Published in final edited form as:

Biochemistry. 2020 September 22; 59(37): 3401–3410. doi:10.1021/acs.biochem.0c00474.

Structural Insights into How Protein Environments Tune the Spectroscopic Properties of a Noncanonical Amino Acid Fluorophore

J. Nathan Henderson¹, Chad R. Simmons¹, Nour Eddine Fahmi¹, Joshua W. Jeffs^{2,3}, Chad R. Borges^{2,3}, Jeremy H. Mills^{1,3}

¹The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, Arizona 85287, United States

²The Biodesign Virginia G. Piper Center for Personalized Diagnostics, Arizona State University, Tempe, Arizona 85287, United States

³School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287, United States

Abstract

Genetically encoded fluorescent noncanonical amino acids (fNCAAs) could be used to develop novel fluorescent sensors of protein function. Previous efforts toward this goal have been limited by the lack of extensive physicochemical and structural characterizations of protein-based sensors containing fNCAAs. Here, we report the steady-state spectroscopic properties and first structural analyses of an fNCAA-containing Fab fragment of the 5c8 antibody, which binds human CD40L. A previously reported 5c8 variant in which the light chain residue Ile^{L98} is replaced with the fNCAA L-(7-hydroxycoumarin-4-yl)ethylglycine (7-HCAA), exhibits a 1.7-fold increase in fluorescence upon antigen binding. Determination and comparison of the apparent pK_a s of 7-HCAA in the unbound and bound forms indicate that the observed increase in fluorescence is not the result of perturbations in pK_a . Crystal structures of the fNCAA-containing Fab in the apo and bound forms reveal interactions between the 7-HCAA side chain and surrounding residues that are

Corresponding Author: Jeremy H. Mills - School of Molecular Sciences and The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, Arizona 85287, United States; Phone: 480-965-7958 jeremy.mills@asu.edu.

J. Nathan Henderson - The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, Arizona 85287, United States

Chad R. Simmons - The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, Arizona 85287, United States

Noureddine Fahmi - The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, Arizona 85287, United States

Joshua W. Jeffs - The Biodesign Virginia G. Piper Center for Personalized Diagnostics, Arizona State University, Tempe, Arizona 85287, United States

Chad R. Borges - School of Molecular Sciences and The Biodesign Virginia G. Piper Center for Personalized Diagnostics, Arizona State University, Tempe, Arizona 85287, United States

SUPPORTING INFORMATION

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.0c00474>.

Detailed experimental methods and additional structural analyses.

ACCESSION CODES

CD40L, P29665; human IgG κ Light Chain P0DOX7; human IgG γ -1 Heavy Chain P0DOX5. All structures reported in this manuscript have been deposited in the PDB: 5c8 wild type at pH 5.5, 6w4w; 5c8* at pH 5.5, 6bjz; 5c8* at pH 9.7, 6w5a; 5c8*/CD40L complex at pH 6.8, 6w9g.

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.biochem.0c00474>

disrupted upon antigen binding. This structural characterization not only provides insight into the manner in which protein environments can modulate the fluorescence properties of 7-HCAA but also could serve as a starting point for the rational design of new fluorescent protein-based reporters of protein function.

Graphical Abstract

Introduction

Fluorescence-based techniques have revolutionized the study of biological systems by providing exquisite sensitivities and resolutions down to the single-molecule level.^{1–3} At the heart of many of these techniques is the specific labeling of a target protein with one or more fluorophores.^{4,5} In the past three decades, a versatile toolkit comprised of several distinct fluorescent dyes and labeling methods has been developed. Multiple approaches were likely required because no individual fluorophore or protein labeling method is universally applicable; furthermore, extensive optimization of the system under study is often necessary. The wealth of published information regarding the benefits and limitations of fluorophores and labeling methods that have previously found extensive use provides a road map for planning new studies. However, in the absence of such information, the widespread adoption of new fluorescent tools may be limited.

One class of fluorophores for which this may be the case consists of the genetically encoded fluorescent noncanonical amino acids (fNCAAs). The side chains of fNCAAs contain small molecule fluorophores (e.g., naphthyl,⁶ dansyl,⁷ acridone,⁸ or coumarin⁹) with excitation and emission maxima in the visible range, which distinguishes them from naturally occurring fluorophores in proteins (e.g., tryptophan). Furthermore, many fNCAAs exhibit sensitivity to their local environments and have chemical footprints only slightly larger than those of canonical amino acids.⁸ Because they are co-translationally incorporated in proteins, placement of fNCAAs is also not limited to sites on the surface or the termini of a target protein as is common with other fluorescent labeling methods. Collectively, these features suggest that fNCAAs should be useful in biological studies but to date, relatively few studies in which fNCAAs play a major role have been communicated. This is likely a result of the fact that until now, no structural characterization of fNCAA-containing proteins used as

reporters of protein function has been carried out. Without such information, the ability to employ fNCAAs in new and innovative ways is limited.

Among the fNCAAs that have been genetically encoded thus far,^{6–9} L-7-hydroxycoumarin-4-ethylglycine [7-HCAA, (Figure 1)] is unique as it is the only photoacid.¹⁰ Specifically, its side chain, 7-hydroxycoumarin (7-HC), contains a phenol whose pK_a drops from $\sim 7.8^9$ in the ground state to ~ 0.4 in the excited state.¹¹ In aqueous environments, excited-state proton transfer (ESPT) from 7-HC to water will occur. Therefore, despite the fact that 7-HC exhibits two absorbance maxima [~ 323 nm for the neutral form and ~ 365 nm for the phenolate form (Figure 1)], emission from the excited state anion at ~ 450 nm is observed in water upon excitation of either band. Alternatively, when ESPT efficiency is reduced or blocked completely, excitation of the neutral 7-HC results predominantly in emission in the ultraviolet (UV) at ~ 380 nm. Because the 7-HCAA amino acid has similar properties to the 7-HC fluorophore, it is possible the protein environment surrounding the 7-HCAA could modulate its photochemical properties in predictable and useful ways. In previous reports, 7-HCAA was used to develop fluorescent sensors of protein-ligand binding,¹² protein-protein interactions,¹³ and post-translational modifications.¹⁴ Despite these promising results, the precise manner in which the protein environment modulates the fluorescence properties of 7-HCAA in these systems remains a matter of speculation. Without additional characterization, it is difficult to predict how use of this versatile fNCAA could be extended.

Here, we report biochemical, spectroscopic, and crystallographic characterizations of a previously reported 7-HCAA-containing Fab that exhibits an increase in fluorescence upon binding its antigen.¹³ Steady-state spectral studies suggested the protein environment surrounding the fNCAA has the ability to alter the pK_a of a phenol on the 7-HCAA side chain and also suggested that this change in pK_a was not the origin of an increase in fluorescence upon protein complex formation. The first structural characterization of an fNCAA-containing protein is also reported and provides insight into the nature of structural rearrangements that occur in response to 7-HCAA incorporation. These data also highlight an unusually short hydrogen bond between a phenol in the fNCAA and a glutamate residue in the unbound protein that could provide an explanation for the observed change in fluorescence upon protein complexation.

Methods

Protein expression and purification

Wild-type 5c8 was cloned into the pBLN200 vector¹⁵ and expressed into the periplasm of Top10 *Escherichia coli* cells. Purification was carried out using a combination of protein G, ion exchange, and size exclusion chromatography. To generate 5c8*, position L98 was mutated to the amber stop codon, TAG. 5c8* expression was carried out using the amber suppression system developed by Schultz et al.¹⁶ Briefly, Top10 *E. coli* were co-transformed with plasmids encoding 5c8* and the pEVOL-CouRS plasmid. pEVOL-CouRS contains an aminoacyl tRNA synthetase (aaRS) with specificity for 7-HCAA and a cognate tRNA that contains an anti-codon loop specific to the amber stop codon. 5c8* was expressed in liquid culture or on a solid agar medium. After purification, the presence of full-length 5c8* was

confirmed by electrospray ionization mass-spectrometry (ESI-MS). Initially, the presence of a contaminant protein that contained tyrosine in place of 7-HCAA was observed upon ESI-MS analysis. Thus, further purification by ion exchange chromatography (IEC) was carried out to improve the purity of 5c8* relative to the tyrosine-containing contaminant. After IEC purification, ESI-MS analysis confirmed that the contaminant protein had been reduced to <5% of the total protein (Figure S1), and a final purification using size exclusion chromatography was carried out prior to spectroscopic or crystallographic characterization studies. See the Supporting Information for detailed descriptions of wild-type 5c8 and 5c8* expression, purification and ESI-MS methods.

A gene encoding the CD40L extracellular domain (residues 116-261) was integrated into the pPIC9K plasmid from the Multi-Copy Pichia Expression Kit (Thermo Fisher), which results in secretion of recombinantly expressed protein. A sequence confirmed pPIC9K-CD40L plasmid was transformed into *Pichia pastoris* using electroporation, and analysis of CD40L integration levels was carried out per the manufacturer's recommendations. Protein expression and purification was carried out as previously described¹³ with minor modifications (see Supporting Information for detailed CD40L expression and purification methods).

Steady state spectroscopy

Absorbance spectra were recorded with a SpectraMax M5 microplate reader (Molecular Devices) in a quartz cuvette (Starna Cell, Inc) at room temperature. Protein concentrations for absorbance assays ranged from 400 to 600 $\mu\text{g/mL}$, while the concentration of 7-HCAA was 5.2 μM in these assays; both protein and small molecule spectra were collected in 100 mM buffer (see below for buffers used in this study) with 150 mM NaCl. Fluorescence spectra were recorded on a NanoLog spectrofluorometer (Horiba Scientific) in a quartz cuvette (Starna Cell, Inc) at 25°C. Protein concentrations in fluorescence assays ranged from 60-130 $\mu\text{g/mL}$ and ~4 μM for 7-HCAA. The pH-dependence of absorbance spectra was assessed in buffers including MES, HEPES, CHES and CAPS; mixtures of these buffers were used to establish the pHs at which the data were collected. Spectra were recorded between 200 and 500 nm and were then normalized to the absorbance at 338 nm, corresponding to the isosbestic point of 7-HCAA. Apparent pK_a values for 7-HCAA, apo 5c8* and 5c8* in complex with CD40L were determined by plotting the intensities at the two major absorption bands (~326 nm and ~371 nm) as a function of pH and fitting (Kaleidagraph, Synergy Software) the data to titration curves of the following form:

$$I(\text{pH}) = C + D / (1 + 10^{pK_a - \text{pH}})$$

where D is the absorption signal dynamic range and C is an offset from the baseline. For comparison of relative fluorescence quantum yields, samples were prepared in 100 mM buffer (HEPES or CAPS) with 150 mM NaCl at pH values approximately one unit above or below the apparent pK_a s and with an absorbance of ~0.05 at the excitation maxima of the relevant 7-HCAA species (323 nm or 373 nm). Data were collected at room temperature on a SpectraMax M5 microplate reader (Molecular Devices) in a quartz cuvette (Starna Cell, Inc).

Protein crystallization

Apo crystals of 5c8 variants and their complexes with CD40L were obtained at room temperature by sitting-drop vapor diffusion under a wide-range of conditions. For the apo 5c8 variant the best crystals grew from 27-33% PEG 3350 and 0.1 M citric acid/Na₃ citrate (pH 3.0-4.0). The 3 μ L crystallization drops contained a 2:1 mixture of protein to reservoir solution at a protein concentration between 3.0-5.6 mg/mL. Diffraction quality crystals of 5c8* in complex with CD40L grew from 16-24% PEG MME 2000, 0.1 M trimethylamine N-oxide (TMANO) and 0.1 M Tris-HCl pH 7.5-9.0 in 3 μ L drops with a 2:1 protein to reservoir ratio at a protein concentration between 1.4-2.2 mg/mL. Prior to being flash frozen in liquid N₂, crystals were soaked for 2 hours in either 40% PEG 3350 and 0.1 M citric acid/Na₃ citrate (pH 5.5) for apo 5c8-WT and 5c8* or 40% PEG MME 2000, 0.1 M Tris-HCl (pH 6.8) and 0.1 M TMANO for the 5c8*/CD40L complex. For the high pH structure of apo 5c8*, the citrate buffer was replaced with 0.1 M CAPS pH 9.7 during the cryoprotection step.

Data collection and structure determination

Diffraction data from the 5c8* apo crystals at pH 5.5 were collected at the Berkeley Center for Structural Biology (BCSB) from the Advanced Light Source (beamline 8.2.2) on an ADSC Q315R detector, 5c8*/CD40L complex diffraction data were measured at the Stanford Synchrotron Radiation Lightsource (beamline 9-2) on a Pilatus 6M detector and all other data were obtained at the Argonne National Laboratory Advanced Photon Source (beamline 19-ID) on a Pilatus 6M detector. Crystals were flash frozen in liquid nitrogen prior to data collection at 100 K. Data were indexed, refined, integrated, and scaled using the HKL2000¹⁷ or MosFLM¹⁸ package. The structure of 5c8-WT was solved by molecular replacement using Phaser¹⁹. Initial searches using the entire 5c8 Fab coordinates from the complex of 5c8 with CD40L (RCSB Protein Data Bank ID 1i9r) as the input model proved unfruitful. Splitting 5c8 in half and searching sequentially for two ensembles, the first consisting of the heavy and light chain constant region and the second the variable domain, gave a reasonable solution. Structures of the subsequent 5c8* apo variants were solved using the 5c8-WT structure as a search model. The structure of the complex of 5c8* with CD40L was determined by molecular replacement using Phaser with PDB entry 1i9r as the search model. All models were refined using Refmac5,²⁰ and model building was carried out with the program Coot.²¹ A chemical description for 7-HCAA was created using Sketcher from the CCP4 program suite,²² and the geometrical restraints used in refinement were taken from the bond lengths and angles reported for the crystal structure of umbelliferone.²³ All structural figures were made with the PyMOL molecular graphics software.²⁴

Results and Discussion

Steady State Absorption and Fluorescence Spectroscopy

In a previous study,¹³ 7-HCAA was incorporated at position 98 in the light chain of an Fab (termed 5c8), that binds CD40L, a homotrimeric protein on the surface of T-cells (Figure 2a,b). Residue Ile^L98 was targeted as a site of 7-HCAA incorporation because of its proximity to the antibody-antigen interface and its lack of involvement in intermolecular hydrogen bonding interactions (Figure 2b). Upon CD40L binding, the fluorescence intensity

of 5c8 Ile^{L98}→7-HCAA (hereafter referred to as 5c8*) increased ~2-fold. However, the fact that no studies that aimed to determine the origin of the increase in fluorescence were carried out might limit the ability to apply these rational design principles to other systems. In an effort to explore this, we developed an initial hypothesis that the protein environments surrounding the 7-HCAA residue at position ^{L98} (termed 7-HCAA^{L98}) in the apo and bound states altered the p*K*_a of the fNCAA, which in turn led to the observed changes in fluorescence.

To test this, we measured absorption (Figure S2) and fluorescence emission spectra [$\lambda_{\text{ex}} = 373$ nm (Figure S3)] of 7-HCAA and 5c8* in the presence and absence of CD40L. Data from these studies fit to titration curves for a single protonatable species (Figure 3) and yielded apparent average p*K*_as of ~8.64 for apo 5c8*, ~8.48 for the 5c8*/CD40L complex and ~7.89 for the 7-HCAA amino acid (Table 1). The preceding p*K*_a values represent averages of values independently determined from fluorescence and absorbance methods as reported in column 3 of Table 1. These data suggest that (1) the local environments surrounding the 7-HCAA^{L98} within 5c8* contribute to an equilibrium shift in favor of neutral 7-HCAA^{L98} relative to 7-HCAA and (2) that 7-HCAA^{L98} is more protonated in apo 5c8* than in the complex. In the original study,¹³ 5c8* was excited exclusively at 316 nm, which corresponds to the neutral/protonated 7-HCAA^{L98}. For an increase in p*K*_a to account for the reported increase in fluorescence¹³, we would expect the p*K*_a in the apo state to be ~0.35 unit lower than in the complex. However, our data indicate that the apparent p*K*_a of 7-HCAA^{L98} in the apo state is slightly higher than in the complex. Therefore, we conclude that the increased fluorescence upon CD40L binding does not arise from modulation of the p*K*_a.

Another possibility is that changes in the local protein environments surrounding 7-HCAA^{L98} in the apo versus bound states of 5c8* may directly affect the emission efficiencies of the fluorophore. To explore this, relative fluorescence quantum yields were determined for 7-HCAA, 5c8* and 5c8*/CD40L complex at pHs corresponding to values approximately one unit below or above the apparent p*K*_as. At pH values below the p*K*_as, the neutral forms of 7-HCAA or 7-HCAA^{L98} should predominate, while at pHs above the p*K*_as, the anionic form of the fNCAA should be favored (Figure 1). At pH values below the p*K*_a, excitation of the neutral form of 7-HCAA^{L98} at 323 nm results in 450 nm emission that is attenuated in both apo and bound samples relative to 7-HCAA albeit substantially more so for apo 5c8* (Figure 4a and Table 1). In contrast, at pH values above the apparent p*K*_as, excitation of the phenolate form of 7-HCAA^{L98} at 373 nm results in 450 nm emission that is essentially equal in intensity for 5c8* and 5c8*/CD40L (Figure 4b and Table 1), and is only slightly attenuated relative to the 7-HCAA in solution (Table 1). These data indicate that: (1) incorporation of 7-HCAA into 5c8 reduces the emission intensity relative to the small molecule irrespective of whether the neutral or anionic form is excited, (2) excitation of neutral 7-HCAA^{L98} gives rise to a ~1.7-fold increase in 5c8* emission intensity in the presence of CD40L (Figure 4a and Table 1) and (3) no change in 450 nm emission is observed upon complexation of CD40L when the anionic form is excited. Thus, non-radiative paths for excited state decay are more favored in apo 5c8* than in the complex.

The steady state emission spectra shown in Figure 4a for samples excited at the absorbance band of the protonated coumarin ($\lambda_{\text{ex}} = 323 \text{ nm}$) exhibit clear shoulders at $\sim 380 \text{ nm}$ indicating minor populations of excited state species emitting from the protonated form. In each sample, substituting H_2O for D_2O decreases the ratio of 450 nm to 380 nm emission (Figure S4), which is a steady-state manifestation of the ESPT kinetic isotope effect. Taking the ratio of the emission band intensities at 450 nm and 380 nm in the presence of H_2O over the same ratio in the presence of D_2O suggests an isotope effect of ~ 2.5 for the 7-HCAA amino acid, 1.5 for apo 5c8* and 1.3 for 5c8*/CD40L complex. The magnitude of this effect is similar that reported for 7-HC (~ 3).¹¹

Crystal structures of 5c8 variants and the 5c8*/CD40L complex

To gain additional insight into how local environments can affect the emission intensities of 7-HCAA^{L98}, we determined the crystal structures of apo 5c8* at pH 5.5 and pH 9.7 and a structure of the CD40L-bound state at pH 6.8. Furthermore, to assess the degree of perturbation caused by incorporation of a fNCAA within a protein, we solved the structure of the unbound, wild-type 5c8 protein (5c8-WT), which has not previously been described. Data collection and refinement statistics for the models are presented in Table S1.

Apo structures

5c8-WT and 5c8* crystallized isomorphously in space group C2 with one Fab protomer per asymmetric unit in conditions containing 0.1 M citrate buffer with pH values from 3-4. Structures solved from crystals in this pH range indicated significant disorder in the Fab CDR loops. However, the crystals remained stable when equilibrated to higher pH in a variety of buffers. A 1.77 Å dataset of 5c8-WT was collected at pH 5.5 and two datasets of uncomplexed 5c8* protein were collected at pH values of 5.5 (1.45 Å resolution) and 9.7 (1.75 Å resolution). In the structure solved at pH 5.5, clear electron density corresponding to 7-HCAA is discernable at position ^{L98} (Figure 5a). Superimposition of the apo 5c8-WT variable domain at pH 5.5 with apo 5c8* at pH 5.5 indicates that the increased bulk of the 7-HCAA^{L98} side chain is primarily accommodated through reorganization of the side chains of 5c8* in the vicinity of the NCAA (Figure 6). However, a 1.8 Å displacement of CDR L3, the CDR loop containing 7-HCAA, appears necessary to accommodate both the increased bulk of the 7-HCAA side chain and the interaction of the 7-HCAA phenol and the carboxylate of Glu^{H50} (Figure 6).

For the apo 5c8* at pH 5.5, the carboxylate of Glu^{H50} is observed in two discrete conformations (Figures 5b and 7). In the primary conformation (55%), the carboxylate oxygen of Glu^{H50} is within 2.5 Å of the coumarin phenol and exhibits good hydrogen bonding geometry. Meanwhile the minor conformation (45%) forms a salt bridge to the neighboring Arg^{H102} residue, which is modeled in a single conformation and appears to form a cation- π interaction²⁵ with the 7-HCAA^{L98} phenol ring. On the opposite side of the coumarin moiety, the carbonyl oxygen of the pyrone ring accepts a hydrogen bond from a well-ordered water molecule, which in turn donates a hydrogen bond to the main chain carbonyl of Asn^{H103} and accepts a hydrogen from the phenolic oxygen of Tyr^{H35} (Figure 5b). The orientation of the 7-HCAA^{L98} side chain is further stabilized through van der Waals interactions with the side chains of Tyr^{H35}, Trp^{H47}, Pro^{L100}, Asn^{H59} and Arg^{H102}.

These structural data suggest that 7-HCAA can be placed within a functional site of a protein in a minimally disruptive fashion and that both hydrophobic packing and hydrogen bonding interactions can be used to stabilize its orientation.

At pH 9.7, the coumarin ring is disordered and the final $2F_o - F_c$ map contains no significant electron density for this residue past the C β carbon (Figure S5). However, the main chain atom positions within the CDR L3 loop remain relatively unperturbed by the pH change (root-mean-square deviation of 0.23 Å for 20 equivalent atoms). Presumably the local disordering results from a repulsive electrostatic interaction between the negatively charged coumarin phenolate and Glu^{H50}.

5c8*/CD40L complex structure at pH 6.8

A structure of the 5c8*/CD40L complex was solved at 1.82 Å resolution and revealed three 5c8* Fabs bound symmetrically around the CD40L trimer, which is consistent with the previously reported wild-type 5c8/CD40L complex structure²⁶ (PDB entry 1i9r). Comparison of the 5c8*/CD40L structure to the wild-type complex structure indicated that substitution of Ile^{L98} with 7-HCAA has minimal effect on the contacts between 5c8 and CD40L. Unlike in the previously reported structure of 5c8/CD40L,²⁶ the electron density maps for the 5c8*/CD40L complex are sufficiently detailed to model an N-linked glycan at position Asn240 in CD40L, though the glycan appears to exist in two distinct conformations. Interactions between the glycan and the Fab are observed in this structure and these interactions as well as the conformational flexibility of the glycan are described in detail in the Supporting Information and illustrated in Figures S6–S8.

Electron density for the 5c8*/CD40L complex is also sufficiently detailed to discern the 7-HCAA^{L98} ethylene linker and the pyrone ring for each monomer within the complex (Figure 5c), while density for the distal phenol ring is present but generally weaker. 7-HCAA^{L98} is observed in two rotameric conformations, modeled at equal occupancies. Moreover, the density comprising the coumarin rings in either rotamer deviates noticeably from planarity suggesting flexibility about χ_3 . Presumably the shape of the cleft where the 7-HCAA^{L98} resides does not stabilize a discreet conformation of the coumarin side chain.

Comparison of the structures of 5c8* in the apo and bound forms suggests that binding of 5c8* to CD40L dramatically alters the position of 7-HCAA^{L98} observed in the apo structure at pH 5.5 with an average displacement of >8.5 Å for the atoms of the coumarin ring. This displacement is the result of a “domino effect” of steric clashes upon complexation (Figure 8). Briefly, CD40L residues Ala130 and His249 occupy positions in the complex that sterically occlude 5c8* residues Tyr^{H33} and Arg^{H102} from their apo conformations. In response to these clashes, Tyr^{H33} moves to a conformation that occludes the Glu^{H50} apo positions, while the 7-HCAA^{L98} apo position is occluded by correlated repositioning of Arg^{H102} and Glu^{H50}. As a result of these rearrangements, the coumarin ring in its two conformations makes van der Waals contacts with residues Asn^{H59}, Phe^{H60}, Asn^{H61} and Glu^{H62} of the heavy chain and CD40L residues Ile204, Thr251 and the N-linked glycan. Additional polar contacts are observed with residues Arg207 of CD40L, Asn^{H59}, Asn^{H61} and Glu^{H62} (Figure 5d and Figure S8).

Structural Basis for Spectroscopic Properties

The crystal structures of 5c8* provide plausible explanations for the spectroscopic data. pH-dependent changes to the 7-HCAA absorbance spectra fit a theoretical model of two interconverting species yielding apparent pK_a s that increase in acidity from 5c8*, to the 5c8*/CD40L complex and finally the free amino acid. A hydrogen bond between the 7-HCAA^{L98} phenol oxygen and the carboxylate of Glu^{H50}, observed in the apo 5c8* structure at pH 5.5 (Figures 5b and 7), presumably stabilizes the coumarin in its neutral form, thus increasing the pK_a relative to the free amino acid (Table 1). A conformational change upon complexation places this phenol, regardless of rotamer, such that it is pointing away from the cleft that surrounds the coumarin ring and towards the acetyl group of the glycan's second *N*-acetylglucosamine (Figures S6–S8). It seems possible that the proximity and flexibility of the CD40L Asn240 N-linked glycan, for which electron density is also weak, might partially occlude the access of water molecules to the coumarin's phenol. Thus, the local environment of the phenol oxygen is likely to be less polar than bulk solvent and could explain the increase in apparent pK_a for the complex compared to the free amino acid (Table 1).

Relative fluorescence quantum yield determinations (Table 1 and Figure 4) suggest that the enhancement of 7-HCAA^{L98} emission upon binding to CD40L¹³ is a consequence of fluorescence being more attenuated in the unbound state than in the complex. From this perspective, the unusually short hydrogen bond between the coumarin phenol and Glu^{H50} appears to be a major contributor to the changes in fluorescence. This interaction dissolves with antigen binding (Figures 5b,d and 8) and within the unbound form itself upon increasing the pH above the apparent pK_a (Figure S5). The fact that both instances are accompanied by increases in quantum yield (Table 1) lend credence to the notion that the 7-HCAA^{L98} environment observed at pH 5.5 in the apo 5c8* structure represents one in which non-radiative excitation decay is competitive.

Electron density maps for Glu^{H50} indicate disorder within this residue at pH 5.5 (Figure 7), while no indication of disorder was found for 7-HCAA^{L98}. Nevertheless, this observation might belie heterogeneity in the ground state electronic configuration of the protonated 7-HCAA^{L98}. The primary Glu^{H50} conformation orients the carboxylate relative to the 7-HCAA^{L98} phenol with good geometry for hydrogen bonding/proton transfer. On the contrary, the secondary Glu^{H50} conformation is not well positioned for hydrogen bonding or proton transfer and instead forms a salt-bridge with Arg^{H102}.

Though no direct evidence is available, Glu^{H50} seems to be the most probable acceptor for the photodissociated proton in the apo structure. The extent to which the proton remains on Glu^{H50} during the lifetime of the fluorescent state, as well as any other structural rearrangements that might occur in the interval, is conjecture. However, if Glu^{H50} remains protonated throughout the excited-state lifetime, this could result in a shift in dipole from Glu^{H50}/Arg^{H102} in the ground state to deprotonated 7-HCAA^{L98}/Arg^{H102} in the excited-state. Moreover, the proximity of the carboxylate side chain presents the possibility that backward proton transfer from Glu^{H50} to 7-HCAA^{L98} in the apo state might well be more efficient than from solvent, as presumably occurs in the 5c8*/CD40L complex.

It has been proposed that quenching by recombination of geminate protons can contribute to fluorescence inefficiencies in photoacids such as 1-naphthol,^{27,28} other hydroxyarenes,^{29,30} and the GFP variant S65T/H148D.³¹ An interesting parallel exists between the structures of apo 5c8* at low pH and GFP S65T/H148D, where the native proton transfer pathway has been short circuited through the introduction of the H148D mutation. Similar to the 7-HCAAL98-GluH50 interaction reported here, the structure of GFP S65T/H148D variant³² reveals a very short ($< 2.4 \text{ \AA}$) hydrogen bond between the chromophore hydroxyl and Asp148 that has been credited with conferring the unusual photophysical properties observed in this variant.

To further explore the contribution of GluH50 to fluorescence inefficiencies within 5c8*, we made the isosteric EH50Q mutation. Measurement of the apparent pK_a for the EH50Q variant revealed a value similar to (but slightly lower than) that of 5c8* (Table 1 and Figure S9). If GlnH50 occupies a conformation similar to that of the parent GluH50 side chain, interaction between the 7-HCAAL98 phenol and the GlnH50 side chain oxygen might have an effect similar to that of the carboxylate of GluH50 in increasing the phenolic pK_a . Such a situation would, at minimum, require rearrangement of the ArgH102 side chain conformation from the position observed in apo 5c8*, as the amide side chain of GlnH50 is incompatible with the salt bridge that is observed between GluH50 and ArgH102 in the parent structure (Figures 5b and 7). We also measured relative quantum yields for 7-HCAA in the anionic and neutral forms of the EH50Q mutant (Table 1). The relative quantum yield of neutral 7-HCAA in the EH50Q was slightly increased relative to that of the parent protein, while a dramatic quenching effect was observed for the anion. The quenching of 7-HCAA at high pH in the EH50Q mutant suggests that unlike the parent protein, the 7-HC side chain in the EH50Q mutant is not solvent-exposed and disordered, but likely adopts a distinct conformation.

Unfortunately, gel filtration analysis indicated that the 5c8* EH50Q mutant does not form a complex with CD40L (Figure S10), which renders spectroscopic characterization of the complex infeasible. Additionally, we have been unable to obtain structural information of apo EH50Q despite repeated attempts. Given our inability to both analyze the EH50Q mutant in the complexed form or obtain its apo structure, it is difficult to speculate how reorganization of the environment surrounding the 7-HCAAL98 residue in this mutant gives rise to the observed data.

Conclusions

It is probable that fNCAAs have not been as widely used as other fluorescent tools because of the limited characterization of proteins in which they have been incorporated. The physicochemical and structural data reported here serve to elucidate a number of features of 7-HCAA that may prove useful for the development of novel fluorescent sensors of protein function. Spectroscopic data indicate that the protein environment can modulate the protonation state of a phenol contained within the 7-HC side chain. Furthermore, the first structural studies of 7-HCAA within a protein provided insight into local structural perturbations that may occur in response to its incorporation; only minimal structural rearrangements were observed in this case. The structural analyses further suggest that

hydrogen bonding interactions between the 7-HC phenol and acidic amino acids may favor non-radiative pathways for relaxation from the excited state and that radiative relaxation can be biased upon disruption of this interaction.

The spectroscopic and structural characterization reported here suggest two ways of altering the fluorescent properties of 7-HCAA that could be used to develop novel sensors. First, protein environments or the presence of an analyte could be used to predictably alter the pK_a of the 7-HCAA phenol. For example, placement of 7-HCAA in the vicinity of acidic functional groups would favor the neutral form of 7-HCAA, while basic functional groups would favor the anion. Alternatively, the excited state dynamics of 7-HCAA could be directly influenced through the introduction or removal of hydrogen bonding interactions to its phenol. While this approach requires greater precision, recently developed computational protein design techniques that include NCAs in the design process^{33–35} could potentially be applied for this purpose.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Peter Schultz for the generous gift of the pEVOL-CouRS synthetase plasmid. The BCSB at the Advanced Light Source is supported in part by the National Institutes of Health (NIH), the National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the U.S. Department of Energy (DOE) under DE-AC02-05CH11231. The SBC (BL 19-ID) of the Advanced Photon Source is a DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the DOE Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institutes of General Medical Sciences (P41GM103393).

Funding

Research reported in this publication was also supported by NIGMS of the National Institutes of Health under award no. 1 R01 GM136996 01 to J.H.M. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS or NIH.

ABBREVIATIONS

fNCAA	fluorescent noncanonical amino acid
7-HCAA	L-(7-hydroxycoumarin-4-yl)ethylglycine
7-HC	7-hydroxycoumarin
ESPT	excited state proton transfer
Fab	fragment antigen-binding
pK_a	acid dissociation constant
λ_{ex}	excitation wavelength

OD	optical density
CDR	complementarity-determining region
PDB	protein data bank
aaRS	amino acyl t-RNA synthetase
ESI-MS	electrospray ionization mass spectrometry
IEC	ion exchange chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MES	2-(N-morpholino)ethanesulfonic acid
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
PEG	polyethylene glycol
PEG MME	poly(ethylene glycol) monomethyl ether
TMANO	trimethylamine N-oxide

REFERENCES

- (1). Moerner WE (2006) Single-molecule mountains yield nanoscale cell images. *Nat. Methods.* 3, 781–782. [PubMed: 16990808]
- (2). Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, and Hess HF (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313, 1642–1645. [PubMed: 16902090]
- (3). Hell SW (2007) Far-field optical nanoscopy. *Science* 316, 1153–1158. [PubMed: 17525330]
- (4). Giepmans BNG, Adams SR, Ellisman MH, and Tsien RY (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312, 217–224. [PubMed: 16614209]
- (5). Speight LC, Samanta M, and Petersson EJ (2014) Minimalist approaches to protein labelling: Getting the most fluorescent bang for your steric buck. *Aust. J. Chem.* 67, 686–700.
- (6). Lee HS, Guo J, Lemke EA, Dimla RD, and Schultz PG (2009) Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* 131, 12921–12923. [PubMed: 19702307]
- (7). Summerer D, Chen S, Wu N, Deiters A, Chin JW, and Schultz PG (2006) A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci.* 103, 9785–9789. [PubMed: 16785423]
- (8). Speight LC, Muthusamy AK, Goldberg JM, Warner JB, Wissner RF, Willi TS, Woodman BF, Mehl RA, and Petersson EJ (2013) Efficient Synthesis and In Vivo Incorporation of Acridon-2-ylalanine, a Fluorescent Amino Acid for Lifetime and Förster Resonance Energy Transfer/Luminescence Resonance Energy Transfer Studies. *J. Am. Chem. Soc.* 135, 18806–18814. [PubMed: 24303933]
- (9). Wang J, Xie J, and Schultz PG (2006) A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* 128, 8738–8739. [PubMed: 16819861]
- (10). Bartok W, Lucchesi PJ, and Snider NS (1962) Protolytic Dissociation of Electronically Excited Organic Acids. *J. Am. Chem. Soc.* 84, 1842–1844.
- (11). Simkovitch R, and Huppert D (2015) Photoprotolytic Processes of Umbelliferone and Proposed Function in Resistance to Fungal Infection. *J. Phys. Chem. B* 119, 14683–14696. [PubMed: 26488429]

- (12). Dean SF, Whalen KL, and Spies MA (2015) Biosynthesis of a novel glutamate racemase containing a site-specific 7-hydroxycoumarin amino acid: Enzyme-ligand promiscuity revealed at the atomistic level. *ACS Cent. Sci.* 1, 364–373. [PubMed: 26539562]
- (13). Mills JH, Lee HS, Liu CC, Wang J, and Schultz PG (2009) A genetically encoded direct sensor of antibody-antigen interactions. *ChemBioChem.* 10, 2162–2164. [PubMed: 19672911]
- (14). Lacey VK, Parrish AR, Han S, Shen Z, Briggs SP, Ma Y, and Wang L (2011) A Fluorescent Reporter of the Phosphorylation Status of the Substrate Protein STAT3. *Angew. Chemie Int. Ed.* 50, 8692–8696.
- (15). Nannenga BL, and Baneyx F (2011) Reprogramming chaperone pathways to improve membrane protein expression in *Escherichia coli*. *Protein Sci.* 20, 1411–1420. [PubMed: 21633988]
- (16). Chin JW, Santoro SW, Martin AB, King DS, Wang L, and Schultz PG (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* 124, 9026–9027. [PubMed: 12148987]
- (17). Otwinowski Z, and Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- (18). Battye TGG, Kontogiannis L, Johnson O, Powell HR, and Leslie AGW (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 271–281. [PubMed: 21460445]
- (19). McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674. [PubMed: 19461840]
- (20). Murshudov GN, Vagin AA, and Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 53, 240–255. [PubMed: 15299926]
- (21). Emsley P, and Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2126–2132. [PubMed: 15572765]
- (22). Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, and Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 235–242. [PubMed: 21460441]
- (23). Ueno K (1985) Structures of umbelliferone and 7-ethoxycoumarin. *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* 41, 1786–1789.
- (24). The PyMOL Molecular Graphics System Version 1.8 Schrödinger LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.
- (25). Gallivan JP, and Dougherty DA (1999) Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9459–9464. [PubMed: 10449714]
- (26). Karpusas M, Lucci J, Ferrant J, Benjamin C, Taylor FR, Strauch K, Garber E, and Hsu Y-M (2001) Structure of CD40 Ligand in Complex with the Fab Fragment of a Neutralizing Humanized Antibody. *Structure* 9, 321–329. [PubMed: 11525169]
- (27). Harris CM (1980) Acid-base properties of 1-naphthol. I Proton-induced fluorescence quenching. *J. Phys. Chem.* 84, 1366–1371.
- (28). Pines E, and Fleming GR (1994) Self quenching of 1-naphthol. Connection between time-resolved and steady-state measurements. *Chem. Phys.* 183, 393–402.
- (29). Pines E, Huppert D, and Agmon N (1988) Geminate recombination in excited-state proton transfer reactions: Numerical solution of the Debye-Smoluchowski equation with backreaction and comparison with experimental results. *J. Chem. Phys.* 88, 5620–5630.
- (30). Tolbert LM, and Solntsev KM (2002) Excited-state proton transfer: From constrained systems to “super” photoacids to superfast proton transfer. *Acc. Chem. Res.* 35, 19–27. [PubMed: 11790085]
- (31). Leiderman P, Genosar L, Huppert D, Shu X, Remington SJ, Solntsev KM, and Tolbert LM (2007) Ultrafast excited-state dynamics in the green fluorescent protein variant S65T/H148D. 3. Short- and long-time dynamics of the excited-state proton transfer. *Biochemistry* 46, 12026–12036. [PubMed: 17918961]

- (32). Shu X, Kallio K, Shi X, Abbyad P, Kanchanawong P, Childs W, Boxer SG, and Remington SJ (2007) Ultrafast excited-state dynamics in the green fluorescent protein variant S65T/H148D. 1. Mutagenesis and structural studies. *Biochemistry* 46, 12005–12013. [PubMed: 17918959]
- (33). Mills JH, Sheffler W, Ener ME, Almhjell PJ, Oberdorfer G, Pereira JH, Parmeggiani F, Sankaran B, Zwart PH, and Baker D (2016) Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl. Acad. Sci. U. S. A.* 113, 15012–15017. [PubMed: 27940918]
- (34). Pearson AD, Mills JH, Song Y, Nasertorabi F, Han GW, Baker D, Stevens RC, and Schultz PG (2015) Trapping a transition state in a computationally designed protein bottle. *Science* 347, 863–867. [PubMed: 25700516]
- (35). Mills JH, Khare SD, Bolduc JM, Forouhar F, Mulligan VK, Lew S, Seetharaman J, Tong L, Stoddard BL, and Baker D (2013) Computational design of an unnatural amino acid dependent metalloprotein with atomic level accuracy. *J. Am. Chem. Soc.* 135, 13393–13399. [PubMed: 23924187]

Figure 1.
Phenol-phenolate acid-base equilibrium of 7-HCAA.

Figure 2.

Graphical representations of the site of 7-HCAA incorporation in 5c8. The crystal structure of the 5c8 Fab and CD40L is shown in panel a in the uncomplexed and bound forms. The chemical structure of 7-HCAA has been superimposed on the proteins at the site of incorporation and is indicated in each case with a blue circle. Panel b shows a zoomed in view of the 5c8-CD40L interface; the site of 7-HCAA incorporation (Ile^L-98) is colored blue. The structure of PDB entry 1i9r was used to generate all images in this figure

Figure 3. 7-HCAA apparent pK_a determinations for the free amino acid (red circles) and for apo (blue squares) or CD40L-complexed 5c8* (green triangles). Curve fits are shown as solid lines of the corresponding color. (a) Plot of the absorbance at 371 nm divided by the 326 nm absorbance as a function of pH. (b) Normalized 450 nm fluorescence emission from 7-HCAA excited at 373 nm and plotted versus pH.

Figure 4. Fluorescence emission spectra demonstrating the relative quantum yields of apo 5c8* (red), the 5c8*/CD40L complex (blue) and the 7-HCAA amino acid (black). (a) $\lambda_{\text{ex}} = 323 \text{ nm}$; $\text{OD}_{323} = 0.05$. (b) $\lambda_{\text{ex}} = 373 \text{ nm}$; $\text{OD}_{373} = 0.05$.

Figure 5.

Structural overview of 5c8* and its CD40L complex. (a) 1.45 Å F_o - F_c omit map of the apo state of 5c8 solved at pH 5.5. The 7-HCAA^L98 side chain is contoured at 5 σ . (b) Stereo image of apo 5c8* coumarin environment. (c) 1.82 Å F_o - F_c omit map of the 5c8*/CD40L complex solved at pH 6.8. The map is contoured at 2 σ . (d) Stereo image of the 5c8*/CD40L coumarin environment with the Fab orientation identical to panel b. Amino acid residues are depicted with ball-and-stick representation of only the side chains with the exception of Asn^H59 where the main chain carbonyl is likewise displayed. Dashed lines indicated hydrogen bonds and the red spheres represent ordered water molecules. The apo state is colored blue. Yellow represents 5c8* in the complex with CD40L, which is colored gray and orange denotes the alternate conformation of the 7-HCAA^L98 side chain in the 5c8*/CD40L complex

Figure 6. Structural superimposition of apo 5c8 structures. The wild-type Fab is colored red and the 5c8* variant is colored blue. The double-headed arrow denotes structural rearrangements within the CDR L3 loop that accompany the introduction of the 7-HCAA^L98 mutation.

Figure 7.
Stereo view of the apo 5c8*coumarin environment illustrating the disorder observed in the Glu^H50 side chain. The $2F_o - F_c$ map is contoured at 1σ .

Figure 8.

Structural rearrangements accompanying complexation of 5c8* with CD40L. The apo structure (blue) is superimposed with the complexed Fab (yellow) and CD40L is shown in gray. The alternate 7-HCAAL98 side conformation within the complex is shown in orange. Red dashed lines indicate steric clashes that occur between atoms of the apo and complex structures. For clarity, only the primary conformation of GluH50 is displayed in the apo

structure. The curved arrow denotes 7-HCAAL98 movement in response to a series of steric clashes that occur upon complexation with CD40L.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Apparent pK_a s and Relative Fluorescence Quantum Yields for 7-HCAA Incorporated into 5c8* and Free in solution.

Sample	pK_a^a	pK_a^b	pK_a^c	$\Phi_{F,323ex}^d$	$\Phi_{F,373ex}^d$
Apo 5c8*	8.77 ± 0.04	8.50 ± 0.06	8.64 ± 0.07	0.36	0.93
5c8*/CD40L	8.58 ± 0.05	8.38 ± 0.06	8.48 ± 0.08	0.60	0.95
Apo 5c8* E ^H 50Q	8.66 ± 0.22	8.43 ± 0.06	8.55 ± 0.23	0.42	0.62
7-HCAA	8.07 ± 0.02	7.70 ± 0.02	7.89 ± 0.03	1.0	1.0

^a Apparent pK_a determined from absorbance spectra.

^b Apparent pK_a determined from fluorescence emission spectra with excitation at 373 nm.

^c Average of apparent pK_a s determined from absorbance and fluorescence emission spectra.

^d Quantum yields relative to the 7-HCAA neutral ($\Phi_{F,323ex}$) or anionic ($\Phi_{F,373ex}$) species, respectively.