For analysis by UPLC-MS, 20 μL of samples were injected to a LCMS-2020 (Shimadzu) with C18 column (Onyx monolithic C18, 50 x 2.0 mm). Samples were eluted over 4 minutes over a linear gradient of acetonitrile from 5% to 20% over the first 3 minutes. For peptide conjugations, angiotensin II (0.5 mM) or peptide mixtures (0.5 mg/mL) were incubated with 5 mM PTAD for 1 hour at room temperature. Peptides were analyzed by UPLC-MS using a Bruker AmaZon SL Ion Trap mass spectrometer (*Bruker* Daltonik GmbH, Germany) in-line with HPLC and ESI source with positive polarity. High resolution analysis of angiotensin II was performed using an LTQ Orbitrap Velos ETD mass-spectrometer (ThermoFisher Scientific, Bremen, Germany).

Protein conjugation. Myoglobin was incubated at 10 μM concentration with 6.6 mM PTAD for 1 hour at room temperature. MALDI-TOF analysis was performed as described above except using a matrix of saturated sinapic acid (Fluka, cat. # 85429). Conjugation of bovine serum albumin for SEC was incubated with PTAD or PTAD-N_3 for 1 hour at room temperature in buffer A (150 mM NaCl, 40 mM Tris-HCl). For DBCO-dye conjugations, PTAD-N_3 and dyes were incubated together with proteins at a 10:1 molar ratio, respectively, for 1 hour at room temperature. Fluorescence was imaged after SDS-PAGE using a Chemidoc MP system (Biorad).

RESULTS

Products made by PTAD reaction to tyrosine. PTAD can react with the phenolic ring of tyrosine. However, previous reports have noted additional products for PTAD reactions with amines, second additions to tyrosine phenols, and a short-lived conjugations to cysteine residues[9a,

[14]. Additional products to PTAD chemistry do not necessarily detract from our goal of measuring changes in solvent accessibility for tyrosine residues, these must be accounted for to interpret data produced by several methods. We confirmed PTAD reactivity by analyzing the product of 3-(4-hydroxyphenyl)propionic acid incubated with PTAD for 15 minutes (**Figure 1A**). To fully react the tyrosine mimic, we added PTAD to the reaction four times and 15 minutes each. Spectra were consistent with previous reports for products of PTAD and the phenolic group of tyrosine or this mimic[9a, 9d].

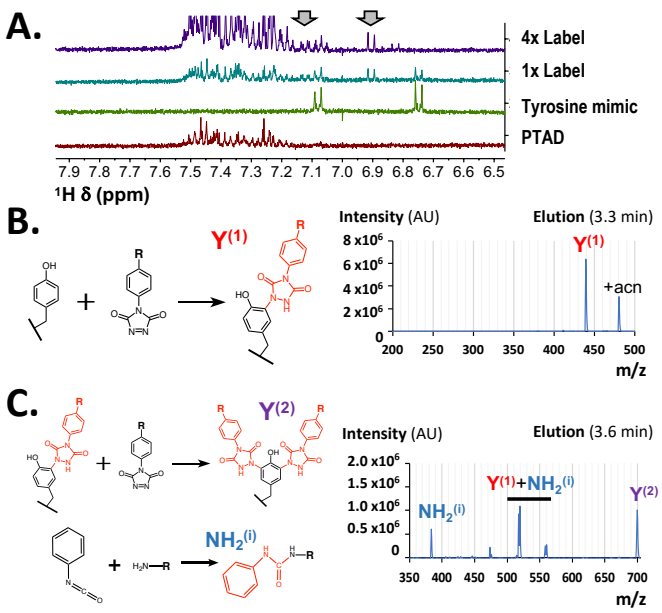


Figure 1. Products of PTAD reaction with tyrosine. **(A)** 1D ^{1}H NMR spectra for PTAD, a tyrosine mimic, a single incubation of the mimic with PTAD for 15 minutes (1x Label), and 4 repeated additions of PTAD to incubate with the mimic (4x Label). Peaks for the products formed are indicated by arrows. **(B)** The conjugation product of 1:1 PTAD to tyrosine at a single *ortho*-position, $Y^{(1)}$, on the phenolic ring was detected by UPLC-MS with the expected m/z of 440 Daltons. **(C)** Additional products are shown by UPLC-MS for PTAD and tyrosine: PTAD conjugated at both *ortho*-positions of the phenolic ring, $Y^{(2)}$, and an isocyanate degradation product

of PTAD conjugated to the primary amine, $NH^{(urea)}$. Also seen in this elution are the +1 source fragments (*) for tyrosine conjugated with one *ortho*-PTAD and one isocyanate, $Y^{(2)}$ + $NH^{(urea)}$. These products were observed unfragmented in the +3 charge state at an earlier elution (see **Table 1**). AU indicates arbitrary units for MS intensities.

We used ultra-performance liquid-chromatography mass spectrometry (UPLC-MS) to detect the products of PTAD reactions with titrating concentrations of tyrosine. For this reaction, reduced PTAD-azide (red·PTAD-N_3) was activated using the oxidant 1,3-dibromo-5,5-methylhydantoin (DBH). A 1:1 molar ratio of PTAD to tyrosine yielded the most pronounced signal of PTAD reacted with the phenolic ring of tyrosine **1** (**Figure 1B**). As the molar ratio for PTAD to tyrosine was increased to 5:1, 10:1, and 50:1, the signals for a single phenolic reaction was diminished until undetectable above noise, as the product pool became dominated with compounds of multiple and sequential conjugations (**Table 1**).

				PTAD-azide / tyrosine			
	charge	**Expect m/z**	**Observed m/z**	**1:1**	**5:1**	**10:1**	**50:1**
PTAD-azide	+1	262	262 ±0.5	2.25 mM	2.25 mM	2.25 mM	2.25 mM
Tyrosine	+1	182	182 ±0.5	2.25 mM	0.45 mM	0.25 mM	0.045 mM
$Y^{(1)}$	+1	440	440 ±1	++	+	+	ND
$Y^{(2)}$	+3[a]	234	236 ±1	++	+	ND	ND
$NH^{(urea)}$	+1	384	384 ±1	+	++	++	++
$Y^{(1)}$+$NH^{(urea)}$	+3[a]	214	215 ±1	+	++	(++)*	(+)*
$Y^{(2)}$+$NH^{(urea)}$	+3	302	304 ±1	+	++	(++)*	(+)*

ND – not detected
*[a] 1+ ion also observed; * Source fragments detected*

Table 1. Products observed by UPLC-MS for PTAD:tyrosine titrations. High or low product signals are indicated by "++" or "+", respectively.

A double-reacted product of two PTADs on the phenolic ring, $Y^{(2)}$, was observed at the 1:1 reaction (**Figure 1C**). The third product was found at the expected mass for a PTAD decomposed to an isocyanate, which can react with the primary amine of the tyrosine, $NH^{(urea)}$ [9a, 15]. The product of the isocyanate reaction with tyrosine was smaller than the sum of PTAD and tyrosine (**Figure 1C**). At 10 or 50-fold molar excess of PTAD over tyrosine, the diminished signals for simple $Y^{(1)}$ and $Y^{(2)}$ products were replaced by signals for high molecular weight products, such as those with the amine also conjugated, $NH^{(urea)}$ (**Figure 1C**, **Table 1**). Changes to relative product abundances was considered to indicate that an excess in PTAD could drive tyrosine to undergo multiple reactions until the largest product of combined phenol and amine conjugations, $Y^{(2)}$+$NH^{(urea)}$, was reached (**Table 1**).

PTAD reactivity with peptides. We proceeded to assess the products of reacting PTAD with the peptide, angiotensin II. Angiotensin II is an 8 amino acid long peptide, NH_2-DRVYIHPF-COOH, that contains the N-terminal amino group, a single tyrosine, and two other ring sidechains, the imidazole group of histidine and the phenyl group of phenylalanine (**Figure 2A**). For this reaction, we used PTAD and did not require the addition of the oxidant, DBH.

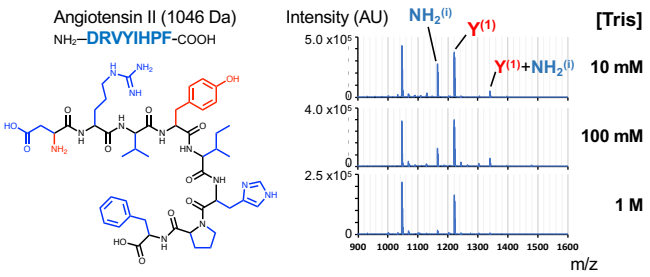


Figure 2. PTAD labeling of angiotensin II. **(A)** Structure for the peptide angiotensin II with the tyrosine and N-terminal amine that can be conjugated by PTAD (MW=175) colored red. **(B)**

Titrating amount of Tris from 10 mM to 1 M reduces or eliminates evidence of isocyanate reaction with amines, $NH^{(urea)}$, seen by MALDI-TOF. AU indicates arbitrary units for MS intensities.

Reports have suggested that the non-specific reaction of the isocyanate decomposition product with primary amines can be scavenged away with 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris) buffer[8a, 9a]. We anticipated that control of this secondary reaction could prove important as modifications to the primary amine of lysine sidechains might prevent trypsin digestion that would allow LC-MS/MS analysis of labeled proteins. Using matrix-assisted laser desorption/ionization and time of flight (MALDI-TOF) mass spectrometry, both the amine and phenolic reacted products were observed from the reaction in 10 mM Tris (pH 7.4) (**Figure 2B**). As the concentration of Tris was increased to 100 mM and 1 M, the non-specific amine conjugate was reduced or undetectable. Nevertheless, at higher Tris concentrations, production of the amine conjugate was dependent on the stoichiometry of PTAD and angiotensin II. In 200 mM Tris, the reaction of PTAD with the tyrosine sidechain was selectively produced until the PTAD concentration reached 50 to 100-fold excess over angiotensin II (**Supplemental Figure 1A**).

Next, we tested the selectivity of PTAD reactivity for a mixture of five peptides whose lengths varied from 2 to 8 amino acids and ranging in mass from 238 to 1046 Da. Each peptide varied the tyrosine position to be at the N-terminus, C-terminus, or internal. Most of the mass intensity for each peptide was shifted to indicate a single conjugated PTAD to tyrosine, $Y^{(1)}$, as seen by UPLC-MS (**Supplemental Figure 1B-D**) or MALDI-TOF (**Supplemental Figure 1E**). Intensities for the isocyanate side reaction with the amine, $NH^{(urea)}$, were relatively small or non-existent, as were the intensities for the double addition of PTAD, $Y^{(2)}$. This panel of peptides revealed no evidence of

overt sequence dependence to PTAD reactivity and reiterated that the isocyanate reaction and double labeling of tyrosine could be mitigated through stoichiometry and an amine scavenger.

Specificity of PTAD labeling of folded or unfolded myoglobin. We next explored whether PTAD could selectively label some or all available tyrosine residues for a small, well-folded protein, myoglobin. We chose cardiac myoglobin from *Equus caballus*. Inspection of the amino acid sequence and crystal structure (PDB: 4NS2) revealed the protein to contain only two tyrosine residues. One tyrosine was well solvent exposed (24% of surface area exposed) and the other buried (1% of surface area exposed). The side chain of lysine has a primary amine and myoglobin contains 20 lysine residues, ranging from 10% to 90% solvent exposed (**Figure 3A**).

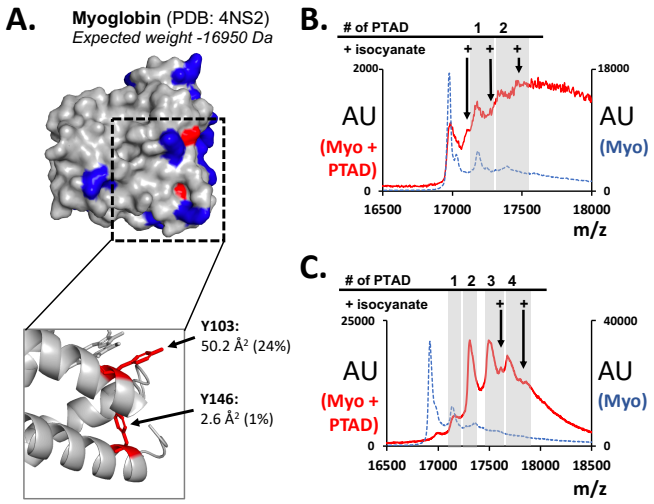


Figure 3. Labeling of folded or unfolded myoglobin. (A) Structure of myoglobin showing the solvent accessible protein surface. Highlighted are the accessible surfaces for myoglobin's two tyrosine (red) and 20 lysine (blue) residues. Inset shows the position of Y103 and Y146 sidechains and their exposed surface area calculated by PyMOL. MALDI-TOF analysis was performed for myoglobin (blue) and myoglobin labeled with PTAD (red). Shaded regions are the expected

masses for myoglobin plus one or more PTAD conjugates and arrows indicate the expected masses for an additional isocyanate reaction. PTAD labeling of folded myoglobin **(B)** yielded up to 2 PTAD additions, with a possible addition of up to one isocyanate to amine conjugation where indicated by "+" symbols. PTAD labeling of myoglobin unfolded with HCl **(C)** yielded up to four PTAD additions, with one addition of one amine conjugation where indicated by "+" symbols. AU indicates arbitrary units for MS intensities.

Conjugation of myoglobin with PTAD under native conditions with 500 mM Tris (pH 7.4) was found to shift the mass (m/z) in MALDI-TOF (**Figure 3B**). Distinct maxima could be discerned at the expected masses for one or two PTAD additions. Maxima in the profile were also observed at the expected masses for the addition of a single reacted amine product or that in combination with a PTAD addition. No maxima were distinguishable at the masses expected for two or more amine products or in combination with PTAD additions. These data suggested that despite myoglobin possessing more than 20 solvent exposed lysine residues, evidence of no more than one amine conjugation could be discerned.

We found that each tyrosine could be doubly conjugated to PTAD, $Y^{(2)}$, when testing PTAD conjugation under acid-denaturing conditions (0.1% HCl). Both tyrosine residues were fully conjugated before the reaction could be quenched by acid or an addition of equimolar free tyrosine. Analysis by MALDI-TOF revealed almost full conversion of myoglobin to have up to four conjugated PTAD molecules (**Figure 3C**). Labeling myoglobin with PTAD-N_3 conjugated to DBCO-Cy3 could also yield up to four Cy3 labels added to myoglobin (**Supplemental Figure 2**). In every myoglobin conjugation, evidence of no more than a single amine conjugate could be

found. The observation of myoglobin with four PTAD additions indicated that doubly labeled myoglobin observed in MALDI-TOF (**Figure 3B**) could result from two PTAD additions to the most solvent exposed tyrosine, Y103.

PTAD labeling preserves protein folding. In order to determine whether tyrosine conjugation could serve as a reliable indicator of protein folding, we inquired whether the PTAD reactions might disrupt or destroy protein structure. We chose bovine serum albumin (BSA) for a model of a well-folded protein. BSA is a 607 amino acid protein (66 kDa) with 21 tyrosine residues that range from 1 to 41% solvent exposed as calculated for its solved structure (PDB: 3V03). BSA is also a highly soluble protein with 60 lysine residues, ranging from 5 to 91% solvent exposed.

We used size exclusion chromatography (SEC) to observe BSA unfolding under titrating concentrations of urea[16]. The midpoint urea concentration to unfold BSA was determined by observing the loss of secondary structure through circular dichroism (CD) spectroscopy, which was determined to be 4.6 ± 0.1 M urea (**Figure 4A**). We observed BSA in its native state elute at the expected volume for a protein of its size, compared to molecular weight standards (**Supplemental Figure 3A**). For concentrations of 4, 6, and 8 M urea, the peak for BSA elution during SEC shifted to earlier volumes and broadened, consistent with a structure that is more extended as it unfolds (**Figure 4B**).

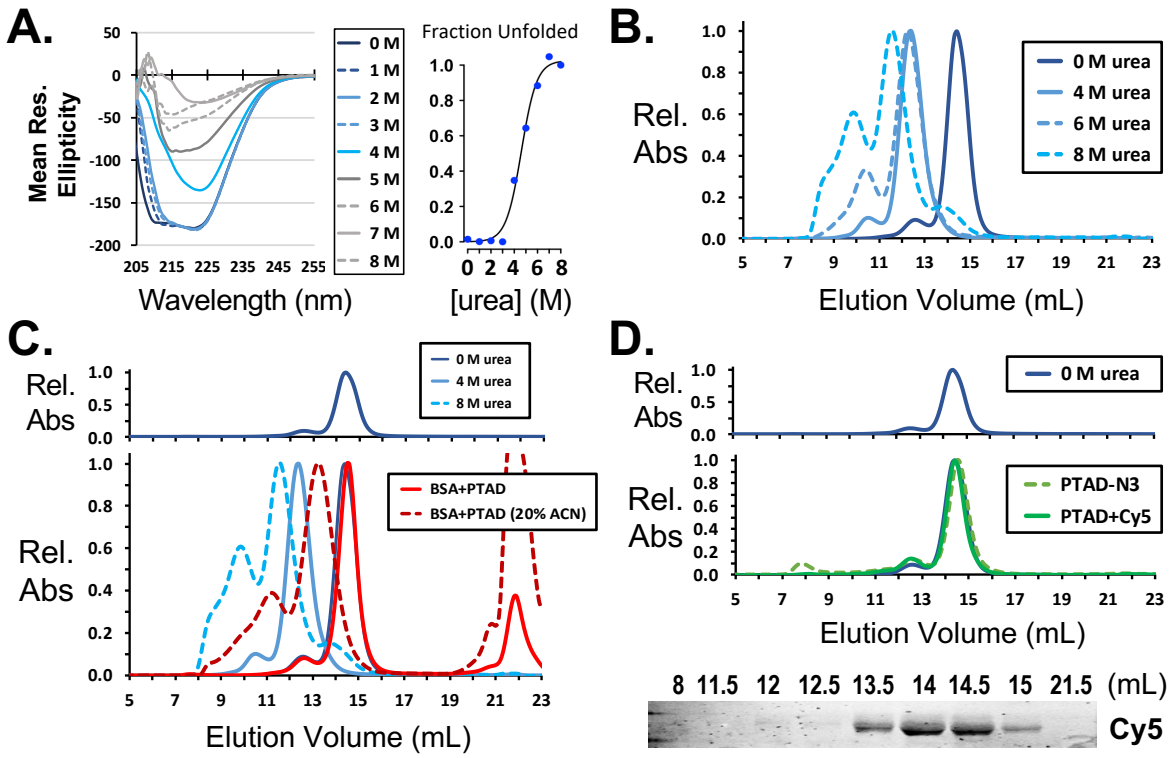


Figure 4. PTAD conjugation does not abolish protein structure. (**A**) The midpoint concentration for urea to unfold BSA was determined to be 4.6 ± 0.1 M urea according to CD spectroscopy. (**B**) Size exclusion chromatography of BSA, measured by UV absorbance and in titrating amounts of urea. BSA elutions were shifted by protein unfolding in urea. "Rel. Abs" in all SEC plots represents relative absorbance measured at 280 nm. (**C**) PTAD conjugated BSA elution (red) completely superimposed over that of folded BSA. Labeling BSA in 20% ACN unfolded the protein and shifted its elution. (**D**) Conjugations of BSA and PTAD-N_3 or PTAD-N_3 clicked with DBCO-Cy5 were indistinguishable from native BSA in SEC. Fluorescence imaging confirmed Cy5 conjugation by SDS-PAGE analysis of eluted SEC fractions.

We then labeled BSA, with 2.25 mM PTAD to 121 μM protein. The effective concentration of the 21 tyrosine residues contained in BSA was calculated to be 0.25 mM, meaning PTAD was at ~10-fold excess over tyrosine. Comparing the SEC profile of native BSA and BSA labeled PTAD, indicated no change in the protein shape (**Figure 4C**, solid red line). Conjugation reactions with stock PTAD dissolved acetonitrile could expose the protein up to 5% acetonitrile/H2O without evidence of unfolding for the protein. However, labeling BSA in 20% acetonitrile/H2O produced a shift and broadening of the elution peak indicated a partial unfolding of the conjugated protein (**Figure 4C**, dashed red line). PTAD also strongly absorbs UV at 280 nm, which resulted in an additional peak at or near the end of the SEC elution. Finally, we labeled BSA with red·PTAD-N_3 in the presence of DBH, or that click-conjugated with DBCO-Cy5. SEC analysis again did not reveal evidence of protein unfolding (**Figure 4D**). Fluorescence imaging after SDS-PAGE of fractions eluted from SEC could confirm that BSA was conjugated with red·PTAD-N_3 and DBCO-Cy5.

Dependence of PTAD reactivity on solvent exposure. We next employed liquid chromatography with tandem mass spectrometry (LC-MS/MS) to quantitatively assess PTAD labeling at each tyrosine position for BSA. The expected outcome was that the ratio of labeled to unlabeled tyrosine at each position would relate directly to the relative solvent exposure in the folded structure. To perturb the levels of solvent exposure for tyrosine residues, we repeated the labeling with PTAD for BSA denatured in 4 M or 8 M urea.

LC-MS/MS identified between 44 and 83 unique peptides for each labeled or unlabeled BSA sample, covering between 58 and 84% of the amino acid sequence for the protein. Of the 21

tyrosine residues in BSA, 20 were observed by LC-MS/MS at least once. We inspected the relative abundance of each unique peptide as a percentage of the total number of peptides detected for each experimental condition (**Figure 5A**). This revealed only small changes between labeled and unlabeled BSA samples, or samples labeled in 0, 4, and 8 M urea and suggested that no peptides had dropped from detection by LC-MS/MS due to PTAD labeling. We quantified the ratio of PTAD-labeled to unlabeled (L/U) for the 13 tyrosine residues observed across all measurements (**Figure 5A**, **Supplemental Figure 4**). Only single additions of PTAD to tyrosine side chains, $Y^{(1)}$, were quantified, since no double additions, $Y^{(2)}$, were found.

For native BSA conjugated with PTAD, L/U ratios were counted for the tyrosine-containing peptides, which included between 39 to 524 observations per replicate (N = 4, **Figure 5B**). The L/U ratios for the tyrosine residues ranged from 0.05 to 1 in the labeled samples. PTAD labeling could be observed increased for residues having the highest solvent exposure, such as Y424 and Y475. Those residues with lower than expected labeling were also noted to be buried and relatively invisible from the protein's surface (**Figure 5C**). Tyrosine residues with high L/U ratios were clearly visible on the surface of BSA.

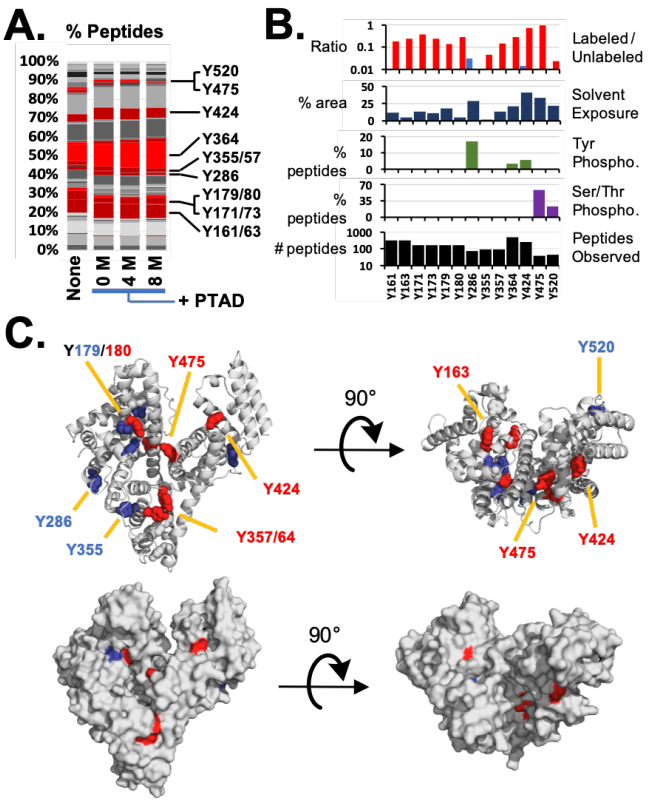


Figure 5. Quantitative analysis of PTAD labeling for BSA. (**A**) Summarizing the relative abundances of all unique peptides measured in LC-MS/MS revealed only small changes for unlabeled or PTAD-labeled BSA in 0, 4, or 8M urea (N = 4 for each condition). Peptides containing tyrosine residues are colored red. Tyrosine residues quantified during analysis across all treatments are listed and the peptides containing these are generally indicated. (**B**) The ratio of labeled to unlabeled residues is shown for BSA in 0M urea (red). Also shown are residues (light blue) called as false positives, meaning PTAD-labeled in unlabeled BSA samples. The computed solvent exposure of for tyrosine residue for BSA (PDB: 3V03) are plotted as % area (dark blue). The amount of phosphorylation the same tyrosine quantified (green) or at serine and threonine (purple) residues is shown as a percentage the total peptides containing the indicated tyrosine. Also shown is the total number of peptides detected in all replicates and containing the tyrosine indicated

(black). Note that the y-axes for only the plots of L/U ratios and numbers of peptides observed are shown in log scale due to the wide range of values included. (**C**) Two views of BSA (PDB: 3V03) are shown with tyrosine residues represented as spheres. Tyrosine residues not detected or quantified are grey (see **Supplemental Figure 3B**) and those shown in part (**B**) subsequently analyzed are colored to indicate whether they were labeled more than expected (red) or less (blue). The same two views for BSA are shown with the solvent accessible surface and the location of those tyrosine residues visible at the surfaces are colored the same as above.

Some residues were rarely or never observed to be PTAD labeled, despite their high solvent accessibility. LC-MS/MS revealed post-translational modifications to tyrosine or other residues in the peptides quantified. Y286 was observed to be phosphorylated with the highest frequency (17% of peptides in 0 M urea), and this residue was never detected conjugated by PTAD (0/174 peptides from all conditions, **Figure 5B**). The high L/U ratio for two other tyrosine residues, Y364 and Y424 (117/524 and 105/252 peptides in 0M urea), was consistent with their calculated solvent accessibility. These were also observed to be phosphorylated but at a much lower incidence (3.3 and 5.7 % of peptides in 0 M urea). Y520 was also rarely observed to be modified with PTAD (4/158 peptides among all conditions), but it was frequently observed that an adjacent threonine residue, T519, was phosphorylated (23.7 % of peptides in 0 M urea). Our solvent accessibility calculation was done for an unmodified BSA structure. We reasoned that phosphorylation of the tyrosine residue itself or a nearby serine or threonine residue may disrupt the local structure to allow PTAD less accessibility to the tyrosine residue than expected by structure analysis.

Local and global differences in structure changes PTAD conjugation. Closer inspection of L/U ratios, particularly for nearby tyrosine residues, could indicate that additional factors influenced PTAD access to conjugate tyrosine than simply solvent accessibility. These analyses were greatly simplified due to the high abundance of tyrosine residues, typical for mammalian proteins like BSA. For this reason, analyses of PTAD conjugations were most often of two or more tyrosine residues within the same peptide, minimizing potential bias in quantification that was not already controlled for by our experiment design.

For native BSA, the residues Y161 and Y163 were observed to have similar levels of PTAD labeling (respectively, 49 and 62 of 326 observations), despite a more than 2-fold difference in solvent accessibility (12 and 5%, respectively, **Figure 6A**). In the structure of BSA, Y163 could be seen oriented such that its hydroxyl and *ortho* positions in the phenolic ring are accessible to solvent. In contrast, the 12% surface accessible for Y163 was the backbone and *meta* positions, while the hydroxyl and *ortho* positions lay buried in the protein. Y171 and Y173 have nearly the same solvent exposure (13 and 11%, respectively), yet Y171 was observed to be conjugated 50% more frequently (44/143 peptides) than Y173 (32/143 peptides, **Figure 6B**). In the structure, both *ortho* positions of Y171 protrude to the protein surface. Only the *meta* and *ortho* positions at one side of Y173 are solvent exposed and the residue resides at the bottom of a narrow pocket in the protein, further limiting access for PTAD to penetrate and bind.

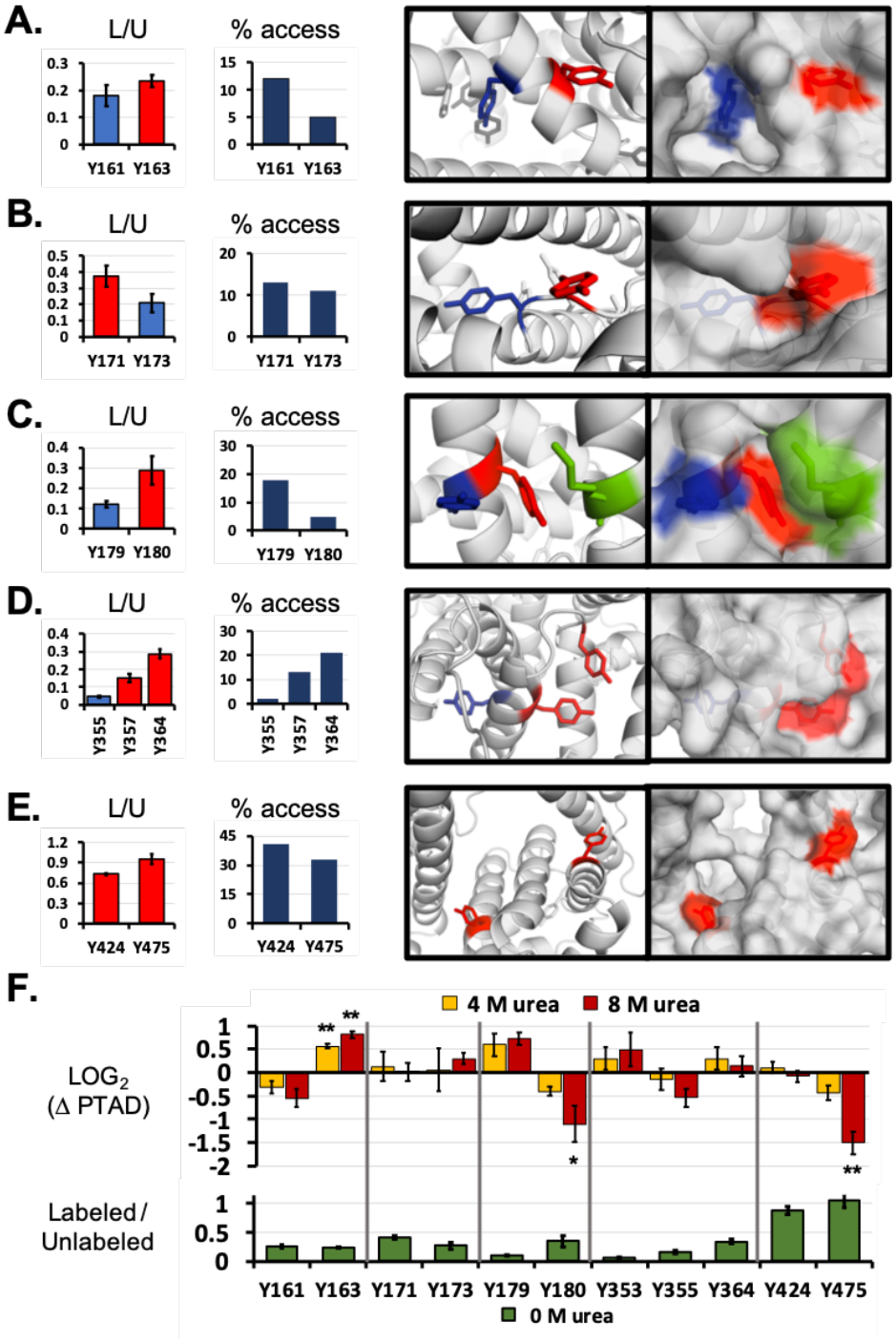


Figure 6. Effects of local and global protein structure to PTAD labeling. (**A–E**) Comparisons of nearby tyrosine residues are shown with the greater labeled residue to unlabeled ratio (L/U) colored red in the left bar graph and structures on the right. Residues with relatively low L/U ratios

compared to neighbors and % accessible surface area are colored blue. The % accessible surface area for each residue is also shown (dark blue bars) for the native structure. Right, tyrosine residues are shown in sticks and their solvent exposure as a surface representation. Included in (**C**), is K187, green, which may contribute to the high L/U for Y180 despite its low % accessible surface area. In (**E**), Y424 and Y475 are both shown in red as having the highest L/U ratios and % solvent exposure the residues analyzed here. (F) The LOG2 of the fold change in PTAD L/U is shown for BSA incubated in 4 M and 8 M urea and normalized to the L/U of native BSA in 0 M urea. Below, the L/U values for each residue in the native BSA structure are shown (green). Error bars are standard error about the mean (N = 4 for all treatments). * $p < 0.05$, and ** is $p < 0.005$, student t-test assuming equal variances.

A dramatic example of the effects of sidechain orientation to PTAD accessibility was Y179 and Y180. These two sidechains are tightly packed against each other in a narrow groove (**Figure 6C**). Y179 is more solvent exposed (18%) and Y180 is among the least solvent exposed in the protein (5%). Yet in the native structure, Y180 was the fourth most conjugated tyrosine detected (36/163 peptides observed) and Y179 was among the least conjugated (20/163 peptides observed). One factor may be the high flexibility suggested by the relatively high B-factors, between 55 and 63 $Å^2$ for Y179 and Y180. In contrast, B-factors for Y161, Y163, Y171, and Y173 were all < 1 $Å^2$. Second, the plane of the ring for Y179 lies at a shallow angle with the protein surface, making the angle of attack for PTAD toward the *ortho* position to be disadvantageous. The ring of Y180 lies nearly perpendicular to the protein surface with the *ortho* position near ideally exposed for PTAD to attack. Last, Y180 may have an advantage that the flexible sidechain of a lysine, K187,

lay less than 6 Å from the *ortho* position, whose basic property may serve to drive the local pH more basic in order to activate the PTAD conjugation to Y180.

The tyrosine residues Y355, Y357, and Y364 were measured to have L/U ratios that closely correlated with their relative percent solvent exposure (L/U = 3/113, 15/101, and 117/407, respectively **Figure 6D**). The small solvent exposure of Y355 was compounded by the occlusion of its *ortho* position and hydroxyl. Y357 and Y364 are located very near each other at the same surface, with Y364 slightly more protruding and Y357 slightly more buried between a rigid bundle of four α-helices. Consistent with their calculated solvent accessibility, Y424 and Y475 were the most solvent exposed tyrosine residues and the most frequently observed to be conjugated.

Lastly, we tested if abolishing the structure in BSA through the addition of urea might redistribute PTAD labeling more evenly across the tyrosine residues in the protein. We quantified the L/U ratios of the 13 residues we had analyzed and for BSA incubated in 4 M or 8 M urea (N = 4 for each treatment). By comparing the LOG2 of the fold change in L/U, we observed a subset of tyrosine residues whose labeling by PTAD was substantially altered by stepwise unfolding of BSA (**Figure 6F**). Residues Y171, Y173, Y364, and Y424 were not found to be substantially changed in their reactivity toward PTAD after unfolding BSA. Y163 was especially opened to more conjugation after unfolding. The least labeled residues in the folded protein, Y179 and Y353, saw increased conjugation. The large disparity of Y179 and Y180 conjugation was abolished once BSA was unfolded. Finally, Y475 enrichment for tyrosine labeling was lost in the unfolded protein.

Since lysine residues are charged and rarely found in the hydrophobic core of a protein, we measured the L/U ratios of the isocyanate produced from PTAD conjugating to lysine residues. For these reaction conditions, $NH^{(urea)}$ additions could be observed for 16 different lysine residues

in BSA (**Supplemental Figure 5A-B**). Unlike tyrosine residues, high concentrations of urea did not change the L/U rations for this modification. We considered this indicated solvent exposure to be fairly uniform for amine groups of lysine residues and unfolding the protein did not dramatically improve the availability of lysine residues for conjugation.

Once the protein is unfolded, the abolishment of disparities in relative PTAD abundance highlights structure to be a significant driver of enrichment or occlusion of PTAD conjugation to tyrosine. Conversely, lysine residues are not expected to differ considerably in solvent exposure, and these were found to be labeled at similar levels for folded or unfolded protein. These findings reveal that relative levels of PTAD detected at tyrosine residues can be used to discriminate the transition of the protein from the unfolded to folded state, as well as to quantify local perturbations in structure.

DISCUSSION

In agreement with earlier reports, we found PTAD to be highly efficient for conjugating tyrosine residues in peptides and proteins. We observed evidence for three types of products: a single phenolic addition to tyrosine, a double addition, and an isocyanate reaction with an amine (**Figure 1**). The side products were abated under controlled conditions and did not interfere with the investigation of tyrosine accessibility in either well-folded or unfolded proteins. Surprisingly, minimal optimization of reaction conditions yielded an easily quantifiable range in PTAD signals that were detected for the most and least accessible residues. This suggests a robustness to this conjugation reaction that may bolster its potential for wider use in probing protein structure-function relationships.

This study found a remarkable range in the level of PTAD labeling of tyrosine that could indicate distinctions in local tertiary structure of proteins (**Figure 5B**). By comparison, unfolding BSA had little or no effect on conjugation of lysine residues, whose charged amines prevent this residue from becoming buried within the protein structure (**Supplemental Figure 5**). This highlights the unique advantage that conjugation to tyrosine can offer to study protein structure. We did not observe clear evidence to predict the influence of primary amino acid sequence to conjugation. Nevertheless, the fact that primary structure can change amino acid solvent accessibility is both known and can be important to interactions of intrinsically disordered proteins[17]. Structural influences on PTAD conjugation that were observed in this study included the orientation of tyrosine residues with respect to the protein surface. Consistent with the ene-like mechanism, the *ortho* and hydroxyl positions of the phenol directed toward the surface was observed to be advantageous (**Figure 6A–B**).

Post-translational modifications, such as phosphorylation in the case of BSA, had the greatest impact on conjugation. The crystal structure available for BSA is not a phosphorylated form; thus, it is not possible to predict whether it is the phosphate itself or the changes caused to the local structure that most alters the ability of a residue to conjugate to PTAD. The percent phosphorylation of the native protein could be higher than measured because of known challenges to detecting unenriched phospho-peptides by LC-MS/MS. For one example, Y520, which lies in a highly flexible region of BSA, this residue might be expected to be more easily targeted than suggested by the crystal structure. However, phosphorylation of T519 would be likely to significantly disrupt the local structure, which can explain the divergence in measured and expected PTAD conjugation.

This study has highlighted the availability of multiple modalities to detect PTAD modifications. Site specific and quantitative analysis for PTAD conjugation can be provided by mass spectrometry approaches. Conjugating BSA with PTAD click modified with a fluorescent dye did not change the fold of the protein observable to SEC, which could encourage development of fluorescence detection as an alternative to MS to quantify protein structure changes. Fluorescence can be a highly sensitive and quantitative method capable of reaching single molecule levels of detection[7a, 8a]. Moreover, the utility of click chemistry is a well-known platform to open new modalities, such as radiolabels, biotin, epitope or affinity tags, and chemical modifications that allow highly sensitive enzyme-linked assays or chromatography[18].

In conclusion, we consider that comparative or quantitative analysis of protein structure by tyrosine conjugation is a feasible approach that has unique advantages apart from conjugations of other protein residues. The first reason is the relative common abundance of tyrosine in proteins. Its amphipathic nature distributes those residues across the divide of the protein surface and hydrophobic core. Last, the functionality of the PTAD chemistry and strength of covalent conjugations to proteins investigated, allows for a large adaptability that can add robustness to this approach. In the future, application of this approach to whole cell proteomics and investigations of tyrosine-rich low complexity proteins should reveal how much this new window into protein biology can uncover.

Corresponding Author

Correspondence should be addressed to Jacob C. Schwartz at jcschwartz@email.arizona.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was funded by the National Institute of Health [NS082376 and R21CA238499] and the American Cancer Society [RSG-18-237-01-DMC] to J.C.S. Research reported in this publication was also supported by the Office of the Director, National Institutes of Health of the National Institutes of Health under award number S10OD013237.

ABBREVIATIONS

BSA, bovine serum albumin; CD, circular dichroism; DBCO, dibenzocyclooctyne; DBH, 1,3-Dibromo-5,5-dimethylhydantoin; LC-MS/MS, liquid chromatography and tandem mass spectrometry; $NH^{(urea)}$, amine conjugated isocyanate; PTAD, 4-Phenyl-3H-1,2,4-triazole-3,5(4H)-dione; PTAD-N_3, 4-(4-(2-Azidoethoxy)phenyl)-1,2,4-triazolidine-3,5-dione, N3-Ph-Ur for e-Y-CLICK; red·PTAD-N_3, reduced PTAD-N_3; SEC, size exclusion chromatography; UPLC-MS, ultra-high pressure liquid chromatography and mass spectrometry; $Y^{(1)}$, tyrosine with single PTAD conjugation; $Y^{(2)}$, tyrosine with two PTAD conjugations.

REFERENCES

[1] aJ. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, *Annu Rev Phys Chem* **1997**, *48*, 545-600; bC. B. Anfinsen, *Science* **1973**, *181*, 223-230.

[2] aL. Wang, M. R. Chance, *Mol Cell Proteomics* **2017**, *16*, 706-716; bS. Toal, R. Schweitzer-Stenner, *Biomolecules* **2014**, *4*, 725-773; cK. A. Dill, J. L. MacCallum, *Science* **2012**, *338*, 1042-1046.

[3] aD. A. Shannon, E. Weerapana, *Curr Opin Chem Biol* **2015**, *24*, 18-26; bE. M. Sletten, C. R. Bertozzi, *Angew Chem Int Ed Engl* **2009**, *48*, 6974-6998.

[4] aJ. Lee, M. Ju, O. H. Cho, Y. Kim, K. T. Nam, *Adv Sci (Weinh)* **2019**, *6*, 1801255; bS. Koide, S. S. Sidhu, *ACS Chem Biol* **2009**, *4*, 325-334.

[5] L. H. Jones, A. Narayanan, E. C. Hett, *Molecular bioSystems* **2014**, *10*, 952-969.

[6] aH. Du, X. Hu, H. Duan, L. Yu, F. Qu, Q. Huang, W. Zheng, H. Xie, J. Peng, R. Tuo, D. Yu, Y. Lin, W. Li, Y. Zheng, X. Fang, Y. Zou, H. Wang, M. Wang, P. S. Weiss, Y. Yang, C. Wang, *ACS Cent Sci* **2019**, *5*, 97-108; bA. A. Bogan, K. S. Thorn, *J Mol Biol* **1998**, *280*, 1-9.

[7] aM. Shadmehr, G. J. Davis, B. T. Mehari, S. M. Jensen, J. C. Jewett, *Chembiochem* **2018**, *19*, 2550-2552; bS. B. Hanay, B. Ritzen, D. Brougham, A. A. Dias, A. Heise, *Macromol Biosci* **2017**, *17*.

[8] aH. Ban, M. Nagano, J. Gavrilyuk, W. Hakamata, T. Inokuma, C. F. Barbas, 3rd, *Bioconjug Chem* **2013**, *24*, 520-532; bH. Ban, J. Gavrilyuk, C. F. Barbas, 3rd, *J Am Chem Soc* **2010**, *132*, 1523-1525.

[9] aD. Alvarez-Dorta, C. Thobie-Gautier, M. Croyal, M. Bouzelha, M. Mevel, D. Deniaud, M. Boujtita, S. G. Gouin, *J Am Chem Soc* **2018**, *140*, 17120-17126; bC. M. Madl, S. C. Heilshorn, *Bioconjug Chem* **2017**, *28*, 724-730; cA. Nilo, M. Allan, B. Brogioni, D. Proietti, V. Cattaneo, S. Crotti, S. Sokup, H. Zhai, I. Margarit, F. Berti, Q. Y. Hu, R. Adamo, *Bioconjug Chem* **2014**, *25*, 2105-2111; dD. M. Bauer, I. Ahmed, A. Vigovskaya, L. Fruk, *Bioconjug Chem* **2013**, *24*, 1094-1101.

[10] J. Wang, J. M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, *Cell* **2018**, *174*, 688-699 e616.

[11] aM. Kato, S. L. McKnight, *Annu Rev Biochem* **2017**; bS. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat Rev Mol Cell Biol* **2017**, *18*, 285-298; cD. S. Protter, R. Parker, *Trends Cell Biol* **2016**, *26*, 668-679; dV. F. Thompson, R. A. Victor, A. A. Morera, M. Moinpour, M. N. Liu, C. C. Kisiel, K. Pickrel, C. E. Springhower, J. C. Schwartz, *Biochemistry* **2018**, *57*, 7021-7032; eJ. C. Schwartz, T. R. Cech, R. R. Parker, *Annu Rev Biochem* **2015**, *84*, 355-379.

[12] R. Schuller, D. Eick, *Trends Biochem Sci* **2016**.

[13] H. Nishi, J. H. Fong, C. Chang, S. A. Teichmann, A. R. Panchenko, *Molecular bioSystems* **2013**, *9*, 1620-1626.

[14] D. Kaiser, J. M. Winne, M. E. Ortiz-Soto, J. Seibel, T. A. Le, B. Engels, *J Org Chem* **2018**, *83*, 10248-10260.

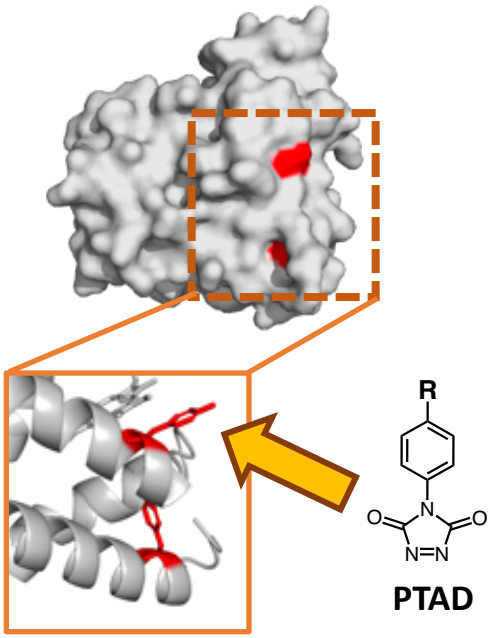
[15] S. Sato, H. Nakamura, *Molecules* **2019**, *24*.

[16] D. R. Canchi, A. E. Garcia, *Annu Rev Phys Chem* **2013**, *64*, 273-293.

[17] aR. Chelli, F. L. Gervasio, P. Procacci, V. Schettino, *Proteins* **2004**, *55*, 139-151; bS. Ahmad, M. M. Gromiha, A. Sarai, *Proteins* **2003**, *50*, 629-635.

[18] aJ. C. Jewett, C. R. Bertozzi, *Chem Soc Rev* **2010**, *39*, 1272-1279; bC. S. McKay, M. G. Finn, *Chem Biol* **2014**, *21*, 1075-1101.

Table of Contents Graphic



For folded proteins, solvent exposure controls the efficiency by which tyrosine residues are available for conjugation. Changes to protein structure can change tyrosine accessibility considerably. Such changes are quantifiable by the extent of tyrosine labeling observed for distinctive folded or conformational states of a protein.