JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JACS

Article

The Magic of Linking Rings: Discovery of a Unique Photoinduced Fluorescent Protein Crosslink

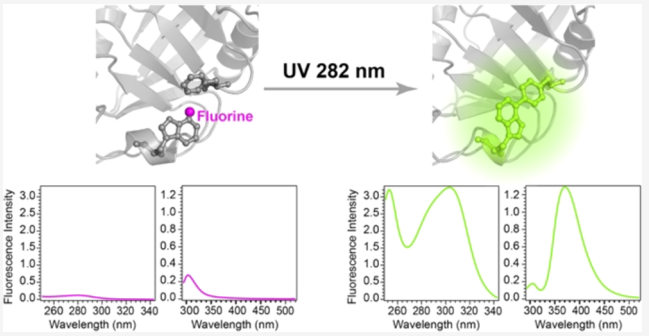
Manman Lu, Dmitri Toptygin, Yufei Xiang, Yi Shi, Charles D. Schwieters, Emma C. Lipinski, Jinwoo Ahn, In-Ja L. Byeon, and Angela M. Gronenborn*

Cite This: J. Am. Chem. Soc. 2022, 144, 10809−10816

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fluorosubstituted tryptophans serve as valuable probes for fluorescence and nuclear magnetic resonance (NMR) studies of proteins. Here, we describe an unusual photoreactivity introduced by replacing the single tryptophan in cyclophilin A with 7-fluoro-tryptophan. UV exposure at 282 nm defluorinates 7-fluoro-tryptophan and crosslinks it to a nearby phenylalanine, generating a bright fluorophore. The crosslink-containing fluorescent protein possesses a large quantum yield of ∼0.40 with a fluorescence lifetime of 2.38 ns. The chemical nature of the crosslink and the three-dimensional protein structure were determined by mass spectrometry and NMR spectroscopy. To the best of our knowledge, this is the first report of a Phe−Trp crosslink in a protein. Our finding may break new ground for developing novel fluorescence probes and for devising new strategies to exploit aromatic crosslinks in proteins.



INTRODUCTION

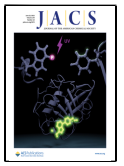
Aromatic amino acids—tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe)—serve as intrinsic fluorescence probes in proteins, offering valuable information on structure and dynamics. Among these three, tryptophan has the largest quantum yield and has been widely used to study protein folding, conformational changes upon ligand binding, and reaction kinetics. Over the last decades, extensive efforts have been devoted to the development of Trp analogues that exhibit improved photophysical properties without appreciably perturbing protein structure. These include azatryptophans,[1−3] cyanotryptophans,[4−6] fluorotryptophans,[7−10] and other indole substituted analogues.[11−13] Singly fluorinated tryptophans, such as 4F-, 5F-, 6F-, and 7F-Trp, with a fluorine atom located at different indole ring positions, display unique fluorescent properties.[7] While 5F- and 6F-Trp possess quantum yields similar to the natural Trp,[11] 4F- and 7F-Trp exhibit significantly reduced quantum yields and negligible fluorescence at room temperature.[7,14,15] Due to the favorable magnetic properties of the ^{19}F isotope, fluorotryptophans are also excellent NMR reporters. Among the different fluorosubstituted tryptophans, 5F-Trp is most widely used and has been exploited as a dual fluorescence and NMR probe.[16,17] 7F-Trp, on the other hand, is less well studied and its photophysical properties were reported only recently.[14,18]

Here, we describe a highly unusual photoreactivity generated by the presence of 7-fluoro-tryptophan in the cyclophilin A (CypA) protein. We made the surprising discovery that the nonfluorescent 7F-Trp, when incorporated into CypA, undergoes a photoinduced defluorination to generate a highly fluorescent fluorophore. We characterized its fluorescence properties, identified its chemical nature, and determined the three-dimensional structure of the modified protein by NMR spectroscopy and mass spectrometry (MS), unveiling an unusual, hitherto unknown Phe−Trp crosslink.

RESULTS

Discovery of a Highly Fluorescent CypA Protein. CypA is a peptidyl prolyl cis−trans isomerase, possessing a single tryptophan (W121) at its active site. We previously demonstrated that different fluorotryptophans can be successfully incorporated into CypA with high efficiency via biosynthesis using fluorosubstituted indoles as precursors.[19] Here, we describe and characterize the unusual properties of CypA with a single fluorine atom at the 7 position of the W121 indole ring (Figure 1A,B). 7F-Trp, as a free amino acid, exhibits dramatically reduced quantum yield compared to nonmodified Trp and is virtually nonfluorescent at room temperature in an aqueous buffer, as observed by us and

Received: February 22, 2022
Published: May 14, 2022



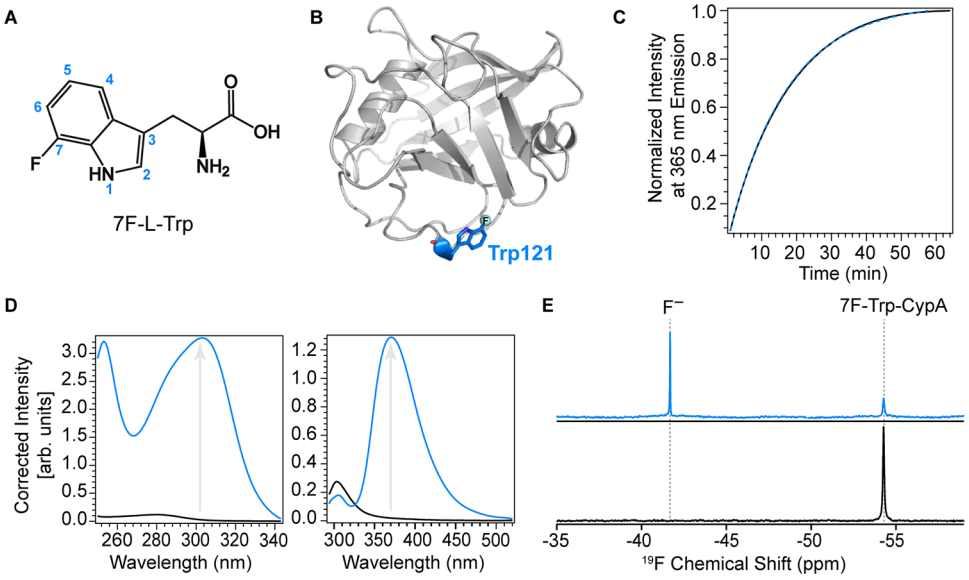


Figure 1. Photoactivation of 7F-Trp-CypA. (A) Chemical structure of 7F-L-Trp. The numbering of the ring atoms is shown in blue. (B) The crystal structure of 7F-Trp-CypA in ribbon representation. Residue W121 is colored blue and shown in stick representation. The fluorine atom is shown as a light cyan sphere. (C) Time dependence of photoactivation determined by fluorescence. Experimental emission intensity (solid black line) was fit to a single-exponential decay function (blue dashed line). (D) Excitation (left) and emission (right) spectra of 7F-Trp-CypA before (black) and after (blue) photoactivation. Excitation spectra were measured with an emission wavelength of 350 nm and emission spectra with an excitation wavelength of 288 nm. Note the significant increase in fluorescence intensities after UV activation, as indicated by the gray arrow. (E) 1D ^{19}F NMR spectra of 7F-Trp-CypA before (black) and after (blue) photoactivation.

others[14] (Figure S1). As a result, not surprisingly, incorporation of 7F-L-Trp into CypA (7F-Trp-CypA) leads to almost complete loss of the intrinsic protein fluorescence, resulting in negligible fluorescence in both excitation and emission spectra in the canonical wavelength range for tryptophan (Figure 1D, black lines). Similar nonfluorescent features have been observed with 4F-Trp, both in the free amino acid as well as when incorporated into proteins.[7,15,20] By silencing the tryptophan fluorescence, tyrosine fluorescence at an emission wavelength (λ_{em}) of 303 nm can be observed (Figure 1D, black lines).

Remarkably, the nonfluorescent 7F-Trp-CypA is photoactivatable, with an unusual highly fluorescent fluorophore being generated by exposure to UV light. To probe the kinetics of the reaction, we performed time-dependent fluorescence experiments, illuminating with light at the excitation wavelength (λ_{ex}) of 282 nm and recording fluorescence emission spectra, monitoring the intensity at 365 nm. The time dependence of the fluorescence intensity follows a single-exponential function, implying a first-order reaction with a rate constant of 0.060 ± 0.001 min^{-1} (Figure 1C). Completion of photoactivation results in a highly fluorescent fluorophore, exhibiting a red-shifted excitation peak at 304 nm with an emission maximum at 373 nm, strikingly different from nonmodified Trp (Figures 1D and S1).

Since small molecule 4-haloindoles can undergo dehalogenation upon light irradiation via photoinduced nucleophilic displacement,[21] we explored whether 7F-Trp-CypA may have lost the fluorine atom during our fluorescence experiments. We collected a series of one-dimensional (1D) ^{19}F NMR spectra for 7F-Trp-CypA prior to and after UV radiation. A substantial decrease in the peak intensity of the initial ^{19}F resonance at −54.26 ppm occurs after UV radiation, and a new resonance at −41.71 ppm appears, consistent with the chemical shift of a free fluoride, providing direct evidence of photoinduced defluorination (Figure 1E). However, defluorination of the indole ring alone cannot account for the unusual high fluorescence, necessitating a search for its origin.

Identification of a Novel Phe−Trp Crosslink in CypA. After UV activation, the modified CypA protein retains its overall native fold, as judged by NMR, since its ^{1}H-^{15}N HSQC spectrum is well dispersed and resembles to a large degree the spectrum before UV radiation (Figure 2A). However, new resonances emerged, and, after close examination, more than one amide resonance per residue appeared to be present for a number of amino acids, suggesting that the sample, after UV radiation, contains a mixture of several species (Figure 2A,B). For example, residues that are located close to Trp121 in the CypA structure, including 60F, 62C, 120E, 121W, and 122L, all exhibit four amide resonances (Figure 2B). One of the species is clearly the initial 7F-Trp-CypA, as evidenced by identical resonance frequencies in the ^{1}H-^{15}N HSQC and 1D ^{19}F spectra before and after photoactivation (Figures 1E and 2A). The new resonances are associated with species that are only present after UV radiation, with one or several of these modified proteins possibly responsible for the unusual bright fluorescence.

In our search toward identifying the chemical nature of the fluorescent moiety, we considered that in the excited state Nε1 (N1) and Cγ (C3) of Trp may become more positively charged, while the Cε3 (C4) and Cζ2 (C7) may become significantly more negatively charged.[22,23] Thus, we hypothesized that the resulting increased electron density at C7 in the excited state promotes the highly electronegative F7 to leave via heterolytic C−F bond cleavage, resulting in a highly reactive C7 position, an ideal target for nucleophilic attack. In aqueous solution, in which we performed the fluorescence experiments, a water molecule may serve as the nucleophile and attack at C7, generating 7-hydroxy-tryptophan (7OH-Trp). Indeed, we experimentally verified this possibility in the

https://doi.org/10.1021/jacs.2c02054
J. Am. Chem. Soc. 2022, 144, 10809−10816

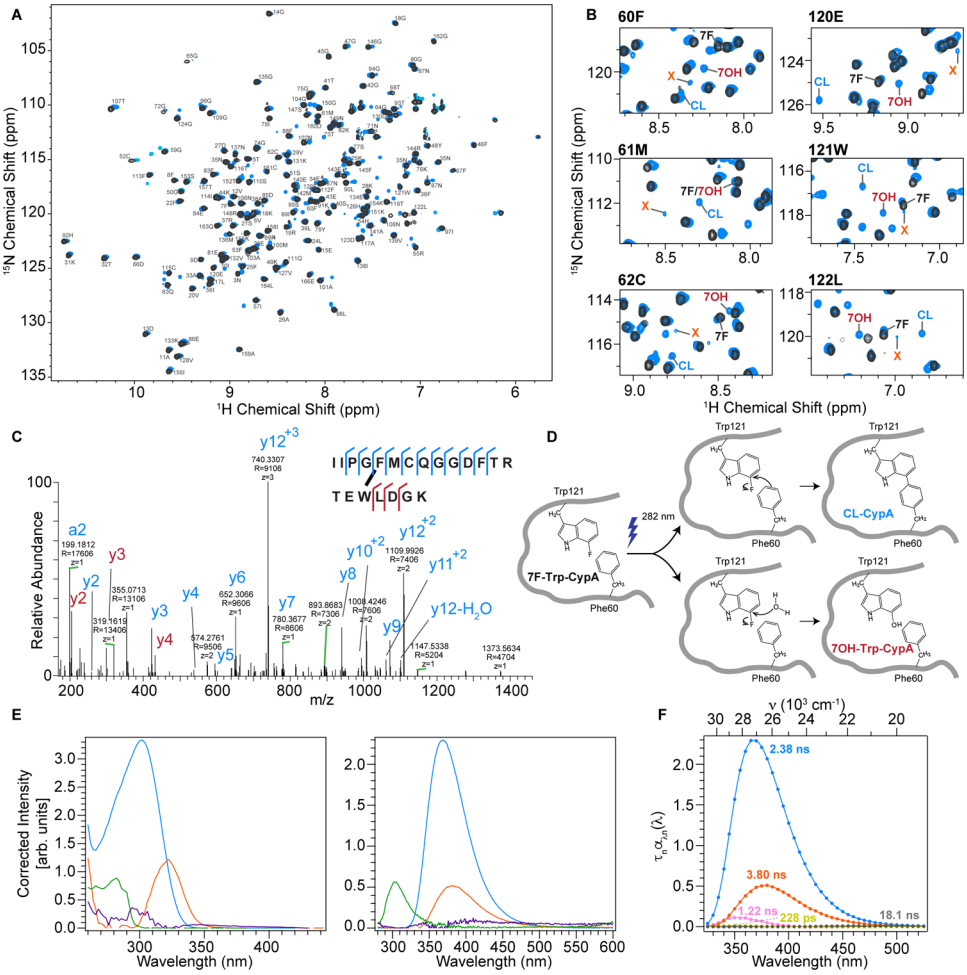


Figure 2. Identification of crosslinked CypA. (A) Superposition of ^{1}H-^{15}N HSQC spectra of U-^{15}N, 7F-Trp- CypA before (black) and after (blue) UV radiation. Backbone amide ^{1}H-^{15}N cross peaks are labeled by residue number and amino acid type. (B) Expansions of ^{1}H-^{15}N HSQC spectral regions around 60F, 61M, 62C, 120E, 121W, and 122L resonances. The four resonances for each position are labeled with CL for the crosslinked CypA and 7F, 7OH, and X for 7F-Trp-CypA, 7-hydroxy-Trp-CypA, and an unknown minor species, respectively. (C) MS/MS spectrum of the crosslinked peptide. The majority of the fragments were identified and mapped. (D) Proposed model for UV activation of 7F-Trp-CypA, resulting in crosslinked CypA (CL-CypA, upper pathway in blue) and 7-hydroxy-Trp-CypA (7OH-Trp-CypA, lower pathway in maroon). (E) Fluorescence excitation (left) and emission (right) spectra derived from SVD of the excitation-emission matrix measured for enriched/purified CL-CypA. Four distinct fluorophores (I–IV) were identified and are colored in blue, orange, green, and purple, respectively. (F) Corrected emission intensity spectra associated with the five exponential terms determined in TCSPC for the enriched/purified sample. Spectra corresponding to lifetimes of 2.38 ns, 3.80 ns, 1.22 ns, 288 ps, and 18.1 ns are colored in blue, orange, pink, sand, and gray, respectively. Intensities obtained by renormalizing $\tau_n\alpha_n$ values using steady-state emission intensities are shown in filled circles. Smooth lines were generated using degree 8 polynomials.

context of the free amino acid by NMR spectroscopy, where 7OH-Trp is clearly detected in 7F-Trp solution after UV radiation (Figure S2A). Since 7OH-Trp is nonfluorescent in aqueous solution at room temperature (Figure S2B), it is likely that 7OH-Trp-CypA is also nonfluorescent and therefore cannot be the source for the unusual high fluorescence. Other candidate nucleophiles in CypA are any electron-rich moieties in the vicinity of C7, such as the aromatic ring of Phe60. Nucleophilic attack by the phenylalanine side chain could generate a crosslink between the Phe and Trp side chains in CypA, resulting in a double-ring extended π orbital system. Such an unusual molecular arrangement could be a plausible origin of the UV-induced high fluorescence.

To test our crosslink hypothesis, we performed tryptic digests on samples of 7F-Trp-CypA prior to and after photoactivation, followed by nanoflow liquid chromatography and high-resolution mass spectrometry (nLC-MS) analysis.[24,25] Comparison of LC/MS spectra before and after UV radiation unambiguously identified a unique crosslinked peptide (56I-69R crosslinked to 119T-125K) solely in the sample after radiation (Figure S3). In addition, tandem mass spectrometry (MS/MS) yielded a high-resolution fragmentation spectrum for this crosslinked peptide and allowed us to map the majority of the fragment ions, identifying the exact crosslinked positions to 60 and 121 (Figure 2C). These findings provide direct and convincing evidence to confirm our crosslink hypothesis. Therefore, we propose the following model for photoactivation of 7F-Trp-CypA, illustrated in Figure 2D. Upon UV radiation at 282 nm, the 7F-indole ring of Trp 121 in CypA in its excited state undergoes photoinduced nucleophilic attack by the nearby phenyl group in Phe60 or/and a water molecule, yielding crosslinked (CL)-CypA and 7OH-Trp-CypA, respectively. In addition, a minor CypA species is also formed during UV radiation, denoted as X-Trp-CypA. The chemical nature of X-Trp-CypA remains unknown because its existence in low amounts in a

https://doi.org/10.1021/jacs.2c02054
J. Am. Chem. Soc. 2022, 144, 10809−10816

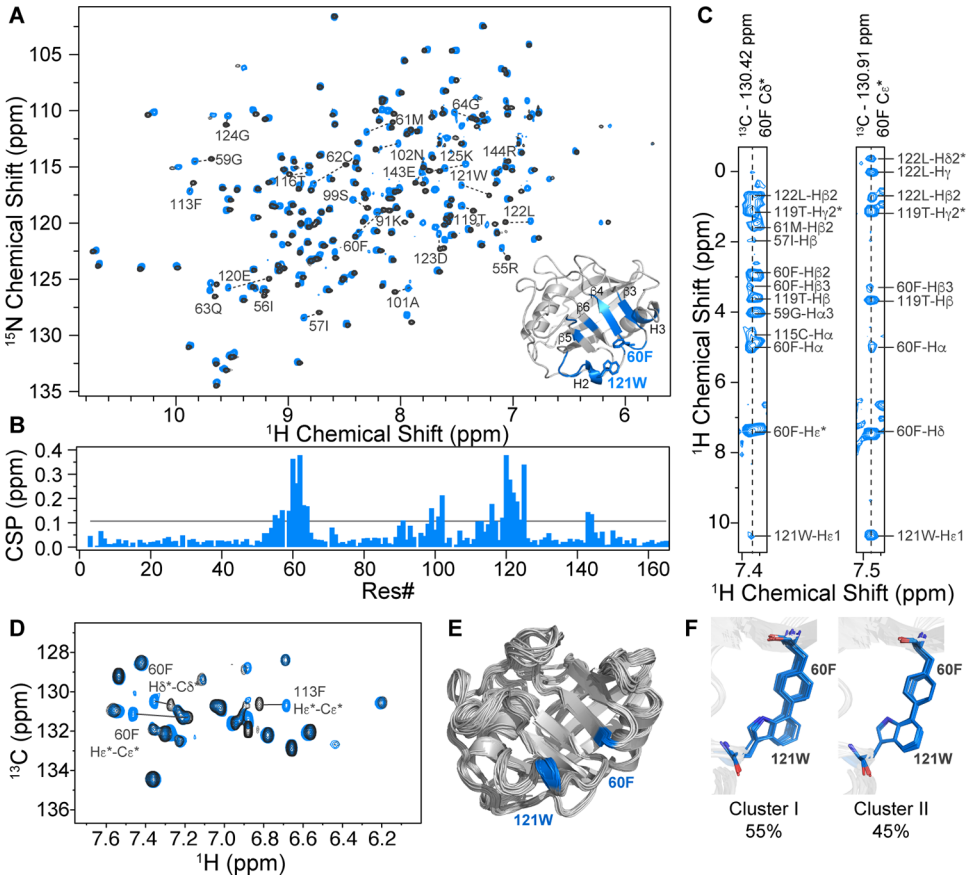


Figure 3. Structure of crosslinked CypA. (A) Superposition of 2D ^{1}H-^{15}N HSQC spectra of U-^{13}C, ^{15}N, 7F-Trp-CypA (black) and enriched/purified U-^{13}C, ^{15}N-CL-CypA (blue). Residues with chemical shift changes larger than 0.107 ppm (1.5 times the standard deviation) are labeled in the spectra by dashed lines and mapped onto the CypA structure (PDB 3K0N). (B) Weighted ^{1}H and ^{15}N chemical shift differences between 7F-Trp-CypA and CL-CypA plotted vs residue number. The weighting factors were 1.00 and 0.14 for ^{1}H and ^{15}N shifts, respectively. The horizontal line at 0.107 ppm represents 1.5 times the standard deviation. (C) Representative strips extracted from the 3D ^{1}H-^{13}C aromatic NOESY-HSQC spectrum of U-^{13}C, ^{15}N-CL-CypA, depicting NOE contacts for Hδ* and Hε* atoms of the F60 residue. (D) Superposition of the 2D ^{1}H-^{13}C aromatic TROSY spectra of U-^{13}C, ^{15}N, 7F-Trp-CypA (black) and enriched/purified U-^{13}C, ^{15}N-CL-CypA (blue). (E) Final ensemble of the 20 lowest-energy structures of CL-CypA shown in ribbon representation. Positions 60 and 121 are colored in blue. (F) The final ensemble possesses two conformations for the crosslinked side chains, cluster I (11 members) and cluster II (9 members). The Trp121-Phe60 motif is shown in blue stick representation.

complex mixture makes detailed structure elucidation difficult. Estimated amounts of 7F-Trp-CypA, 7OH-Trp-CypA, CL-CypA, and X-Trp-CypA in the final mixture are ~26, 32, 29, and 14%, respectively, based on resonance intensities in the ^{1}H-^{15}N HSQC spectrum (Table S1). This is in excellent agreement with the ESI-MS data of the sample after UV irradiation, which contains a new peak at 18256.7 Da, about 20 Da less than the original peak at 18276.2 Da of unmodified 7F-Trp-CypA (Figures S4A and S4B). This observed mass is consistent with the predicted mass for CL-CypA. The masses of 7F-Trp-CypA and 7OH-Trp-CypA differ by only 2 Da and are unresolved in the experimental spectrum, with a single peak at 18275.0 Da (Figure S4B).

Isolation of Crosslinked CypA (CL-CypA). Since UV irradiation resulted in a mixture of CypA variants for all assessed conditions, it was necessary to develop an approach for purifying CL-CypA from the reaction mixture to pursue a more detailed characterization of its properties. To that end, we exploited the binding of wild-type (WT) CypA to the HIV-1 capsid (CA) protein.[26] CypA interacts with the CA protein in solution in a 1:1 stoichiometry, albeit with only 10−20 mM affinity.[27,28] Fortuitously, the two crosslinking amino acids, 60F and 121W, are located in the CA-binding site of CypA, facing each other and guarding the entrance to the binding pocket.[29,30] Based on a structural model,[29,30] we inferred that the crosslink at 60F-121W would completely block the entry of CA to the binding site and abolish the CA-CypA interaction, while minor modifications, such as fluoro- or hydroxy substitutions in the 121W side chain, may be tolerated and would not influence CA binding significantly. This drastic difference in CA binding was exploited to isolate CL-CypA from the UV-irradiated mixture. Adding an excess amount of CA protein resulted in the formation of CA/7F-Trp-CypA and CA/7OH-Trp-CypA complexes and permitted the separation and isolation of CL-CypA via size exclusion chromatography. Although it proved impossible to generate 100% pure CL-CypA, a significantly enriched sample was obtained, containing approximately 80 ± 5% CL-CypA, as demonstrated by ESI-MS (Figure S4C). Furthermore, the enriched/purified sample exhibited NMR spectra with only one set of intense resonances, those belonging to CL-CypA, whereas cross peaks associated with 7F-Trp-CypA, 7OH-Trp-CypA, and X-Trp-CypA were either missing or close to the noise level,

https://doi.org/10.1021/jacs.2c02054
J. Am. Chem. Soc. 2022, 144, 10809−10816

suggesting that the quantity of any of these CypA species is below 5% (Figure 3A).

Fluorescence Characterization. To determine the number of fluorophores in the photoactivated sample and to evaluate whether CL-CypA harbors the brightest fluorophore, we employed singular value decomposition (SVD) of the excitation−emission matrix.[31] Steady-state fluorescence intensity was measured for all combinations of 69 excitation and 321 emission wavelengths, generating a 321 × 69 matrix. SVD of this matrix yielded five singular components above the noise level in the unpurified UV-radiated mixture and four singular components above the noise level in the enriched/purified CL-CypA sample (for details, see the Supporting Information). To transition from the SVD basis spectra to the excitation and emission spectra of the actual fluorophores, we exploited the fact that the excitation and emission spectra cannot have negative intensities and developed an algorithm subject to this constraint (see the Supporting Information). Unique excitation and emission spectra of the fluorophores recovered by this algorithm are provided in Figures 2E and S5A. The brightest fluorophore (fluorophore I, blue trace), with an excitation peak at 302 nm and an emission peak at 369 nm, is the only fluorophore that displays increased fluorescence in the enriched/purified sample compared to the unpurified mixture (Table S2). We therefore assigned this fluorophore to the Phe−Trp crosslink in CL-CypA. For a second fluorophore (fluorophore II, orange trace), with excitation and emission peaks at 322 nm and 381 nm, respectively, the fluorescence is only slightly reduced in the enriched/purified sample, implying that this species is probably a modified product of CL-CypA or 7OH-Trp-CypA. In addition, the different CypA species in the UV-illuminated sample also exhibited tyrosine fluorescence (fluorophore III, green trace), with a characteristic excitation maximum at 282 nm and an emission peak at 303 nm. Notably, the measured tyrosine fluorescence intensities in the enriched/purified and unpurified samples are approximately proportional to the corresponding total protein concentrations (Table S2). Furthermore, a minor species (fluorophore IV, purple trace) with an excitation peak around 296 nm and an emission peak around 415 nm is seen, whose intensity is significantly less in the enriched/purified sample. A very minor red-emitting species (fluorophore V, gray trace), with an excitation peak around 366 nm and an emission peak around 474 nm, is only detected in the unpurified sample. These latter two minor species contribute very little to the overall total fluorescence signal and are barely above the noise level. (Figures 2E, S5A and Table S2).

Fluorescence decays were measured by time-correlated single photon counting (TCSPC) at 41 emission wavelengths for both the enriched/purified and the unpurified samples. Global fitting of all TCSPC data required five exponential terms for a statistically adequate fit. Emission spectra associated with individual exponential terms were obtained, as described previously.[32] The highest intensity spectrum is associated with a lifetime of 2.38 ± 0.01 ns and is attributed to the Phe−Trp crosslink in CL-CypA, given its emission peak wavelength of 368 nm and the increased relative fluorescence intensity in the enriched/purified sample (Figures 2F and S5B). Similarly, the spectrum associated with a lifetime of 3.80 ± 0.02 ns exhibits a characteristic peak emission wavelength of 378 nm and must originate from fluorophore II (Figures 2F and S5B). The spectra associated with the lifetimes of 1.22 ± 0.04 and 0.228 ± 0.011 ns (Figures 2F and S5B) most likely result from a time-dependent red shift (TDRS) in the emission spectra of the major fluorophores, Phe−Trp and fluorophore II. The nanosecond TDRS is possibly associated with the relaxation of the polar protein environment around the solvatochromic fluorophore, such as Trp, or its derivatives.[8] The red-shifted spectrum (emission peak at 474 nm) associated with the lifetime of 18.1 ± 0.39 ns resembles one of the spectra of fluorophore V obtained by SVD of the unpurified sample (Figure S5C), and it is probably due to a short-lived transient product generated during 7F-Trp photoactivation. Importantly, the emission spectra of CL-CypA, determined using two independent methodologies, SVD and TCSPC, and for two different samples, enriched/purified and unpurified, are in excellent agreement with one another (Figure S5D). Therefore, all fluorescence data are consistent with our observations by NMR and MS and together strongly support our proposed model (Figure 2D).

The quantum yield (η) of the enriched/purified CypA (80 ± 5% CL-CypA) in aqueous buffer at λ_{ex} = 305 nm is 0.395 ± 0.003 (see the Supporting Information). Since the small amounts of 7OH-Trp-CypA and 7F-Trp-CypA that are present also absorb light at 305 nm but have low quantum yields, we posit that for 100% pure CL-CypA the quantum yield is even greater than 0.395. Compared to tryptophan (η = 0.13) and phenylalanine (η = 0.024) alone,[33] the Phe−Trp crosslink in CypA clearly has a much higher quantum yield. This, together with the short lifetime (τ = 2.38 ns), is indicative of a high radiative decay rate ($k_r = \eta/\tau$) and, consequently, a large-magnitude transition dipole moment μ, possibly due to the extended π-orbital system in the Phe−Trp crosslink. Since both k_r and the extinction coefficient (ε) are directly proportional to $|\mu|^2$, the extinction coefficient of the Phe−Trp crosslink in CypA must be high, in excellent agreement with the high fluorescence intensity for CL-CypA.

Atomic-Resolution Structure of Crosslinked CypA. The enriched/purified CL-CypA sample exhibits NMR spectra with only one set of intense resonances for each residue, enabling high-resolution structure determination by NMR (Figure 3A). Most amide cross peaks in the 2D ^{1}H-^{15}N HSQC spectrum of CL-CypA exhibit very similar frequencies as those of 7F-Trp-CypA, implying a similar overall protein fold. However, for some cross peaks, differences were readily observed (Figure 3A,B), a few of which could not be confidently and unambiguously assigned solely based on similarities in the 2D spectra. We therefore acquired a complete set of 3D experiments, including HNCA, HNCO, HN(CO)CA, H(CCO)NH, and C(CO)NH spectra, on a U-^{13}C,^{15}N-labeled CL-CypA sample, to unequivocally assign both backbone and side chain resonances (BMRB 30977). Residues that are associated with chemical shift differences between CL-CypA and 7F-Trp-CypA 1.5 times larger than the standard deviation were mapped onto the structure of CypA protein (PDB ID 3K0N;[34] Figure 3A,B). Gratifyingly, the most strongly affected residues are located at or near residues 60F and 121W, where the crosslink occurs (Figure 3A,B). Additional resonances with notable differences belong to residues throughout the four-strand β-sheet structure (β3−β6), reaching out all the way to helix 3 (H3) (Figure 3A).

Although we were able to identify the Phe60-Trp121 crosslink by MS/MS, the identity of the bond connecting the two aromatic rings remained to be determined. This was achieved by NMR. The ring Cδ*, Cε*, Hδ*, and Hε* resonances of Phe60 were assigned using 2D ^{1}H-^{13}C aromatic

https://doi.org/10.1021/jacs.2c02054
J. Am. Chem. Soc. 2022, 144, 10809−10816

TROSY and 3D ^{1}H-^{13}C aromatic NOESY-HSQC spectra (Figure 3C,D), and since no aromatic proton resonance was connected to the same spin system, we established that the Cζ atom of the Phe60 aromatic ring has no hydrogen atom attached. This indicates that the Trp121 ring is crosslinked at the Cζ, the para-position of Phe60, with a covalent bond between C7 of Trp121 and Cζ of Phe60. This result is in line with expectations for an isolated phenyl ring, where the methylene group serves as an activator, promoting ortho- and para-substitutions.[35] In the present scenario, the ortho-substitution is unlikely due to steric considerations, leaving the para-position as the only favorable site for the crosslink.

The final NMR structure determination of CL-CypA required the generation of a special topology parameterization for the para-Phe−Trp crosslink, with a bond of 1.54 Å length between Cζ of Phe and C7 of Trp in Xplor-NIH.[36] The coordinates of the 7F-Trp-CypA crystal structure (unpublished data) were used as the initial model, edited with the newly defined para-Phe−Trp topology. A total of 2611 unambiguously assigned $^{1}H-^{1}H$ nuclear Overhauser effects (NOEs) were extracted from 3D ^{1}H-^{15}N NOESY-HSQC and 3D ^{1}H-^{13}C aromatic NOESY-HSQC spectra and used to derive distance restraints, divided into short-, medium-, and long-range bins. These distance restraints, together with 262 backbone dihedral angle restraints derived from chemical shifts using TALOS−N,[37] were used for structure calculation in Xplor-NIH (Table S3). The final ensemble of the 20 lowest-energy structures is well-defined and exhibits pairwise atomic backbone root-mean-square deviations (RMSDs) of 0.9 ± 0.2 Å (Figure 3E and Table S3). CL-CypA retains the overall structure of 7F-Trp-CypA or WT-CypA, as suggested from the similarities in the ^{1}H-^{15}N HSQC spectra. Pairwise backbone RMSD values between CL-CypA and 7F-Trp-CypA or CL-CypA and WT-CypA are 1.31 ± 0.11 and 1.29 ± 0.12 Å, respectively (Figure S6). Interestingly, two distinctive local orientations between the two linked aromatic rings were identified (Figure 3F). The first cluster represents 55% of the entire ensemble, with an average dihedral angle between the two aromatic rings of 37.8°, while the second cluster (45%) exhibits a dihedral angle of 23.9° (Figure 3F).

DISCUSSION

Fluorine, nearly absent from biology, has emerged as an attractive tool in protein engineering, not only acting as a background-free probe for structure and dynamics studies by NMR but also serving as a unique modification to fine-tune the properties of molecules.[38−40] Although fluorine-containing amino acids, especially Trps, are widely used probes in spectroscopy and are often assumed to cause minimal structural changes,[41,42] this may not always be the case. This is dramatically revealed here for 7F-Trp-CypA, which generated an unusual and potent photoreactivity. Indeed, during the course of our studies, aimed at extending our work beyond CypA, we observed the photoinduced formation of 4- or 7-hydroxy-Trp, as well as their corresponding photo-irradiated products, not only in the free 4F- and 7F-Trp amino acids but also in several proteins we investigated (Figures S2 and S7), highlighting that extra caution and careful analytical characterization are essential to ensure sample integrity when investigating 4F- or 7F-Trp-labeled proteins. However, the aromatic crosslink between the Phe−Trp side chain in CypA is a very specific and highly unusual chemical modification. Not only do the rings have to be close in space, but it is also very likely that a defined, explicit orientation is required for nucleophilic attack and C−C bond formation. In that regard, we investigated several proteins that were structurally and functionally studied in our laboratory over the last decades, such as the cyanobacterial *Oscillatoria Agardhii agglutinin* (OAA) and the dimer of the C-terminal domain of the HIV-1 capsid protein (CA CTD, a L151F variant), all of which possess Trp and Phe residues close in space in their three-dimensional structure, either naturally or strategically designed by mutagenesis. However, for none of these proteins could we generate a photoinduced Phe−Trp crosslink (Figure S7), implying that the precise local structural arrangement of the Phe and Trp side chains in CypA is unique and essential for linking these aromatic rings. Therefore, we examined proteins in the Protein Data Bank (PDB, www.wwpdb.org)[43] that possess Trp−Phe pairs in similar configurations as observed in CypA (see the Supporting Information). 589 potential candidates were carefully evaluated, and the *de novo*-designed protein PS1[44] was selected to evaluate whether our finding is more general and applicable to other proteins. PS1, a four-helix bundle, had been created as a small molecule binder, comprising a tightly packed folded core and an under-packed binding region.[44] It contains a single Trp (W68) in the flexible binding region with a Phe (F17) nearby. Although the W68-F17 pair in the apo state of PS1 (PDB ID 5TGW) adopts a suboptimal conformation, the holo-PS1 (PDB ID 5TGY) emerged as a top candidate from our PDB search. Its W68-F17 pair is positioned very similar with respect to each other as the W121-F60 pair in CypA (Figures S8A and S8B). We therefore hypothesized that the flexibility in the apo state binding region, without the bound ligand, may offer enough plasticity to permit W68 to crosslink to F17. We prepared U-^{15}N, 7F-Trp-apo-PS1 and, after photoactivation, observed a new fluorophore, exhibiting an excitation peak at 312 nm and an emission peak at 355 nm (Figures S8C and S8D), associated with a −20 Da mass change (Figures S8E and S8F). This implies that, indeed, a W69-F17 crosslinked PS1 had been generated. Furthermore, we explored whether a crosslink could be produced between Trp and a Tyr instead of a Phe. To this end, we prepared the F60Y CypA variant. Similar to WT-CypA, photoactivation of 7F-Trp F60Y CypA also resulted in a highly fluorescent species with an excitation peak at 311 nm and an emission peak at 388 nm, most likely originating from a Trp−Tyr crosslink (Figure S9A, S9B). This was corroborated by a loss of 20 Da in the mass spectrum of this modified CypA (Figures S9C and S9D). Interestingly, the Trp-Tyr crosslinked CypA is generated faster than its Trp−Phe counterpart, exhibiting a first-order rate constant of 0.103 ± 0.001 min^{-1} (Figure S9A).

CONCLUSIONS

The discovery of a highly fluorescent fluorophore generated by photoinduced crosslinking between Trp and Phe or Trp and Tyr side chains in a protein, associated with a single fluorine atom in the 7 position of the indole ring of Trp is unprecedented. It offers a novel and controllable approach for synthesizing and exploiting bright and modular designer proteins or protein tags. The fact that these highly fluorescent fluorophores can be generated in aqueous media under physiological conditions opens exciting avenues in a variety of fields. For example, it may be possible to design and place photoactivatable fluorescent probes into proteins in a highly selective manner for fluorescence studies using spectroscopy,

https://doi.org/10.1021/jacs.2c02054
J. Am. Chem. Soc. 2022, 144, 10809−10816

Förster resonance energy transfer, or fluorescence imaging. We also envision that it will be possible to improve the original Trp−Phe/ Trp−Tyr crosslink if the emission wavelengths can be shifted from 369/388 nm into visible range, enabling wide applications in microscopy.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c02054.

The supporting information includes materials and methods; additional fluorescence spectra of Trp, F-Trps, WT-CypA, and 7OH-Trp; nLC-MS chromatography and spectra of trypsin-digested 7F-Trp-CypA prior to and after UV radiation; ESI-MS spectra of 7F-Trp containing proteins before and after photoactivation; fluorescence characterization of 7F-Trp-CypA after photoactivation; superpositions of the structures of crosslinked CypA, 7F-Trp-CypA, and WT-CypA; photoactivation of free amino acids 4F- and 7F-Trp and 4F- and 7F-Trp containing proteins; summary of NMR restraints and structural statistics; singular values determined by the SVD method; and reduced global $\chi 2$ values obtained by simultaneously fitting the TCSPC data (PDF)

Accession Codes

The atomic coordinates and NMR restraints of crosslinked CypA have been deposited in the Protein Data Bank under accession codes PDB 7TA8. NMR chemical shifts of crosslinked CypA have been deposited in the Biological Magnetic Resonance Data Bank under accession code 30977.

■ AUTHOR INFORMATION

Corresponding Author

Angela M. Gronenborn − *Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0001-9072-3525; Email: amg100@pitt.edu

Authors

Manman Lu − *Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-4156-4975

Dmitri Toptygin − *Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, United States;* orcid.org/0000-0001-5800-861X

Yufei Xiang − *Department of Cell Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* Present Address: Center of Protein Engineering and Therapeutics, Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States

Yi Shi − *Department of Cell Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* Present Address: Center of Protein Engineering and Therapeutics, Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States

Charles D. Schwieters − *Computational Biomolecular Magnetic Resonance Core, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, United States*

Emma C. Lipinski − *Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States*

Jinwoo Ahn − *Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States*

In-Ja L. Byeon − *Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15261, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c02054

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Colin W. Wilburn, Kumaran Baskaran, and Jeffrey Hoch for generous help with the PDB search; all members of the Gronenborn laboratory for discussion; Doug Bevan for computer support; Michael J. Delk for maintenance of the NMR facility; and Teresa Brosenitsch for editorial assistance. This work was supported by the National Institutes of Health (NIH; P50AI150481 to A.M.G. and R35GM137905 to Y.S.) and the National Science Foundation (NSF; CHE-1708773 to A.M.G.). C.D.S. was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health.

■ REFERENCES

(1) Merkel, L.; Hoesl, M. G.; Albrecht, M.; Schmidt, A.; Budisa, N. Blue Fluorescent Amino Acids as in Vivo Building Blocks for Proteins. *ChemBioChem* **2010**, *11*, 305−314.

(2) Lepthien, S.; Hoesl, M. G.; Merkel, L.; Budisa, N. Azatryptophans Endow Proteins with Intrinsic Blue Fluorescence. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16095−16100.

(3) Smirnov, A. V.; English, D. S.; Rich, R. L.; Lane, J.; Teyton, L.; Schwabacher, A. W.; Luo, S.; Thornburg, R. W.; Petrich, J. W. Photophysics and Biological Applications of 7-Azaindole and Its Analogs. *J. Phys. Chem. B* **1997**, *101*, 2758−2769.

(4) Hilaire, M. R.; Ahmed, I. A.; Lin, C. W.; Jo, H.; DeGrado, W. F.; Gai, F. Blue Fluorescent Amino Acid for Biological Spectroscopy and Microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 6005−6009.

(5) Talukder, P.; Chen, S. X.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. Cyanotryptophans as Novel Fluorescent Probes for Studying Protein Conformational Changes and DNA-Protein Interaction. *Biochemistry* **2015**, *54*, 7457−7469.

(6) Markiewicz, B. N.; Mukherjee, D.; Troxler, T.; Gai, F. Utility of 5-Cyanotryptophan Fluorescence as a Sensitive Probe of Protein Hydration. *J. Phys. Chem. B* **2016**, *120*, 936−944.

(7) Minks, C.; Huber, R.; Moroder, L.; Budisa, N. Atomic Mutations at the Single Tryptophan Residue of Human Recombinant Annexin

V: Effects on Structure, Stability, and Activity. *Biochemistry* **1999**, *38*, 10649−10659.
(8) Toptygin, D.; Gronenborn, A. M.; Brand, L. Nanosecond Relaxation Dynamics of Protein GB1 Identified by the Time-Dependent Red Shift in the Fluorescence of Tryptophan and 5-Fluorotryptophan. *J. Phys. Chem. B* **2006**, *110*, 26292−26302.
(9) Liu, T.; Callis, P. R.; Hesp, B. H.; de Groot, M.; Buma, W. J.; Broos, J. Ionization Potentials of Fluoroindoles and the Origin of Nonexponential Tryptophan Fluorescence Decay in Proteins. *J. Am. Chem. Soc.* **2005**, *127*, 4104−4113.
(10) Broos, J.; Maddalena, F.; Hesp, B. H. In Vivo Synthesized Proteins with Monoexponential Fluorescence Decay Kinetics. *J. Am. Chem. Soc.* **2004**, *126*, 22−23.
(11) Lotte, K.; Plessow, R.; Brockhinke, A. Static and Time-Resolved Fluorescence Investigations of Tryptophan Analogues - a Solvent Study. *Photochem. Photobiol. Sci.* **2004**, *3*, 348−359.
(12) Alexander Ross, J. B.; Szabo, A. G.; Hogue, C. W. V. Enhancement of Protein Spectra with Tryptophan Analogs: Fluorescence Spectroscopy of Protein-Protein and Protein-Nucleic Acid Interactions. *Methods Enzymol.* **1997**, *278*, 151−190.
(13) Hogue, C. W. V.; Rasquinha, I.; Szabo, A. G.; Macmanus, J. P. A New Intrinsic Fluorescent-Probe for Proteins - Biosynthetic Incorporation of 5-Hydroxytryptophan into Oncomodulin. *FEBS Lett.* **1992**, *310*, 269−272.
(14) Nandy, T.; Mondal, S.; Singh, P. C. Fluorine Induced Conformational Switching and Modulation in Photophysical Properties of 7-Fluorotryptophan: Spectroscopic, Quantum Chemical Calculation and Molecular Dynamics Simulation Studies. *J. Photochem. Photobiol.* **2020**, *3*−*4*, No. 100011.
(15) Hott, J. L.; Borkman, R. F. The Non-Fluorescence of 4-Fluorotryptophan. *Biochem. J.* **1989**, *264*, 297−299.
(16) Winkler, G. R.; Harkins, S. B.; Lee, J. C.; Gray, H. B. Alpha-Synuclein Structures Probed by 5-Fluorotryptophan Fluorescence and ^{19}F NMR Spectroscopy. *J. Phys. Chem. B* **2006**, *110*, 7058−7061.
(17) Sun, X.; Dyson, H. J.; Wright, P. E. Fluorotryptophan Incorporation Modulates the Structure and Stability of Transthyretin in a Site-Specific Manner. *Biochemistry* **2017**, *56*, 5570−5581.
(18) Nandy, T.; Singh, P. C. Photophysical Properties of Noncanonical Amino Acid 7-Fluorotryptophan Sharply Different from Those of Canonical Derivative Tryptophan: Spectroscopic and Quantum Chemical Calculations. *J. Phys. Chem. B* **2021**, *125*, 6214−6221.
(19) Lu, M.; Ishima, R.; Polenova, T.; Gronenborn, A. M. ^{19}F NMR Relaxation Studies of Fluorosubstituted Tryptophans. *J. Biomol. NMR* **2019**, *73*, 401−409.
(20) Bronskill, P. M.; Wong, J. T. Suppression of Fluorescence of Tryptophan Residues in Proteins by Replacement with 4-Fluorotryptophan. *Biochem. J.* **1988**, *249*, 305−308.
(21) Yang, N. C. C.; Huang, A.; Yang, D. D. H. Photodehalogenation of 4-Haloindoles. *J. Am. Chem. Soc.* **1989**, *111*, 8060−8061.
(22) Toptygin, D.; Woolf, T. B.; Brand, L. Picosecond Protein Dynamics: The Origin of the Time-Dependent Spectral Shift in the Fluorescence of the Single Trp in the Protein GB1. *J. Phys. Chem. B* **2010**, *114*, 11323−11337.
(23) Muiño, P. L.; Callis, P. R. Hybrid Simulations of Solvation Effects on Electronic-Spectra - Indoles in Water. *J. Chem. Phys.* **1994**, *100*, 4093−4109.
(24) Xiang, Y. F.; Sang, Z.; Bitton, L.; Xu, J. Q.; Liu, Y.; Schneidman-Duhovny, D.; Shi, Y. Integrative Proteomics Identifies Thousands of Distinct, Multi-Epitope, and High-Affinity Nanobodies. *Cell Syst.* **2021**, *12*, 220−234.
(25) Shi, Y.; Pellarin, R.; Fridy, P. C.; Fernandez-Martinez, J.; Thompson, M. K.; Li, Y. Y.; Wang, Q. J.; Sali, A.; Rout, M. P.; Chait, B. T. A Strategy for Dissecting the Architectures of Native Macromolecular Assemblies. *Nat. Methods* **2015**, *12*, 1135−1138.
(26) Luban, J.; Bossolt, K. L.; Franke, E. K.; Kalpana, G. V.; Goff, S. P. Human-Immunodeficiency-Virus Type-1 Gag Protein Binds to Cyclophilin A and B. *Cell* **1993**, *73*, 1067−1078.
(27) Yoo, S. H.; Myszka, D. G.; Yeh, C. Y.; McMurray, M.; Hill, C. P.; Sundquist, W. I. Molecular Recognition in the HIV-1 Capsid/Cyclophilin A Complex. *J. Mol. Biol.* **1997**, *269*, 780−795.
(28) Schaller, T.; Ocwieja, K. E.; Rasaiyaah, J.; Price, A. J.; Brady, T. L.; Roth, S. L.; Hue, S.; Fletcher, A. J.; Lee, K.; KewalRamani, V. N.; Noursadeghi, M.; Jenner, R. G.; James, L. C.; Bushman, F. D.; Towers, G. J. HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. *PloS Pathog.* **2011**, *7*, No. e1002439.
(29) Dorfman, T.; Weimann, A.; Borsetti, A.; Walsh, C. T.; Gottlinger, H. G. Active-Site Residues of Cyclophilin A Are Crucial for Its Incorporation into Human Immunodeficiency Virus Type 1 Virions. *J. Virol.* **1997**, *71*, 7110−7113.
(30) Gamble, T. R.; Vajdos, F. F.; Yoo, S. H.; Worthylake, D. K.; Houseweart, M.; Sundquist, W. I.; Hill, C. P. Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid. *Cell* **1996**, *87*, 1285−1294.
(31) Toptygin, D.; Savtchenko, R. S.; Meadow, N. D.; Brand, L. Homogeneous Spectrally- and Time-Resolved Fluorescence Emission from Single-Tryptophan Mutants of IIA^{Glc} Protein. *J. Phys. Chem. B* **2001**, *105*, 2043−2055.
(32) Toptygin, D. Analysis of Time-Dependent Red Shifts in Fluorescence Emission from Tryptophan Residues in Proteins. *Methods Mol. Biol.* **2014**, *1076*, 215−256.
(33) Chen, R. F. Fluorescence Quantum Yields of Tryptophan and Tyrosine. *Anal. Lett.* **1967**, *1*, 35−42.
(34) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. Hidden Alternative Structures of Proline Isomerase Essential for Catalysis. *Nature* **2009**, *462*, 669−673.
(35) Clayden, J.; Greeves, N.; Warren, S., *Organic Chemistry*; Oxford University Press, 2012.
(36) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR Molecular Structure Determination Package. *J. Magn. Reson.* **2003**, *160*, 65−73.
(37) Shen, Y.; Bax, A. Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N. *Methods Mol. Biol.* **2015**, *1260*, 17−32.
(38) Salwiczek, M.; Nyakatura, E. K.; Gerling, U. I.; Ye, S.; Koksch, B. Fluorinated Amino Acids: Compatibility with Native Protein Structures and Effects on Protein-Protein Interactions. *Chem. Soc. Rev.* **2012**, *41*, 2135−2171.
(39) Swallow, S. Fluorine in Medicinal Chemistry. *Prog. Med. Chem.* **2015**, *54*, 65−133.
(40) Berger, A. A.; Voller, J. S.; Budisa, N.; Koksch, B. Deciphering the Fluorine Code-the Many Hats Fluorine Wears in a Protein Environment. *Acc. Chem. Res.* **2017**, *50*, 2093−2103.
(41) Tobola, F.; Lelimousin, M.; Varrot, A.; Gillon, E.; Darnhofer, B.; Blixt, O.; Birner-Gruenberger, R.; Imberty, A.; Wiltschi, B. Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *ACS Chem. Biol.* **2018**, *13*, 2211−2219.
(42) Welte, H.; Zhou, T.; Mihajlenko, X.; Mayans, O.; Kovermann, M. What Does Fluorine Do to a Protein? Thermodynamic, and Highly-Resolved Structural Insights into Fluorine-Labelled Variants of the Cold Shock Protein. *Sci. Rep.* **2020**, *10*, No. 2640.
(43) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10*, 980.
(44) Polizzi, N. F.; Wu, Y.; Lemmin, T.; Maxwell, A. M.; Zhang, S.-Q.; Rawson, J.; Beratan, D. N.; Therien, M. J.; DeGrado, W. F. De Novo Design of a Hyperstable Non-Natural Protein−Ligand Complex with Sub-Å Accuracy. *Nat. Chem.* **2017**, *9*, 1157−1164.

https://doi.org/10.1021/jacs.2c02054
J. Am. Chem. Soc. 2022, 144, 10809−10816