Supporting Information for

Site-specific Protein Photochemical Covalent Attachment to Carbon Nanotube Side Walls and its Electronic Impact on Single Molecule Function.

Suzanne K. Thomas[†a], W. David Jamieson[$a], Rebecca Gwyther[‡a], Ben J, Bowen[‡a], Harley L. Worthy[‡a], Adam Beachey[†], J. Emyr Macdonald[†], Martin Elliott[†*], Oliver K Castell[$*], & D. Dafydd Jones[‡*].

[a] Joint first authors.

† School of Physics and Astronomy, Cardiff University, Cardiff, UK. CF24 3AA.

‡ School of Biosciences, Cardiff University, Cardiff, UK. CF10 3AX

$ School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff, UK. CF10 3NB

* Corresponding authors. DDJ, jonesdd@cardiffi.ac.uk; OKC, castello@cardiff.ac.uk; ME, martin.elliott@astro.cf.ac.uk

Supporting Information

Supporting Methods

Computational design of azF Computational design of azF variants. To model the incorporation of p-azido-L-phenylalanine in the target protein at a desired position, the residue was first made de novo in Avagadro [1]. The azF structure was then optimised and partial charges calculated using RESP ESP charge Derive (RED) software [2-3] combined with GAMESS-US [4]. Bond angle and dihedral parameters were calculated with the ACPYPE pipeline [5]. The parameterised PyMOL pickle file (pymol residue file) was incorporated into PyMOL via the Swiss-side chains plugin [6]. PyMOL was then used to mutate each of the residues individually. To ensure there were no structural clashes, models we subjected to energy minimisation and a brief (~100 ns) molecular dynamics run using GROMACS [7]. Parameters generated from RED and ACPYPE were used to modify the AMBER 99sb forcefield to tolerate the incorporation of azF.

Protein engineering. The GFP mutants used for this paper were all created using whole plasmid PCR, as described previously [8]. The TEM^{87azF} and cyt $b_{562}{}^{50azF}$ mutants were constructed using the primers sets given below in Table S1. PCR was performed with NEB Phusion polymerase according to manufacturer's directions. The primers used to generate the TAG mutations and the calculated Tm for the primers are listed in Table S1. PCR products were then purified using the QIAgen PCR purification kit. Purified PCR products were recircularised with T4 PNK and Quick ligase (both NEB). DNA sequences were verified by sequencing provided by Eurofins Genomics and are presented in the Supporting Information.

Table S1. Primers used to create amber stop codons mutants with amber stop codon highlighted in **red**. Tm's for each primer pair are in final column

Mutant	Template Plasmid	Primers (Forward – Reverse) 5' -3'
GFP^{132azfF} [a]	pBAD-sfGFP WT	**TAG** GAT GGC AAC ATT CTG GGT CAT AAC TGG TTT AAA ATC AAT ACC TTT CAG TTC AAT GCG GTT C
GFP^{204azF} [a]	pBAD-sfGFP WT	**TAG** AGC GTT CTG AGC AAA GAT CCG AAT G GGT GCT CAG ATA ATG ATT ATC CGG CAG CA
cyt $b_{562}{}^{50azF}$	pBAD-cyt$_{b562}$	ACA GCC CGG AAA GAC GCG TTA AC CCG GTG ATT T**CT A**TT CGA GCT TCG
BL^{87azF}	pET24a-TEM [b]	CGT GTT GAC GCC **TAG** CAA GAG CAA CT GG ATA ATA CCG CGC CAC ATA GCA GAA CTT
BBP^{41azF}	pET-BLIP-II [c]	TAT ACA TAT G**TA G**AC CTC GGT GGT G TCT CCT TCT TAA AGT TAA ACA AAA TTA TTT CTA

[a] First created by Reddington et al [9]. [b] First reported by Sosa-Peinado et al [10] and kindly donated by Tim Palzkill [11-12]. [c] Kindly donated by Tim Palzkill [13]

Determined gene sequences of the variants used in this study. The TAG mutations are highlighted in **red**.

> pBAD-sfGFP 132TAG
ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGTGATGTGAATGGCCATA
AATTTAGCGTTCGTGGCGAAGGCGAAGGTGATGCGACCAACGGTAAACTGACCCTGAAATTTATTTGCACCACCGGTAA
ACTGCCGGTTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGCGTTCAGTGCTTTAGCCGCTATCCGGATCATATG
AAACGCCATGATTTCTTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGTACCATTAGCTTCAAAGATGATGGCA
CCTATAAAACCCGTGCTGAAGTTAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGTATTGATTTTAA
A**TAG**GATGGCAACATTCTGGGTCATAAACTGGAATATAATTTCAACAGCCATAATGTGTATATTACCGCCGATAAACAG
AAAAATGGCATCAAAGCGAACTTTAAAATCCGTCACAACGTGGAAGATGGTAGCGTGCAGCTGGCGGATCATTATCAGC
AGAATACCCCGATTGGTGATGGCCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACCCAGAGCGTTCTGAGCAAAGA
TCCGAATGAAAAACGTGATCATATGGTGCTGCTGGAATTTGTTACCGCCGCGGGCATTACCCACGGTATGGATGAACTG
TATAAAGGCAGCCACCATCATCATCACCATTAA

> pBAD-sfGFP 204TAG
ATGGTTAGCAAAGGTGAAGAACTGTTTACCGGCGTTGTGCCGATTCTGGTGGAACTGGATGGTGATGTGAATGGCCATA
AATTTAGCGTTCGTGGCGAAGGCGAAGGTGATGCGACCAACGGTAAACTGACCCTGAAATTTATTTGCACCACCGGTAA
ACTGCCGGTTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGCGTTCAGTGCTTTAGCCGCTATCCGGATCATATG
AAACGCCATGATTTCTTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGTACCATTAGCTTCAAAGATGATGGCA
CCTATAAAACCCGTGCTGAAGTTAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGTATTGATTTTAA
AGAAGATGGCAACATTCTGGGTCATAAACTGGAATATAATTTCAACAGCCATAATGTGTATATTACCGCCGATAAACAG
AAAAATGGCATCAAAGCGAACTTTAAAATCCGTCACAACGTGGAAGATGGTAGCGTGCAGCTGGCGGATCATTATCAGC
AGAATACCCCGATTGGTGATGGCCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACC**TAG**AGCGTTCTGAGCAAAGA
TCCGAATGAAAAACGTGATCATATGGTGCTGCTGGAATTTGTTACCGCCGCGGGCATTACCCACGGTATGGATGAACTG
TATAAAGGCAGCCACCATCATCATCACCATTAA

>pET24a-TEM 87TAG
ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCCACCCAGAAACGCTGG
TGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCT
TGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGT
GTTGACGCC**TAG**CAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAG
AAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA
CTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT
GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAA
CGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAA
AGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGG
CAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAA

> pBAD-cyt b_{562} 50TAG
ATGGCAGATCTTGAAGACAATATGGAAACCCTCAACGACAATTTAAAAGTGATCGAAAAAGCGTGTAACGCGGCGCAAG
TCAAAGACGCGTTAACGAAGATGCGCGCCGCAGCGCTGGATGCGCAAAAAGCAACGCCGCCGAAGCTCGAA**TAG**AAATC
ACCGGACAGCCCGGAAATGAAAGATTTCCGCCACGGTTTCGACATTCTGGTCGGTCAGATTGACGACGCGCTGAAGCTG
GCAAATGAAGGTAAAGTAAAAGAAGCGCAGGCTGCTGCAGAGCAACTGAAAACGACCCGCAACGCCTATCACCAGAAGT
ATCGTTAA

>pET-BLIP-II 41TAG
ATG**TAG**ACCTCGGTGGTGGCCTGGGGCGGTAACAATGACTGGGGTGAAGCTACCGTGCCGGCCGAAGCGCAGAGTGGTG
TGGATGCAATTGCAGGTGGTTATTTTCATGGGCTGGCACTGAAAGGGGGTAAAGTACTGGGCTGGGGTGCAAATCTGAA
CGGGCAGCTGACAATGCCGGCGGCGACCCAGAGCGGCGTTGATGCTATCGCGGCGGGCAATTATCACTCTCTGGCTCTG
AAAGATGGGGAAGTGATTGCTTGGGGCGGTAACGAGGATGGCCAAACTACGGTGCCGGCCGAGGCCCGTTCCGGTGTAG
ATGCTATTGCGGCAGGCGCTTGGGCGAGCTACGCGCTGAAAGACGGCAAAGTGATCGCCTGGGGTGATGATTCCGACGG
TCAGACCACCGTGCCGGCGGAAGCCCAGTCGGGTGTGACCGCGCTGGATGGTGGTGTGTATACCGCGCTGGCAGTAAAA
AACGGTGGTGTTATTGCGTGGGGGGATAATTATTTTGGCCAGACCACAGTGCCGGCGGAGGCTCAGTCCGGGGTGGATG
ATGTTGCAGGCGGCATCTTTCACAGCCTGGCGCTGAAAGATGGTAAAGTTATTGCGTGGGGCGATAATCGCTATAAACA
AACCACAGTTCCAACCGAGGCGCTGAGTGGCGTGTCGGCCATTGCTTCAGGTGAATGGTATAGCCTGGCTCTGAAAAAT
GGTAAAGTAATTGCGTGGGGTAGCAGCCGCACCGCGCCTAGCTCCGTCCAATCGGGGGTGAGTTCCATTGAAGCCGGTC
CGAACGCCGCTTACGCACTGAAAGGTGGGAGCGGTTCTGGCCATCATCACCATCATCATTAA

Protein Expression and purification. The pBAD-based expression plasmids housing the GFP and cyt b_{562} TAG mutants were co transformed with pDULE-azF [14] into *E. coli* Top10 cells (ThermoFisher) and spread onto LB Agar plates supplemented with carbenicillin and tetracycline (50 and 25 µg/mL respectively). Single colonies were used to inoculated 1L cultures of arabinose auto-induction media [15] supplemented with 50 µg/mL carbenicillin and 25 µg/mL tetracycline. Expression cultures were grown overnight at 37 °C. One hour after inoculation cultures were supplemented with *p*-azido-L-phenylalanine to a final concentration of 1 mM.

The pET-based expression plasmids housing the BL (TEM β-lactamase) and BBP (BLIP-II) TAG mutants were co transformed with pEVOL-pAzF [16] into *E. coli* BL21 (DE3) cells (New England Biolabs) and spread onto LB Agar plates supplemented with kanamycin and chloramphenicol (30 and 35 µg/mL respectively). Single colonies were used to inoculate 250 mL cultures in 2×YT media supplemented with 30 µg/mL kanamycin and 35 µg/mL chloramphenicol. Expression cultures were grown at 37°C until reaching an OD_{600} of 0.4 then induced with 1 mM IPTG and 0.1% w/v arabinose. Expression cultures were then grown for 24 hours at 18 °C.

From this point cultures were kept in the dark to prevent premature photolysis of the azide group. Cells were harvested via centrifugation at 5000 x *g* for 20 mins. Cultures expressing proteins variants were resuspended in 50 mM Tris-HCl, pH8.0 (20 mL per litre of expression culture). Purity of protein samples at each stage were assessed by SDS-PAGE.

For GFP, cyt b_{562} and BBP variants, resuspended cell mixtures were lysed using a French Press exerting 20,000 psi of pressure. Soluble protein was harvested by centrifugation at 25,000 x g for 30 mins. For GFP variants, lysates were first purified by nickel affinity chromatography, as described previously [8-9]. Samples were loaded onto a 5 mL HisTrap HP Column (GE Healthcare), equilibrated in binding buffer (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole pH8.0), using an ÄKTAprime chromatography system. The column was then washed with 20 column volumes of binding buffer before being eluted in a single step with 250 mM imidazole. Eluted protein was concentrated to ~2.5 mL using a Vivaspin 10 kDa MWCO spin concentrator (GE Healthcare) and desalted into 50 mM Tris HCl (pH 8.0) using a PD10 desalting column (GE Healthcare) according to the manufacturer's instructions. Purity of protein was assessed by SDS-PAGE. The cyt b_{562}^{50azF} [17-18] was purified by procedures described previously.

Periplasmic contents containing wt BL or the BL^{87azF} variant were extracted from *E. coli* by cold osmotic shock then desalted into 50 mM Tris HCl (pH 8.0) as above. The BL sample

was then loaded onto a 5mL HiTrap Q HP anion exchange column (GE) and eluted over a 0-100 mM NaCl gradient using an ÄKTApurifier FPLC system. The BL sample was further purified and desalted by size exclusion chromatography using a Superdex 200 10/300 GL column (GE) with 50 mM Tris HCl (pH 8.0).

Cell lysates containing histidine-tagged BBP^{41azF} were first purified by cobalt affinity chromatography using a home-made gravity column containing 5mL HisPur cobalt resin (Thermofisher) equilibrated in binding buffer (50 mM Tris-HCl, 5 mM imidazole pH8.0). The column was then washed with 10 column volumes of binding buffer before being eluted in a single step with 20mL 100 mM imidazole, then concentrated and desalted as above.

SWCNT preparation and protein attachment. A solution of single wall carbon nanotubes was prepared from powder (Sigma Aldrich: single walled diameter 0.7-1.4nm, carbon >90%, ≥77% carbon as SWCNT). The powder was dispersed into a solution 1% w/v SDS (Sigma Aldrich, molecular biology grade). The solution was sonicated using a probe sonicator in bursts of 10 seconds for five iterations, followed by 30 seconds for three iterations and finally 60 seconds for five iterations. After this treatment the grains of SWCNT can be visually confirmed to have decreased in density, thus implying a good dispersion within the solution. The SWCNT solution was then divided into 750µl aliquots and placed into a bench top microfuge for 30 minutes at 14,000rpm. Once a pellet had formed at the bottom of the solution, the supernatant was removed and used for subsequent preparations.
The prepared SWCNT solution (20 µl) was drop cast onto the surface of a plasma treated glass coverslip and allowed to dry for 30 minutes. Any solution remaining on the surface was washed away under flowing DI water followed by a 10 second wash with flowing ethanol (Fischer Chemicals, analytical reagent grade). The coverslip was then submerged in ethanol for 1 hour to remove any surfactant or unbound tubes from the surface. Finally, the coverslip was given a final rinse with flowing DI water to remove the ethanol, and dried under nitrogen stream.
In order to bind protein to the carbon nanotubes on the coverslip substrate, samples were prepared within a nitrogen-flushed glove box to minimise the effect of oxidation. A humidity monitor was used to verify that the glove box was thoroughly flushed. The SWCNT samples were placed on a specially designed stage within a small (also nitrogen flushed) lock box inside the glove box. This allowed the 305 nm LED mounted in the lid of the lock box to be positioned directly above the sample. Protein solution (20 µl of 10-100 nM samples in 50 mM

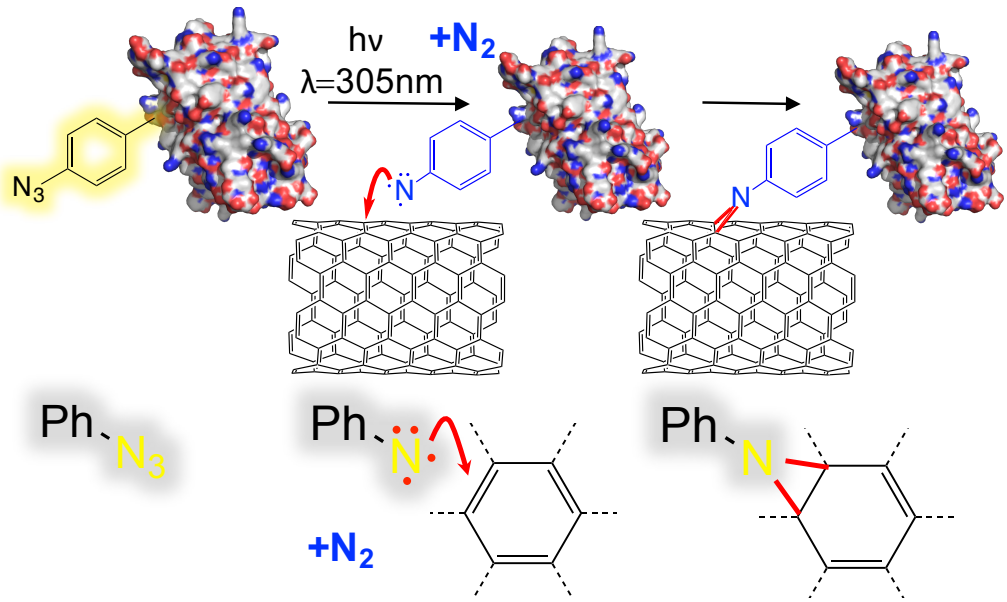
Tris HCl pH 8) was drop cast onto the surface of the sample and the lock box closed. For illumination experiments, the 305 nm LED with an intensity of $18Wm^{-2}$ was switched on and the sample left for 5 minutes. The sample was then removed from the stage and rigorously washed for 30 seconds under flowing DI water. Control samples were included that did not undergo illumination prior to washing. The sample was finally dried and removed from the low humidity for mounting and imaging.

AFM imaging. Samples were studied using tapping mode AFM (Digital Instruments Multimode SPM - Nanoscope IIIa with Nanoscope basic extender module) to produce topographical images in which the heights and distribution of the bound proteins were analysed. The images acquired were analysed in Gwyddion image analysis software (http://gwyddion.net) [19]. Here, the cross sections of the proteins on top of the SWCNTs were recorded and compared with the height of the SWCNT it was bound to by finding the difference between their bi-lateral minima. An average height and standard deviation in height was acquired for each protein variant by analysing large numbers of protein structures.

TIRF microscopy and analysis*.* Single molecule imaging was performed using a custom built total internal reflection fluorescence (TIRF) microscope based on a Nikon Ti-U inverted microscope and Andor iXon ultra 897 EMCCD camera. Illumination was provided by a Ventus 473nm DPSS laser with a power output of 100mW. Laser coupling into the microscope was achieved via a custom built optical circuit (components were sourced from Thorlabs, Chroma and Semrock) followed by a single mode fibre-optic launch. Laser power at the microscope stage averaged at $5.8\mu W/\mu m^2$. The total internal reflection illumination angle was generated using a combination of fibre-optic micro-positioning and a high numerical aperture TIRF objective (Nikon, CFI Apochromat TIRF 60X oil, NA1.49). The Excitation and fluorescence emission wavelengths were separated using a dichroic mirror with a 488nm edge (Chroma zt488rdc-xr). Emitted wavelengths were further filtered using a 500nm edge long pass filter (Chroma hhq500lp) and a 525nm band pass filter (Chroma et525/50m). Acquisitions were controlled using the Andor Solis software package. Frame exposure times were set to 60ms and an EM gain of 250 was used. Information was saved in Andor's proprietary .sif format to preserve file associated metadata.
SWCNT deposition and protein attachment is similar to that used for AFM imaging. Coverslips used for TIRF imaging were plasma treated to remove contaminants, then coated in 0.01mg/ml SWNTs in SDS solution. Samples were then rinsed with deionised water and soaked in ethanol for 1 hour, followed by a rinse under a flow of ethanol and dried in a flow of N_2. Protein solutions were deposited on the coverslips whilst they were examined under

the microscope. Photo-cross linking to CNTs was also performed on the microscope stage using the 305nm LED with an intensity of 18Wm^{-2} for 5 minutes before rinsing on the stage with deionised water. Protein solutions diluted to single molecule concentrations (as determined by TIRF microscopy), in 50mM Tris HCl pH 8 were deposited on the coverslips whilst they were examined under the microscope. Photo-cross linking to CNTs was also performed on the microscope stage using the 305nm LED with an intensity of 18Wm-2 for 5 minutes before rinsing on the microscope stage with deionised water.

Single molecule imaging data were processed and analysed using ImageJ [20] and Matlab (R2012a) (MathWorks U.S.A.). 32 bit floating point TIFF image stacks were used throughout. The first acquisition frame was removed from all image sequences to account for latency of shutter opening by the camera TTL trigger. All images were processed to normalise for spatial variation in intensity profile of the laser illumination using a reference image look-up of relative spatial illumination intensity mapping the laser illumination created from a Gaussian blurred (20 pixel radius) median z-projection of a fluorescent image stack. The resulting image stack was then corrected for temporal laser intensity fluctuations to minimise the noise in extracted traces. This was achieved by quantifying fluctuations in the global image background and scaling the corresponding frame accordingly, relative to the mean (i.e. this meant dividing each frame by a value, between 0.98 and 1.02 corresponding to the relative laser intensity of any given frame). Practically, this was achieved by removing bright fluorescent spots, defined as any pixel with an intensity greater than 0.05 standard deviations above the median pixel intensity of that frame. Identified pixels were assigned a value equal to the median pixel intensity, effectively erasing them to give a background only image stack. Each frame was scaled relative to the mean intensity of all frames (all pixels) and used to create a temporal lookup table of relative frame to frame laser power fluctuations. This enabled correction of the main image stack. Background counts were subtracted by the pixel-wise subtraction of time averaged median pixel intensity of a background region of interest. Spots were detected in either single frames or projected images using the ImageJ plugin trackmate [21], integrated in the FIJI [22] distribution of ImageJ, prior to frame-wise intensity extraction. The resultant intensity traces were plotted and analysed in Matlab.



Scheme S1. Photochemical insertion of a protein via genetically encoding phenyl azide chemistry.

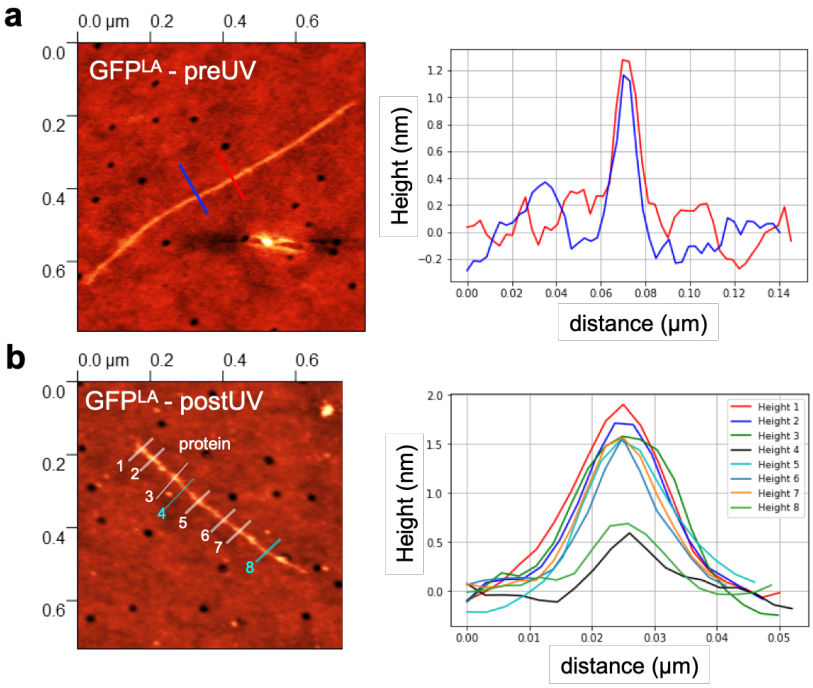


Figure S1. AFM analysis of GFP^{LA}. Each panel presents a representative AFM image of protein (100 nM) deposited on coverslips containing SWCNTs together with height analysis of representative areas. (a) Control sample without illumination (preUV) and after washing; (b) samples after illumination at 305 nm (postUV) and washing. There are clear differences between sample pre- and post-illumination, including distinctive spots with heights and widths larger than of the SWCNT, indicative of protein attachment only on irradiation. The right hand panel in (a) represents the height analysis of two separate regions as indicated in the AFM image. The right-hand panel in (b) represents height analysis across the length of the tube. The heights in the plot key correspond the equivalent areas numbered on the AFM image. Heights 4 and 8 (cyan) in the AFM image in correspond to bare tubes between protein attached areas (white). The heights represent a trajectory perpendicular to the long axis of the tube. The measured heights are lower than expected based on the crystal structure but it should be noted that the proteins are unlikely to be uniformly attached at the apex so fully protrude along their long axis perpendicular to the surface and the tube (see Figure 2 for schematic representation). Together with the fact that the azF is placed in a flexible loop [23], a true height profile of the protein alone relative to the tube will be difficult to measure. Furthermore, AFM is known to underestimate heights of soft matter such as biomolecules [24-25]. The overall difference in width of the protein spots compared to SWCNT is an also important independent validation of protein binding.

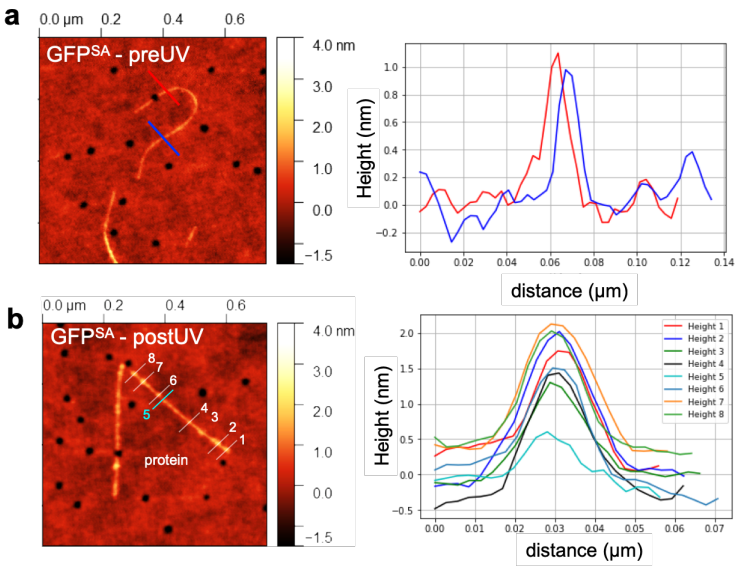


Figure S2. AFM analysis of GFP^{SA}. Each panel presents a representative AFM image of protein (100 nM) deposited on coverslips containing SWCNTs together with height analysis of representative areas. (a) Control sample without illumination (preUV) and after washing; (b) samples after illumination at 305 nm (postUV) and washing. There are clear differences between sample pre- and post-illumination, including distinctive spots with heights and widths larger than of the SWCNT, indicative of protein attachment only on irradiation. The right hand panel in (a) represents the height analysis of two separate regions as indicated in the AFM image. The right-hand panel in (b) represents height analysis across the length of the tube. The heights in the plot key correspond the equivalent areas numbered on the AFM image. Height 5 (cyan) in the AFM image in (b) correspond to bare tubes between protein attached areas (white). The heights represent a trajectory perpendicular to the long axis of the tube The measured heights are lower than those based the crystal structure but it should be noted that the proteins are unlikely to be uniformly attached at the apex of the tube so a true height profile of the protein alone relative to the tube will be difficult to measure; a schematic representation of the relative heights is shown in Figure 2 in the main text. AFM also underestimates heights of soft matter such as biomolecules [24-25]. The overall difference in width of the protein spots compared to SWCNT is an also important independent validation of protein binding.

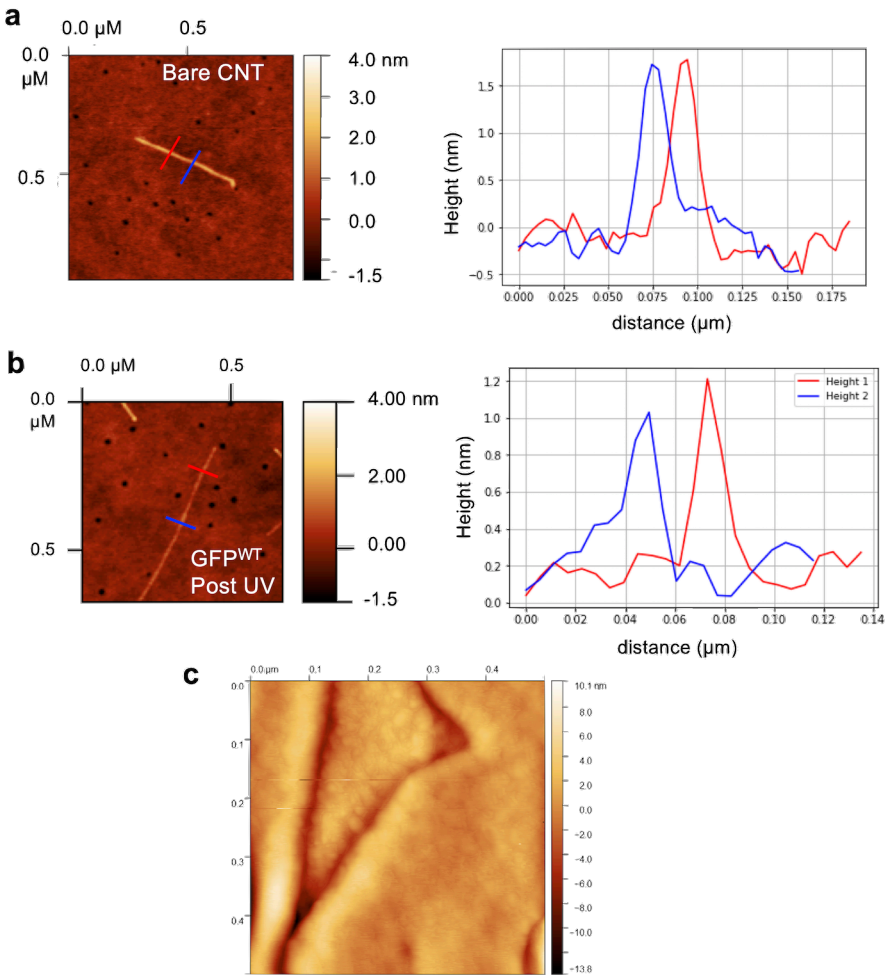


Figure S3. Representative AFM images and height analysis of (a) bare SWCNT and (b) wild-type GFP^{WT} (no azF incorporated) after illumination and washing. The height profiles shown in the right-hand panels represent the equivalent coloured the lines in the AFM images. In the GFP^{WT} sample, protein was applied (100 nM) and subjected to 305 nm illumination as for the azF containing proteins. There is no evidence of permanent protein binding given both the lack of defined spots on the SWCNTS together with heights and width profiles (compare with Figures S1 and S2). (c) AFM image of the application of GFP^{WT} (100 nM) to SWCNTs laid down on surface **without** rinsing with water. A complete blanket of protein is observed with no CNTs distinguishable.

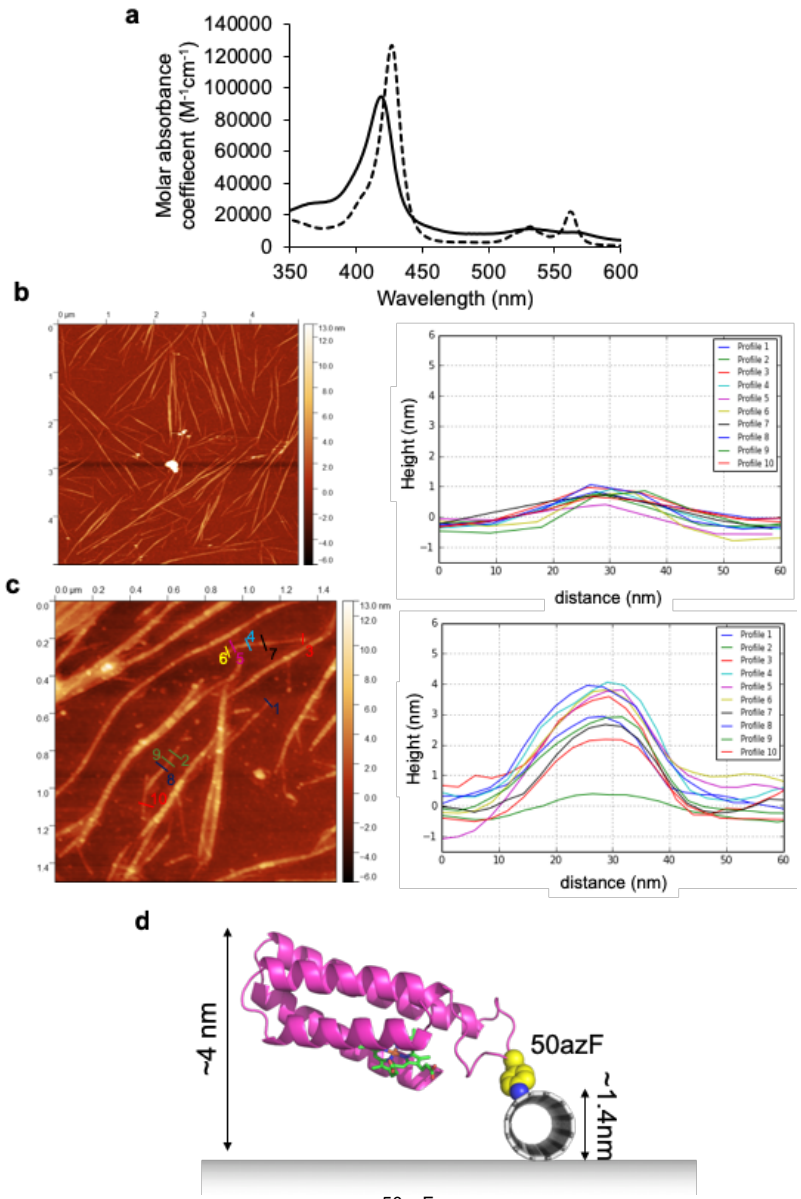


Figure S4. Analysis of cyt b_{562}^{50azF}. (a) Absorbance spectra of oxidised and reduced cyt b_{562}^{50azF} performed as described previously [17, 26-27]. (b-c) AFM analysis. Each panel presents a representative AFM image of protein (10 nM) deposited on coverslips containing SWCNTs together with height analysis of representative areas. (b) Controls samples without illumination and after washing. (b) samples after illumination at 305 nm and washing. In both (b) and (c), the left and right panels represent the AFM image and corresponding height profiles across the CNTs, respectively. There are clear differences between sample pre- and post-illumination, including distinctive spherical spots with heights and widths larger than of the CNTs, indicative of protein attachment only on irradiation. In (c), profile 9 represents a region of CNT with no apparent protein bound. (d) A schematic representation of the protein attached to the CNT by the introduced azF residue, including relative theoretical heights of protein, CNT and surface. It should be noted that the proteins are unlikely to be uniformly attached at the apex and so protruding vertically up with respect to the surface; a true height profile of the protein alone relative to the tube will thus be difficult to measure. AFM also underestimates heights of soft matter such as biomolecules [24-25]. The overall difference in width of the protein spots compared to CNT is an also important independent validation of protein binding.

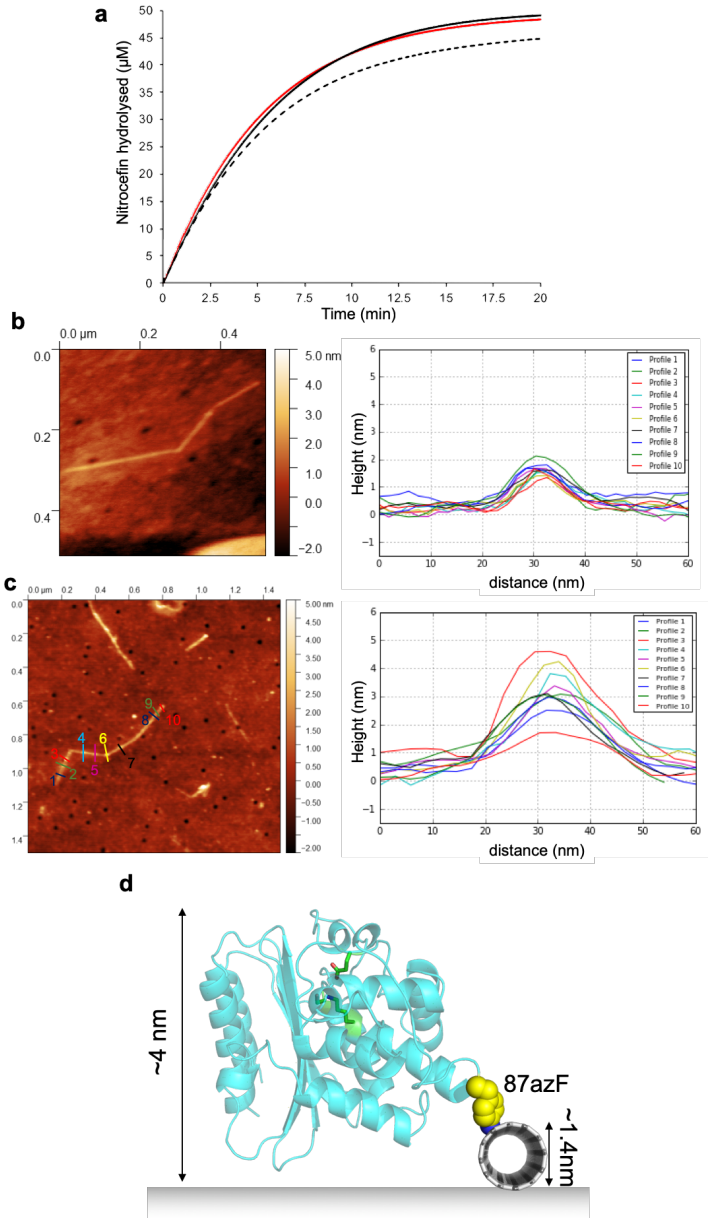


Figure S5. Analysis of TEM^{87azF}. (a) Activity of the TEM^{87azF} variant before (black line) and after (black dashed line) 30 min of UV exposure. The assay was performed as described previously [28]. Change in absorbance at 485 nm was monitored. The wild-type TEM (red line) is shown for comparison. (b-c) AFM analysis of TEM^{87azF}. Each panel presents a representative AFM image of protein (10 nM) deposited on coverslips containing SWCNTs together with height analysis of representative areas. (b) Controls samples without illumination (preUV) and after washing. (c) samples after illumination at 305 nm (postUV) and washing. In both (b) and (c), the left and right panels represent the AFM image and corresponding height profiles across the CNTs, respectively. There are clear differences between sample pre- and post-illumination, including distinctive spherical spots with heights and widths large than of the CNT, indicative of protein attachment only on irradiation. In (c), profile 10 represents a region of CNT with no apparent protein bound. (d) A schematic representation of the protein attached to the CNT by the introduced azF residue, including relative theoretical heights of protein, CNT and surface. It should be noted that the proteins are unlikely to be uniformly attached at the apex and so protruding vertically up with respect to the surface; a true height profile of the protein alone relative to the tube will thus be difficult to measure. AFM also underestimates heights of soft matter such as biomolecules [24-

[25]. The overall difference in width of the protein spots compared to CNT is an also important independent validation of protein binding.

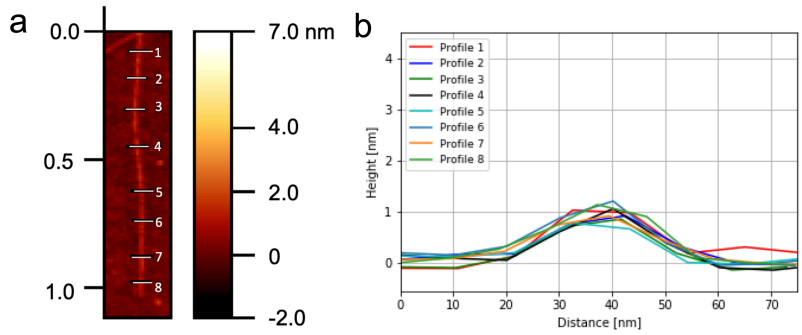


Figure S6. AFM analysis of BBP^{41azF} binding before illumination UV light. (a) Representative AFM image of BBP (10 nM) deposited on coverslips containing SWCNTs together with (b) height analysis of representative areas. No evidence of permanent BBP binding was observed.

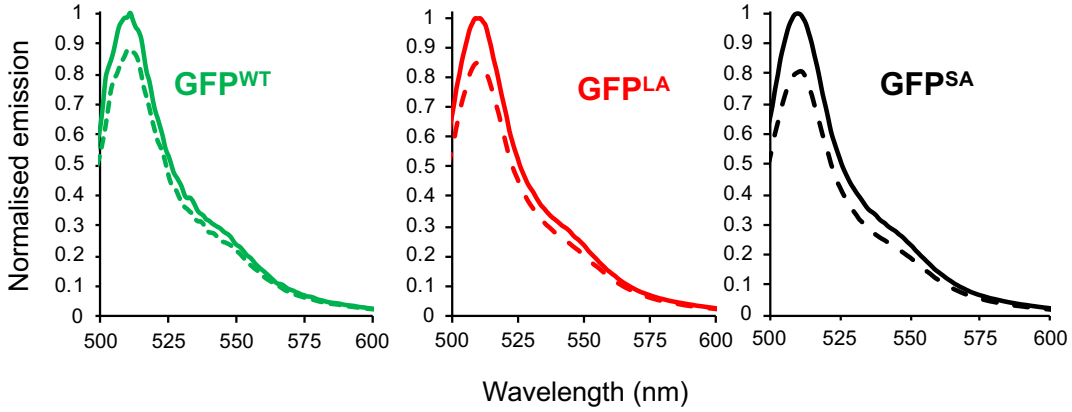


Figure S7. The effect of UV exposure on GFP fluorescence: GFP^{WT} (the starting GFP variant; green trace), GFP^{LA} (red trace) and GFP^{SA} (black trace). Solid and dashed lines represent fluorescence emission before and after 1 hr UV exposure, respectively. UV light was applied via a handheld lamp and bulk steady state fluorescence was measured as described previously [8, 28].

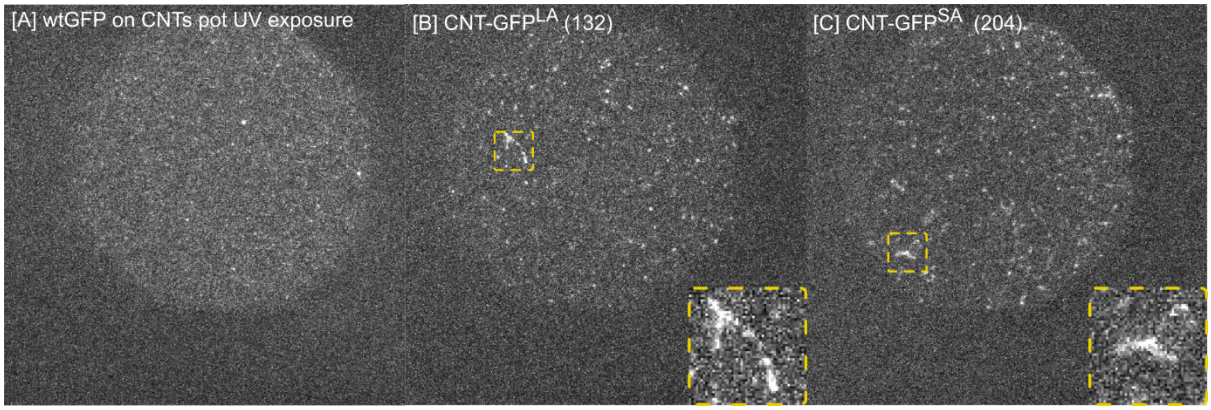


Figure S8. Single molecule imaging of GFP^{WT}, GFP^{LA} and GFP^{SA} with CNTs following UV treatment. GFP variants were added to SWCNTs on glass coverslips and UV irradiated to initiate azF mediated CNT grafting, before washing in-situ to remove excess GFP and

imaging via TIRF (see methods). (a) GFP^{WT} (wtGFP) is not observed to undergo UV initiated binding to CNTs, with no fluorescently decorated tube-like structures visible. AzF containing proteins (b) GFP^{LA} and (c) GFP^{SA} variants are observed to decorate CNTs, with fluorescent tube-like structures visible via TIRF imaging (see also Figure 2), demonstrating the dependency on azF incorporation for UV-initiated SWCNT grafting. For improved signal to noise in visual presentation, each panel (a-c) represents the mean pixel intensity of the first ten frames of image acquisition. See also Supporting Movies S1 and S2 for supporting control experiments.

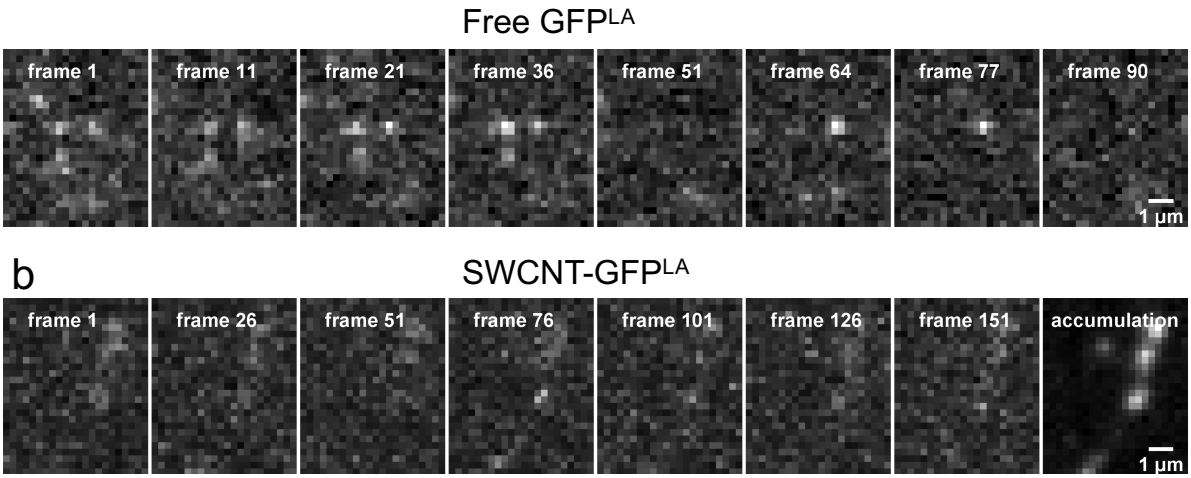
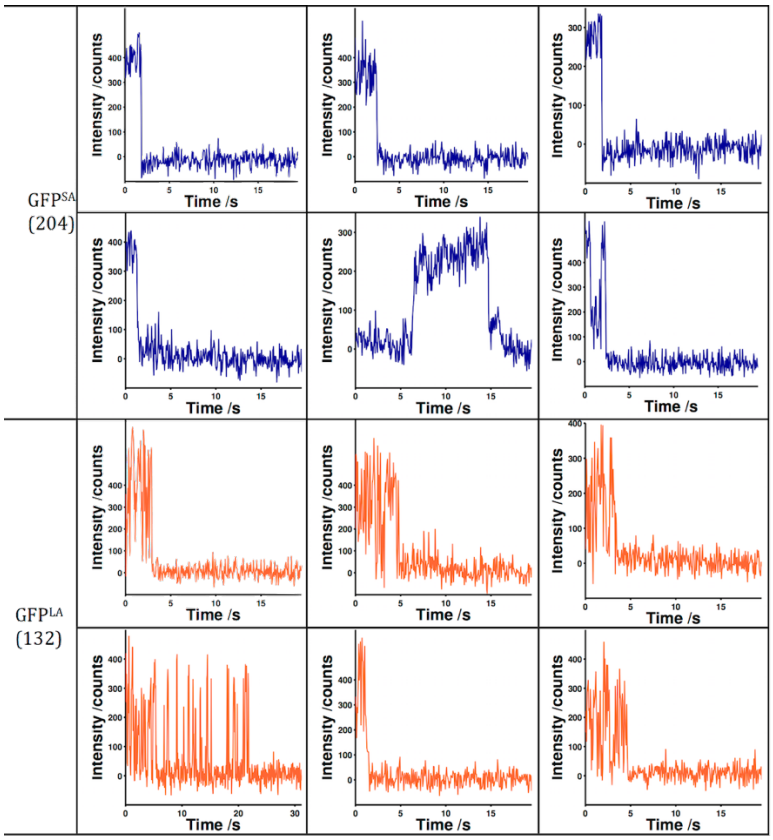


Figure S9.TIRF imaging time course of GFP^{LA} free (a) and CNT attached (b). (a) Single free GFP^{LA} are observed that stochastically photobleach. (b) SWCNT-GFP^{SA} is less bright and undergoes significant intensity fluctuations. Summing frames reveals the linear structure of the decorated SWCNTs.



Continued on next page

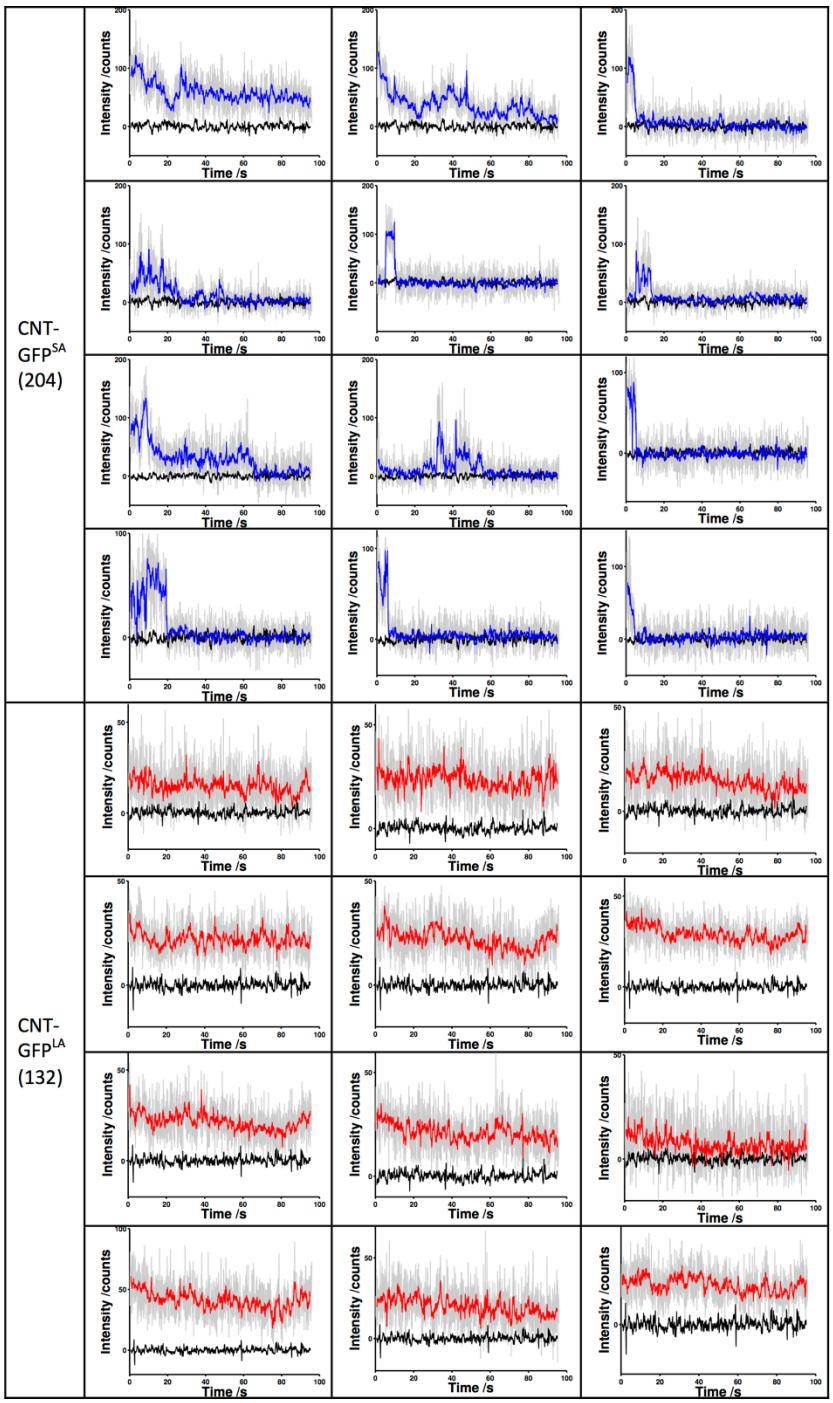


Figure S10. Further representative intensity traces of GFP^{SA} and GFP^{LA} free in solution and SWCNT bound (to accompany figure 2). Single free GFP^{SA} are observed that stochastically photobleach. GFP^{SA} and GFP^{LA} free in solution are of comparable brightness and photobleach with a single step. GFP^{LA} undergoes significant blinking, this is observed less frequently in GFP^{SA}. CNT attachment of GFP^{SA} and GFP^{LA} reduces fluorescence intensity, prolonging lifetime and inducing fluctuation. CNT-GFP^{LA} is less bright and more resistant to photobleaching than CNT-GFP^{SA}. (Raw data (grey), Chung Kennedy filtered data (blue/red) and Chung Kennedy filtered background (black) intensity traces are shown for CNT attached complexes). Some instances of step-like photobleaching or blinking are observed in CNT-GFP^{SA}, but not in CNT-GFP^{LA}. Intensity fluctuations throughout the time course of the experiment, in addition to stochastic activation are reasoned to contribute to the non-exponential decay kinetics observed in ensemble analysis (Figure 2 i&j).

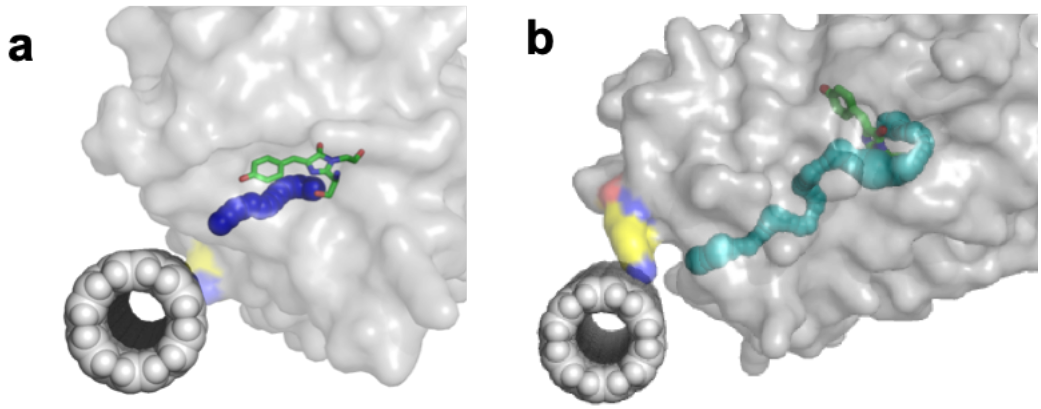


Figure S11. Internal tunnels (calculated using CAVER[29]) for GFP^{SA} (a, blue) and GFP^{LA} (b, cyan) from the functional centre, the chromophore, to the SWCNT sidewall. The chromophore is coloured green and showed in the stick representation. The surface of the protein is shown in grey and as transparent so that the internal tunnels and chromophore can be visualised.

Movie S1. SVid1. GFP^{SA} comparison of fluorescence in the presence and absence of CNTs and UV irradiation. The top row of panels shows the post processed video (original sequence). For each condition, the bottom panels is the frame-by-frame cumulative average pixel intensity of the original sequence (e.g frame 2=mean of frames 1 & 2 from the corresponding top panel. Frame 100=mean of frames 1-100). (a and b) Free GFP^{SA} [GFP204] on coverslip; (c and d) GFP^{SA} [GFP204] and SWCNT on coverslip but no UV irradiation before washing; (e and f) GFP^{SA} [GFP204] and SWCNT on coverslip after UV irradiation and washing.

Movie S2. SVid2. GFP^{LA} comparison of fluorescence in the presence and absence of CNTs and UV irradiation. The top row of panels shows the post processed video (original sequence). For each condition, the bottom panels is the frame-by-frame cumulative average pixel intensity of the original sequence (e.g frame 2=mean of frames 1 & 2 from the corresponding top panel. Frame 100=mean of frames 1-100). (a and b) Free GFP^{LA} [GFP132] on coverslip; (c and d) GFP^{LA} [GFP132] and SWCNT on coverslip but no UV irradiation before washing; (e and f) GFP^{LA} [GFP132] and SWCNT on coverslip after UV irradiation and washing.

Movie S3. SVid3: GFP^{SA} (204) video to accompany montage in Figure 2a, cropped region of interest of select single molecules (17fps).

Movie S4. SVid4: CNT-GFP^{SA} (204) video to accompany montage in Figure 2b, cropped region of interest of select single decorated SWCNT (17fps).

Supporting References

1. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform 4*, 17.
2. Dupradeau, F. Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W., and Cieplak, P. (2010). The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys 12*, 7821-7839.
3. Dupradeau, F. Y., Cezard, C., Lelong, R., Stanislawiak, E., Pecher, J., Delepine, J. C., and Cieplak, P. (2008). R.E.DD.B.: a database for RESP and ESP atomic charges, and force field libraries. *Nucleic Acids Res 36*, D360-7.
4. Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., et al, (1993). General atomic and molecular electronic structure system. *Journal of Computational Chemistry 14*, 1347-1363.
5. Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes 5*, 367.
6. Gfeller, D., Michielin, O., and Zoete, V. (2013). SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Res 41*, D327-32.
7. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., et al, (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 845-854.
8. Reddington, S. C., Rizkallah, P. J., Watson, P. D., Pearson, R., Tippmann, E. M., and Jones, D. D. (2013). Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence. *Angew Chem Int Ed Engl 52*, 5974-5977.
9. Reddington, S. C., Tippmann, E. M., and Jones, D. D., Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry (2012). *Chemical communications 48*, 8419-8421.
10. Sosa-Peinado, A., Mustafi, D., and Makinen, M. W. (2000) Overexpression and biosynthetic deuterium enrichment of TEM-1 beta-lactamase for structural characterization by magnetic resonance methods. *Protein Expr Purif 19*, 235-245.
11. Marciano, D. C., Brown, N. G., and Palzkill, T. (2009). Analysis of the plasticity of location of the Arg244 positive charge within the active site of the TEM-1 beta-lactamase. *Protein Sci 18*, 2080-2089.
12. Marciano, D. C., Pennington, J. M., Wang, X., Wang, J., Chen, Y., Thomas, V. L., Shoichet, B. K., and Palzkill, T. (2008). Genetic and structural characterization of an L201P global suppressor substitution in TEM-1 beta-lactamase. *J Mol Biol 384*, 151-164.
13. Brown, N. G., Chow, D. C., Sankaran, B., Zwart, P., Prasad, B. V., and Palzkill, T. (2011). Analysis of the binding forces driving the tight interactions between beta-lactamase inhibitory protein-II (BLIP-II) and class A beta-lactamases. *J Biol Chem 286*, 32723-32735.

14. Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A., and Mehl, R. A. (2010). Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry 49*, 1667-1677.
15. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif 41*, 207-234.
16. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G., Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J Am Chem Soc 124*, 9026-9027.
17. Della Pia, E. A., Chi, Q., Jones, D. D., Macdonald, J. E., Ulstrup, J., and Elliott, M. (2011). Single-molecule mapping of long-range electron transport for a cytochrome b_{562} variant. *Nano letters 11*, 176-182.
18. Della Pia, E. A., Macdonald, J. E., Elliott, M., and Jones, D. D. (2012). Direct binding of a redox protein for single-molecule electron transfer measurements. *Small 8*, 2341-2344.
19. Nečas, D., and Klapetek, P. (2012). Gwyddion: an open-source software for SPM data analysis. *Central European Journal of Physics 10*, 181-188.
20. Schneider, C. A., Rasband, W. S., Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods 9*, 671-675.
21. Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods 5*, 695-702.
22. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods 9*, 676-682.
23. Arpino, J. A., Rizkallah, P. J., and Jones, D. D. (2012). Crystal structure of enhanced green fluorescent protein to 1.35 a resolution reveals alternative conformations for glu222. *PloS One 7*, e47132.
24. Fuentes-Perez, M. E., Dillingham, M. S., and Moreno-Herrero, F. (2013). AFM volumetric methods for the characterization of proteins and nucleic acids. *Methods 60*, 113-121.
25. Santos, S., Barcons, V., Christenson, H. K., Font, J., and Thomson, N. H. (2011). The intrinsic resolution limit in the atomic force microscope: implications for heights of nano-scale features. *PLoS One 6*, e23821.
26. Arpino, J. A., Czapinska, H., Piasecka, A., Edwards, W. R., Barker, P., Gajda, M. J., Bochtler, M., and Jones, D. D. (2012). Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. *J Am Chem Soc 134*, 13632-13640.
27. Jones, D. D., Barker, P. D. (2004). Design and characterisation of an artificial DNA-binding cytochrome. *Chembiochem 5*, 964-971.
28. Hartley, A. M., Worthy, H. L., Reddington, S. C., Rizkallah, P. J., and Jones, D. D. (2016). Molecular basis for functional switching of GFP by two disparate non-native post-translational modifications of a phenyl azide reaction handle. *Chemical Science* 7, 6484-6491.
29. Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., et al (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **2012,** *8*, e1002708.