# Biosynthesis of the RiPP Trojan Horse Nucleotide Antibiotic Microcin C is Directed by the N-Formyl of the Peptide Precursor

Shi-Hui Dong‡,§, Alexey Kulikovsky‡,†,Δ, Inna Zukher†, Paola Estrada‡, Svetlana Dubiley†,Δ, Konstantin Severinov†,Δ,∏,*, and Satish K. Nair

**Electronic Supplementary Information**

## Materials and Methods

**General materials and methods.** Oligonucleotides were purchased from Integrated DNA Technologies. Restriction endonucleases, DNA polymerases, T4 DNA polymerase, and HiFi DNA AssemblyMater Mix were obtained from New England Biolab. Amicon Ultra-4 centrifugal filters were ordered from MILLIPORE. Chemicals were ordered from Sigma Aldrich. All synthetic peptides were ordered from GenScript. *E. coli* DH5α and *E. coli* Rosetta 2 cells were used as host for cloning and plasmid propagation, and host for protein expression, respectively. Expression vectors were obtained from Addgene. All polymerase chain reactions (PCR) were carried out on a Mastercycler thermal cycler (Eppendorf). DNA sequencing was performed by ACGT, Inc.

**Molecular biology techniques.** The gene encoding MccB was amplified by PCR with a previously published construct as template.[1] The purified DNA was cloned into pET His6 TEV LIC cloning vector using HiFi DNA Assembly Mater Mix for protein overexpression. Site-directed mutagenesis was performed by the QuikChange method (Agilent) or HiFi DNA Assembly method. All primers used in this study were listed in **Table S2**. All clones were sequenced to verify the DNA sequences.

**MccB protein overproduction and purification.** The MccB clone was transformed into chemically competent *E. coli* Rosetta 2 cells for overproduction. Starter cultures (each with 6 mL LB) were grown overnight and used to inoculate into 2 L LB containing ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL). Cultures were grown at 37 °C with vigorous shaking until the OD600 reached ~0.6 before cooled down in ice water for 15 min. Protein expression was induced by the addition of 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and cultures were shaken for an additional 18 h at 18 °C and 200 rpm. Cell pellets were harvested by centrifugation at 4 °C, resuspended with 40 mL suspension buffer (500 mM NaCl, 10% glycerol, 20 mM Tris, PH 8.0). Harvested cells were lysed by sonication, and the lysates were clarified by centrifugation at 4 °C. The clear supernatant was loaded onto a 5 mL immobilized metal ion affinity resin column (Hi-Trap Ni-NTA, G.E. Healthcare) pre-equilibrated with binding buffer (1 M NaCl, 5% glycerol, 20 mM Tris, PH 8.0). The column was washed with 50 mL of 12% elution buffer (1 M NaCl, 250 mM imidazole, 20 mM Tris, PH 8.0), and then eluted by a linear gradient to 100% elution buffer. The fractions were analyzed by Coomassie-stained SDS-PAGE gel (12 %). Fractions containing MccB protein were combined and dialyzed with 100 μg of TEV protease against dialysis buffer (300 mM NaCl, 10% glycerol, 20 mM Tris-HCl, pH 7.5) overnight at 4 °C. The dialyzed protein was then

passed through Hi-Trap Ni-NTA column, and the His6 tagged free MccB protein was collected for further purification by size-exclusion chromatography (Superdex Hiload 200 16/60, GE Healthcare) using an isocratic gradient buffer composed of 100 mM KCl, 20 mM HEPES, pH 7.5. Purified proteins were concentrated, and the final concentration was quantified by Bradford analysis (Thermo Scientific). All the MccB variants were purified the same way, and all the purified protein samples were flash froze in liquid nitrogen with small aliquots.

**HPLC analysis of MccB reactions.** The reactions using *N*-formylated MccA and desformyl MccA were carried out in same condition, which contains 1 μM MccB, 50 μM peptide, 3 mM ATP, 100 mM NaCl, 10 μM $MgCl_2$, 20 mM Tris PH 8.0. The reactions were quenched at different time points by adding formic acid to a final concentration of 0.1%, and then spun down to remove any precipitation. All reactions were analyzed by HPLC equipped with a C18 column (Agilent Eclipse Plus, 6 x 100mm, 3 μm), and the mobile phase gradient was 0% to 5% B in 1 min and then 5% to 15% B in 10 min (A: water with 0.1% formic acid; B: acetonitrile with 0.1% formic acid).

**Purine nucleoside phosphorylase (PNP)-based kinetic analysis.** The kinetic parameters for the peptide substrates were determined by a previously reported PNP-coupled assay.[2] To obtain kinetic parameters for MccB with *N*-formylated MccA, the peptide concentration was varied (0–40 μM) in the reaction with 0.25 μM of MccB, 5 mM ATP, 5 mM $MgCl_2$, 5 mM DTT, 0.5 U/mL of PNP, and 0.5 U/mL of Inorganic pyrophosphatase (IPP). The kinetic parameters for MccB with desformyl MccA was acquired using same reaction mixture with the difference of MccB concentration (5 μM) and peptide concentration (0–600 μM). The kinetic parameters for MccB E26A variant with *N*-formylated MccA was obtained using the same reaction mixture except that the MccB E26A protein concentration was 1 μM and the peptide concentration varied from 0 to 50 μM. All reactions were initiated by the addition of corresponding protein, and then reaction progress was monitored by absorbance changes at 360 nm on an Epoch Microplate Spectrophotometer (BioTek). Initial rates of phosphate production were calculated based on initial linear absorbance change and the extinction coefficient of the resulting purine analog (11,000 $M^{-1}$ $cm^{-1}$). Reactions were carried out in triplicate and regression analysis was used to obtain the kinetic parameters by curve-fitting in GraphPad Prism 5 (GraphPad).

**Crystallization and data collection of MccB.** The concentrated MccB protein (10 mg/mL) was used for crystallization by the hanging drop vapor diffusion method. Protein was incubated with 2 mM of ligands (ATP, *N*-formylated MccA, and $MgCl_2$) for 30 min before mixing with precipitant solution in a 1:1 ratio (v/v). The precipitant solution consisted of 0.1 M Magnesium formate, 15% w/v Polyethylene glycol 3350, and 6% w/v 1,6-Hexanediol. The MccB crystals showed up after 2 days incubation at 9 °C, and reached their largest size at 4-7 days. The crystals were flash frozen in liquid nitrogen after dipping into the precipitant solution with 35% Polyethylene glycol 3350 as cryo protection. All diffraction data were collected at LS-CAT (Sector 21, Advanced Photon Source, Argonne National Labs, IL) using MX-300 or Eiger 9M detectors. All data were integrated and scaled using either HKL2000[3] or XDS[4].

Crystallographic phases were determined by the molecular replacement method as implemented in Phaser[5] using the coordinates of previously published MccB structure (PDB: 3H5R).[6] For each structure, iterative model building was carried out using phenix-refine[7] and further improved by manual fitting and adjustment using COOT.[8] Cross-validation, using 5% of the data for the calculation of the free R factor,[9] was utilized throughout model building process in order to monitor building bias. The stereochemistry of all of the models was routinely monitored throughout the course of refinement using PROCHECK.[10] Relevant data collection and refinement parameters are provided in **Table S1**. The coordinates for the MccB structures can be accessed under PDB codes XXX.

**Fluorescein-labeling of peptides.** Synthetic *N*-formylated MccA and desformyl MccA with 3 additional residues (GGK) at C-terminus (**fMRTGNAN**GG**K** and **MRTGNAN**GG**K**, with bold black indicating the MccA sequence and bold red indicating where the free amine was labeled) were used to react with fluorescein isothiocyanate (FITC). The labeling reaction of *N*-formylated peptide was conducted identically to a previous study,[11] while the labeling reaction of desformyl peptide was conducted with a revised condition that the peptide and FITC ratio was 1:1.2 instead of 1:5. Both FITC labeled peptides were purified by HPLC equipped with a C18 column (YMC-Pack ODS-A, 250 x 4.6 mm, 5 μm), and the mobile phase gradient was 5% to 70% B in 30 min (A: water with 10 mM ammonium bicarbonate PH 8.0; B: acetonitrile). The identity and purity of the FITC labeled peptide were confirmed by Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) using a Bruker UltrafleXtreme instrument (Bruker Daltonics) in

reflector positive mode at the University of Illinois School of Chemical Sciences Mass Spectrometry Laboratory. The pure FITC labeled peptide fractions were combined and dried using a FreeZone 2.5 Liter Benchtop Freeze Dry System (LABCONCO), and re-dissolved in binding buffer (100 mM KCl, 20 mM HEPES, pH 7.5) to serve as a stock, which was stored in -80 °C. The concentration of FITC labeled peptide was calculated based on its UV absorption at 492 nm and FITC extinction coefficient of 75,000 $M^{-1}$ $cm^{-1}$ using NanoDrop One$^c$ Microvolume UV-Vis Spectrophotometer (Thermo Scientific).

**Fluorescence polarization (FP) binding assay.** Labeled peptide binding to MccB and its variants in binding buffer was assessed by equilibrium FP at 25 °C in a non-binding-surface 384-black-well polystyrene microplate (Corning) using a Synergy H1 Hybrid Multi-Mode Reader (BioTek) equipped with a green FP filter. Data was recorded with Gen5 software. For each titration, protein was serially diluted into binding buffer, mixed with 20 nM FICT labeled peptide, and equilibrated for 30 min at 25 °C in darkness covered with opaque tape. Dissociation constant ($K_d$) values were calculated from the 50% saturation point using a dose-response curve fit in GraphPad Prism 5 (GraphPad) with three independent titrations. Background fluorescence from the proteins alone was subtracted from the fluorescence polarization signal obtained with the fluorophore.

**Competition FP binding assay.** The protein concentrations used for competition FP binding assay were determined based on each FP binding assay. In general, unlabeled peptides were serially diluted into binding buffer, mixed with 20 nM FITC labeled peptide and corresponding protein, and incubated for 30 min at 25 °C in darkness covered with opaque tape. Data was obtained on the same instrument and analyzed with the same software as above. $IC_{50}$ values were calculated from the 50% saturation point using a dose-response curve fit in GraphPad Prism 5 (GraphPad) with three independent titrations. $K_i$ values for unlabeled peptides were calculated from experimentally-determined IC50 values using the following formula (see ref 12 for additional details):[12]

$$K_i = \frac{I_{50}}{1 + \frac{P_o}{K_d} + \frac{L_{50}}{K_d}}$$

**Table S1. Data collection, phasing and refinement statistics**

| **PDB code** | XXX |
|---|---|
| **Data collection** | |
| Space group | $P2_12_12_1$ |
| Unit cell dimensions (Å) | 55.1, 76.1, 131.9, 90, 90, 90 |
| Resolution (Å) | 65.9 – 1.7 (1.725 – 1.7) |
| Total reflections | 836,773 |
| Unique reflections | 59,837 |
| $R_{sym}$ (%)[1] | 0.124 (1.117) |
| $I/\sigma(I)$ [1] | 14.2 (2.3) |
| Completeness (%)[1] | 96.1 (93.7) |
| Redundancy | 14.0 (14.5) |
| CC (1/2) | 0.997 (0.785) |
| **Refinement** | |
| Resolution (Å) | 65.9 – 1.7 |
| No. reflections | 56,811 |
| $R_{work}$ / $R_{free}$[2] | 15.0/19.7 |
| Number of atoms | |
| Protein | 5,412 |
| Ligand (**4**) | 152 |
| Solvent | 445 |
| B-factors | |
| Protein | 21.4 |
| Ligand (**4**) | 20.6 |
| Solvent | 28.0 |
| R.m.s deviations | |
| Bond lengths (Å) | 1.203 |
| Bond angles (°) | 0.019 |

1. Highest resolution shell is shown in parenthesis. 2. R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

**Table S2. FP binding data of wild type MccB**

| **MccB** | Probe | Competitor | $K_d$ (μM) | $IC_{50}$ (μM) | $K_i$ (μM) | [Protein] (μM) |
|---|---|---|---|---|---|---|
| WT | fMccA-FITC | fMccA | 3.58 ± 0.69 | 25.40 ± 2.93 | 9.13 | 5 |
| WT | fMccA-FITC | MccA | 3.58 ± 0.69 | > 5000 | NA | 5 |
| WT | MccA-FITC | fMccA | 378.70 ± 130.0 | 121.90 ± 32.63 | 52.27 | 100 |
| WT | MccA-FITC | MccA | 378.70 ± 130.0 | > 5000 | NA | 100 |

**Table S3 Primers used for generating clones of MccB and its variants.** All sequences provided 5' to 3', "F" indicates a forward primer, "R" indicates reverse primer.

| Primer name | Primer sequence |
|---|---|
| MccB-Coli-LIC-F | TACTTCCAATCCAATGCAatggattatatattgggtcgctatgtcaaaattg |
| MccB-Coli-LIC-R | TTATCCACTTCCAATGTTATTAcattctatttccacacacagagcaaac |
| MccB-E26A-F | gtggaggcggaaaggCgcaatatgttgagg |
| MccB-E26A-R | cctcaacatattgcGcctttccgcctccac |
| MccB-E26D-F | ggtggaggcggaaaggaTcaatatgttgaggatttagcgc |
| MccB-E26D-R | gcgctaaatcctcaacatattgAtcctttccgcctccacc |
| MccB-E26Q-F | gttggtggaggcggaaagCAAcaatatgttgaggatttag |
| MccB-E26Q-R | ctaaatcctcaacatattgTTGctttccgcctccaccaac |
| MccB-S212A-F | gctgatatatgggttgttGcAgcagatcatccattt |
| MccB-S212A-R | aaatggatgatctgcTgCaacaacccatatatcagc |
| MccB-F217A-F | tctgcagatcatccaGCAaatctgatcaactggg |
| MccB-F217A-R | cccagttgatcagattTGCtggatgatctgcaga |
| MccB-I243A-F | gggtatgttaatgatattgcagtctttggtcc |
| MccB-I243A-R | ggaccaaagactgcTGCatcattaacataccc |
| MccB-V264A-F | gtgccaaaaagtcgCggcagatctatatgg |
| MccB-V264A-R | ccatatagatctgccGcgacttttggcac |
| MccB-Y268A-F | cgtggcagatctaGCAggttcagaaaagg |
| MccB-Y268A-R | ccttttctgaaccTGCtagatctgccacg |
| MccB-Y268F-F | cgtggcagatctaTTTggttcagaaaagg |
| MccB-Y268F-R | ccttttctgaaccAAAtagatctgccacg |
| MccB-H333G-F | tggtcagatgaaataaaaataGGAtcccaaaatatgg |
| MccB-H333G-R | ccatattttgggaTCCtatttttatttcatctgacca |
| MccB-H333A-F | tggtcagatgaaataaaaataGCAtcccaaaatatgg |
| MccB-H333A-R | ccatattttgggaTGCtatttttatttcatctgacca |

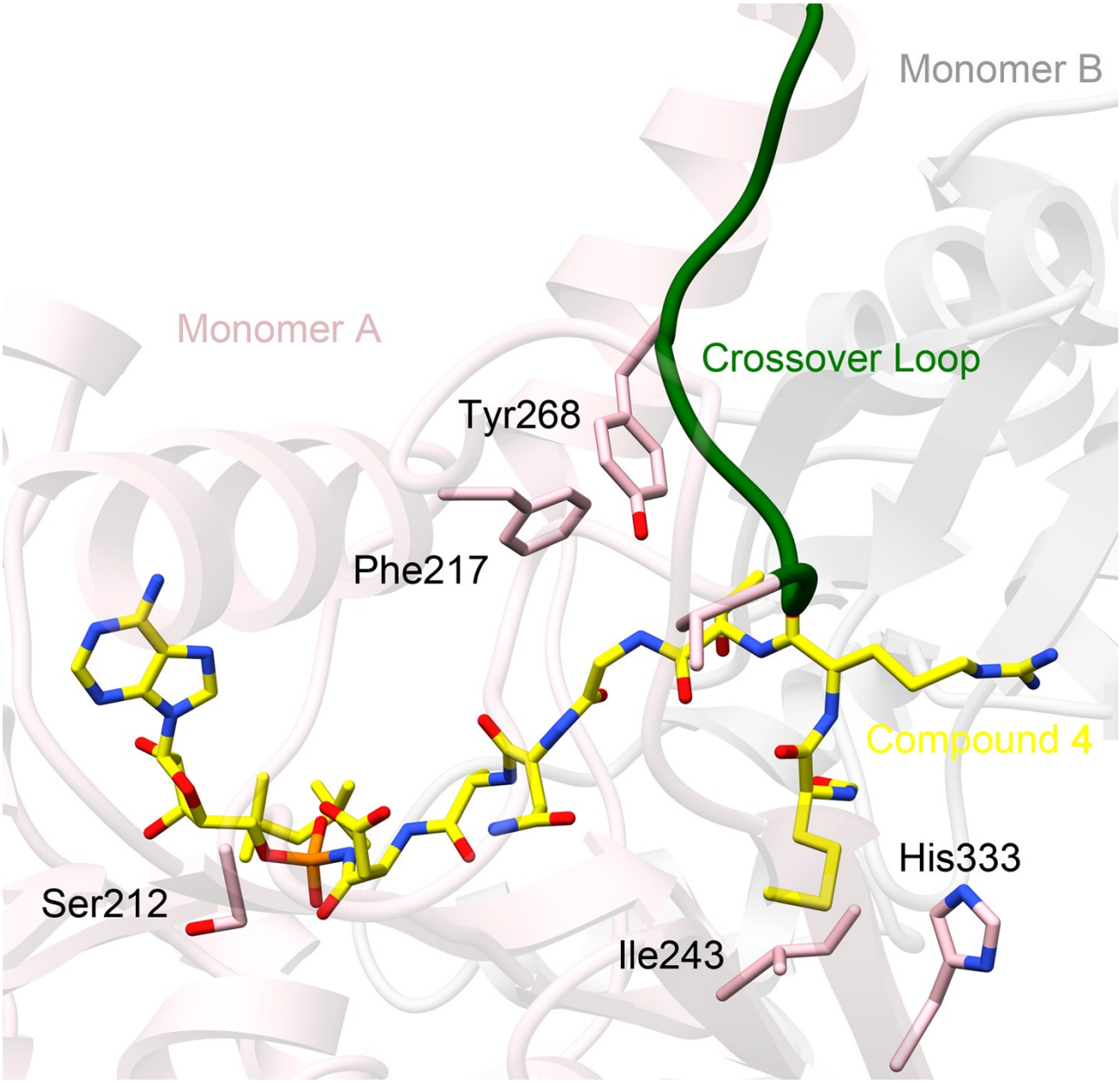


**Figure S1. Structure of MccB showing the binding site of compound 4.** One monomer is colored in gray and the other in pink. The "crossover loop" region is shown in green and compound **4** in yellow. The side chains of residues involved in interactions listed in Table 1 are shown and labeled.

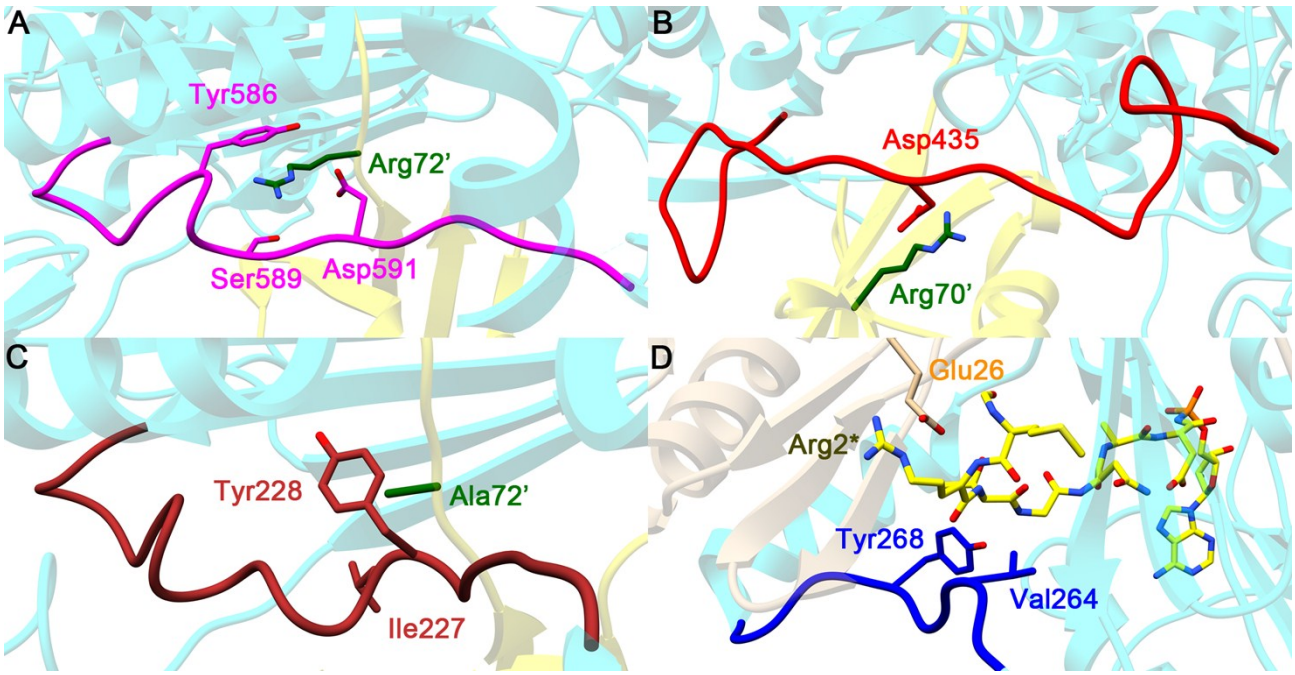


**Figure S2. Structures of ubiquitin-like protein activating enzymes (E1) around "crossover loop" region.** The ubiquitin-like protein and *N*-formylated MccA were shown in yellow. (**A**) Closeup view of the structure of Uba1 in complex with ubiquitin (PDB code: 3CMM).[13] Arg72 from ubiquitin interacts with Tyr586, Ser589, and Asp591 from the Uba1 "crossover loop" shown in magenta. (**B**) Closeup view of the structure of Sae2 in complex with SUMO (PDB code: 1Y8R).[14] Arg70 from SUMO interacts with Asp435 from the Sae2 "crossover loop" shown in red. (**C**) Closeup view of the structure of Uba3 in complex with NEDD8 (PDB code: 2NVU).[15] Ala72 from NEDD8 interacts with Ile227 and Tyr228 from the Uba3 "crossover loop" shown in brown. (**D**) Closeup view of the structure of MccB in complex with *N*-formylated MccA. The newly observed loop (Lys262 through Lys271; shown in blue) in MccB that corresponding to "crossover loop" in ubiquitin-like protein activating enzymes aligns the bound peptide substrate.

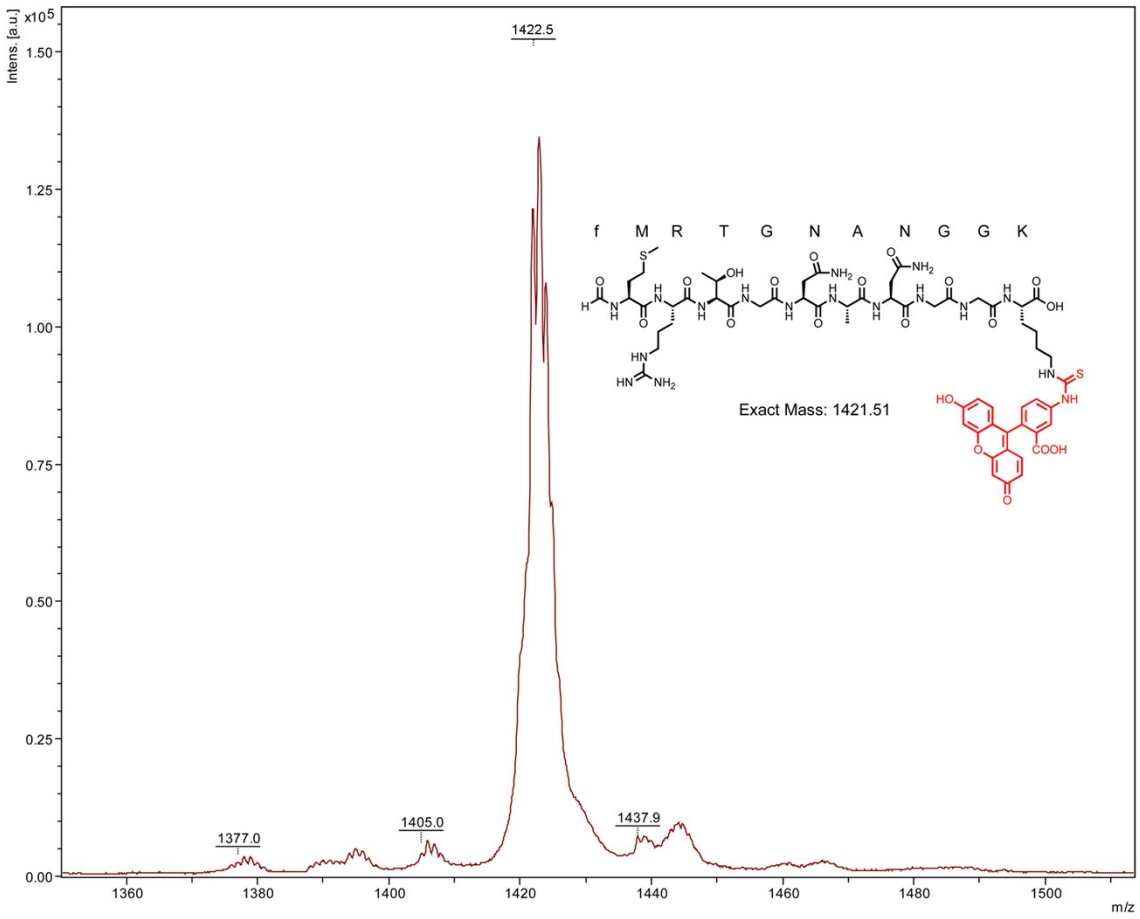


**Figure S3. The MALDI-TOF-MS analysis of FITC labeled *N*-formylated MccA peptide with additional GGK at C-terminus.** The chemical structure and the theoretical mass of the labeled peptide were shown with FITC part in red, and the peptide sequence was marked on the top.

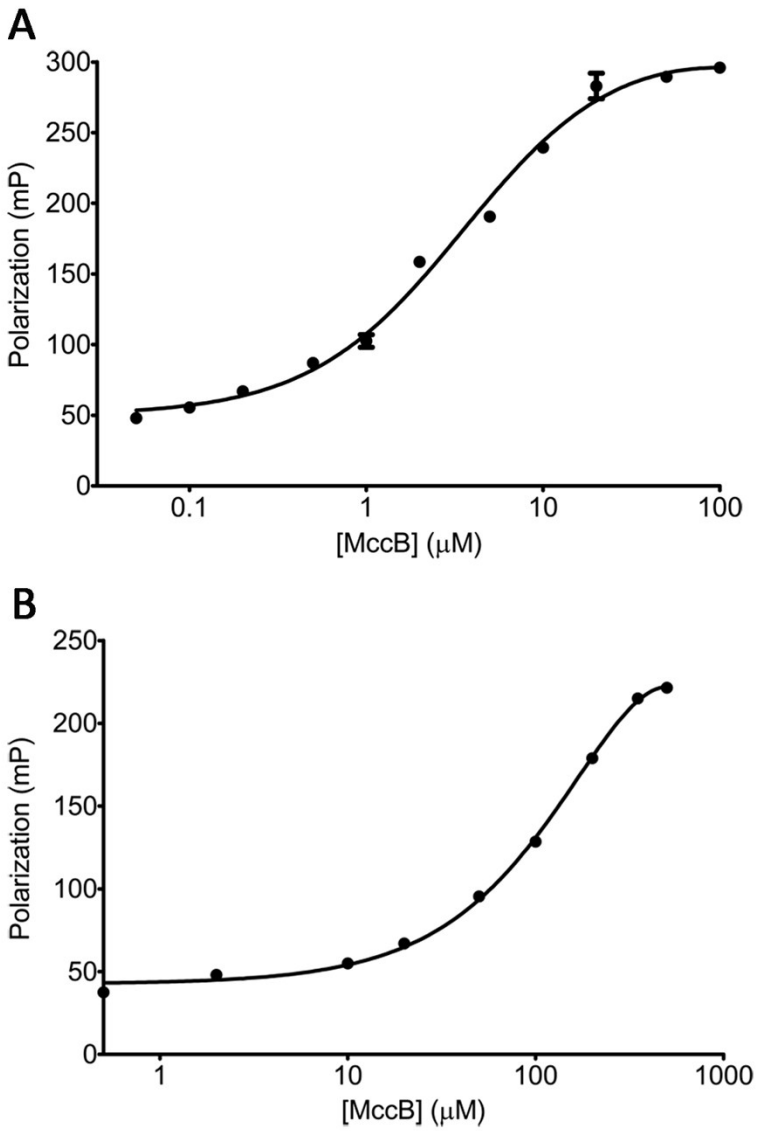


**Figure S4. FP binding curves of MccB using FITC labeled *N*-formylated MccA (A) and FITC labeled desformyl MccA (B) as probe, respectively.** Error bars represent standard deviation of the mean (n = 3).

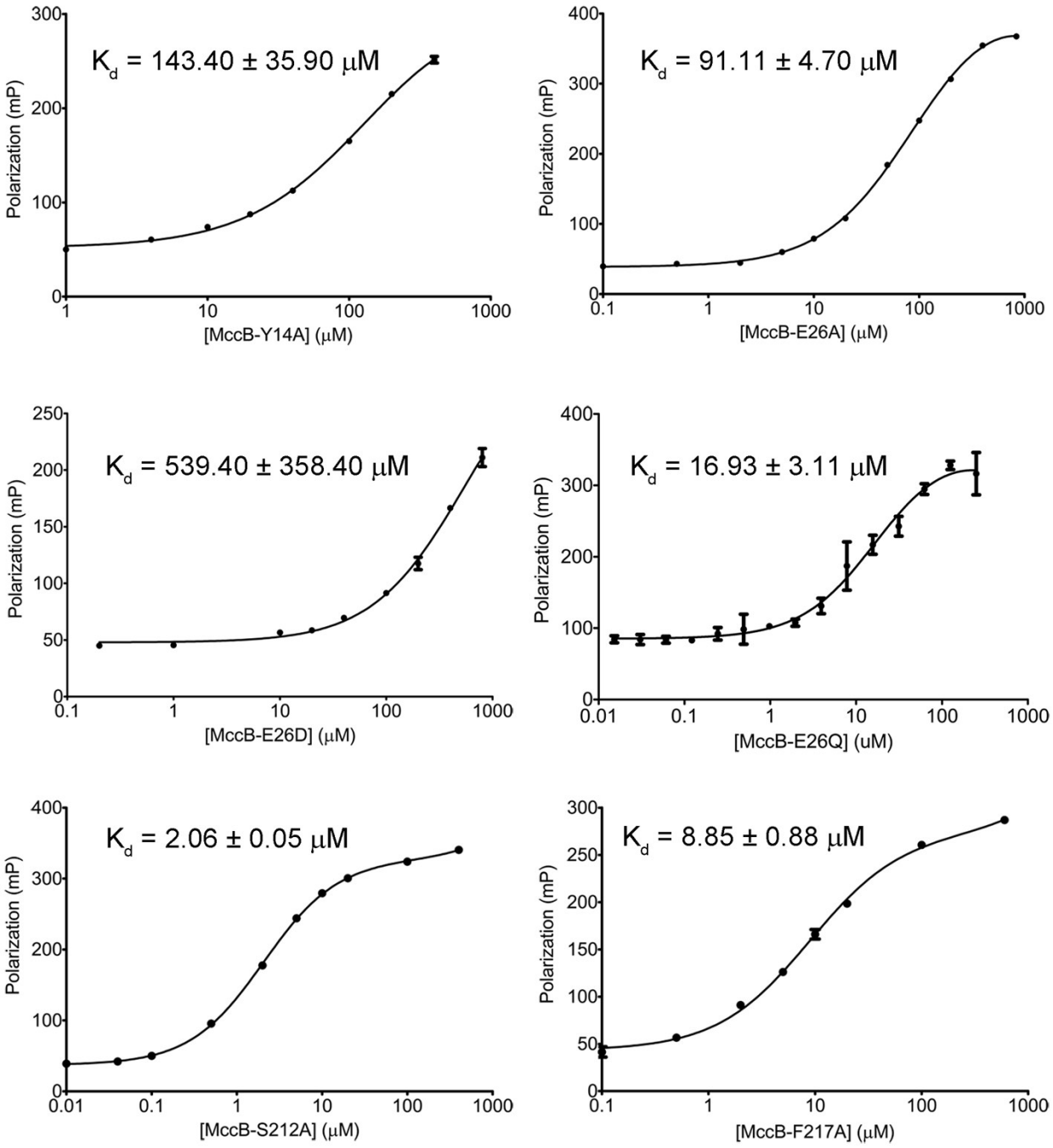


**Figure S5. FP binding curves and $K_d$ of MccB variants using FITC labeled *N*-formylated MccA peptide.** The variants include Y14A, E26A, E26D, E26Q, S212A, and F217A. The identity of each variant of MccB was labeled below the X-axis, and the corresponding $K_d$ was shown on the top of each binding curve. Error bars represent standard deviation of the mean (n = 3).

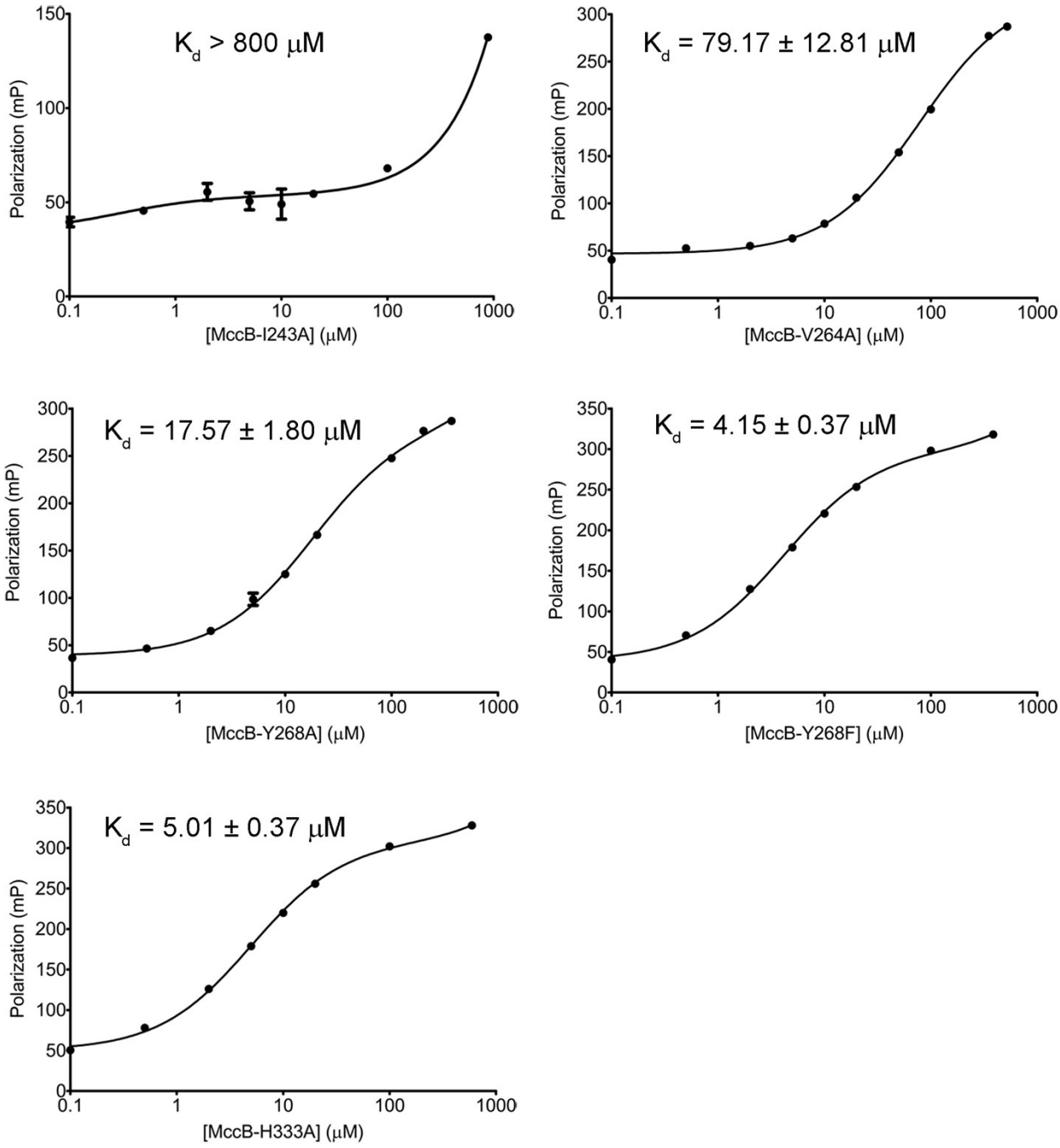


**Figure S6. FP binding curves and $K_d$ of MccB variants using FITC labeled *N*-formylated MccA peptide.** The variants include I243A, V264A, Y268D, Y268F, and H333A. The identity of each variant of MccB was labeled below the X-axis, and the corresponding $K_d$ was shown on the top of each binding curve. Error bars represent standard deviation of the mean (n = 3).

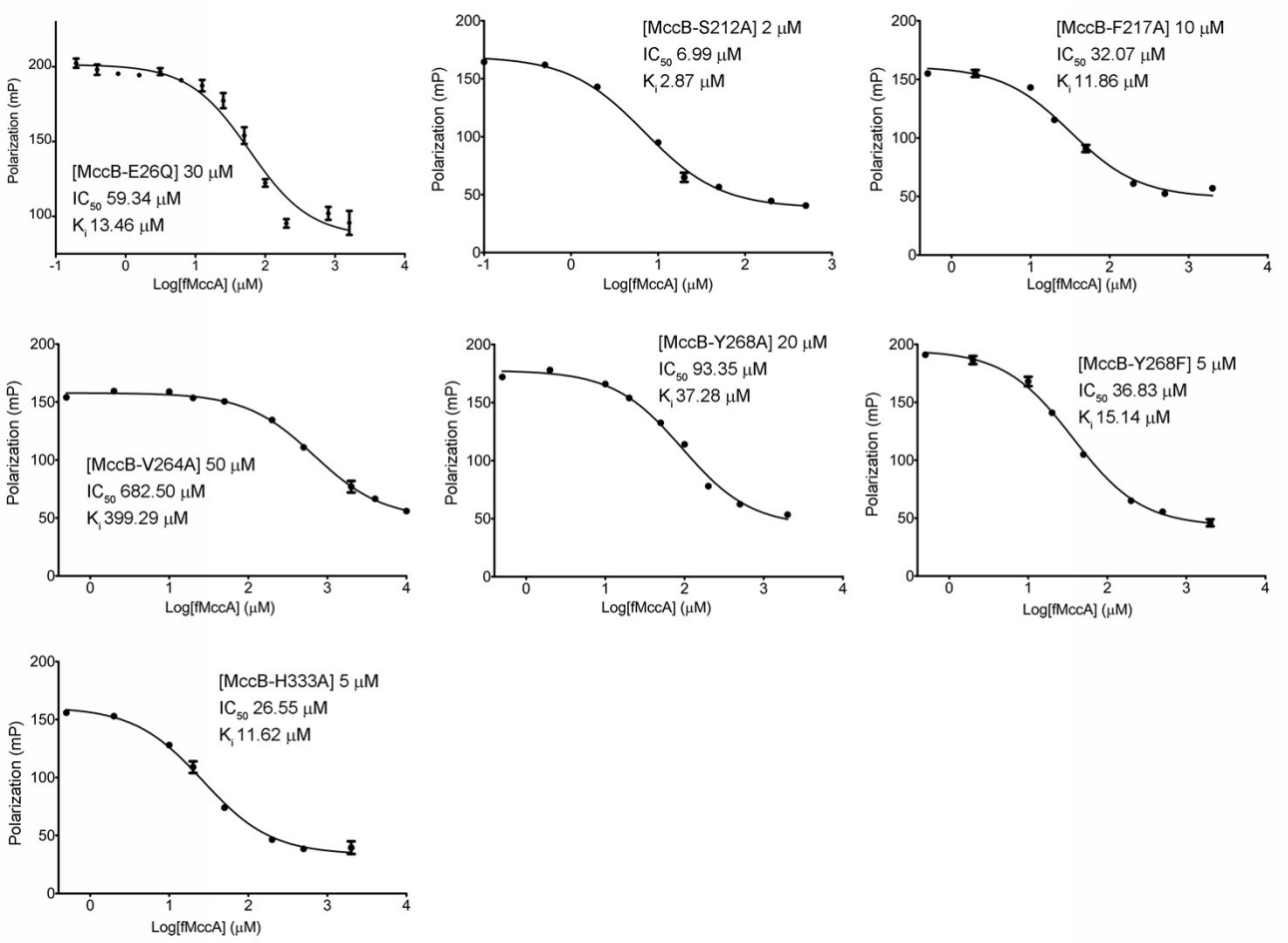


**Figure S7. Competition FP binding curves of MccB variants using FITC labeled *N*-formylated MccA peptide as probe.** *N*-formylated MccA peptide was used for the competition assay. The identity of each variants of MccB and the corresponding parameters were shown on the top of each binding curves. Error bars represent standard deviation of the mean (n = 3).

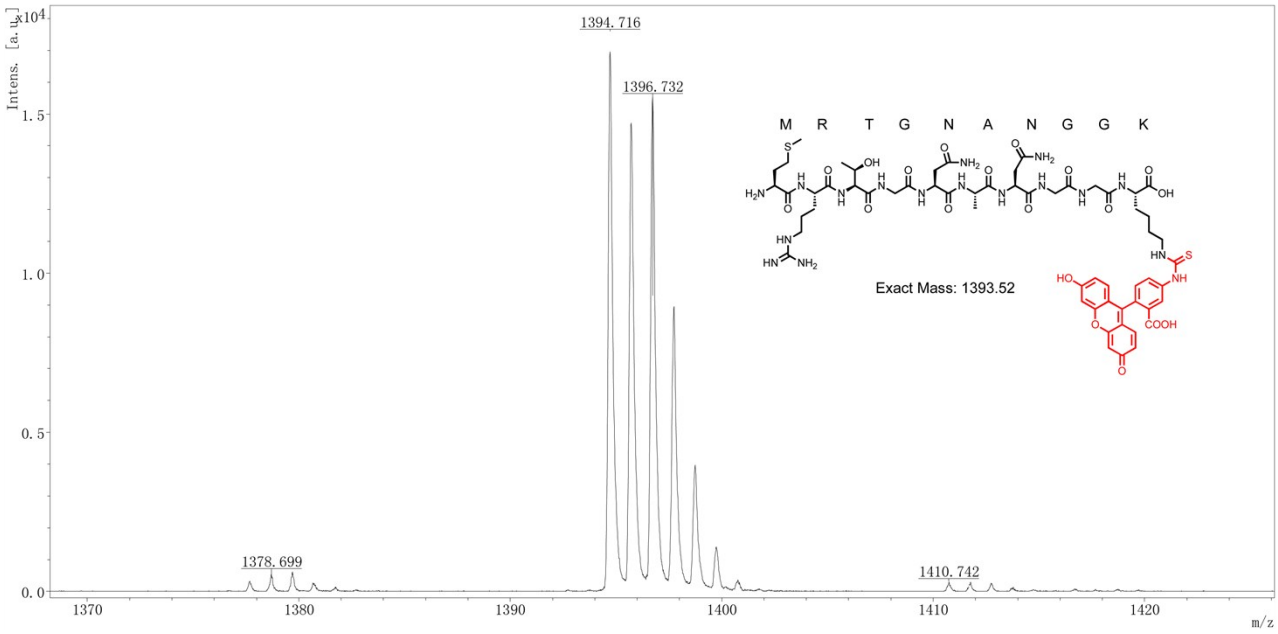


**Figure S8. The MALDI-TOF-MS analysis of FITC labeled desformyl MccA peptide with additional GGK at C-terminus.** The chemical structure and the theoretical mass of the labeled peptide were shown with FITC part in red, and the peptide sequence was marked on the top.

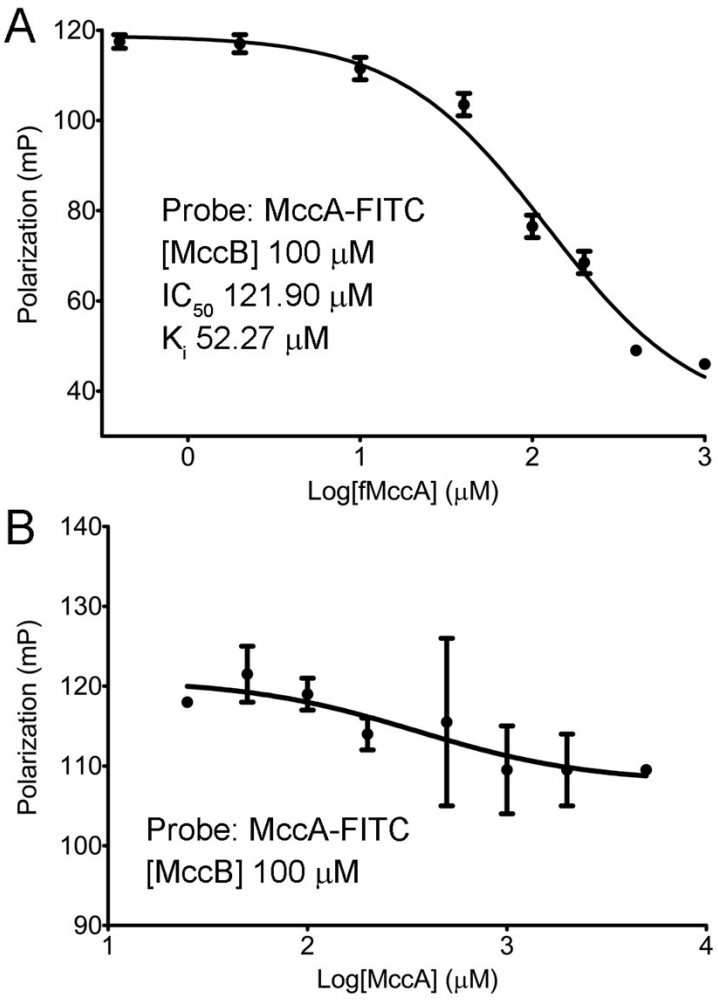


**Figure S9. The competition FP binding curves of MccB.** A, *N*-formylated MccA competition FP binding curve against FITC labeled desformyl MccA; B, desformyl MccA competition FP binding curve against FITC labeled desformyl MccA. Error bars represent standard deviation of the mean (n = 3).

## REFERENCES


1. O. Bantysh, M. Serebryakova, K. S. Makarova, S. Dubiley, K. A. Datsenko and K. Severinov, *MBio*, 2014, **5**, e01059-01014.

2. M. R. Webb, *Proc Natl Acad Sci U S A*, 1992, **89**, 4884-4887.

3. Z. Otwinowski, D. Borek, W. Majewski and W. Minor, *Acta Crystallogr A*, 2003, **59**, 228-234.

4. W. Kabsch, *Acta Crystallogr D Biol Crystallogr*, 2014, **70**, 2204-2216.

5. A. J. McCoy, *Acta Crystallogr D Biol Crystallogr*, 2007, **63**, 32-41.

6. C. A. Regni, R. F. Roush, D. J. Miller, A. Nourse, C. T. Walsh and B. A. Schulman, *EMBO J*, 2009, **28**, 1953-1964.

7. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams, *Acta Crystallogr D Biol Crystallogr*, 2012, **68**, 352-367.

8. P. Emsley and K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 2126-2132.

9. A. T. Brunger, *Methods Enzymol*, 1997, **277**, 366-396.

10. R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton, *J Biomol NMR*, 1996, **8**, 477-486.

11. Z. Zhang, G. A. Hudson, N. Mahanta, J. I. Tietz, W. A. van der Donk and D. A. Mitchell, *J Am Chem Soc*, 2016, **138**, 15511-15514.

12. Z. Nikolovska-Coleska, R. Wang, X. Fang, H. Pan, Y. Tomita, P. Li, P. P. Roller, K. Krajewski, N. G. Saito, J. A. Stuckey and S. Wang, *Anal Biochem*, 2004, **332**, 261-273.

13. I. Lee and H. Schindelin, *Cell*, 2008, **134**, 268-278.

14. L. M. Lois and C. D. Lima, *EMBO J*, 2005, **24**, 439-451.

15. D. T. Huang, H. W. Hunt, M. Zhuang, M. D. Ohi, J. M. Holton and B. A. Schulman, *Nature*, 2007, **445**, 394-398.