# **Electronic Supplementary Information**

# Structure-Guided Synthesis of a Protein-Based Fluorescent Sensor for Alkyl Halide

Myeong-Gyun Kang,<sup>†</sup><sup>[a]</sup> Hakbong Lee,<sup>†</sup><sup>[b],[c]</sup> Beom Ho Kim,<sup>[a]</sup> Yerkin Dunbayev,<sup>[a]</sup> Jeong Kon Seo,<sup>[d]</sup> Changwook Lee,<sup>[b],[c],\*</sup> and Hyun-Woo Rhee<sup>[a],\*</sup>

<sup>[a]</sup> Department of Chemistry, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 44919, Korea

<sup>[b]</sup> Department of Biological Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 44919, Korea

<sup>[c]</sup>Cell Logistics Research Center, Gwangju Institute of Science and Technology, Gwangju 61005, Korea

<sup>[d]</sup>Central Research Facilities (UCRF), Ulsan National Institute of Science and Technology (UNIST), Ulsan, 44919, Korea

\*To whom correspondence should be addressed: rhee@unist.ac.kr, changwook@unist.ac.kr

**†** These authors equally contributed to this work

### Materials and methods



Scheme S1. Synthesis of D1-HTL and D2-HTL

**Probe synthesis** To synthesize Dansyl-PEG2 (**D2**), 2-(2-aminoethoxy)ethanol (40 mg, 0.39 mmol, Sigma Aldrich) was added into dansyl chloride (100 mg, 0.37 mmol, Sigma Aldrich) dissolved in dichloromethane (anhydrous, Alfa aesar). The reaction was allowed to proceed at room temperature for 6 hours. After reaction, the crude product was dried by rotary evaporator and purified by silica gel column chromatography. Purified product was stored in -20°C without any solvent. Yield: 122 mg (97%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 8.538 (d, J = 8.4 Hz, 1H), 8.290 (d, J = 8.8 Hz, 1H), 8.247 (d, J = 7.2 Hz, 1H), 7.561 (t, J = 8.0 Hz, 1H), 7.515 (t, J = 8.0 Hz, 1H), 7.185 (d, J = 7.6 Hz, 1H), 5.517 (t, J = 7.0 Hz, 1H), 3.568 (t, J = 4.4 Hz, 2H), 3.380 (t, J = 4.8 Hz, 2H), 3.309 (t, J = 4.4 Hz, 2H), 3.117 (q, J = 5.2 Hz, 2H), 2.881 (s, 6H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): 152.00, 134.90, 130.43, 129.89, 129.61, 129.44, 128.31, 123.18, 118.76, 115.22, 72.08, 69.09, 61.56, 45.40, 43.06. HR-ESI-MS: m/z calcd. for, 339.1373; measured: 339.1367.

To synthesize Dansyl-PEG2 HaloTag ligand (**D2-HTL**), Dansyl-PEG2 (D2, 95 mg) was added into dried NaH (10 mg, 0.336 mmol, Sigma Aldrich) dissolved in 2:1 dried THF and dried DMF at 0°C. After stirring at 0°C for 30 min, 1-lodo-6-chlorohexane (103.5 mg, 0.42 mmol, Alfa aesar) was added to the solution. The reaction mixture was stirred overnight and was extracted with ethyl acetate (3 x 5ml), washed with cold water and brine. The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated. The crude product was purified by silica gel column chromatography. Purified product was stored in -80°C in DMSO. Yield: 29.7 mg (23.2%).<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 8.534 (d, J = 8.4 Hz, 1H), 8.229 (d, J = 6.4 Hz, 1H), 8.210 (d, J = 9.2 Hz, 1H), 7.540 (t, J = 8.0 Hz, 1H), 7.512 (t, J = 8.0 Hz, 1H), 7.176 (d, J = 7.6 Hz, 1H), 3.642 (s, 2H), 3.588 (t, J = 7.2 Hz, 4H), 3.489 (t, J = 4.4 Hz, 2H), 3.343 (t, J = 6.4 Hz, 2H), 3.260 (t, J = 7.2 Hz, 2H), 2.878 (s, 6H), 1.488 (m, 2H), 1.417 (m, 2H), 1.150 (m, 2H), 0.990 (m, 2H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): 151.78, 134.92, 130.37, 130.14, 129.99, 129.92, 127.95, 123.08, 119.46, 115.13, 72.39, 68.52, 61.66, 47.05, 45.66, 45.41, 44.73, 32.26, 27.45, 26.17, 25.54. HR-ESI-MS: m/z calcd. for, 457.1922; measured: 457.1919.

To synthesize Dansyl-PEG1 (**D1**), ethanol amine (67.8 mg, 1.11 mmol, Sigma Aldrich) was added into dansyl chloride (150 mg, 0.555 mmol, Sigma Aldrich) dissolved in dichloromethane (anhydrous, Alfa aesar). The reaction was allowed to proceed at room temperature for 6 hours. After reaction, the crude product was dried by rotary evaporator and purified by silica gel column chromatography. Purified product was stored in -20°C without any solvent. Yield: 155.6 mg (95.5%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 8.550 (d, J = 8.4 Hz, 1H), 8.290 (d, J = 8.8 Hz, 1H), 8.250 (d, J = 7.2 Hz, 1H), 7.569 (t, J = 8.0 Hz, 1H), 7.522 (t, J = 8.0 Hz, 1H), 7.187 (d, J = 7.6 Hz, 1H), 5.314 (br, 1H), 3.607 (t, J = 4.8 Hz, 2H), 3.042 (q, J = 5.6 Hz, 2H).<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): 147.32, 129.60, 125.89, 125.16, 124.97, 124.80, 123.80, 118.41, 113.77, 110.52. HR-ESI-MS: m/z calcd. for, 295.1111; measured: 295.1105

To synthesize Dansyl-PEG1 HaloTag ligand (**D1-HTL**), Dansyl-PEG1 (D1, 98 mg, 0.33 mmol) was added into dried NaH (11.9 mg, 0.4 mmol, Sigma Aldrich) dissolved in 2:1 dried THF and dried DMF at 0°C. After stirring at 0°C for 30 min, 1-lodo-6-chlorohexane (123.2 mg, 0.5 mmol, Alfa aesar) was added to the solution. The reaction mixture was stirred overnight and was extracted with ethyl acetate (3 x 5ml), washed with cold water and brine. The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated. The crude product was purified by silica gel column chromatography. Purified product was stored in -80°C in DMSO. Yield: 32.3 mg

(23.5%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 8.545 (d, J = 8.4 Hz, 1H), 8.280 (d, J = 8.8 Hz, 1H), 8.190 (d, J = 7.6 Hz, 1H), 7.538 (m, 2H), 7.182 (d, J = 7.6 Hz, 1H), 3.719 (t, J = 5.6 Hz, 2H), 3.420 (q, J = 6.4 Hz, 4H), 3.346 (t, J = 7.6 Hz, 2H), 1.621 (m, 2H), 1.528 (m, 2H), 1.301 (m, 2H), 1.167 (m, 2H).<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): 151.89, 134.47, 130.56, 130.08, 129.83, 128.21, 123.11, 119.17, 115.25, 60.73, 49.42, 48.34, 45.39, 44.80, 32.30, 28.14, 26.28, 25.76, HR-ESI-MS: m/z calcd. for, 413.1660; measured: 413.1654.

Cloning and Protein Production. The DNA fragments encoding the residues 2–297 of wild type HaloTag were amplified by PCR, and cloned into a modified pET-Duet vector including a TEV protease cleavage site. The constructs used in this study are listed in supplementary table 1. For protein production, the plasmids were transformed into Escherichia coli BL21 (DE3) cells. The cells grown at 37°C to an OD<sub>600</sub> 0.7, then shifted to 18°C, and proteins were induced by adding 0.2 mM Isopropyl β-D-1thiogalactopyranoside (IPTG). Cells were harvested using centrifugation after 20 hours post induction, and resuspended in a lysis buffer comprising 25 mM sodium phosphate pH 7.8, 400 mM sodium chloride, and 10 mM imidazole. After sonication, lysed cells were clarified using centrifugation for 60 min at 32,000 g. The supernatant was applied onto Ni<sup>2+</sup>-IMAC affinity column equilibrated with the binding buffer consisting of 25 mM sodium phosphate pH 7.8, 400 mM NaCl, and 10 mM imidazole. The proteins were eluted with binding buffer supplemented with 400 mM imidazole. Histidine tags were removed using TEV protease, followed by dialysis against 25 mM Tris pH 7.8, 150 mM NaCl, and 5 mM β-mercaptoethanol. The dialyzed proteins were passed through Ni<sup>2+</sup> affinity column again to remove histag and TEV protease. The proteins were further purified by gel filtration (Superdex 200 16/600 column, GE Healthcare), eluted in buffer with 25 mM Tris pH 7.5, 150 mM NaCl, and 5 mM DTT. For HaloTag-D2-HTL ligand complex, the purified proteins were mixed with D2-HTL ligand in a 1:3 molar ratio (protein:ligand) for 5 hours, and subjected to another round of size-exclusion purification under same condition to remove extra D2-HTL. The complex was concentrated to approximately 15 mg/ml and flash frozen in liquid nitrogen for storage. All mutants were generated using a PCR based site directed mutagenesis and the mutations were confirmed by DNA sequencing.

**Crystallization and Structure Determination.** Crystals of apo-<sup>M175C</sup>HaloTag were obtained by the hanging-drop method in the well solution containing 25% PEG 20K, 100 mM Tris pH 8.2, 200 mM magnesium chloride, and 5% tertiary butanol at 20°C. The initial crystals of <sup>M175C</sup>HaloTag conjugated with D2-HTL ligand were grown in the crystallization solution comprising 1.8 M ammonium sulfate and 100 mM citric acid pH 5.5 at room temperature. Because the crystal of holo-<sup>M175C</sup>HaloTag were grown as a needle cluster, the micro-seeding method was performed to obtain the diffraction quality crystals. Microseeds generated by Seed Bead Kit (HR2-320, Hampton Research) were added to crystallization drop comprising protein and reservoir solution. By microseeding, the plate shape crystals were obtained in the improved reservoir solution containing 1.3 M ammonium sulfate and 100 mM citric acid pH 5.5 at room temperature. For X-ray diffraction experiments, the crystals of apo-<sup>M175C</sup>HaloTag and holo-<sup>M175C</sup>HaloTag were transferred to cryo-solutions consisting of crystallization solution plus 20% glycerol and then flash-frozen in liquid nitrogen. Diffraction data were collected at beamline 7A of the Pohang Accelerator Laboratory (PAL) and processed using HKL2000<sup>1</sup>. Space groups of the apo-<sup>M175C</sup>HaloTag and holo-<sup>M175C</sup>HaloTag are P4<sub>3</sub>2<sub>1</sub>2 (a, b=62.642, c=164.192 Å), and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (a=44.393, b=72.355, c=174.942 Å), respectively. The crystal structures were determined by molecular replacement with the Phaser program<sup>2</sup>, using the apo-HaloTag structure (PDB ID: 4KAF) as a reference model. The model building and refinement were performed by Coot<sup>3</sup> and Phenix<sup>4</sup> program, respectively. All structural figures were generated by Pymol (https://www.pymol.org). Crystallographic data were summarized in Table S2.

**Spectral Measurements of Alkyl Halides Assay.** 200  $\mu$ M D1-HTL, D2-HTL, HaloTag protein and alkyl halides were prepared as a stock solution. Without alkyl halides reaction, 1  $\mu$ l of D1-HTL or D2-HTL was added into 96-well black plate (SPL, Korea). 198  $\mu$ l PBS (pH 7.4) and 1ul HaloTag protein were added sequentially. For alkyl halides assay, 1  $\mu$ l HaloTag protein and 1  $\mu$ l alkyl halide were reacted in 197  $\mu$ l of PBS solution in the black plate for 10 min. After 10 min reaction, 1  $\mu$ l of D1-HTL or D2-HTL was added into the solution. Alkyl halides assay was analyzed by Spectramax i3x (Molecular Devices).

Alkyl Halides Assay in Human Serum. 1  $\mu$ l of 75mM alkyl halides were mixed with 74  $\mu$ l human serum (Sigma Aldrich, H4522). Every 10 min, 1  $\mu$ l human serum mixtures were added with 1:1000 diultion into 1  $\mu$ M<sup>M175P</sup>HaloTag solution in PBS and incubated for 10 min. Then, 1  $\mu$ l of 1 mM D1-HTL was added into mixtures. After 5 min, D1-HTL fluorescence was obtained by Spectramax i3x (Molecular Devices).

Mass Sample Preparation. 1 mM 1-Bromohexane and 1-Bromoheptane was incubated in human serum (containing 2mg serum protein) for 1 hr. 100 µg of alkyl halide-treated human serum was loaded on 10% SDS-PAGE gel and run by 15 mA current for 1 hr. Gels were cut into 1mm x 1mm size by knife and washed with water 3 x 5 min. Then water was replaced by 100 mM ammonium bicarbonate (ABC, Sigma Aldrich, A6141) buffer and incubated for 5 min. Next 1:1 of 100 mM ABC and acetonitilie was incubated 2 x 5 min and it's replace by 100% acetonitile and incubated for 5 min. 100 mM ABC, 1:1 solution, and 100% acetonitlile was again incubated for 5 min, respectively. After removing of 100% acetonitlie, gels were dried by speed-vac for 10 min. 10 mM of dithiothreitol (DTT) in 100 mM ABC solution was added to dried gel and incubated at 56 °C for 1 hr. DTT solution was replaced by 55 mM iodoacetamide (IAM, Sigma Aldrich, I1149) in 100 mM ABC solution for 30 min with dark condition. After alkylation by IAM, gels were washed by 100 mM ABC, 1:1 solution, and 100% acetonitlile 2 x 5 min, respectively. Gels were again dried by speed-vac and 25 ng/µl concentration of trypsin was incubated overnight. Peptides were eluted by 1:2 mixture of 5% formic acid and 100% acetonitrile solution. Eluted peptides were dired with speed-vac and again eluted with 0.1% formic acid for purification by desalting step.

Desalting and LC MS/MS A home-made column was used for purification of eluted sample. The end of a 200-µL Eppendorf tip was blocked with a 3M Empore C8 disk (3M Bioanalytical Technologies, 2214) and ~10mg of activated charcoal (Sigma Aldrich, C5510) was added. The column was activated by sequential centrifugation at 2000 × g with 150  $\mu$ L of 3 % acetonitrile/0.1 % formic acid (v/v), 150  $\mu$ L of 100 % acetonitrile, and 150 µL of 3 % acetonitrile/0.1 % formic acid (v/v). Then, the samples in the 0.1 % formic acid were added to the column. The column was washed using 150 µL of 0.1 % formic acid (v/v) two times. The column was then eluted with 140 µL of 70 % acetonitrile/0.1 % formic acid (v/v) twice and then with 200 µL of 100 % acetonitrile. The eluted fraction was dried in a speed-vac and peptides were analyzed on LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific) equipped with a nanoelectrospray ion source. A C18 reverse-phase HPLC column (500 mm × 75 µm i.d.) was used to separate the peptide mixture using a 2.4-17.6% acetonitrile/0.1% formic acid gradient for 120 min at a flow rate of 300 nL/min. For MS/MS analysis, precursor ion scan MS spectra (m/z 400 - 2000) were acquired using the Orbitrap spectrometer at a resolution of 60 K at 400 m/z with an internal lock mass. The 20 most intensive ions were isolated and fragmented in the linear ion trap by collisionally induced dissociation (CID).

entry	$\lambda_{ex}^{a}$	$\lambda_{em}^{b}$	Solvent	Second order rate constant (M <sup>-1</sup> s <sup>-1</sup> )	Quantum Yield <sup>c</sup>	Extinction Coefficient (M <sup>-1</sup> cm <sup>-1</sup> )
D1-HTL	316	568	PBS	N/A	0.0015	429
D2-HTL	328	572	PBS	N/A	0.0011	411
D2-HTL+ <sup>WT</sup> HaloTag	275	508	PBS	0.053 x 10 <sup>4</sup>	0.14	11550
D2-HTL+ <sup>M175P</sup> HaloTag	274	507	PBS	0.175 x 10 <sup>4</sup>	0.17	13200
D1-HTL+ <sup>WT</sup> HaloTag	276	508	PBS	0.10 x 10 <sup>4</sup>	0.23	14850
D1-HTL+ <sup>M175P</sup> HaloTag	274	507	PBS	1.14 x 10 <sup>4</sup>	0.41	15000

# Table S1. Photophysical properties of Dansyl HaloTag ligands

<sup>a</sup> excitation wavelength (nm) <sup>b</sup> emission wavelength (nm)

<sup>c</sup> Quantum yields were measured by comparison of quantum yield of dansyl chloride<sup>5</sup>

Name (expected size)	Features	Promotor/ Vector	Details
HaloTag wild type (34.1kDa)	Histag-TEV- <i>BamHI-</i> HaloTag7(wild type) -Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175C) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175C)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175H) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175H)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175K) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175K)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175L) (34.1kDa)	Histag-TEV- <i>BamHI-</i> HaloTag7(M175L)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175N) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175N)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175P) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175P)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175T) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175T)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175W) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175W)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS
HaloTag (M175Y) (34.1kDa)	Histag-TEV <i>-BamHI-</i> HaloTag7(M175Y)-Stop codon- <i>HindIII</i>	T7/ pET-duet	HaloTag7 (PDB ID: 4KAF): 2-297 aa; Histag: HHHHHH TEV: ENLYFQS

### Table S2. Plasmids Information for HaloTag mutants.

Table S3. X-ray Data Collection and Refinement Statistics					
	M175CHaloTag	M175CHaloTad/D2-HTI			
Data set:	Native	Native			
PDB accession #	5Y2X	5727			
X-ray source	Beamline 7A PAI	Beamline 7A PAI			
Temperature (K)	100	100			
Space group:	P4 <sub>2</sub> 2 <sub>4</sub> 2	P2,2,2,			
Cell parameters a, b, c (Å)	62.642.62.642.164.192	44.393.72.355.174.942			
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00			
Data processing					
Wavelength (Å)	1.00000	1.00000			
Resolution (Å)	50-2.02	50-2.27			
R <sub>merge</sub> (%) <sup>a</sup>	9.4 (33.4)	5.6 (7.3)			
Ι/σ	26.4 (5.4)	56.1 (38.6)			
Completeness (%)	99.5 (90.8)	98.4 (88.0)			
Redundancy	6.1 (5.5)	10.4 (9.6)			
Measured reflections	423819	275846			
Unique reflections	22255	49251			
Refinement statistics					
Data range (Å)	34.44-2.02	34.73-2.27			
Reflections	22178	26291			
Nonhydrogen atoms	2366	4730			
Water molecules	259	318			
R.m.s. $\Delta$ bonds (Å) <sup>b</sup>	0.005	0.012			
R.m.s. $\Delta$ angles (°) <sup>b</sup>	1.022	1.367			
R-factor (%) <sup>c</sup>	16.3	15.8			
R <sub>free</sub> (%) <sup>c, d</sup>	19.2	21.5			
Ramachandran plot, residues in					
Most favored regions (%)	95.9	95.7			
allowed regions (%)	4.1	4.3			
Disallowed regions (%)	0	0			

\*Highest resolution shell is shown in parenthesis.

<sup>a</sup>  $R_{merge} = 100 \times \sum_{h} \sum_{i} |I_i(h) - \langle I(h) \rangle | / \sum_{h} \langle I(h) \rangle$ , where  $I_i(h)$  is the *i*th measurement and  $\langle I(h) \rangle$ is the weighted mean of all measurement of *I*(h) for Miller indices h.

<sup>b</sup>Root-mean-squared deviation (r.m.s.  $\Delta$ ) from target geometries.

 $^{c}$  R-factor = 100 x  $\Sigma|F_{P} - F_{P(calc)}|/\Sigma$   $F_{P}.$   $^{d}$  R<sub>free</sub> was calculated with 5% of the data.

#### **Supplementary Results**



Figure S1. Fluorescence Excitation and Emission Spectra of D1-HTL. (A) Fluorescence excitation spectra of D1-HTL and D2-HTL (1  $\mu$ M) before and after addition of wild-type (WT) HaloTag or <sup>W175P</sup>HaloTag (1  $\mu$ M). Fluorescence emission wavelength was fixed at 520 nm. Maximum fluorescence excitation wavelength was observed at 280nm. (B) Fluorescence emission spectra of D1-HTL(1  $\mu$ M) upon addition of <sup>W175P</sup>HaloTag (from 0 to 2  $\mu$ M). The fluorescence was measured after 30min incubation. (C) Fluorescence intensity change (F/Fo) of D1-HTL and D2-HTL at 515 nm (excitation at 280 nm) upon addition of <sup>W175P</sup>HaloTag and <sup>WT</sup>HaloTag, respectively. F<sub>o</sub> value indicates fluorescence emission intensity of D1-HTL or D2-HTL without addition of HaloTag proteins.



Figure S2. Time-course measurement of D2-HTL with HaloTag Mutants. (A) scheme of HaloTag mutants (M175X) screening assay with D2-HTL. (B) Time-course measurement of the fluorescence response of D2-HTL (1  $\mu$ M) by the addition of each HaloTag mutant (1  $\mu$ M); The excitation wavelength is 280 nm and emission wavelength is 515 nm.



**Figure S3: In-gel fluorescence image of HaloTag mutants with D1-HTL.** HaloTag mutants proteins and **D1-HTL** (each 50 pmol) were incubated for 10min at RT and the protein-ligand complex were analyzed by 10% SDS-PAGE. Wild-type, M175P, M175C, M175H, M175K, M175L-HaloTag (upper) and Wild-type, M175P, M175N, M175T, M175W, M175Y (below) were analyzed separately. In-gel fluorescence was analyzed by AF430 channel in Chemidoc (Syngene, UK). Aesterisk marks indicates unbound D1-HTL. Same amount of protein addition was validated by Coomassie Blue stain (right) with the same gel.



**Figure S4. Primary and secondary structure of HaloTag.** (A) The sequences of <sup>M175C</sup>HaloTag are represented with the secondary structure elements determined by crystal structure. Blue arrows and yellow cylinders indicate the *do-strands and v-helices*, respectively. The residue highlighted by red and black box indicates the mutation site (M175C) and conjugation site (D106) by **D2-HTL**. The red and black dots represent residues that are involved in the interaction with **D2-HTL** ligand, and residues coordinating chloride ion, respectively. (B) Figure shows the topology diagrams of the holo-<sup>M175C</sup>HaloTag with helices and strands represented as cylinders and arrows, respectively. The HaloTag protein consists of fourteen *v*-helices and eight *do-strands*. The black and red star indicate the conjugating residues of **D2-HTL** ligand and mutation site (M175) used in this study, respectively.



**Figure S5: F144 and M175 of HaloTag act as a gate for D2-HTL conjugation.** Figure shows the molecular diagrams of holo-<sup>M175C</sup>HaloTag (A), apo-<sup>M175C</sup>HaloTag (B), and wild-type HaloTag (C) to compare the structural differences at dansyl ring binding sites among three proteins. The D2-HTL coordinated by <sup>M175C</sup>HaloTag is shown cyan stick model in (A). The two structural variations for C175 and M175 in apo-HaloTag are indicated in (B) and (C), respectively.



**Figure S6. Surface filling model of holo-HaloTag complex.** This surface filling model highlighting the van der waals interaction between dansyl ring of D2-HTL (cyan) and the side chains of F144, C175 (A) or P175 (B) in HaloTag pocket. Note that the structure of P175 is from putative modeling based on the crystal structure of holo-Halo tag







**Figure S8. MS/MS spectra of alkylated peptides.** MS/MS spectra of alkylated peptides of human serum albumin (HAS) by 1-bromohexane and 1-bromoheptane were shown.



**Figure S9.** Time-course detection of consumed alkyl halides (e.g. 1-bromohexane, 1bromoheptane) by human serum albumin (HSA, 20  $\mu$ g/ $\mu$ l) using D1-HTL (1  $\mu$ M) and <sup>M175P</sup>HaloTag (1  $\mu$ M). 1mM alkyl halides were pre-incubated with 20  $\mu$ g/ $\mu$ l human serum albumin and the pre-incubated solution (0-60 min) was 1000x diluted in hybrid probe solution of D1-HTL (1  $\mu$ M) and <sup>M175P</sup>HaloTag (1  $\mu$ M). The dansyl fluorescence was measured after 10min incubation at room temperature (excitation at 280 nm/emission at 510 nm).



Figure S10: Alkyl halides detection in tap water. Time course measurement of fluorogenic response of D1-HTL (0.1  $\mu$ M) upon addition of <sup>M175P</sup>HaloTag or a premix of M175PHaloTag and various alkyl halide (0.1  $\mu$ M each) in tap water. The excitation wavelength is 280 nm and the emission wavelength is 515 nm.

/g19		88888886666676777777 888888866666666666	666		-40
Parameter	Value		$\forall$		
1 Data File Name					l l
2 Title	PROTON_01				
3 Comment	Mg19				-35
4 Origin	Varian				
5 Owner					-
5 Site					
7 Spectrometer	vnmrs				-30
3 Author					
9 Solvent	cdcl3				
10 Temperature	25.0				F
11 Pulse Sequence	s2pul				
12 Number of Scans	32				-25
13 Receiver Gain	30				
14 Relaxation Delay	1.0000				
15 Pulse Width	0.0000				
16 Acquisition Time	2.5559				
17 Acquisition Date	2016-10-18T 19:17:18				-20
18 Modification Date	2016-11-08T 23:24:16				
19 Spectrometer Freque	ancy 399.89				-
20 Spectral Width	6410.3				
21 Lowest Frequency	-805.8				-15
22 Nucleus	1H				
23 Acquired Size	16384				
24 Spectral Size	32768				T I
					-10
					-50
		e fe 🛛 🖢 🖡			-
		///////			
		1.4 1.1	۲	4 75 5 1	
		2.06	10.1	222222	
			1 1		
1 10 10	11 10	9 8 7	6 5	4 3 2 1 0	1 -1 -2

Figure S11: <sup>1</sup>H NMR spectrum (400 MHz, CDCl<sub>3</sub>) of D2

CARBON_01 Mo19		6.0	5,22 2,22 2,22 2,22 2,22 2,22 2,22 2,22	20 38	06 06	190
maro		12	2222222222	-61.	45	F180
Parameter	Value		1 1 1		17	[170
1 Data File Name						[170
2 Title	CARBON_01					-160
3 Comment	Mg19					
4 Origin	Varian					-150
5 Owner 6 Site						140
7 Spectrometer	vnmrs					140
8 Author						-130
9 Solvent	cdcl3					
10 Temperature	25.0					-120
11 Pulse Sequence	s2pul					
12 Number of Scans	2000					
13 Receiver Gain	30					-
14 Relaxation Delay	1.0000					-100
15 Pulse Width	0.0000					-
16 Acquisition Time	1.3107					-90
1/ Acquisition Date	2016-10-181 02:36:01					
18 Modification Date	2016-11-081 23:24:14					-80
19 Spectrometer Frequer	100.50					
21 Lowest Frequency	-1420.4					-70
22 Nucleus	190					t and
23 Acquired Size	32768					-60
24 Spectral Size	65536					1
						-50
						40
						-30
						-20
						-10
	daya Manazin (kelenar darik program direk dari kalar) (kalar program direk program direk program direk program	alanda sa jeng ng pang		Alexandra and a second second and an and a second	and a second	umperioral production and the second
						-
230 220 210 20	0 190 180 170 1	60 150	140 130 120 110 100 9	90 80 70 60	50 40 30 20 1	0 0 -10
			t1 (ppm)			

Figure S12: <sup>13</sup>C NMR spectrum (100 MHz, CDCl<sub>3</sub>) of D2



Figure S13: <sup>1</sup>H NMR spectrum (400 MHz, CDCl<sub>3</sub>) of D2-HTL



Figure S14: <sup>13</sup>C NMR spectrum (100 MHz, CDCl<sub>3</sub>) of D2-HTL

Mg23_second		\$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	<b>T C C</b>	8688838	-240
vlg23			فرفريهم	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	
Parameter	Value		¥	¥ ¥P	-2300
1 Data File Name	Value				-2200
2 Title	Mg23_second				-21 00
3 Comment	Mg23				1.000
4 Origin	Varian				-2000
5 Owner					-1900
o Sile 7 Spectrometer	uppore			1 1	-1800
8 Author	wiiiiiis		(		170
9 Solvent	cdcl3				-1700
10 Temperature	25,0				-1600
11 Pulse Sequence	s2pul				-150
12 Number of Scans	32				1.00
13 Receiver Gain	30				-1400
14 helaxation belay 15 Pulse Width	0.0000				-1300
16 Acquisition Time	2 5559				-1200
17 Acquisition Date	2016-09-09T15:09:20				-
18 Modification Date	2016-11-08T23:24:35				-1100
19 Spectrometer Freq	uency 399,89				-1000
20 Spectral Width	6410,3				Looo
21 Lowest Frequency	-805,8				- 500
22 Nucleus 22 Acquired Size	16204				-800
24 Spectral Size	32768				-700
Er opoordi oleo	02100				t con
					-500
					-400
				1 1	-300
					-200
			À		
		<u> </u>	Ť	<u> </u>	10
		2.06	1.02	6.02	L_201
					F-20

Figure S15: <sup>1</sup>H NMR spectrum (400 MHz, CDCl<sub>3</sub>) of D1



Figure S16: <sup>13</sup>C NMR spectrum (100 MHz, CDCl<sub>3</sub>) of D1

Perameter         Value           2 Tils         PROTON_01           3 Comment         Mg24           4 Origin         Valien           5 Owner         Sconter           6.Site         Sconter           7 Tipe Sectorneter         vmm's           8 Author         Sconter           9 Solvent         Cocil           11 Pulse Secuence         s2pul           12 Number of Scans         32           13 Receiver Cain         42           14 Relaxation Delay         1.0000           15 Pulse Widm         0.0000           15 Pulse Widm         0.0000           15 Acquisition Time         2.559           17 Acquisition Date         2016/01-1872.55:13           19 Solectimet Frequency         30.89           20 Socetal Widm         4013           21 Lowest Frequency         40.84           22 Nucleus         11 H           23 Spectral Widm         40.34           24 Spectral Size         22769           25 Nucleus         11 Pulse Scout Action Stree           24 Spectral Size         22769	PROTON_01 Mg24		222 222 222 222 222 222 222 222	222.444.444.444.444.444.444.444.444.444	-5000
1 Data File Name       2 Trite       PROTONLO1       1         3 Comment       Mg24       4       1	Parameter	Value			
2 Tite         PROTON_01         Image: Constant Mp24           3 Commant         Mp24         Origin         Varian           5 Owner         6.Site         Solvert         Image: Constant Mp24           6 Site         Solvert         Image: Constant Mp24         Image: Constant Mp24           7 Spectrometer         virian         Image: Constant Mp24         Image: Constant Mp24           9 Solvent         cdd3         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24           9 Solvent         cdd3         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24           9 Solvent         cdd3         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24           11 Pulse Sequence         s2.00         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24           12 Number OStant         s2.00         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24           13 Receiver: Grade Constant         s2.00         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24           14 Relaxation Date         20 Spectral Mp34         S2 Spectral Mp34         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24         Image: Constant Mp24	1 Data File Name				F
3 Comment         Mg24           4 Origin         Varian           5 Owner	2 Title	PROTON_01			45.00
4 Origin         Varian           5 Owner	3 Comment	Mg24			-4500
5 Owner         6 Site         7 Spectrometer         vmms         7 Spectrometer         vmms           8 Author         9 Solvent         cdcl3         / //         ////         ////         ///         ////	4 Origin	Varian			
6 Site 8 Autor 9 Solventi colci3 10 Temperature 10 Temperature 11 Pulse Sequence s 20u 11 Pulse Sequence s 20u 12 Number of Scans 32 13 Receiver Gain 42 14 Relaxation Delay 1.0000 15 Pulse Width 0.0000 16 Acquisition Time 2.5559 19 Spectrometer Frequency 399.89 20 Spectral Width 64103 21 Lovest Frequency 405.8 22 Nucleus 11H 23 Acquired Size 16384 24 Spectral Size 32768 14 Spectral Size 32768 15 Size 16384 24 Spectral Size 32768 15 Size 32768 16 Size 32768 17 Size 32768 17 Size 32768 18 Size 32 Size 3	5 Owner				
7 Spectrometer       vnmris         9 Solvent       cdcl3         10 Temperature       25.0         11 Pulse Sequence       s2pul         12 Number of Scans       32         13 Receiver Gain       42         14 Relaxation Delay       1.0000         15 Pulse Width       0.0000         16 Acquisition Time       2.558         17 Acquisition Time       2.558         18 Acquisition Time       2.558         19 Spectrometer Frequency       399.80         20 Spectral Width       6410.3         21 Lowest Frequency       405.8         22 Nucleus       14         23 Acquired Size       3884         24 Spectral Size       32706	6 Site				-4000
8 Author	7 Spectrometer	vnmrs	1		
9 Solvent         cdcl3         ////////////////////////////////////	8 Author				-
10 Temperature       25.0       1.7.7       7.7.7	9 Solvent	cdcl3			
11 Pulse Sequence     \$2pul       12 Number of Scans     32       13 Receiver Gain     42       14 Relaxation Delay     1.0000       15 Pulse Width     0.0000       16 Acquisition Time     2.5559       17 Acquisition Date     2016/11-18723:54:13       18 Modification Date     2016/11-08723:24:25       19 Spectrometer Frequency     399.88       20 Spectral Width     6410,3       21 Lowest Frequency     653.8       22 Nucleus     1H       23 Acquired Size     32768	10 Temperature	25.0			
12 Number of Scans       32         13 Receiver Gain       42         14 Relaxation Delay       1,0000         15 Pulse Width       0,0000         16 Acquisition Time       2,5559         17 Acquisition Date       2016-01-16T23:54-13         18 Modification Date       2016-01-16T23:54-13         18 Modification Date       2016-01-16T23:24-25         19 Spectrometer Frequency       399.89         20 Spectral Width       6410.3         21 Lowest Frequency       -605.8         22 Nucleus'       1H         23 Acquired Size       16384         24 Spectral Size       32766	11 Pulse Sequence	s2pul			
13 Receiver Gain       42       4	12 Number of Scans	32			
14 Relaxation Delay       1.0000         15 Pulse Width       0.0000         16 Acquisition Time       2.5559         17 Acquisition Date       2016-101-18723-54:13         18 Modification Date       2016-11-08723-24:25         19 Spectrometer Frequency       980.89         20 Spectral Width       6410.3         21 Lowest Frequency       -605.8         22 Nucleus       1H         23 Acquired Size       16384         24 Spectral Size       32768         14 Agent of the state       14 Agent of the state         19 Agent of the state       14 Agent of the state         19 Spectral Width       6410.3         24 Spectral Size       32768	13 Receiver Gain	42			2000
15 Pulse Width       0.0000         16 Acquisition Time       2.5539         17 Acquisition Date       2016-11-08T23:24:25         19 Spectral Width       6410/3         21 Lowest Frequency       405.8         22 Nucleusi       1H         23 Acquired Size       16384         24 Spectral Size       16384         24 Spectral Size       1640         11 Unwest Frequency       400         11 Lowest Frequency       400         11 Lowest Frequency       400         12 Lowest Frequency       400         12 Lowest Frequency       400         12 Lowest Frequency       400         13 Acquired Size       16384         14 Spectral Size       16384         15 Provide Size       16384         16 Provide Size       16384         16 Provide Size       16384         16 Provide Size       16384         16 Provide Size       16384         17 Provide Size       16384         18 Provide Size       178         19 Provide Size       178         19 Provide Size       188         19 Provide Size       188         10 Provide Size       188 <t< td=""><td>14 Relaxation Delay</td><td>1.0000</td><td></td><td></td><td></td></t<>	14 Relaxation Delay	1.0000			
16 Acquisition Time       2.5559         17 Acquisition Date       2016-01-18T23:54:13         18 Modification Date       2016-11-08T23:24:25         19 Spectrometer Frequency       399.89         20 Spectral Width       6410.3         21 Lowest Frequency	15 Pulse Width	0.0000			
17 Acquisition Date       2016-01-18T23:54:13         18 Modification Date       2016-01-18T23:24:25         19 Spectral Width       6410,3         21 Lowest Frequency       -905.8         22 Nucleus       1H         23 Acquired Size       15394         24 Spectral Size       32768	16 Acquisition Time	2,5559			
18 Modification Date       2016+11-08T23:24:25         19 Spectral Width       6410(3)         21 Lowest Frequency       405,8         22 Nucleus       1H         23 Acquired Size       16384         24 Spectral Size       32768         10       10         11       10         11       10         11       10         11       10         12       10         13       10         14       10         15       10         16       10         17       10         18       10         19       10         10       10         10       10         11       10         12       10         13       10         14       10         15       10         16       10         17       10         18       10         19       10         10       10         10       10         10       10         10       10         10       10	17 Acquisition Date	2016-01-18T 23:54:13			-2500
19 Spectrometer Frequency 399.89       20 Spectral Width       6410.3         21 Lowest Frequency       -805.8       22 Nucleus         21 Lowest Frequency       -805.8       -         22 Nucleus       11 H       -         23 Acquired Size       16384       -         24 Spectral Size       32768       -         10 Mither       -       -         11 Mither       -       -         12 Nucleus       -       -         13 Acquired Size       16384         24 Spectral Size       32768         11 Mither       -         12 Mucleus       -         13 Mither       -         14 Spectral Size       32768         15 Mither       -         16 Mither       -         17 Mither       -         18 Mither       -         19 Mither<	18 Modification Date	2016-11-08T 23:24:25			
20 Spectral Width 6410,3 21 Lowest Frequency -005,8 22 Nucleus 1H 23 Acquired Size 15394 24 Spectral Size 32769	19 Spectrometer Freque	ancy 399.89			-
21 Lowest Frequency     -805,8       22 Nucleus     1H       23 Acquired Size     16394       24 Spectral Size     32768	20 Spectral Width	6410.3			
11         14           12         1000           14         1000           14         1000           15         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           14         1000           15         1000           15         1000           15         1000           15         1000           15         1000           16         1000           17         1000           16         1000           17         1000           16         1000           17         1000           16         1000           17         1000           16         1000           17	21 Lowest Frequency	-905.8			
23 Acquired Size 23 Acquired Size 24 Spectral Size 32769	22 Nucleus	1H			
	23 Acquired Size	16384			1
	24 Spectral Size	32768			1500
	24 0000141 0120	GEFOO			-1500
					ľ
					ŀ
					-500
			e et 1 1		
			<u>4 27</u> # 4	X 27 1 27 77	-
			2.0 110	2 7 7 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	
	4 12 12	11 10			
4 13 12 11 10 3 0 7 0 5 4 3 2 1 0 -1 -2 f1[nnm]	4 13 12	11 10	3 0 / 5 f1(nnm	54321U-1	-2

Figure S17: <sup>1</sup>H NMR spectrum (400 MHz, CDCl<sub>3</sub>) of D1-HTL

CARBON_01			N 98 85 5 N 9			100
Mg24		15	2 2 3 3 3 3 3 3 3	50. Z 19. 4	20230000000000000000000000000000000000	-130
Parameter	Value	Ī	14-775	ĬĬ	IV SSV	
1 Data File Name						-120
2 Title	CARBON_01					
3 Comment	Mg24					
4 Origin	Varian					
5 Owner						
6 Site						100
7 Spectrometer	vnmrs					-100
8 Author						-
9 Solvent	cdcl3					Lon
10 Temperature	25.0					
11 Pulse Sequence	s2pul					
12 Number of Scans	2000					-80
13 Receiver Gain	30					
14 Relaxation Delay	1.0000					
15 Pulse Width	0.0000					-70
16 Acquisition Time	1.3107					
17 Acquisition Date	2016-10-18T03:55:44					
18 Modification Date	2016-11-08T23:24:26					-60
19 Spectrometer	100.56					
Frequency						
20 Spectral Width	25000,0					-50
21 Lowest Frequency	-1439,4					
22 Nucleus	13C					40
23 Acquired Size	32768					-40
24 Spectral Size	65536					
						-30
						-20
						-10
n menering and a state of the	i i sudi na kajanjar 1 ja jugi Unita janjan kajanjar kaja kaja kaja kaja kaja kaja kaja k	an n'i Marine ( initia de la Constanti de la C	an far sam fan de fan fan de fan	enen internitiva, nontras variante anticipation de la construction de la construcción de la construcción de la	a log til De prodesse ken verken det og blev kryf højle de sek skyldere han stære og sek sek sek sek sek sek s E	AND IN THE AND A DESCRIPTION OF A DESCRIPTION OF A DESCRIPTION OF A DESCRIPT OF A DESCRIPTION OF A DESCRIPA DESCRIPTION OF A DESCRIPTION OF A DESCRIPTION OF A
						-
230 220 210 20	190 180 170 1	60 150	140 130 120 110 100 90	80 70 60 5		0 -10
200 220 210 20	55 ,50 100 170 1	00 100	f1 (ppm)		, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	5 10

Figure S18: <sup>13</sup>C NMR spectrum (100 MHz, CDCl<sub>3</sub>) of D1-HTL

#### **Reference in Supporting Information**

(1) Otwinowski, Z.; Minor, W. In *Methods Enzymol.*; Charles W. Carter, Jr., Ed.; Academic Press: 1997; Vol. Volume 276, p 307.

(2) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658.

(3) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. D* **2010**, *66*, 486.

(4) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D* **2010**, *66*, 213.

(5) Keyur, D. B.; Hrishikesh, S. G.; Bharat, A. M.; Disha, J. V/; Debdepp, M.; Vinod K. J. *J Fluoresc*. 2012, 22, 1493