Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification

Marcus Jäger, Eyal Nir and Shimon Weiss

Protein Sci. 2006 15: 640-646; originally published online Feb 1, 2006;
Access the most recent version at doi:10.1110/ps.051851506

Supplementary data
"Supplemental Research Data"
http://www.proteinscience.org/cgi/content/full/ps.051851506/DC1

References
This article cites 58 articles, 24 of which can be accessed free at:
http://www.proteinscience.org/cgi/content/full/15/3/640#References

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

Notes

To subscribe to Protein Science go to:
http://www.proteinscience.org/subscriptions/

© 2006 Cold Spring Harbor Laboratory Press
FOR THE RECORD

Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification

MARCUS JÄGER,1 EYAL NIR,1 AND SHIMON WEISS1,2,3
1Department of Chemistry and Biochemistry, 2Department of Physiology, and 3Molecular Biology Institute, University of California at Los Angeles (UCLA), Los Angeles, California 90095, USA

(RECEIVED September 16, 2005; FINAL REVISION November 22, 2005; ACCEPTED November 28, 2005)

Abstract
An often limiting factor for studying protein folding by single-molecule fluorescence resonance energy transfer (FRET) is the ability to site-specifically introduce a photostable organic FRET donor (D) and a complementary acceptor (A) into a polypeptide chain. Using alternating-laser excitation and chymotrypsin inhibitor 2 as a model, we show that chemical labeling of a unique cysteine, followed by enzymatic modification of a reactive glutamine in an N-terminally appended substrate sequence recognition tag for transglutaminase (TGase) affords stoichiometrically D-/A-labeled protein suitable for single-molecule FRET experiments. Thermodynamic data indicate that neither the presence of the TGase tag nor D/A labeling perturbs protein stability. As the N terminus in proteins is typically solvent accessible, a TGase tag can (in principle) be appended to any protein of interest by genetic engineering. Two-step chemical/ enzymatic labeling may thus represent a simple, low-cost, and widely available strategy for D/A labeling of proteins for FRET-based single-molecule protein folding studies, even for non-protein-experts laboratories.

Keywords: protein labeling; transglutaminase; fluorescence resonance energy transfer; single-molecule spectroscopy; alternating laser excitation; fluorescence-aided molecular sorting

Supplemental material: see www.proteinscience.org

Fluorescence resonance energy transfer between a single donor (D) and a single acceptor (A) (single-pair or sp-FRET) is a sensitive method for monitoring conformational dynamics in biomolecules (Selvin 2000; Weiss 2000; Deniz et al. 2001; Ha 2001, 2004; Haran 2003). The FRET efficiency (E) depends on the D/A distance (R) according to $E = [1 \pm (R/R_0)^6]^{-1}$. The Förster radius $R_0$ is a characteristic constant that corresponds to R at which $E = 50\%$ (Clegg 1992). For typical single-molecule fluorophores, $R_0 \sim 4-6$ nm, allowing the use of sp-FRET as a distance ruler in the 2-8 nm range (Stryer and Haugland 1967).

As cysteines (Cys) are underrepresented in proteins, D/A labeling can be achieved by sequential modification of two engineered Cys in a polypeptide chain. However, unless the dye accessibility of the thiol side chains differs drastically, the first added fluorophore can react with either of the two Cys, resulting in mixtures of D-/A-labeled molecules and their dye-permutated A/D analogs. Unless such...
mixture can be chromatographically separated (Ratner et al. 2002), this may lead to sample heterogeneity, as the photophysical properties of the conjugated fluorophore may depend on local environment (charge, pH, or hydrophobicity) (Moerner and Orrit 1999; Brasselet and Moerner 2000).

Methodologies to improve labeling specificity include total chemical synthesis of proteins (Jia et al. 1999; Deniz et al. 2000; Talaga et al. 2000); specific chemical modification/laboring of certain N-terminal residues (Gaertner et al. 1992; Geoghegan and Stroh 1992; Zhang and Tam 1996; Schuler and Pannell 2002; Tolbert and Wong 2002); segmental labeling/assembly of proteins using intein, chemical, or chemoenzymatic ligation (Kaiser et al. 1989; Dawson et al. 1994; Liu and Tam 1994; Muir et al. 1998); the use of protein–protein interactions to protect a Cys in a protein–protein interface (Jäger et al. 2005); in vitro transcription/translation with tRNAs aminoacylated with noncanonical amino acids (Kiga et al. 2002; Olejnik et al. 2005); or (most elegantly, but not yet generally available) the incorporation of nonnatural amino acid analogs in vitro (or in vivo) using genetically engineered orthogonal tRNA/tRNA synthetase pairs (Noren et al. 1989; Wang et al. 2002; Chin et al. 2003).

Enzymatic post-translational modification of proteins is an alternative approach toward site-specific labeling. Transglutaminase (TGase) catalyzes the acyl transfer reaction between the γ-carboxyamide group of an acceptor glutamine (Gln) residue and a primary amine donor (Folk 1980). Although the sequence requirements that render a Gln into a TGase substrate are not fully understood, tissue TGase seems to exhibit a very stringent recognition for the primary amine of Alexa647 maleimide (A488 hereafter). A substrate sequence-recognition tag (Pro-Lys-Pro-Gln-Gln-Phe) for tissue TGase was appended at the N terminus of CI2. TGase will catalyze the acyl transfer reaction between the primary amine of Alexa647 and the γ-carboxyamine moiety of the reactive Gln (Gln4) in the TGase tag (Sato et al. 1996; Taki et al. 2004). The three internal Glns (Gln22, Gln28, Gln 59) in CI2 (Fig. 1B), although partially solvent-exposed, are embedded in a rigid tertiary structure context and should not be recognized as substrates for tissue TGase.

**Results and Discussion**

**Model system and proposed labeling scheme**

Chymotrypsin inhibitor 2 (CI2) (Fig. 1A) is a 64-residue, single-domain serine protease inhibitor and has served as a model system for experimental and theoretical protein folding studies (Daggett et al. 1996; Ladurner et al. 1998; Deniz et al. 2000; Kazmirski et al. 2001). A unique Cys residue (Cys40) was introduced at a solvent-exposed position into CI2 for site-specific labeling with Alexa488 maleimide (A488 hereafter). A substrate sequence-recognition tag (Pro-Lys-Pro-Gln-Gln-Phe) for tissue TGase was appended at the N terminus of CI2. TGase will catalyze the acyl transfer reaction between the primary amine of Alexa647 and the γ-carboxyamine moiety of the reactive Gln (Gln4) in the TGase tag (Sato et al. 1996; Taki et al. 2004). The three internal Glns (Gln22, Gln28, Gln 59) in CI2 (Fig. 1B), although partially solvent-exposed, are embedded in a rigid tertiary structure context and should not be recognized as substrates for tissue TGase.

**Labeling nomenclature**

In the following, the Cys40 variant of CI2 without the N-terminal TGase tag is dubbed CI2-C40, and the tagged variant is abbreviated TG-CI2-C40. CI2, singly labeled with A488 at Cys40, is named CI2-(A488)C40 (no TGase tag) or TG-CI2-(A488)C40 (with TGase tag). D-/A-labeled CI2 is abbreviated (A647)TG-CI2-(A488)C40, with A647 attached to Gln4 in the appended N-terminal TGase tag and A488 conjugated to Cys40.

**Labeling specificity of TGase**

A prerequisite for the proposed labeling scheme (Fig. 1A) is that TGase modifies the solvent-exposed, reactive Gln in the N-terminal TGase tag (Gln4) but none of the semixposed intrinsic Glns (Fig. 1B). To verify that only Gln4 in the appended N-terminal TGase tag is reactive, we incubated CI2-(A488)C40 or TG-CI2-(A488)C40 with a 10-fold stoichiometric excess of A647. The extent of A647 labeling was estimated from absorption spectra taken with the protein solution after removal of nonreacted A647 (extensive washing on a 3-kDa-cutoff membrane until the absorbance of the flowthrough at 647 nm was negligible). As TGase does not contribute to absorption in the visible range, it was not chromatographically removed at this stage. As shown in Figure 1C, TG-CI2-
A488 exhibited two absorption bands with maxima at 488 and 647 nm (typical for A488 and A647, respectively). Only the A488 peak was detectable for CI2-(A488)C40, lacking the N-terminal TGase tag. We conclude that TGase labeling is specific for Gln4 in the flexible TGase tag; none of the three semiexposed intrinsic Glns is recognized as a substrate for TGase.

Labeling stoichiometry

Labeling stoichiometry was confirmed by matrix-assisted laser-desorption ionization (MALDI) mass spectrometry. A summary of the mass spectrometric analysis is given in Table 1. Raw mass spectra are provided as supplementary material.

Single-molecule FRET

Alternating-laser excitation (ALEX) is a novel single-molecule spectroscopic technique that allows fluorescence-aided molecular sorting (FAMS) of freely diffusing molecules by calculation of two ratiometric values (Kapanidis et al. 2004, 2005; Jäger et al. 2005; Lee et al. 2005). The traditional FRET efficiency \( E = \frac{F_{\text{em}}^{\text{Dexc}} + \gamma F_{\text{em}}^{\text{Aexc}}}{F_{\text{Dexc}} + F_{\text{Aexc}}} \) reports on the \( D/A \) distance; the distance-independent stoichiometry ratio \( S = \frac{F_{\text{Dexc}}}{F_{\text{Dexc}} + F_{\text{Aexc}}} \) reports on the relative \( D/A \) stoichiometry.

Figure 1. (A) Labeling scheme of Chymotrypsin inhibitor 2 (CI2), represented in surface-accessible area mode (gray). The unique Cys40 is indicated by a solid black dot and labeled (C40). The added flexible N-terminal TGase tag is depicted by a solid black line with the reactive Gln indicated by a solid black dot and labeled (Q4'). A488-maleimide (A488) and A647-cadaverine (A647) are shown as open circles. (B) Location and solvent accessibility of the internal Glns (dark gray) in CI2 (light gray). (C) Labeling specificity of TGase. Absorbance spectra obtained with CI2-(A488)C40 (solid line) or TG-CI2-(A488)C40 (dotted line) after incubation with a 10-fold stoichiometric excess of A647 in the presence of TGase. Unreacted A647 was removed by filtration prior to acquisition of spectra. For comparison, spectra are normalized to the absorbance at 488 nm.

Table 1. Probing of labeling specificity and labeling stoichiometry by matrix-assisted-laser-desorption-ionization mass spectrometry

<table>
<thead>
<tr>
<th>Protein/fluorophore</th>
<th>Theoretical mass (Da)</th>
<th>Experimental mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A488(^a)</td>
<td>720.20</td>
<td></td>
</tr>
<tr>
<td>A647(^b)</td>
<td>911.0(^c)</td>
<td></td>
</tr>
<tr>
<td>CI2-C40</td>
<td>7245.5(^d)</td>
<td>7235.6</td>
</tr>
<tr>
<td>TG-CI2-C40</td>
<td>7971.4(^e)</td>
<td>7960.5</td>
</tr>
<tr>
<td>TG-CI2-(A488)C40</td>
<td>8691.6</td>
<td>8679.4</td>
</tr>
<tr>
<td>(A647)TG-CI2-(A488)C40</td>
<td>9602.6</td>
<td>9619.0</td>
</tr>
</tbody>
</table>

\(^a\) Alexa-488 maleimide (http://www.probes.com).
\(^b\) Alexa-647 cadaverine (A647-NH-CO-(CH\(_2\)\(_5\)-NH\(_2\) (http://www.probes.com).
\(^c\) Estimated from A647-succinimidylester (1020 Da) (http://www.idtdna.com).
\(^d\) Molecular weight of CI2-C40 with the initiation-Met (sequence: MKTE-...).
\(^e\) Molecular weight of TG-CI2-C40 without the initiation-Met (sequence: PKPQFMKT-...).
$P_{\text{Dexc}}^\text{Dem}$ is the $D$-excitation-based $D$ emission, $P_{\text{Dexc}}^\text{Em}$ is the $D$-excitation-based $A$ emission, $P_{\text{Dexc}}$ is a sum of $D$-excitation-based emissions, $P_{\text{Aexc}}$ is a sum of $A$-excitation-based emissions, and $\gamma$ is a detection correction factor. Previous studies showed that $\gamma \approx 1$ (Ha et al. 1999), and $\gamma = 1$ was assumed in this study.

The expected location of $D$-only, $A$-only, and $D/A$ subpopulations in a 2D $E$-$S$ histogram is schematically depicted in Figure 2A. Experimental $E$-$S$ histograms obtained with $(A647)$TG-CI2-(A488)C40 are depicted in Figure 2, B–D. At (or below) 2.0 M GdnCl, CI2 is folded and a single subpopulation with mean $E \approx 0.9$ and mean $S \approx 0.55$ is observed (Fig. 2B). A high $E$ ratio indicates that the interdye distance is much smaller than the Förster radius ($R_0$) for the A488/A647 dye pair ($\approx 56$ Å) (http://www.probes.com), consistent with a C-Cys40 and the N-terminal methionine in CI2 lacking the TGase tag (Radisky and Koshland 2002). Essentially no A488-only or A647-only subpopulations are visible, indicating that unreacted TG-CI2-(A488)C40 has been removed by ion-exchange chromatography (IEX) (Fig. 1A, step 2) and that fluorophore photobleaching during sample preparation/data acquisition is negligible. Addition of denaturant leads to unfolding of CI2, as indicated by the emergence of a second subpopulation with lower mean $E$ ($E \approx 0.45$, corresponding to a $D/A$ distance of $\approx 55$ Å assuming the experimentally determined $R_0 = 56$ Å) (Fig. 2C,D). At 3.6 M GdnCl (midpoint of unfolding), native and unfolded subpopulations are approximately equally populated. Complete unfolding is observed at (or above) 5 M GdnCl.

Figure 2E depicts 1D $E$ histograms obtained with $(A647)$TG-CI2-(A488)C40 at denaturant concentrations close to or within the unfolding transition region (2–5 M GdnCl), where folded and unfolded conformers coexist. The FAMS capabilities of ALEX were exploited to construct these histograms by projection of only the region of the 2D $E$-$S$ histogram onto the $E$-axis that contained photons from $D/A$-labeled molecules (selected area indicated by dashed black boxes in Fig. 2B–D; Kapanidis et al. 2004, 2005; Jäger et al. 2005; Lee et al. 2005). This option eliminates the contaminating $D$-only (and $A$-only) contribution to the 1D FRET efficiency $E$ histogram, which becomes more pronounced in the presence of high concentrations of denaturant (Denz et al. 1999; Schuler et al. 2002; McCarney et al. 2005). Although not required here, ALEX-FAMS also allows accurate mean-$E$ determinations for distances outside the dynamic FRET range ($E < 0.1$, due to $D/A$ colocalization). For clarity, only the $0.2 < E < 1.0$ range is shown in Figure 2E. The solid lines represent fits of the $E$ histograms to Gaussian functions.

The disappearance of the high-$E$ subpopulation under strongly denaturing conditions is further single-molecule evidence that the internal Glns in CI2 (Fig. 1B) are not recognized as substrates by TGase. Assuming that the

**Figure 2.** Single-molecule FRET. (A) Schematic 2D $E$–$S$ histogram showing the expected location of $A488$-only CI2, $A647$-only CI2, and $A647$/$A488$-labeled CI2 (folded protein, high-$E$, denatured protein, lower $E$). (B–D) Measured $E$–$S$ histogram for $A647$/$A488$-labeled CI2 at 2 M GdnCl (B), 3.7 M GdnCl (midpoint of folding) (C), and 5 M GdnCl (D). The area of the $E$–$S$ histograms selected for projection of the 1D $E$ histograms (shown in E) is indicated by a dashed black box. (E) 1D $E$ histograms of $A647$/$A488$-labeled CI2 at various concentrations of GdnCl. The solid red lines are fits of the data to Gaussian functions.
sequence bracketed by Cys40 and any of the three internal Glns can be approximated by a random coil in >5 M denaturant (Kazmirski et al. 2001), the root-mean-square value of the distance r between the Cα atoms of Cys40 (attachment site for A488) and an internal Gln (potential acceptor site for A647) is given by \( \langle r \rangle_{0}^{1/2} = n^{1/2} \cdot l \) (where n is the number of residue separating the two fluorophore chromophores and l is the bond length). With \( l = 3.8 \, \text{Å} \) (the length of a peptide bond in the trans-configuration), \( n = 12 \) (Gln28), 18 (Gln28), or 19 (Gln59) and \( R_0 = 56 \, \text{Å} \) for the A647/A488 dye pair, this would lead to \( r = 13.1–16.5 \, \text{Å} \), corresponding to a mean FRET efficiency \( E \) of \(~1\), even under conditions in which CI2 is fully denatured.

**Thermodynamic analysis**

For TGase labeling to be useful for protein folding studies, it needs to be shown that the presence of the moderately hydrophobic N-terminal TGase tag (Pro-Lys-Pro-Gln-Gln-Phe) does not significantly perturb the folding energy landscape of CI2. To investigate this, we studied the equilibrium unfolding of CI2-C40 (no TGase tag) and TG-CI2-C40 (TGase-tagged) by monitoring the change in intrinsic Trp fluorescence (Fig. 3A, B). The effect of \( D/A \) labeling on CI2 stability was addressed by monitoring the unfolding of (A647)TG-CI2-(A488)C40 via A488-sensitized A647 emission \( \left( \frac{F_{\text{D/A}}}{F_{\text{A/A}}} \right) \) as described elsewhere (Jäger et al. 2005; Fig. 3C). All three samples unfolded cooperatively and normalized denaturation curves are essentially superimposable (Fig. 3D). Fitting the denaturation curves to a two-state unfolding model (Santoro and Bolen 1988) yielded the following thermodynamic parameters (\( \Delta G_{N,D} \) is the free energy of folding; \( m_{eq} \) is the unfolding cooperativity): (1) CI2-C40: \( \Delta G_{N,D} = 31.0 \pm 0.6 \, \text{kJ/mol} \), \( m_{eq} = 8.2 \pm 0.2 \, \text{kJ/mol M}^{-1} \); (2) TG-CI2-C40: \( \Delta G_{N,D} = 30.4 \pm 0.4 \, \text{kJ/mol} \), \( m_{eq} = 8.4 \pm 0.1 \, \text{kJ/mol M}^{-1} \); (3) (A647)TG-CI2-(A488)C40: \( \Delta G_{N,D} = 30.2 \pm 0.8 \, \text{kJ/mol} \), \( m_{eq} = 8.2 \pm 0.2 \, \text{kJ/mol M}^{-1} \).

Also depicted in Figure 3D are data points obtained from Gaussian fitting of the histogram series shown in Figure 2E (solid black circles). The fraction of denatured protein \( (F_D) \) at each GdnCl concentration was calculated from the integrated area of the folded (high-\( E \)) and denatured (low-\( E \)) subpopulations. The single molecule data reproduce the ensemble data that were obtained by assuming the validity of a two-state unfolding scenario. We conclude that neither the presence of the N-terminal TGase tag nor labeling of the tagged protein with single A647/A488 FRET-fluorophore pair significantly perturbs the folding thermodynamics of CI2.

**Concluding remarks**

Using CI2 as a model, we have shown that chemical labeling of a unique Cys, followed by site-specific enzymatic modification of a reactive Gln in an N-terminally TGase substrate recognition tag, affords site-specifically \( D/A \)-labeled CI2 for FRET-based single-molecule studies. A unique Cys in CI2 avoids sample heterogeneity due to the formation of \( A/-/D \)-labeled and \( D/-/A \)-labeled molecules, which is possible in conventional sequential labeling of two Cys in a polypeptide chain (Ratner et al. 2002). Mass spectrometry in conjunction with fluorescence-aided single molecular sorting verified the high substrate specificity of tissue TGase. In agreement with previous studies (Sato et al. 1996; Taki et al. 2004), only the solvent-exposed Gln in the flexible N-terminally appended TGase tag was labeled; none of the semiburied internal Glns has been modified. Even though this result may not be representative for other proteins, the frequency of Gln residues (3.7%) is below the statistically expected value, and in most cases, it should be possible to replace reactive internal Glns with near-isosteric substitutes (e.g., Asn, Glu) without compromising protein stability or function. Equilibrium unfolding experiments revealed that the TGase tag (although slightly hydrophobic) does not significantly affect protein stability. While this again may not be representative for all proteins, it should be mentioned that similar observations were made with Acyl-CoA binding protein (Kragelund et al. 1999) and Protein L (Scalley et al. 1997; M. Jäger and S. Weiss, unpubl.). Two-step chemical/enzymatic labeling of proteins may thus represent an attractive
alternative to more costly or labor-intensive labeling strategies by affording a simple and widely accessible method for site-specific labeling of proteins for single-molecule FRET studies, even for non-protein-expert's laboratories.

Materials and methods
Alexa488 maleimide (A488) and Alexa647 cadavenerine (A647) were purchased from Molecular Probes. Bovine pig liver transglutaminase (TGase) was from Sigma. An expression plasmid for C12 (Glu76Ala/Lys53Arg double mutant) was a gift from Dr. Daniel Koshland (UC Berkeley). Two variants of C12 are discussed in this study. Variant 1 has a unique Cys at position 40 (dubbed C12-C40), while variant 2 carries an additional N-terminal TGase substrate sequence recognition tag (dubbed TG-C12-C40).

Protein expression and purification of C12-C40 and TG-C12-C40 were performed as described elsewhere (Radisky and Koshland 2002; Radisky et al. 2003). Labeling of C12-C40 or TG-C12-C40 with A488 was performed in 20 mM sodium phosphate (pH 7.0, dye/protein ratio = 2:1) for 4 h at 25°C in the dark. Unreacted dye was removed from the solution by repeated concentration/dilution (in 5 mM Tris at pH 8.0) of the protein solution on a Centricon YM3 membrane (Millipore) until the absorbance of the flowthrough fraction at 488 nm (emission maximum of A488) was negligible. Purification of C12-(A488)C40 from trace amounts of unreacted protein was accomplished by ion-exchange chromatography (IEX) using a MonoQ 5/5-column (Amersham-Pharmacia). A linear salt gradient (0–500 mM sodium chloride in 5 mM Tris at pH 8.0, 45 min, 2 mL/min) was used [nonlabeled C12 did not bind to the MonoQ-resin under these conditions; elution of C12-(A488)C40 occurred at ~80 mM NaCl]. Enzymatic labeling of TG-C12-(A488)C40 (10 μM) with A647 was performed in 100 mM Tris (pH 8.0), 1 mM A647, 10 mM CaCl2, and 0.5 unit of TGase for 24 h at 25°C in the dark in a total volume of 1 mL. Unreacted A647 was removed by gel filtration, and (A647)TG-C12-(A488)C40 was purified from unreacted TG-C12-(A488)C40 and TGase by IEX [(A647)TG-C12-(A488)C40 eluted at ~180 mM NaCl]. Yields of double-labeled C12-(A488)A647-C40 were determined to allow alternating-laser excitation (ALEX) using a two-laser excitation source (488 nm Ar+ laser, 634 nm diode-laser) was used. All measurements were carried out in 20 mM sodium phosphate (pH 6.3) and 100 μg/mL bovine serum albumin. Protein concentration was 40 PM. The 488-excitation power was 100 μW, and the 647-excitation power was 45 μW. Only bursts containing >180 photons were retained for further analysis.

Acknowledgments
This work was funded by NIH grant no. 1R01-GM65382 (to S.W.) and the NSF. The Center for Biophotonics, an NSF Science and Technology Center, is managed by the University of California, Davis, under Cooperative Agreement no. PHY0120999. E.N. is supported by the Human Frontier Science Program (HFSP).

References


