Supporting Information

Genetic Incorporation of a Metal-ion Chelating Amino Acid into Proteins as a Biophysical Probe

Hyun Soo Lee,[†] Glen Spraggon,[‡] Peter G. Schultz*[,†,‡] and Feng Wang*[,†]

[†]*Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037*

[‡]*Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, California 92121*

Materials and Methods

General. All chemicals and DNA oligomers were obtained from commercial sources and used without further purification. ^{1}H NMR spectra were recorded on a Bruker DRX-500 instrument with chemical shifts recorded relative to tetramethylsilane. Protein mass spectra were acquired at the Scripps Center for Mass Spectrometry (La Jolla, CA).

Synthesis of 8-methoxy-3-methylquinoline. To a stirred solution of *o*-anisidine (9.0 g, 73 mmol) and sodium iodide (0.15 g, 1.0 mmol) in sulfuric acid (70%, 26 mL) at 110 °C was added methacrolein (10 mL, 170 mmol) over 5 h. After 1 h at 110 °C, the reaction mixture was cooled to room temperature, poured into 1 M aqueous sodium carbonate solution (450 mL) and extracted with CH_2Cl_2. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica (CH_2Cl_2) to afford 8-methoxy-3-methylquinoline (10 g, 80%) as a brownish solid. ^{1}H-NMR (500 MHz, $CDCl_3$) : δ 2.57 (s, 3H), 4.15 (s, 3H), 7.04 (d, 1H), 7.36 (m, 1H), 7.49 (t, 1H), 7.94 (d, 1H), 8.84 (d, 1H). LC-MS (ESI) calcd for $C_{11}H_{11}NO$ (M+1) 174.1, obsd. 174.2.

Synthesis of 3-methylquinolin-8-ol. 8-Methoxy-3-methylquinoline (10 g, 58 mmol) was dissolved in HBr (48%, 100 mL) and the solution was heated to reflux for 30 h. The reaction mixture was cooled to room temperature, neutralized with 3 M sodium hydroxide and 1 M sodium carbonate and extracted with CH_2Cl_2. The organic layer was

dried over $MgSO_4$ and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica (CH_2Cl_2) to afford 3-methylquinolin-8-ol (8.9 g, 97%) as a brownish solid. ^{1}H-NMR (500 MHz, $CDCl_3$) : δ 2.61 (s, 3H), 7.12 (d, 1H), 7.33 (d, 1H), 7.49 (dd, 1H), 7.99 (d, 1H), 8.71 (d, 1H). LC-MS (ESI) calcd for $C_{10}H_9NO$ (M+1) 160.1, obsd. 160.1.

Synthesis of 3-methylquinolin-8-yl acetate (2). 3-Methylquinolin-8-ol (8.0 g, 50 mmol) was dissolved in acetic anhydride (70 mL) and the solution was stirred for 30 min at 130 °C. The reaction mixture was cooled to room temperature and concentrated under reduced pressure. The residue was dissolved in EtOAc and washed with saturated aqueous $NaHCO_3$ solution. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure to give 3-methylquinolin-8-yl acetate (9.9 g, 98%) as a brownish solid. ^{1}H-NMR (500 MHz, CD_3OD) : δ 2.59 (s, 6H), 7.44 (dd, 1H), 7.57 (t, 1H), 7.71 (dd, 1H), 8.00 (d, 1H), 8.85 (d, 1H). LC-MS (ESI) calcd for $C_{12}H_{11}NO_2$ (M+1) 202.1, obsd. 202.1.

Synthesis of diethyl 2-acetamido-2-((8-acetoxyquinolin-3-yl)methyl)malonate (3). A solution of 3-methylquinolin-8-yl acetate (4.0 g, 20 mmol), NBS (3.6 g, 20 mmol) and AIBN (0.32 g, 2.0 mmol) in CCl_4 (150 mL) was heated to reflux for 12 h. The solvent was removed under reduced pressure. The residue was dissolved in CH_2Cl_2 and washed with saturated aqueous $NaHCO_3$ solution. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure to give the crude brominated product. A mixture of diethyl acetamidomalonate (4.3 g, 20 mmol) and sodium hydride (0.80 g, 20 mmol, 60% in mineral oil) in dry DMF (100 mL) was stirred for 30 min at 0 °C. To the solution was added the crude product in dry DMF (20 mL) at 0 °C and the mixture was stirred for 1 h at room temperature. The reaction mixture was diluted with EtOAc (300 mL) and washed with 10% aqueous sodium thiosulfate solution (2 × 200 mL). The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. The crude product was purified by flash silica gel column chromatography (30% EtOAc in hexane) to give 2-acetamido-2-((8-acetoxyquinolin-3-yl)methyl)malonate (4.3 g, 52%) as a brownish solid ^{1}H-NMR (500 MHz, CD_3OD) : δ 1.39 (t, 6H), 2.14 (s, 3H), 2.58 (s, 3H), 3.95 (s, 2H), 4.38 (m, 4H),

6.66 (s, 1H), 7.50 (dd, 1H), 7.60 (t, 1H), 7.71 (dd, 1H), 7.90 (d, 1H), 8.66 (d, 1H). ^{13}C-NMR (500 MHz, CD_3OD) : δ 14.5, 21.5, 23.5, 35.7, 63.4, 67.5, 121.9, 125.9, 127.0, 129.5, 129.6, 137.1, 140.8, 147.8, 152.4, 167.6, 169.9, 170.1. LC-MS (ESI) calcd for $C_{21}H_{24}N_2O_7$ (M+1) 417.2, obsd. 417.2.

Synthesis of 2-amino-3-(8-hydroxyquinolin-3-yl)propanoic acid dihydrochloride (1). 2-acetamido-2-((8-acetoxyquinolin-3-yl)methyl)malonate (4.2 g, 10 mmol) in 12 M HCl was heated to reflux for 6 h. The reaction mixture was concentrated under reduced pressure to give the product (2.3 g, 99%) as a yellowish powder in the HCl salt form. ^{1}H-NMR (500 MHz, D_2O) : δ 3.62 (d, 2H), 4.38 (t, 1H), 7.47 (d, 1H), 7.72 (d, 1H), 7.77 (t, 1H), 8.98 (m, 2H). ^{13}C-NMR (500 MHz, D_2O) : δ 33.3, 54.3, 117.2, 119.6, 128.7, 130.2, 130.3, 131.7, 143.8, 147.4, 147.7, 171.7. LC-MS (ESI) calcd for $C_{12}H_{12}N_2O_3$ (M+1) 233.1, obsd. 233.2.

Genetic Selection of the mutant synthetase specific for HQ-Ala. *E. coli* DH10B containing the pREP(2)/YC plasmid[1] was used as the host strain for the positive selection. Cells were transformed with the pBK-3D library[2], recovered in super optimal broth with catabolite repression (SOC) for 1 h, then plated on GMML-agar plates supplemented with kanamycin (50 μg/mL), chloramphenicol (35 μg/mL), tetracycline (12 μg/mL) and HQ-Ala (1 mM). The plates were incubated at 37 °C for 60 hours and surviving cells were pooled and pBK plasmids were extracted and purified by agarose gel electrophoresis. The isolated plasmids were then transformed into electrocompetent cells harboring the negative selection plasmid pLWJ17B3[2], which were recovered in SOC for 1 h and then plated on LB-agar plates containing arabinose (0.2% wt/vol) and ampicillin (100 μg/mL). The plates were then incubated at 37 °C for 12 hours and the pBK plasmids from the surviving clones were isolated as described above. The library was then carried through a subsequent round of positive selection, followed by a negative selection and a final round of positive selection. 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 GMML plates. One plate was supplemented with tetracycline (12 μg/mL), kanamycin (50 μg/mL), chloramphenicol (50 μg/mL) and HQ-Ala (1 mM). The other

plate was identical but contained no HQ-Ala. After 48 h incubation at 37 °C, 15 synthetase variants which survived only on the plate with HQ-Ala were selected for further analysis. DNA sequence analysis of the selected mutants is summarized in Table 1.

Expression and purification of Z-domain protein. Plasmid pLEIZ-K7TAG[2] was co-transformed with pBK-HQ-3D4 into *E. coli* DH10B. Cells were amplified in LB medium supplemented with kanamycin (40 ug/mL) and chloramphenicol (35 ug/mL). Starter culture (2.5 mL) was used to inoculate 100 mL of glycerol minimal medium[4] with 0.05% glucose and 1 mM IPTG. Cells were grown at 37 °C for 36 h and harvested by centrifugation. The protein was purified by Ni-NTA affinity chromatography under native conditions according to the manufacturer's protocol (Qiagen, Madison, WI).

Fluorescence measurements. Measurements were performed on an Aminco SLM 8100 spectrofluorimeter (SLM Instruments Inc., Rochester, NY) at room temperature with excitation at 400 nm. Fluorescence of the mutant Z-domain protein at 10 μM (200 μL) was measured in 10 mM Tris (pH 7.6) buffer with 200 mM NaCl and different concentrations of $ZnSO_4$ (0, 10, 30, 50, 80 and 140 μM) and was uncorrected.

Expression and purification of TM0665. Site-directed mutagenesis was used to introduce the F22TAG mutation into the TM0665 gene in the plasmid, pBAD-TM0665, which expresses the C-terminal hexahistidine tagged wild type TM0665. The mutant *Mj*TyrRS gene, HQ-3D4, was sub-cloned from pBK-HQ-3D4 into the NdeI and PstI sites of pSup-BpyRS-6TRN[5] to generate pSup-HQ-3D4. The plasmid, pBAD-TM0665-F22TAG, was co-transformed with pSup-HQ-3D4 into *E. coli* DH10B. Cells were amplified in LB media supplemented with ampicillin (100 μg/mL) and chloramphenicol (35 μg/mL). Starter culture (25 mL) was used to inoculate 1 L of glycerol minimal media[4] with 0.05% glucose and 0.2% arabinose. Cells were grown at 37 °C for 20-24 h and harvested by centrifugation. The protein expressed was purified by Ni-NTA affinity chromatography under native conditions according to the manufacturer's protocol.

Crystallization of TM0665_HQ-Ala. Crystallization of the HQ-Ala TM0665 mutant was accomplished by the hanging-drop vapor diffusion method [28]. Hanging drops containing 2 µl of protein solution at 10 mg/ml and 2 µl of buffer (0.1 M Tris pH 8.5, 45-50% PEG 400, 200-300 mM NaCl, 2 mM $ZnSO_4$) were equilibrated at 20 °C in Linbro plates against 1 ml of the same buffer in the reservoir.

Data Collection and Processing. Diffraction data from a single crystal was collected at 121 K using a cryo-protection solution consisting of a reservoir solution (0.1 M Hepes pH 7.5, 14-16% isopropanol, 200-300mM NaCl, 10 mM $MgCl_2$) with the addition of 30% PEG 400. Crystals diffracted to 2.1 Å at beamline Beamline 5.0.2, Advanced Light Source, Berkeley, CA. Diffraction data was collected at peak wavelengths (wavelength of 1.2815Å) corresponding to the maximum value of $\delta f''$ as determined from a fluorescent scan. In all, 120º of data were collected in two 30º segments using an inverse beam strategy in 5º wedges (Table S2). The data were integrated and reduced using HKL2000. Crystals of TM0665 belong to the space group $P4_22_12$ with cell dimensions a = b = 135.31 Å, c = 74.93 Å, $\alpha = 90°$, $\beta = 90°$, and $\gamma = 90°$, with two molecules in each asymmetric unit.

Structure determination and model refinement. SAD phasing was carried out with the program SOLVE in PHENIX (http://www.phenix-online.org/). The data of peak wavelength was included into the calculation within the resolution range 2.1 to 20 Å. One heavy atom site in each asymmetric unit was found. The overall figure of merit is 0.34. The electron density map was further improved by RESOLVE program. A clearly interpretable electron density map was calculated using FFT in CCP4 suite. The initial model was built by an automatic modeling building program, TEXAL. Once again, in the CCP4 suite the structure was refined with REFMAC rigid body refinement program, and manual rebuilding was performed using XtalView. The model was further refined with cycles of model building and REFMAC restrained refinement to an R_{factor} of 20.7% and an R_{free} of 24.4%. During the final cycles of the refinement, water molecules were added into peaks above 3-σ of the F_0-F_c electron density maps that were within hydrogen-bonding distance from appropriate protein atoms. Atomic coordinates and structure

factors of TM0665_HQ-Ala have been deposited in the Protein Data Bank (PDB code:XXX)

References

1. Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* **2005**, *128*, 338–339.
2. Xie, J.; Liu, W.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2007**, *46*, 9239–9242.
3. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56–61.
4. Studier, F. W. *Protein Expr. Purif.* **2005**, *41*, 207–234.
5. Ryu, Y.; Schultz, P. G. *Nat. Methods* **2006**, *3*, 263–265.

Ref. 10 in the main text;

Lesley, S. A.; Kuhn, P.; Godzik, A.; Deacon, A. M.; Mathews, I.; Kreusch, A.; Spraggon, G.; Klock, H. E.; McMullan, D.; Shin, T.; Vincent, J.; Robb, A.; Brinen, L. S.; Miller, M. D.; McPhillips, T. M.; Miller, M. A.; Scheibe, D.; Canaves, J. M.; Guda, C.; Jaroszewski, L.; Selby, T. L.; Elsliger, M. A.; Wooley, J.; Taylor, S. S.; Hodgson, K. O.; Wilson, I. A.; Schultz, P. G.; Stevens, R. C. *Proc Natl Acad Sci U S A*, **2002**, *99*, 11664–11669.