Supporting Information

The Incorporation of a Photoisomerizable Amino Acid into Proteins in *E. coli*

Mohua Bose, Dan Groff, Jianming Xie, Eric Brustad and Peter G. Schultz

Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, Genomics Institute of The Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA

Materials and Methods

General. All chemicals were obtained from commercial sources and used without further purification. ^{1}H NMR spectra were recorded on a Bruker AMX-400 instrument with chemical shifts reported relative to either residual $CHCl_3$ (7.25 ppm), residual HOD (4.80 ppm), or residual CH_3OD (3.30 ppm). Mass spectra were acquired at the Scripps Center for Mass Spectrometry (La Jolla, CA) and Mass Spectrometry Center, Stanford University (Palo Alto, CA).

Synthesis of N-(*tert*-Butoxycarbonyl)-L-phenylalanine-4'-azobenzene (1). L-N-*tert*-Butoxycarbonyl-*p*-aminophenylalanine (1 g, 3.6 mmol) was dissolved in glacial acetic acid (200 mL) at room temperature. Nitrosobenzene (578 mg, 5.4 mmol) was added to this solution in one portion and allowed to stir for 8 hours. The reaction mixture was then quenched with satd. $NaHCO_3$ solution (300 mL) and extracted with ethyl acetate (3x 100 mL). The organic layers were then combined, dried (anhydrous $MgSO_4$) and concentrated on a rotary evaporator. The crude material was then purified by silica gel column chromatography (CH_2Cl_2-MeOH. 90:10). The final product was obtained as an orange solid (900 mg, 68%); ^{1}H NMR (400 MHz, CD_3OD) δ 1.30 (s, 9 H), 2.93 (dd, J = 9, 14

Hz, 1 H), 3.19 (dd, J = 5, 14 Hz, 1 H), 4.33 (dd, J = 5, 9 Hz, 1 H), 7.35 (d, *J* = 8 Hz, 2 H), 7.44 (m, 3 H), 7.77 (d, *J* = 8 Hz, 2 H), 7.82 (d, *J* = 8 Hz, 2 H); HRMS (ESI) calcd for $C_{20}H_{23}N_3O_4Na$ (M+ Na^+) 392.1586, obsd. 392.1578.

Synthesis of L-Phenylalanine-4'-azobenzene (AzoPhe) (1a). To a solution of **1** (546 mg, 1.5 mmol) in dioxane (6 mL) at 0°C was added 4N HCl (2 mL) and the mixture was allowed to stir for 3 hours. The mixture was concentrated on a rotary evaporator and thoroughly dried under vacuum (Δ, P_2O_5 sidearm) to provide the final product as an orange solid (380 mg, 94%); ^{1}H NMR (400 MHz, D_2O) δ 2.46 (dd, *J* = 1, 10 Hz, 1 H), 2.83 (dd, *J* = 1, 12 Hz, 1 H), 3.23 (dd, J = 10, 12 Hz, 1 H), 7.00 (m, 5 H), 7.32 (d, *J* = 6 Hz, 2 H), 7.38 (d, *J* = 6 Hz, 2 H); HRMS (ESI) calcd for $C_{15}H_{14}N_3O_2$ $(M-H)^-$ 268.1092, obsd. 268.1083.

Synthesis of Phenylthiohydantoin- phenylalanine-4'-azobenzene (PTH-AzoPhe) (2). The amino acid **1a** (200 mg, 0.74 mmol) was dissolved in pyridine-water (1:1; 5 mL) and the pH was adjusted to 8.6 using 2N NaOH. The mixture was stirred at 40°C and phenylthiocyanate (200 μL, 1.48 mmol) was added dropwise. After 1 hour, the mixture was extracted with toluene (3x 5 mL), the aqueous layer was cooled over ice and acidified to pH~1 by the addition of 2N HCl. The resultant precipitate was filtered and dried (*in vacuo*) to provide phenylthiocarbamyl-L-phenylalanine-4'-azobenzene (PTC-AzoPhe) (239 mg crude). The PTC-AzoPhe was then dissolved in 1N HCl (2.5 mL) and the solution was heated at 80°C for 10 minutes and then cooled to ambient temperature. The reaction mixture was then extracted with ethyl acetate (3x5 mL), the organic layers combined and dried ($MgSO_4$). Flash chromatography (CH_2Cl_2-MeOH. 90:10) of the crude material provided PTH-AzoPhe (50 mg, 18%); ^{1}H NMR (400 MHz, $CDCl_3$) δ 3.18

(dd, *J* = 7, 14 Hz, 1 H), 3.45 (dd, *J* = 4, 14 Hz, 1 H), 4.60 (dd, *J* = 4, 7 Hz, 1 H), 7.12 (app d, *J* = 6 Hz, 1 H), 7.25 (m, 2 H), 7.40-7.55 (m, 7 H), 7.93 (app d, *J* = 8 Hz, 4 H); HRMS (ESI) calcd for $C_{22}H_{18}N_4OSNa$ ($M+Na^+$) 409.1099, obsd. 409.1106.

Plasmids and cell lines used. Plasmid pBK-lib2 encodes a library of *M. jannaschii* tyrosyl tRNA synthetase (TyrRS) mutants randomized at residues Tyr32, Leu65, Phe108, Gln109, Asp158 and Leu162 as well as a Kn^r marker; plasmid pREP(2)/YC encodes $MjtRNA_{CUA}^{Tyr}$, the chloramphenicol acetyltransferase (CAT) gene with a TAG codon at residue 112, the GFP gene under control of the T7 promoter, a mutant T7 RNA polymerase with a TAG codon at residues 1 and 107, and a Tet^r marker; plasmid pLWJ17B3 encodes $MjtRNA_{CUA}^{Tyr}$ under the control of the *lpp* promoter and *rrnC* terminator, and the barnase gene (with three amber codons at residues 2, 44 and 65) under the control of the *ara* promoter, and an Amp^r marker; plasmids pBAD/JYAMB-75TAG-Myo, pBAD/JYAMB-TAG-Myo encode the mutant sperm whale myoglobin gene (with an amber mutation at residue 75 or 4) with an arabinose promoter and *rrnB* terminator, the $MjtRNA^{tyr}_{CUA}$ with an *lpp* promoter and *rrnC* terminator, and a tetracycline resistance marker; plasmids pBAD/JYAMB-125TAG-CAP, pBAD/JYAMB-71TAG-CAP encode the *E. coli* catabolite activator protein (CAP) gene with amber mutations corresponding to amino acids 125 and 71; *E. coli* strain GeneHog®-Fis, cells with genotype F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) 80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ*rps*L(Str^R) *nup*G, *fis*::Tn*7* (DHFR)

Genetic Selection of the Mutant Synthetase specific for Azphe. pBK-lib consisting of 10^9 TyrRS independent clones was constructed using the overlap PCR method. *Pfu* Ultra from Stratagene was used for all PCRs with the manufacturer's protocol. *E. coli* DH10B

harboring the pREP(2)/YC plasmid was used as the host strain for the positive selection. Cells were transformed with the pBK-lib2 library, recovered in SOC for 1 h, then washed twice in the cold with glycerol minimal media with leucine (GMML) before plating on GMML-agar plates supplemented with kanamycin, chloramphenicol, tetracycline and AzoPhe at 50 μg/mL, 60 μg/mL, 12 μg/mL and 1 mM respectively. The plates were incubated at 37^{o}C for 60 hours and surviving cells were pooled and plasmid DNA was extracted and purified by agarose gel electrophoresis. Reisolated DNA was then transformed into electrocompetent cells harboring the negative selection plasmid pLWJ17B3, which were recovered for 1 h in 1 mL of LB and then plated on LB-agar plates containing arabinose (0.2% wt/vol) and ampicillin (50 μg/mL). The plates were then incubated at 37^{o}C for 12 hours and pBK-lib2 DNA for the surviving clones was recovered as described above. The library was then carried through a subsequent round of positive selection (with AzoPhe and a higher chloramphenicol concentration of 80 μg/mL), followed by a negative selection and a final round of positive selection (with chloramphenicol at both 100 and 120 μg/mL). At this stage, 96 individual synthetase clones were selected and each was suspended in 100 μL of GMML in a 96-well plate, and replica-spotted on two sets of GMML plates. One set of GMML-agar plates was supplemented with tetracycline (12 μg/mL), kanamycin (50 μg/mL) and chloramphenicol at concentrations of 80, 100 and 120 μg/mL with the unnatural amino acid, AzoPhe, present at 1 mM. The other set of plates were identical but contained no AzoPhe. Additionally, chloramphenicol concentrations used were 0, 10, 20 and 40 μg/mL. After 48 h incubation at 37^{o}C, 10 synthetase variants were selected for further analysis. Cells harboring these mutant synthetases survived to 120 μg/mL chloramphenicol in the

presence of 1 mM AzoPhe and to 20 μg/mL chloramphenicol in the absence of the unnatural amino acid. DNA sequence analysis of 10 mutants revealed a single synthetase sequence (Gly32, Glu65, Ala108, Glu109, Gly158, His162).

Expression of Mutant Myoglobin and Catabolite Activator Protein (CAP). Plasmid pBAD/JYAMB-75TAG-Myo was cotransformed with pBK vector expressing AzoPheRS (pBK-AzoPheRS) into GeneHog®-Fis *E coli* cells. Cells were amplified in LB media (5 mL) supplemented with kanamycin (40 μg/mL) and tetracycline (24 μg/mL), and washed (twice) with PBS. The starter culture (8 μL) was used to inoculate 150 mL of liquid GMML supplemented with appropriate antibiotics and AzoPhe (1 mM). Cells were then grown at 37°C to an OD_{600} of 0.5 and protein expression was induced by the addition of 0.02% arabinose. After another 8 hours of growth, cells were harvested by centrifugation. The mutant protein (myoglobin with AzoPhe at position 75) was purified by Ni-NTA affinity chromatography under non-denaturing conditions. Samples were desalted, eluted in water and subjected to MALDI-TOF mass spectrometric analysis: expected M_{avg} 18535.172; observed M_{avg} 18534. ESI analysis gave an observed M_{avg} 18535.

Mutant sperm whale myoglobin with AzoPhe at position 4 was similarly expressed from the plasmid pBAD/JYAMB-4TAG-Myo. The first six amino acids from this protein were sequenced on an Applied Biosystems PROCISE 494HT Protein Sequenator at UCSD, Division of Biological Sciences Protein Sequencing Facility. PTH-AzoPhe was used as the standard. The HPLC profile of the PTH-amino acids was analyzed at 269 nm with a gradient of solvent A (3.5% tetrahydrofuran in water containing 20 mL of ABI's Premix buffer concentrate, 900 μL of

1% acetone in water, 75 μL of TFA, 100 μL of 1M KH_2PO_4 per 960 mL of the solvent A) to solvent B (12% isopropanol in acetonitrile).

To demonstrate that AzoPhe could be isomerized after incorporation into a protein, Catabolite Activator Protein (CAP) of *E. coli* having a Leu 125 to AzoPhe mutation was expressed from pBAD/JYAMB-125TAG-CAP in 2YT with 0.002% arabinose. The yield of this mutant CAP was 1.5-1.75 mg/liter of cell culture. UV-visible absorption spectra of both the dark-adapted (predominantly *trans*) and the irradiated samples of mutant CAP were then taken on a double beam UV-visible spectrophotometer (Uvikon 933 from Kontron Instruments). A high pressure mercury lamp (500 W, from Spectra Physics, Mt. View, CA) equipped with a long pass optical filter (320 nm from Spectra Physics, Mt. View, CA) and interchangeable narrow bandpass interference filters (334 nm and 420 nm; band width $\pm$ 5 nm from Edmund Optics, NJ) was used for irradiation of the AzoPhe mutant CAP at 25^{o}C.

Photomodulation of protein–DNA interactions in vitro. A) *Isolation of E. coli lactose promoter segment containing the primary CAP binding site*.[1] Plasmid pUC19 propagated in *E. coli* Dh10B was purified using the QIAfilter Plasmid Mega Kit. A 214 bp lactose promoter fragment containing both the primary and secondary CAP binding sites was isolated by *Hin*f1 restriction digest of pUC19 and preparative polyacrylamide gel electrophoresis. Further digestion with *Hpa*II of the 214 bp fragment, followed by separation on a 10% polyacrylamide-TBE gel (40 cm x 20 cm-length x width) afforded the 118 bp lactose promoter fragment containing only the CAP primary binding site. The118 bp fragment was treated with calf intestinal alkaline phosphatase (CIP) and

purified using the Qiagen Nucleotide Removal Kit. The 5'termini of the 118 bp fragment were labeled with ^{32}P according to the method of Maxam and Gilbert.[2]

B) *Mutant CAP protein expression.* A CAP mutant having an Ile 71 to Azophe mutation was expressed similarly from plasmid pBAD/JYAMB-71TAG-CAP (that had been cotransformed into GeneHog®-Fis *E coli* cells containing plasmid pBK-AzoPheRS) in (2YT) with 0.002% arabinose. The purified protein was exchanged into 50 mM sodium phosphate buffer (plus 300 mM NaCl, pH 8) and the protein concentration determined spectrophotometrically ($\varepsilon_{280} = 3.5 \times 10^4\ M^{-1}\ cm^{-1}$/CAP dimer).[3]

C) *Formation of CAP-DNA complexes and Analysis.* 160 nM of mutant or wild-type CAP was added to 33 nM of 5' ^{32}P-labeled *lac* promoter fragment containing the primary CAP binding site. This mixture was incubated at 25°C for 1 hour. The binding reaction buffer consisted of 10 mM Tris (pH 7.5), 50 mM NaCl, 500 μM EDTA, 500 μM DTT, 1 mM $MgCl_2$, 4% glycerol and 20 μM cAMP. After incubation, this reaction was mixed with loading buffer and loaded directly into the wells of a 6% TBE-PAGE gel that has been pre-run (30 minutes, 150 V) with the running buffer (0.089 M Tris base, 0.089 M boric acid, and 0.002 M EDTA, pH 8.3) with 20 μM cAMP. Gel electrophoresis of the binding reaction samples were carried out at 150 V for 45 minutes. On completion, the gel was blotted free of excess buffer, wrapped in a plastic wrap and then exposed to a storage phosphor screen for 1 hour. Following this, the screen was imaged on a Storm® Phosphorimager System (from Molecular Dynamics).

For gel-shift assays with the AzoPhe mutant, samples of the mutant CAP (Ile 71 AzoPhe) in 50 mM phosphate buffer, pH 8, were irradiated at 334 nm at 0°C for 40 minutes prior to binding with the lactose promoter fragment.

To determine the binding constant[4] of CAP (wild-type or mutant CAP71AzoPhe) to its primary binding site on the lac promoter, a typical three-fold serial dilution of the protein was made from a starting concentration of $1.6x10^{-6}$ M. The lac promoter fragment and cAMP were maintained at a constant 33 nM and 20 μM, respectively. For each dilution step, the binding reaction of CAP to the DNA fragment was allowed to reach equilibration and analyzed by gel electrophoresis as described previously. From the relative intensities of the radioactive bands in the electromobility shift assay, the fraction of bound DNA (PC/P_0) and free DNA (P/P_0) was determined, where P_0 represents the total DNA, PC represents the DNA bound to CAP, P represents the free DNA upon DNA-CAP equilibration, and C_0 and C stand for total protein and unbound protein respectively. A linear regression analysis of the plot of log ([PC]/[P]) versus log[C], provides the binding constant of the CAP-DNA interaction (K_b) from the interpolation on the x-axis.

Photomodulation of protein–cAMP interactions in vitro

Wt or CAP I71AzoPhe at 160 nM was added to 20 μM ^{3}H-labled cyclic adenosine monophosphate (cAMP; 762 cpm/pmol; Amersham-GE Healthcare, Picataway, NJ), 1 mM EDTA, 0.4 M KCl, 1 mM DTT, and 40 mM Tris-HCl, pH 8.0 at 0° C[5]. 50 μg of casein was added as a carrier and after five minutes, protein was precipitated by the addition of 600 μl saturated $(NH_4)_2SO_3$, and allowed to sit at 0° C for ten minutes. The CAP casein mixture was pelleted by centrifuging for 10 minutes at 20800 x *g* at 0° C. The supernatant was removed by aspiration, and the pellet was rinsed with 500 μl of saturated $(NH_4)_2SO_3$, which was removed and then discarded. The pellet was resuspended in 500 μl of water, allowed to stand for 5 minutes and then added to 5

ml Ecolume scintillation cocktail (ICN, Costa Mesa, CA). Radioactivity was measured using a Tri-Carb 2100 liquid scintillation analyzer from Packard (Meridian, CN). To account for non-specific cAMP binding, a blank was prepared using only casein. The measured radioactivity for each CAP sample was adjusted by subtracting out the radioactivity in the casein only sample.

References

1. Hudson, J. M.; Fried, M. G. *J Bacteriol* **1991,** *173*, 59-66.
2. Maxam, A. M.; Gilbert, W. *Proc Natl Acad Sci U S A* **1977,** *74*, 560-4.
3. Anderson, W., Schneider, A. B., Emmer, M., Perlman, R., Pastan, I. *J Biol. Chem.* **1971,** *246*, 5929-37.
4. Fried, M. G.; Crothers, D. M. *J Mol Biol* **1984,** *172*, 241-62.
5. Donoso-Pardo, J.L.; Turner, P. C.; King, R. W. *Eur. J. Biochem.* **1987**, *168*, 687-694

Supporting Figures and Figure Legends

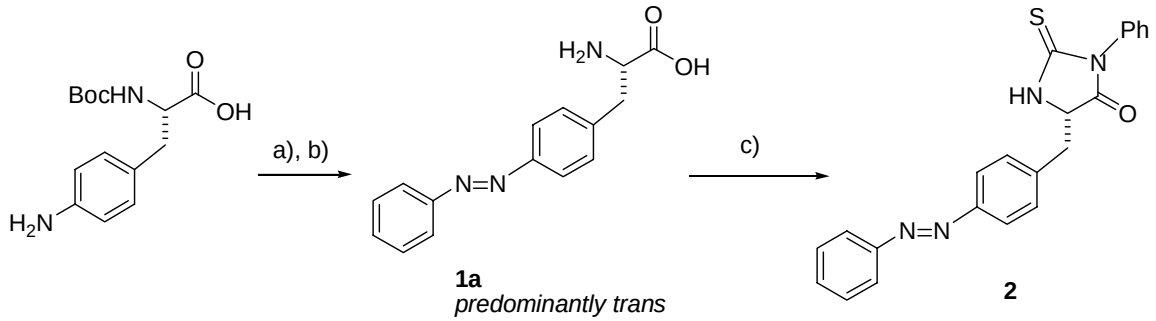

Figure S1. Synthesis of AzoPhe (1a), AzoPhe-PTH (2).
a) PhNO, glacial acetic acid, room temperature; b) **1**, 4N HCl, dioxane, 0^{o}C; c) i) PhNCS, pyridine-H_2O, 40^{o}C; ii) 1N HCl, 80^{o}C.

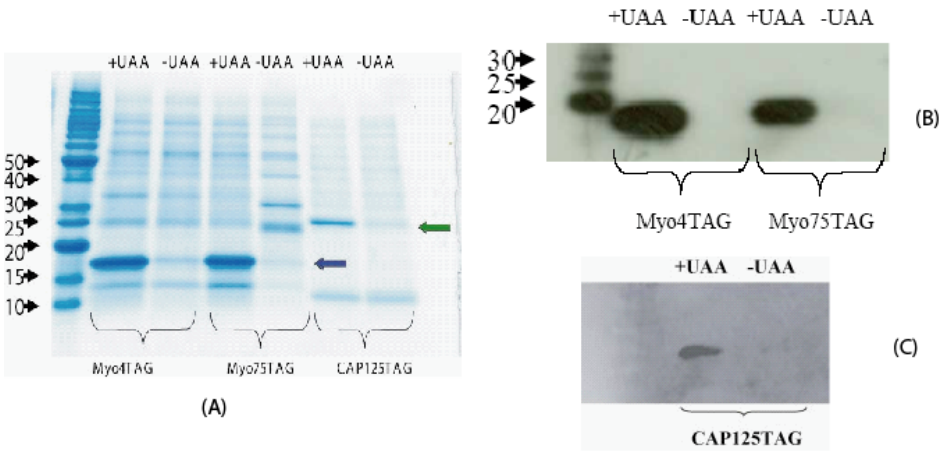

Figure S2. Characterization of the mutant proteins. (A), (B) and (C) Expression of mutant myoglobins (Myo4TAG and Myo75TAG, indicated by blue arrow) and mutant CAP (CAP125TAG, indicated by green arrow) in the presence (+UAA) and absence (-UAA) of 1 mM unnatural amino acid (UAA= AzoPhe), detected with both Gelcode Blue® and anti-His6 antibody after Ni-NTA purification.

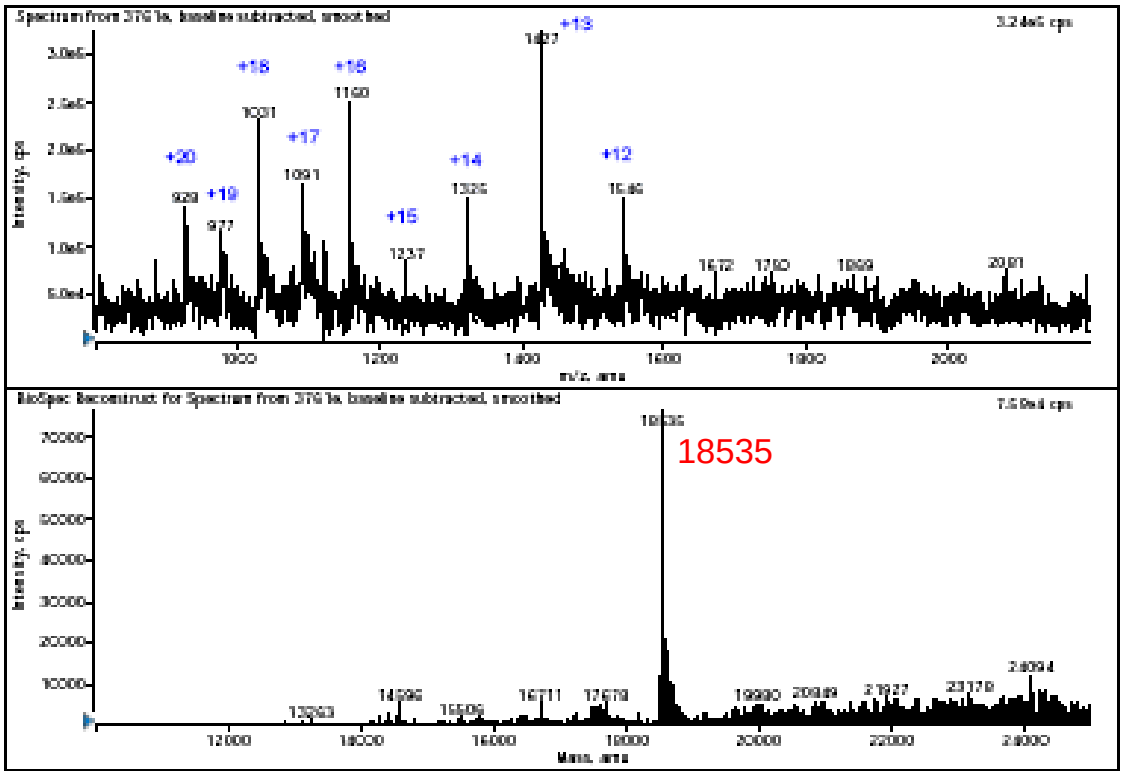

Figure S3. ESI mass spectrum of mutant myoglobin expressed with AzoPhe at position 75.

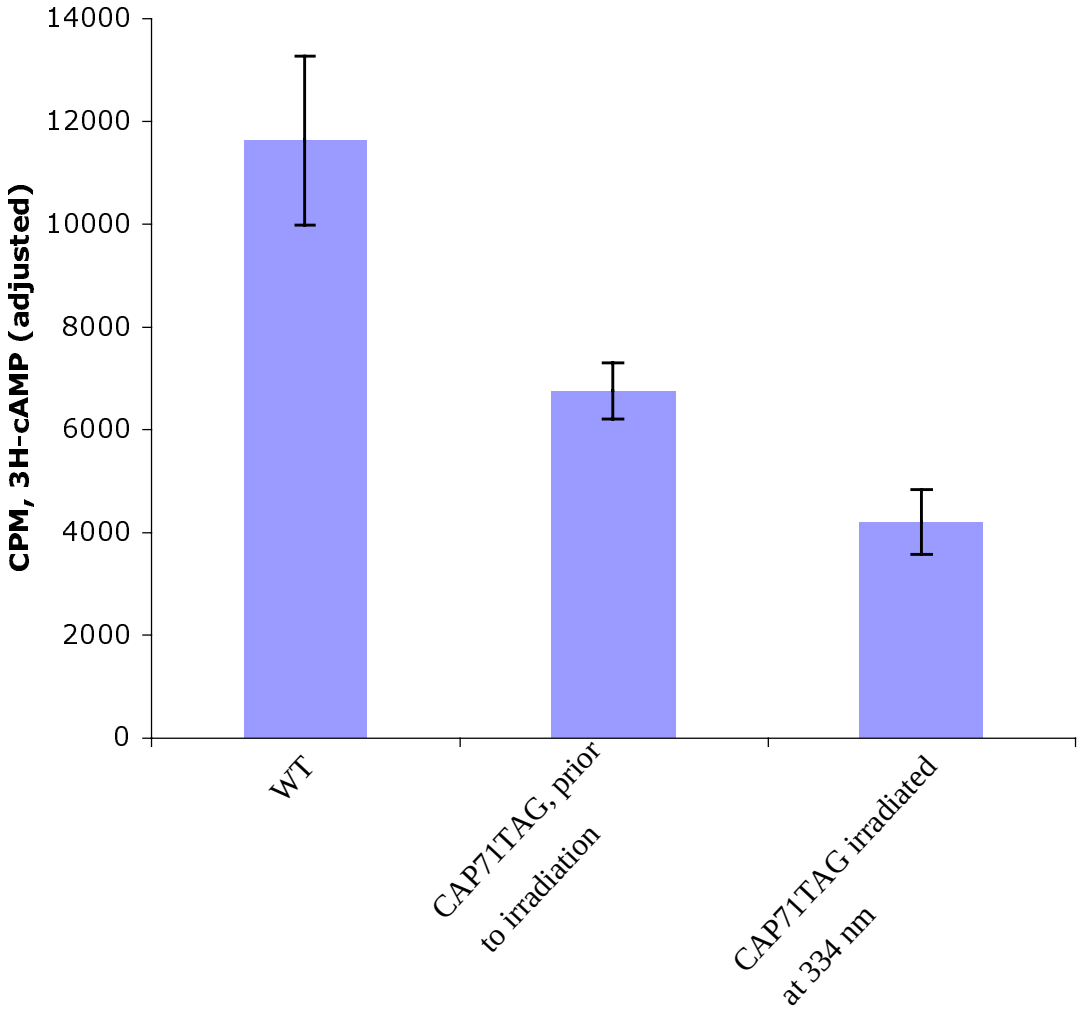

Figure S4. cAMP binding of CAP mutants. Wild-type and CAP71TAG, 160 nM; 1 mM EDTA, 50 µg casein, 20 µM ^{3}H-cAMP (762 cpm/pmol), 40 mM Tris-HCl, 0.4 M KCl, 1 mM DTT, pH 8.0 buffer.