SUPPLEMENTARY MATERIAL

Development of a Dehalogenase-Based Protein Fusion Tag Capable of Rapid, Selective and Covalent Attachment to Customizable Ligands

Lance P. Encell*[,#,1], Rachel Friedman Ohana[#,1], Kris Zimmerman[1], Paul Otto[1], Gediminas Vidugiris[1], Monika G. Wood[1], Georgyi V. Los[1,3], Mark G. McDougall[2], Chad Zimprich[1], Natasha Karassina[1], Randall D. Learish[1,4], Robin Hurst[1], James Hartnett[1], Sarah Wheeler[1], Pete Stecha[1], Jami English[1], Kate Zhao[1,5], Jacqui Mendez[1], Hélène A. Benink[1], Nancy Murphy[1], Danette L. Daniels[1], Michael R. Slater[1], Marjeta Urh[1], Aldis Darzins[1,6], Dieter H. Klaubert[2], Robert F. Bulleit[1], and Keith V. Wood[1]

[1]*Promega Corporation, Madison, WI, USA;* [2]*Promega Biosciences Incorporated, San Luis Obispo, CA, USA*

MATERIALS

Mammalian cell lines were from ATCC; Dulbecco's modified essential medium (DMEM), F12, and fetal bovine serum (FBS) were from Invitrogen; 24-well plates were from Nalge Nunc International, and LT1 transfection reagent was from Mirus Bio. Goat anti-p53 IgG was from R&D Systems; mouse anti-p65 IgG was from BD Biosciences; rabbit anti-Src IgG was from ABR Affinity BioReagents; secondary Cy5 IgGs were from GE Healthcare. Cell-free expression systems were from Promega. Chemicals were from Sigma-Aldrich unless otherwise noted. Vectors and other reagents were from Promega unless otherwise noted.

METHODOLOGIES

Synthesis of Chloroalkane Ligands and Intermediates

6-chlorohexan-1-aminium chloride was prepared by refluxing tert-butyl 6-hydroxyhexylcarbamate in carbon tetrachloride in the presence of triphenylphosphine. The isolated alkyl chloride was treated with HCl (gas) in ether at 0 °C yielding the desired hydrochloride salt.

N-(6-chlorohexyl)-fluorescein-5-carboxamide. To a stirring solution of fluorescein-5(6)-carboxy succinimidyl ester (50 mg, 1.0 x 10^{-4} mol) and 6-chlorohexan-1-aminium chloride[2] (35 mg, 2 x 10^{-4} mol) in 3 ml dry DMF (stored over molecular sieves) was added diisopropylethylamine (350 µl, 2 x 10^{-3} mol). The reaction mixture was allowed to stir for 12 h at room temperature then subjected to preparative HPLC purification. Two separate product peaks were isolated as orange solids. Yield (5 isomer): 20 mg, 41%. ^{1}H NMR (CD_3OD): δ = 8.43 (s, 1H, Ar-4), 8.15 (dd, 1H, Ar-6), 7.31 (d, 1H, Ar-7), 6.71 (s, 2H, 1',8'Xan), 6.69 (s, 2H, 4',5'Xan), 6.56 (dd, 2H, 2',7'Xan), 3.57 (t, 2H, C$\underline{H_2}$-N), 3.44 (t, 2H, C$\underline{H_2}$-Cl), 1.79 (m, 2H, -C$\underline{H_2}$-), 1.68 (m, 2H, -C$\underline{H_2}$-), 1.48 (m, 4H, -C$\underline{H_2}$-) ppm. MS: m/z calculated for $C_{27}H_{23}ClNO_6^-$: 492.12(100%), 493.12(29.6.%). Found: 492. 15, 493.29.

N-(6-(3-chloropropylamino)-6-oxohexyl)-fluorescein-5 (and 6)-carboxamide. To a stirring solution 6-(fluorescein-5(6)-carboxamido)hexanoic acid, succinimidyl ester (50 mg, 8.5 x 10^{-5} mol) and 3-chloroproplyamine hydrochloride (22 mg, 1.7 x 10^{-4} mol) in 3 ml dry DMF (stored over molecular sieves) was added diisopropylethylamine (300 µl, 1.7 x 10^{-3} mol). The reaction mixture was allowed to stir for 12 h at room temperature and then subjected to preparative HPLC purification. A single product peak was isolated yielding an orange solid consisting of both 5 and 6 isomers. Yield: 46 mg, 81%. ^{1}H NMR (5 isomer, 57%, CD_3OD): δ = 8.41 (s, 1H, Ar-4), 8.12 (dd, 1H, Ar-6), 7.30 (d, 1H, Ar-7), 6.68 (d, 2H, 4',5'Xan), 6.61 (dd, 2H, 1',8'Xan), 6.54 (dt, 2H, 2',7'Xan), 3.56 (t, 2H, C$\underline{H_2}$-N), 3.53 (t, 2H, C$\underline{H_2}$-N), 3.43 (t, 2H, C$\underline{H_2}$-Cl), 2.22 (t, 2H, C$\underline{H_2}$-C(O)), 2.15 (t, 2H, C$\underline{H_2}$-C(O)), 1.68 (m, 4H, -C$\underline{H_2}$-), 1.56 (m, 2H, -C$\underline{H_2}$-), 1.42 (m, 2H, -C$\underline{H_2}$-) ppm; (6 isomer, 43%, CD_3OD): 8.10 (dd, 2H, Ar-4,5), 7.60 (s, 1H, Ar-7), 6.68 (d, 2H, 4',5'Xan), 6.61 (dd, 2H, 1',8'Xan), 6.54 (dt, 2H, 2',7'Xan), 3.56 (t, 2H, C$\underline{H_2}$-N), 3.53 (t, 2H, C$\underline{H_2}$-N), 3.43 (t, 2H, C$\underline{H_2}$-Cl), 1.93 (t, 2H, C$\underline{H_2}$-C(O)), 1.89 (t, 2H, C$\underline{H_2}$-C(O)), 1.68 (m, 4H, -C$\underline{H_2}$-), 1.56 (m, 2H, -C$\underline{H_2}$-), 1.42 (m, 2H, -C$\underline{H_2}$-) ppm. MS: m/z calculated for $C_{30}H_{30}ClN_2O_7^+$: 565.2(100%), 566.2(33.1%), 567.2 (32.2%). Found: 565.3, 566.3, 567.3.

Tert-Butyl 2-(2-(4-chlorobutoxy)ethoxy)ethylcarbamate. To tert-butyl 2-(2-hydroxyethoxy)ethylcarbamate (3.5 g, 17.3 mmol) in 20 ml of DMF cooled in an ice bath was added a 60% oil infusion of sodium hydride (0.865 g, 21.6 mmol). The chilled reaction mixture was allowed to stir for

*Address correspondence to this author at the Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711, USA; Tel: 608-274-1181; Fax: 608-298-4818; E-mail: lance.encell@promega.com
#These authors contributed equally to this work.
Present address:
[3]Thermo Fisher Scientific, Life Science Research-Cellomics, Pittsburg, PA, USA
[4]Cellular Dynamics International, Madison, WI, USA
[5]GenScript USA Incorporated, Piscataway, NJ, USA
[6]Gas Technology Institute, Des Plaines, IL, USA

1875-3973/12

2 h. The slurry was then transferred by use of a wide bore canula into a 250 ml ice cooled flask containing 1-chloro-4-iodobutane (2.65 ml, 21.6 mmol) in 5 ml dry DMF under nitrogen. The resulting mixture was stirred for 4 h on ice. The reaction mixture was then stripped of DMF, redissolved in dichloromethane and the resulting suspension filtered through celite. The clear solution was next evaporated in the presence of silica gel and subjected to flash chromatography (30% ethyl acetate in heptane). A clear oil was collected. Yield: 2.2 g, 43%. ^{1}H NMR ($CDCl_3$): δ = 4.95 (bs, 1H exchangeable, N<u>H</u>), 3.58(m, 10H, C<u>H</u>$_2$-O/N), 3.32 (t, 2H, C<u>H</u>$_2$-Cl), 1.85 (m, 2H, -C<u>H</u>$_2$-), 1.74 (m, 2H, -C<u>H</u>$_2$-), 1.44 (d, 9H, -C<u>H</u>$_3$) ppm. MS: m/z calculated for $C_{13}H_{27}ClNO_4^+$: 296.2(100%), 297.2(14.5%), 298.2(32%). Found: 296.2, 297.1, 298.1.

2-(2-(4-Chlorobutoxy)ethoxy)ethylammonium chloride. t-Butyl 2-(2-(4-chlorobutoxy)ethoxy)ethylcarbamate (1 g, 3.4 mmol) in 20 ml of Et_2O was cooled in an ice bath. HCl (g) was slowly bubbled into the stirring solution for 15 min where upon the flask was stoppered. The reaction mixture was stirred cold for an addition 45 min, after which all volatiles were evaporated under reduced pressure. An off-white waxy solid was isolated in quantitative yield. ^{1}H NMR ($CDCl_3$): δ = 8.34 (bs, 3H exchangeable, N<u>H</u>$_3^+$), 3.58(m, 10H, C<u>H</u>$_2$-O/N), 3.25 (bs, 2H, C<u>H</u>$_2$-Cl), 1.85 (bs, 2H, -C<u>H</u>$_2$-), 1.74 (bs, 2H, -C<u>H</u>$_2$-) ppm. ^{13}C NMR ($CDCl_3$): δ = 70.58 (-<u>C</u>H$_2$-O), 70.07(-<u>C</u>H$_2$-O), 45.05 (-<u>C</u>H$_2$-Cl), 40.06 (-<u>C</u>H$_2$-NH_3^+), 29.43 (-<u>C</u>H$_2$-), 26.93 (-<u>C</u>H$_2$-) ppm. MS: m/z calculated for $C_8H_{19}ClNO_2^+$: 196.1(100%), 197.1(9.1%), 198.1(32.4%).

Found: 196.2, 197.0, 198.1.

N{2-[2-(4-Chlorobutoxy)ethoxy]ethyl}-fluorescein-5(and 6)-amide. To a stirring solution of fluorescein-5(and 6)-carboxy succinimidyl ester (26 mg, 5.5 x 10^{-5} mol) and 2-(2-(4-chlorobutoxy)ethoxy)ethylammonium chloride (19 mg, 8.2 x 10^{-5} mol) in 1 ml dry DMF (stored over molecular sieves) was added diisopropylethylamine (140 µl, 8.2 x 10^{-4}). The reaction mixture was allowed to stir for 12 h at room temperature. Preparatory HPLC yielded two separable peaks which were isolated as orange solids. Yield (5 isomer): 10 mg, 33%. ^{1}H NMR (CD_3OD): δ = 8.45 (s, 1H, Ar-4), 8.16 (dd, 1H, Ar-6), 7.31 (d, 1H, Ar-7), 6.78 (d, 2H, 1',8'Xan), 6.68 (d, 2H, 4',5'Xan), 6.57 (dd, 2H, 2',7'Xan), 3.66 (m, 8H, C<u>H</u>$_2$-O), 3.52 (m, 4H, C<u>H</u>$_2$-N/Cl), 1.79 (m, 2H, -C<u>H</u>$_2$-), 1.69 (m, 2H, -C<u>H</u>$_2$-) ppm. MS: m/z calculated for $C_{29}H_{29}ClNO_8^+$: 554.2(100%), 555.2(32.4%), 556.2(38.25%). Found: 554.6, 555.2, 556.0. Yield (6 isomer): 15.5 mg, 51%. ^{1}H NMR (CD_3OD): δ = 8.10 (s, 2H, Ar-4,5), 7.65 (s, 1H, Ar-7), 6.82 (d, 2H, 1',8'Xan), 6.68 (d, 2H, 4',5'Xan), 6.59 (dd, 2H, 2',7'Xan), 3.55 (m, 10H, C<u>H</u>$_2$-O/N), 3.42 (m, 2H, C<u>H</u>$_2$-Cl), 1.73 (m, 2H, -C<u>H</u>$_2$-), 1.61 (m, 2H, -C<u>H</u>$_2$-) ppm. MS: m/z calculated for $C_{29}H_{29}ClNO_8^+$: 554.2(100%), 555.2(32.4%), 556.2(38.25%). Found: 554.7, 555.2, 556.1.

2-[2-(6-chloro-hexyloxy)-ethoxy]-ethyl-amine. A 60% dispersion of sodium hydride in mineral oil (1.77 g, 44.2 mmol) was added into a 50 ml centrifuge tube. Heptane (~20 ml) dried over molecular sieves was added, the capped tube was vortexed, and centrifuged for 5 min. After stopping, excess solvent was decanted and the above process repeated two more times. The remaining solid was dried under high vacuum for 30 min. The dry sodium hydride was suspended in dry DMF (30 ml) and this suspension transferred to a 100 ml flask under nitrogen. The flask was cooled by means of an ice bath and 2-(2-aminoethoxy)ethanol (3.7 ml, 36.8 mmol) was added slowly via syringe. The chilled reaction mixture was allowed to stir for 45 min. The slurry was transferred by use of a wide bore canula into a 250 ml ice-cooled flask containing 1-chloro-6-iodohexane (10 g, 40.5 mmol) in 30 ml dry DMF under nitrogen. The resulting mixture was stirred for 4 h on ice. The suspension was filtered through celite and quickly stripped of DMF. The residue was dissolved in methylene chloride and washed with a saturated solution of bicarbonate, followed by two washes with water. The organic layer was dried with sodium sulfate, filtered, and the methylene chloride quickly removed by rotoevaporation. During solvent removal the bath temperature was maintained at less than 30 ^{o}C. The resultant oil was cooled in a bath containing dry ice and isopropanol and left under high vacuum for 2 h. Yield: 5.3 grams, 23.7 mmol, 64% yield).

N{6-[2-(2-(6-chlorohexyloxy)ethoxy)ethylamino]-6-oxohexyl}-fluorescein-5(and 6)-amide. To a stirring solution of 6-(fluorescein-5(6)-carboxamido)hexanoic acid, succinimidyl ester (50 mg, 8.5 x 10^{-5} mol) and 2-[2-(6-chlorohexyloxy)-ethoxy]-ethyl-ammonium chloride (40 mg, 1.7 x 10^{-4} mol) in 3 ml dry DMF was added diisopropylethylamine (300 µl, 1.7 x 10^{-3}). The reaction mixture was allowed to stir for 12 h at room temperature. Preparatory HPLC yielded two separable peaks which were isolated as light orange solids. Yield (5 isomer): 36 mg, 61%. ^{1}H NMR (CD_3OD): δ = 8.44 (d, 1H, Ar-4), 8.12 (dd, 1H, Ar-6), 7.30 (d, 1H, Ar-7), 6.78 (d, 2H, 1',8'Xan), 6.67 (d, 2H, 4',5'Xan), 6.57 (dd, 2H, 2',7'Xan), 3.54 (m, 8H, C<u>H</u>$_2$-O), 3.44 (m, 4H, C<u>H</u>$_2$-N), 3.34 (t, 2H, C<u>H</u>$_2$-Cl), 2.23 (t, 2H, C<u>H</u>$_2$-C(O)), 1.69 (m, 4H, -C<u>H</u>$_2$-), 1.56 (m, 4H, -C<u>H</u>$_2$-), 1.39 (m, 6H, -C<u>H</u>$_2$-) ppm. MS: m/z calculated for $C_{29}H_{29}ClNO_8^+$: 695.3(100%), 696.3(40.9%), 697.3(32.3%). Found: 695.5, 696.5, 697.2. Yield (6 isomer): 16.5 mg, 28%. ^{1}H NMR (CD_3OD): δ = 8.09 (d, 2H, Ar-4,5), 7.65 (s, 1H, Ar-7), 6.83 (d, 2H, 1',8'Xan), 6.67 (d, 2H, 4',5'Xan), 6.58 (dd, 2H, 2',7'Xan), 3.54 (m, 8H, C<u>H</u>$_2$-O), 3.44 (m, 4H, C<u>H</u>$_2$-N), 3.34 (t, 2H, C<u>H</u>$_2$-Cl), 2.16 (t, 2H, C<u>H</u>$_2$-C(O)), 1.69 (m, 4H, -C<u>H</u>$_2$-), 1.56 (m, 4H, -C<u>H</u>$_2$-), 1.39 (m, 6H, -C<u>H</u>$_2$-) ppm. MS: m/z calculated for $C_{29}H_{29}ClNO_8^+$: 695.3(100%), 696.3(40.9%), 697.3(32.3%). Found: 695.4, 696.3, 697.1.

dhaA Cloning and Vectors

Wild type *dhaA* was subcloned from pET-3a [1, 2] into the T7 promoter-driven GST fusion vector, pGEX-5X3 (GE Healthcare), using *Sal*I and *Not*I. A sequence encoding the FLAG peptide (DYKDDDDK) was introduced to the 3' end of *dhaA* using *Age*I and *Eco*47III. The pGEX-5X3-derived linker peptide between GST and DhaA contained a Factor Xa cleavage site for the removal of GST from the fusion following affinity purification. For expression in mamma-

lian cells under the CMV promoter, *dhaA* or mutant versions were subcloned from pGEX-5X3 into pCI-neo (Promega) using *Sal*I and *Not*I. Modified versions of bacterial T7-based Flexi vectors (Promega) were constructed to enable convenient transfer of HT2 or subsequently derived variants to either terminus of different fusion partners. pF1K(+) was the initial vector encoding variants at the N-terminus of specific partners. This enabled their convenient transfer to different partners (also in the context of pF1K(+) using *Sgf*I and *Ssp*I or transfer to the vector, pFN2m, to result in fusions where HT2 or subsequent variants were located at the C-terminus of a partner protein. The same sequence could be moved to the vector pFN2m (or pFN2mA) as the C-terminal partner sequence using *Sgf*I and *Pme*I. Variant sequences could be moved from pFN2m-based plasmids to pF1K(+) using *Sgf*I and *Ssp*I.

Expression in CHO-K1 Cells

Cells were maintained in DMEM/F12 media supplemented with 10% FBS at 37 ^{o}C/5% CO_2. Cells were plated on 24-well plates at a density of 25 x 10^3 cells/cm^2 in growth medium and allowed to grow to ~85% confluency (24–48 h). Cells were transiently transfected with H272F (pCI-neo) using LT1 reagents as previously described [3]. For TMR-ligand incubations with cells, media was replaced with growth media containing the ligand and incubated at 37 ^{o}C/5% CO_2. Cells were washed with PBS and processed for SDS-PAGE and fluorescence analysis.

GST-based Affinity Purification

GST fusions were purified from bacterial lysates using Glutathione Sepharose 4 Fast Flow Resin (GE Healthcare) as previously described [3]. The GST tag was removed by treatment with Factor Xa according to the manufacturer's protocol (GE Healthcare).

Variants in the context of pFN2mA (GST) were overexpressed overnight at 25 ^{o}C by autoinduction in *E. coli* KRX [4]. Induced cultures were harvested and processed for affinity purification (Glutathione Sepharose 4 Fast Flow, GE Healthcare) as previously described [3]. Purification was carried out using an AKTA Explorer FPLC with a XK16 column (GE Healthcare), a wash buffer consisting of PBS (pH 6.8) + 300 mM NaCl + 1 mM DTT), and an elution buffer consisting of 50 mM Tris (pH 8) + 50 mM glutathione. Desired fractions were dialyzed (20 mM HEPES pH 7.5) overnight and purified further using Source 30Q resin (GE Healthcare) and FPLC. 20 mM HEPES (pH 7.5) was used as the wash buffer, and target protein was eluted (NaCl gradient (0–0.5 M) in 20 mM HEPES, pH 7.5; fractions tested by gel), dialyzed (PBS + 20% glycerol + 1 mM DTT), and quantitated by Bradford assay (Pierce). For liberating target proteins from GST the same purification protocol was followed, however the glutathione elution step was replaced by proteolytic cleavage of resin-bound GST fusion protein using 1,000 units of ProTEV (Promega) in 50 mM HEPES (pH 7.3) containing 1 mM DTT and 0.5 mM EDTA. ProTev was subsequently removed during the Source 30Q binding and washing steps.

Cellular Imaging of HT2

CHO-K1 cells were maintained as described in the main text. Cells were transiently transfected with HT2 (pCI-neo) using LT1 reagents based on the manufacturer's protocol. To label cells, media was replaced with growth media containing the TMR-ligand and incubated at 37 ^{o}C/5% CO_2. Cells were washed with PBS and fixed using 4% paraformaldeyde and then imaged by inverted confocal microscopy (Olympus FV500; 543 nm Ar/Kr laser; 570 nm emission filter).

Capturing HT2 to a Chloroalkane Surface

An ELISA-based approach was developed and used to determine whether HT2 could be captured to an immobilized chloroalkane. Briefly, a 96-well streptavidin coated plate (Reacti-Bind HBC; Pierce) was incubated with 3 μM PEG Biotin-ligand (300 pmol used in order to saturate the streptavidin; ~100 pmol) in PBS + 0.1% CHAPS + 0.5% BSA (PBSCB) for 2 h at 25 ^{o}C. The plate was emptied, washed 4x with PBS + 0.1% Tween 20 (PBST), and then incubated with affinity purified HT2 (diluted in PBSCB) for 2.5 h at 25 ^{o}C. The plate was next emptied, washed 4x with PBST, and incubated with rabbit anti-HaloTag pAB (Promega) diluted 1:50,000 in PBSCB for 1 h at 25 ^{o}C. The plate was emptied, washed 4x with PBST, and incubated with an anti-rabbit IgG-HRP conjugate (diluted in PBSCB) for 1 h at 25 ^{o}C. Finally, the plate was emptied, washed 8x with PBST, developed with 3,3′,5,5′-tetramethylbenzidine (TMB) and 0.2 M H_2SO_4 (stop solution), and read at 450 nm using a SpectraMAX PLUS reader (Molecular Devices).

Random Mutagenesis

Error prone PCR was performed using GeneMorph II (Agilent) under conditions that produced on average 2–3 mutations per kb. For the first round of mutagenesis, the target sequence, codon-optimized HT2(GNF) was amplified using 100 ng of template DNA and 22 cycles. The PCR was digested with *Sgf*I and *Eco*ICR1 and ligated into the vector pF1K(+) (upstream of Rluc). For the second round of mutagenesis, HT3 was amplified using 10 ng target template and was run for 22 cycles. The PCR was digested with *Nco*I and *Ssp*I and ligated into the vector pFN2m downstream of Rluc, firefly luciferase (Fluc), or Id. Random substitutions at the C-terminus of HT6 (positions 291–295) and in the C-HT7 linker were introduced using a cassette-based approach [5]. The C-terminus library was ligated into pFN2m-Id-HT6 using *Bss*HII and *Bam*HI, and then variants of interest were moved to the N-terminal tag vector, pF1K(+), upstream of a particular partner sequence using *Nco*I and *Acc*III. C-terminus variants were screened for improved functional expression by labeling samples to completion with the TMR-ligand (20 μM, 1 h at 25 ^{o}C) and determining the level of active fusion protein by SDS-PAGE and fluorescence scanning. C-HT7 linker cassettes were introduced to pFN2m (variants fused downstream of Id) using either *Xho*I or *Hin*dIII and *Bam*HI. Variants were screened for improved linker properties using bacterial lysates.

Mass Spectrometry Analysis of Labeled H272F, HT2, and HT7

Molecular weight determination of proteins was performed at the Mass Spectrometry Facility (Biotechnology Center, University of Wisconsin-Madison) using a 4800 MALDI TOF-TOF mass spectrometer (Applied Biosystems). 150 μg of GST affinity purified protein (in 300 μl of 20 mM Tris-HCl, pH 7.5) was incubated with or without 100 μM TMR-ligand (molar excess) for 1 h at 25 ^{o}C, and reactions precipitated with ice-cold acetone. Samples were then washed in 50% acetone and a portion of the generated pellet spotted directly onto an Opti-TOF™ 384-well plate (Applied Biosystems), re-crystallized with 1μl of matrix (10 mg/ml α-cyano-4-hydroxycinnamic acid in acetonitrile/H_2O /TFA (70%:30%:0.2%)) and run in positive linear mode for MW determination.

For peptide mapping by "in liquid" digestion, proteins were resolubilized in 20 μl of 6 M urea (in 100 mM NH_4HCO_3, pH 8.5) and then 2 μl ProteaseMax™ (Promega) added. Next, protein was reduced (5 mM DTT, 10 min, 35 ^{o}C) and then alkylated (9 mM iodoacetamide, 2 min, RT). Trypsin (3.6 μg in 25 mM NH_4HCO_3; Sequence Grade, Promega) was added and the digestions incubated for 1 h at 35 ^{o}C, followed by a second addition of trypsin (2 μg) and additional incubation for 2 h. Reactions (60 μl) were quenched by the addition of 2.5% TFA (60 μl) and freezing at -20 ^{o}C. Peptides were cleaned up and concentrated (C18 SPEC-PLUS™-PT pipette tips, Varian) and then eluted with ACN/H_2O/TFA (70%:25%:0.2%), dried by speed-vac, and reconstituted in 30 μl 0.1% formic acid. 30 μg was then analyzed by μLC-MS/MS using a Micromass Q-TOF2 Hybrid Quadrupole/Orthogonal Time of Flight Mass Spectrometer (Waters Corp.) to locate the TMR moiety. Chromatography of peptides prior to mass spectral analysis was accomplished using reverse phase HPLC (C18) from which the eluted species were directly micro-electrosprayed. Columns were made using lengths of fused silica tubing (365 μm OD, 100 μm ID) with pulled tips (1 μm orifice) that were packed to 12 cm with Zorbax Eclipse XDB-C18 (Agilent), 5 μm, 300 Å pore size media. An Agilent 1100 series HPLC delivered solvents A: 0.1% (v/v) formic acid in water, and B: 95% (v/v) acetonitrile, 0.1% (v/v) formic acid at either 1 μL/min, to load sample, or 150-200 nL/min, to elute peptides over a 180 min 10% (v/v) B to 70% (v/v) B gradient. Voltage was applied upstream of the column through a platinum wire electrode into the fluid path via a PEEK T-junction. As peptides eluted from the HPLC-column/electrospray source, MS/MS spectra were collected from 400 to 2200 m/z; redundancy was limited by dynamic exclusion. Collision energy profiles were empirically pre-determined for different peptide charge states. MS/MS data were converted to pkl file format using Micromass Protein Lynx Global Server v2.1.5 (Waters Corp.). Resulting pkl files were used to search a user defined database with construct sequences and common contaminant proteins using Mascot Search Engine (Matrix Science) with methionine oxidation, glutamic and aspartic acid deamidation and TMR modification (714.326 Da for modified Asp residue) as variable modifications. Putative modifications identified by Mascot were confirmed using manual assignments of MS/MS spectra.

Computational Structure Models For DhaA + Asn^{272} and HT7

Molecular modeling and visualization of HT2 containing Asn at position 272 (Fig. **S3**) was performed using Insight II software (Accelrys Software Inc.). The structure model for HT7 was built using Discovery Studio software (Accelrys Software Inc.). In generating the HT7 model we first built (with DS MODELER) a homology model of HT6 based on the X-ray crystallographic structure of *Rhodococcus* DhaA (PDB code 1BN6). Then the HT7 amino acid changes (Table **S1**) were introduced to the C-terminus of the HT6 model. Energy minimization was performed on residues 291-297 (no constraints) and exposed residue side chains within 12 Å of the C-terminus (harmonic constraint on backbone atoms).

Expression and Labeling of Protein in HeLa Cells

HeLa cells were maintained in DMEM/F12 media containing FBS (10%) at 37 ^{o}C and 5% CO_2. Cells at ~85% confluency (24–48 h) were transfected with a P65-HT or P65-HT7 construct (pF4-based, CMV promoter, Promega) using LT1 reagent. For cells to gel analysis, growth media was replaced with media containing 0.2 μM TMR-ligand and the cells incubated for 2 h (37 ^{o}C, 5% CO_2). Cells were washed and then lysed in SDS gel loading buffer, boiled for 5 min, analyzed by SDS-PAGE, and scanned for fluorescence (E_{ex}/E_{em} = 532/580 nm) using a Typhoon 9400 and ImageQuant software from Amersham Biosciences. For imaging, growth media was replaced with media containing 5 μM TMR-ligand and the cells incubated for 15 min (37 ^{o}C, 5% CO_2). Cells were washed and imaged using inverted confocal microscopy (Olympus FV500; 543 nm Ar; 570 nm emission filter).

Vector Transfers

Bacterial expression vectors containing a T7 promoter were suitable for use with the TNT® T7 or SP6 Coupled Reticulocyte Lysate Systems. For use with the TNT SP6 High-Yield Protein Expression System (wheat germ), fusion partners from modified bacterial vectors were transferred to modified pFN19 vectors (SP6 promoter, Promega) using *Sgf*I and *Pme*I. For expression in mammalian cells, fusion partners were also transferred to pF4 CMV vectors (Promega) using *Sgf*I and *Pme*I.

Expression and Labeling of Protein in Cell-Free Lysates

In vitro expression of HT2 or HT7 (optimized linker) fused to the N-terminus of different partner proteins was performed using either the TNT T7 Coupled Reticulocyte Lysate System or the TNT SP6 High-Yield Protein Expression System (wheat germ) according to the manufacturer's protocol. Expression reactions were labeled with 1 μM TMR-ligand for 1 h, analyzed by SDS-PAGE, and scanned for fluorescence as previously described for the HeLa lysates.

Alternative Expression Protocol in Bacteria (Autoinductions)

Inductions were carried out in a similar fashion to the bacterial screen described in the main text, except that the volumes were scaled up to 2 ml and the media used was LB. In addition, 0.05% glucose was used instead of 0.025% to

provide a similar delay for the onset of induction in richer media.

Expression and Analysis of Src, Luc, p65, and p53 as HT7-Tagged and Untagged Protein

These sequences were expressed in the absence of tag using the vectors pF1 (*E. coli*), pF4 (mammalian cells), or pTs1 (TNT SP6 High-Yield/wheat germ). HT7 fusions were expressed using the vectors pFN18 for *E. coli*, pFN21 (N-terminal HT7) or pFC14 (C-terminal HT7) for mammalian cells, or pFN19 (N-terminal HT7) or pFC20 (C-terminal HT7) for TNT SP6 High-Yield/wheat germ. Bacterial expression was carried out using an autoinduction protocol (0.2% rhamnose, 0.05% glucose; 16 h at 25 ^{o}C) previously described [4]. For mammalian cell expression experiments, HEK-293T cells maintained in DMEM media containing FBS (10%) were grown to 70–80% confluency (24 h) and transfected with the appropriate construct using LT1 reagent. For the TNT SP6 High-Yield/wheat germ experiments, 8 μg plasmid DNA was added to 50 μl cell-free lysates and incubated for 2 h at 25 ^{o}C. Expression for all systems was examined by SDS-PAGE and quantitated by fluorescent Western analysis using a Typhoon 9400 scanner (E_{ex}/E_{em} = 633/670 nm) and ImageQuant software (GE Healthcare).

Effect of pH on HT7 Activity

The activity of affinity purified HT7 (see main text for overexpression and purification procedures) at different pH values was determined from the initial rate of binding to the FAM-ligand using FP. Reactions (16.5 nM protein, 7.5 nM FAM-ligand) were performed using a buffer system suitable for maintainig constant ionic strength of 150mM from pH 4–9 [6]. pH adjustments were made using acetic acid and tetramethyl ammonium hydroxide.

Effect of NaCl on HT7 Activity

The activity of affinity purified HT7 in the presence of NaCl was determined from the initial rate of binding to the FAM-ligand using FP (E_{ex}/E_{em} = 485/535 nm; Tecan GENios Pro reader). Reactions between HT7 and ligand were monitored over time in 20 mM HEPES (pH 7.5) containing varying concentrations of NaCl.

Effect of Detergents and Other Additives on HT7 Binding Activity

Affinity purified HT7 was exposed to a variety of compounds commonly used as additives in biochemical reactions. In the presence of the additives, HT7 activity was determined from the initial rate of binding to the FAM-ligand using FP (16.5 nM protein, 7.5 nM FAM-ligand).

RESULTS

Molecular Structure Model for Asn^{272} Substitution

We built a molecular model of HT2 containing Asn at position 272, and compared it to previously generated models/structures (Fig. **S3**).

Unlike His^{272}, Asn^{272} was not in close enough proximity to the water molecule for activation. The Asn^{272} substitution did, however, re-establish the hydrogen bond with the nearby Glu^{130}, an interaction that exists in DhaA but is not possible (because of H272F) in HT2. Because of the presumed structural importance of the hydrogen bond to Glu^{130} [2, 7], regaining this interaction may have played an important role in improving stability and expression. In addition to the hydrogen bonding potential, Asn fills a similar geometric space as the original His found in DhaA. This, like the hydrogen binding to Glu^{130}, was not the case with Phe^{272} found in HT2, so it is reasonable to speculate that the structure being provided by the space filling found with His and Asn may be conducive to a more stable molecule.

Linker Development

An effective linker between two proteins should be flexible and long enough to maintain the maximal structural integrity of the two proteins. Another important feature is the presence of a protease recognition site, so that a protein of interest can be liberated from a tag, i.e. purified. The linkers used in the data presented thus far (N-3, C-1; Table **S2**) contained a flexible spacer sequence combined with a modified TEV protease recognition site: Glu(P6)-Asn-Leu-Tyr-Phe-Gln(P1)-Ala(P1′); Note–the Ala replaces native Ser at the P1′ position in order to present an *Sgf*I restriction site for cloning. It is well known that linker sequence and context can have a dramatic impact on the expression of functional fusion protein [8-11]. We therefore tried to engineer a novel linker sequence that would function well in multiple expression systems and ultimately make HT7 an even more robust tag. For N-terminal fusions (HT7 at N-terminus), we first examined shortening linker N-3 by replacing the $(Ser\text{-}Gly_4)_3$ with a single $Ser\text{-}Gly_5$. This linker, N-1 (Table **S2**), provided improved expression and reduced non-specific cleavage for certain fusions such as HT7-Id (Fig. **S5**; compare Fig. **9E**, main text), however it was not cut efficiently by TEV protease. We were concerned that the hydrophobic nature of a three (or four) amino acid stretch Ala(P1′)-Ile-Ala-Met (Met may or may not be present for a protein of interest) could interfere with efficient proteolysis, and therefore introduced rational changes to this region of the linker. Ala(P1′) was changed back to Ser, and then we examined the insertion of various combinations of the hydrophilic residues, Asp and Asn, between the P1′ site and a recreated Ala-Ile-Ala (maintains the *Sgf*I cloning site). Using expression and cleavage efficiency as criteria for identifying improved sequences, we determined that the presence of an Asp-Asn dipeptide downstream of the TEV P1′ Ser offered the best expression and TEV cleavage efficiency (data not shown). We also explored the TEV recognition sequence for optimization. Asn(P5) was changed to Asp based on a previous report showing Asp provided faster TEV cleavage [12]; and finally, the upstream spacer region of the linker (Gly_5) was replaced with part of the natural upstream sequence found in TEV, Glu-Pro-Thr-Thr (Glu replaced native Ile to maintain an *Xho*I cloning site; Pro-Thr-Thr is native to TEV). This final optimized linker sequence, N-HT7 (Table **S2**), provided further improved expression as well as efficient TEV protease cleavage (Fig. **S5**).

Many of the beneficial changes to the N-terminal linker were applied to the parental C-terminal linker, C-1 (Table **S2**) in an attempt to improve on the poor cleavage efficiency

(Fig. **S5**). The Glu-Pro-Thr-Thr sequence was inserted immediately upstream of the TEV site Glu(P6); and TEV site residues Asn(P5) and Ala(P1′) were replaced with Asp and Ser, respectively. The final modification was at the Ser-Gly_5 spacer region downstream of P1′, where we created a library of random substitutions and screened for expression, non-specific cleavage, and TEV protease cleavage efficiency. Using this process, we identified the tripeptide sequence Asp-Asn-Asp, which was incorporated into the final linker (C-HT7; Table **S2**). As shown in Fig. **S4**, this linker provided improved expression and essentially complete cleavage. In addition to improved expression in bacteria, we found that the engineered linkers provided similar benefits to fusions expressed in both cell-free and mammalian cell systems (data not shown).

Characterization of HT7 Expression in HeLa Cells

P65 fusions to the N-terminus of HT2, HT3, or HT7, i.e. p65-HT2, p65-HT3, p65-HT7 (non-optimized linker, pF4) were expressed in HeLa cells, labeled *in vivo* with the TMR-ligand, and then lysed and analyzed by SDS-PAGE/fluorescence scanning (Fig. **S6**). HT7 provided ~6-fold more functional expression compared to HT2. Note the presence of what are believed to be degradation products. These products represent <5% of the total labeled protein, and in subsequent experiments with the optimized linker (C-HT7) these products were essentially undetectable. In a separate experiment, cells expressing either p65-HT2 or p65 HT7 were labeled with the TMR-ligand and imaged (Fig. **S5**). Cells expressing HT7 were significantly brighter than the HT2 cells, presumably due to elevated levels of the fusion protein.

Characterization of HT7 Expression in Wheat Germ Cell-Free Lysates

HT2, HT3, HT6, and HT7 were fused to the N-terminus of 6 different partner proteins and expressed using wheat germ lysates. Lysates were labeled to completion with TMR-ligand (2 μM ligand, 1 h, 25 ^{o}C) and analyzed by SDS-PAGE/fluorescence scanning. All of the variants showed improved expression with respect to HT2. An example is shown in Fig. (**S7**), where HT7 fused to T antigen provided ~20-fold more functional protein than HT2.

Summary of HT7 Expression Improvements in Different Expression Systems

To capture the scope of the general expression benefit provided by HT7 and the optimized linkers, we have summarized the improvements observed for different fusion partners, different orientations (N or C), and different expression systems not shown in the main text (Table **S3**). These data indicate a broad range of improvements provided by HT7 using different scenarios, with *E. coli* experiencing the most consistent level of high magnitude improvement. However, rabbit reticulocyte lysates showed some of the greatest improvements with specific protein partners. More importantly, these results indicate that although the magnitude of the expression benefit may vary for different partners, orientation, and expression systems, the general benefit of HT7 was realized for all scenarios examined.

Effect of HT7 on the Expression of Src, Luc, p65, and p53

Proteins were expressed as downstream fusions (*E. coli*) or both downstream and upstream fusions (HEK-293T cells, Wheat Germ lysates) to HT7 and examined for yield by fluorescent Western analysis. For each protein, the yield was greater than what was observed for the non-tagged control protein. The range of improvements is summarized in Table **S4**. These results demonstrate the potential for HT7 to enhance the expression of proteins that may be difficult to produce at significant levels without the assistance of an effective expression tag.

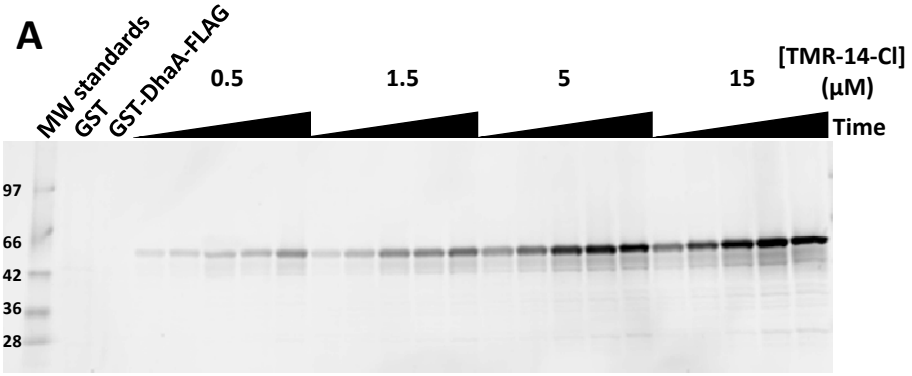
Effect of pH and NaCl on HT7 Binding

We examined the effect of pH on binding kinetics and determined that the reaction rate was maximal from pH 5–7 (Fig. **S10**). The kinetics suffered slightly above pH 8, but were reduced much more dramatically when the pH was below 5. We also examined the impact of NaCl on the binding reaction. The rate of the reaction between HT7 and the FAM and TMR ligands was measured by FP in the presence of up to 2 M NaCl. The results indicated a general NaCl-dependent increase in binding kinetics for the FAM-ligand (Fig. **S11**). This effect may have been due to a reduction in the electrostatic interactions between the negative patches near the entrance of the tag binding tunnel and the negatively charged FAM. We did not observe a similar enhancement in binding to the TMR-ligand in the presence of NaCl, but rather an inhibitory effect (data not shown). This is likely due to increased hydrophobicity of the TMR-ligand in the presence of NaCl.

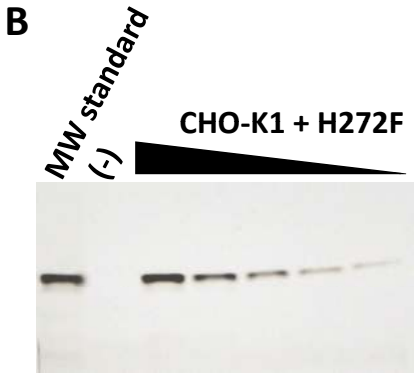
Effect of Additives on HT7 Binding

HT7 was exposed to a variety of different detergents and other common additives and then examined for TMR-ligand binding kinetics to determine the effects of a particular compound on activity. The results of these experiments are summarized in Tables **S5** and **S6**. It should be noted that compounds causing reduced kinetics may slow down a reaction but not necessarily change the amount of protein that eventually binds to ligand. Most notable with regard to detergents was the negative impact of Tween 20, IGEPAL CA-630, and DTAB on activity. Tween 20 and DTAB inhibited binding activity at concentrations below and above their critical micelle concentrations (CMC). IGEPAL CA-630 (0.005%) showed no impact on activity. Other detergents, such as Triton X-100, provided faster binding kinetics (1.5–2-fold). It may also be of interest that Mn^{+}, Mg^{+}, and Zn^{+} provided an improvement to binding kinetics (1.5–2-fold). EDTA by itself did not impact HT7 activity; however it could potentially inhibit activity in situations where divalent cations are present or added to enhance ligand reactivity. Finally, both formaldehyde and para-formaldeyde were tolerated for short periods of time, suggesting that these cross-linking agents can be used in the presence of HT7 prior to ligand binding. In summary, non-charged or zwitterionic compounds appear to be more tolerated than other classes of detergents [13].

SUPPLEMENTARY MATERIAL FIGURES AND TABLES

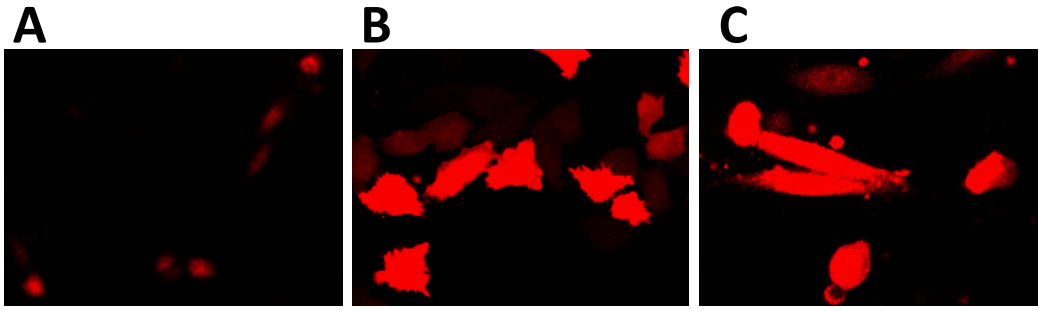


Concentration and time-dependent labeling of H272F. Fluorescence image (E_{ex}/E_{em}=532/580 nm) of SDS-PAGE showing the concentration (0.5–15 μM TMR-ligand) and time (10, 30, 60, 90, 120 min) dependent labeling of bacterial lysates containing overexpressed GST-H272F (62 kDa). No product was detected from 2 h incubations between control lysates (GST, 28 kDa; or GST-DhaA, 62 kDa) and 15 μM TMR-ligand.

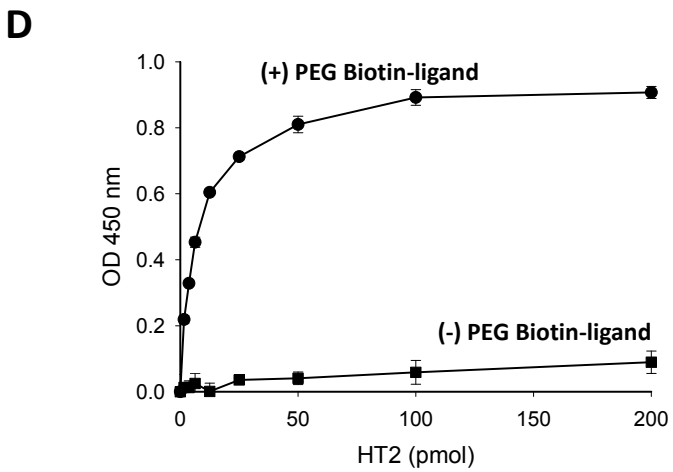


Fluorescence image (E_{ex}/E_{em}=532/580 nm) of SDS-PAGE showing the TMR-ligand labeling of CHO-K1 cells expressing H272F. Lane 1=marker (affinity-purified H272F, 34 kDa). No product was detected from incubations with non-transfected cells ((-), lane 2), while the lysates from cells transfected with H272F (lanes 3–7, 1:2 serial dilutions) demonstrate labeled protein of the expected size.

Fig. (S1).

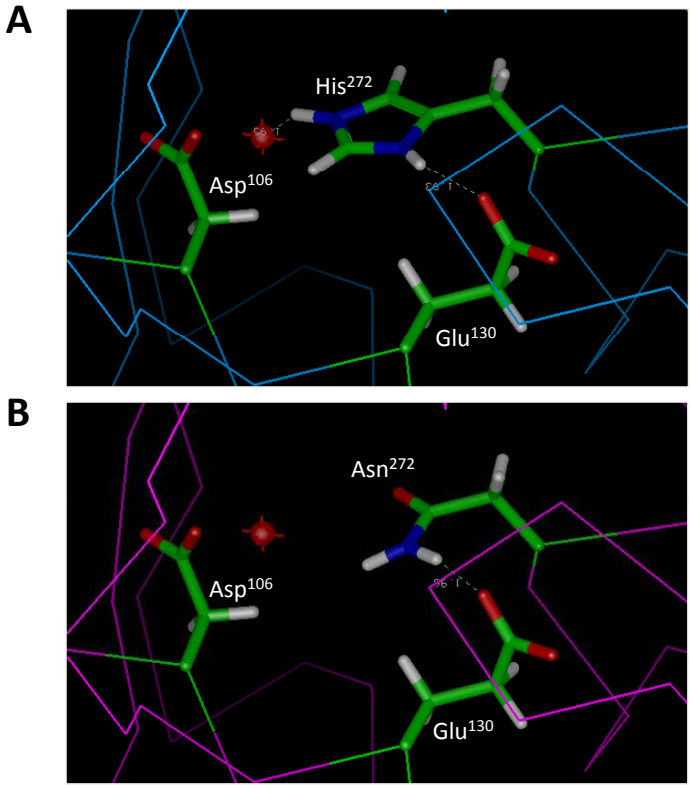


Fixed cell imaging of CHO-K1 cells expressing HT2. CHO-K1 cells transiently transfected with H272F (panel **A**) or HT2 (panels **B**, **C**) were labeled with the TMR-ligand for 30 min at 37 °C, washed 3x with PBS, and fixed prior to imaging with an Olympus FV500 confocal microscope (543 nm Ar/Kr laser + 570 nm emission filter). Panel **A**: H272F, 5 μM ligand; Panel **B**: HT2, 5 μM ligand; Panel **C**: HT2, 0.2 μM ligand.



Capturing HT2 to a surface. A microtiter plated coated with streptavidin was used to immobilize chloroalkane ligand, which was then used to capture affinity purified HT2. The dose dependent capture of HT2 was measured using an anti-HaloTag pAB (n=3).

Fig. (S2).



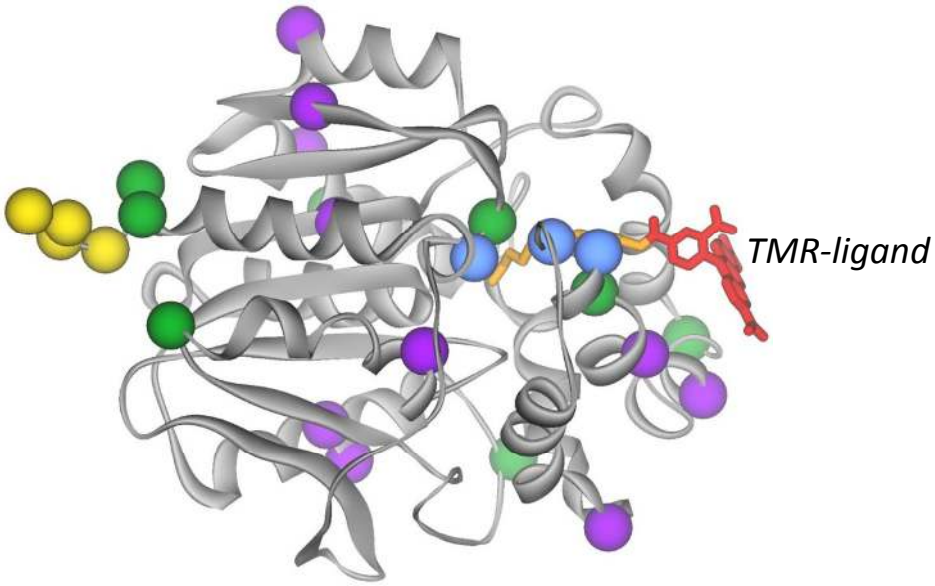
Modeling position 272. Panel **A** indicates hydrogen bonding between His^{272} and the lone water molecule, while panel **B** suggests that Asn is not close enough to the water molecule to form such a bond. Both His^{272} and Asn^{272} fill similar geometric space and can form a hydrogen bond with Glu^{130} that is considered important for structural integrity. Note that Phe^{272} (not shown), does not fill a similar space as His and Asn, and is unable to form a hydrogen bond with Glu^{130}.

Fig. (S3).

Table S1. Summary of Amino Acid Substitutions Introduced to DhaA

Variant	Amino Acid Substitutions
HT2	K175M, C176G, H272F, Y273L
HT3	**HT2** + S58T, D78G, A155T, A172T, A224E, F272N, P291S, A292T, (Gln^{294}, Tyr^{295})[a]
HT6	**HT3** + L47V, Y87F, L88M, C128F, E160K, A167V, K195N, N227D, E257K, T264A
HT7	**HT6** + Q294E, Y295I, (Ser^{296}, Gly^{297})[b]

[a] Gln^{294} and Tyr^{295} were added to the C-terminus of HT3 for the purpose of introducing an *Ssp*I cloning site.
[b] Ser^{296} and Gly^{297} were added to the C-terminus of HT7 for the purpose of introducing an *Acc*III cloning site.



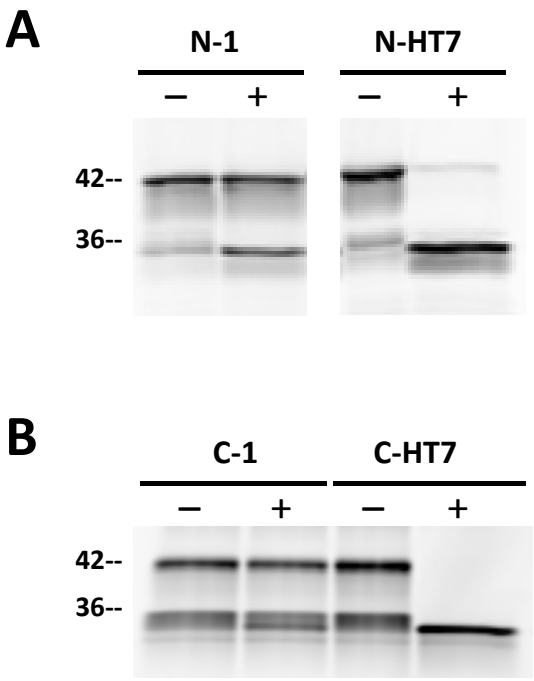
HT7 structure model. The positions of the 25 amino acid changes in HT7 are highlighted in the structure model as follows: HT2, blue; HT3, green; HT6, violet; and HT7, yellow. Note H272N (base of binding tunnel) is shown as green; as this site was originally changed from His to Phe (HT2) and then changed again to Asn (HT3). The approximate location of the TMR ligand is shown (based on a model of HT with bound TMR ligand).

Fig. (S4).

Table S2. Linker Sequences

Linker[a]	Sequence[b]	Length (# of residues)
N-3	(Ser-Gly-Gly-Gly-Gly)$_3$-**Glu-Asn-Leu-Tyr-Phe-Gln-Ala**-Ile-Ala	24
N-1	Ser-Gly-Gly-Gly-Gly-Gly-**Glu-Asn-Leu-Tyr-Phe-Gln-Ala**-Ile-Ala	15
N-HT7	Glu-Pro-Thr-Thr-**Glu-Asp-Leu-Tyr-Phe-Gln-Ser**-Asp-Asn-Ala-Ile-Ala	16
C-1	Val-Ser-Leu-**Glu-Asn-Leu-Tyr-Phe-Gln-Ala**-Ser-Gly-Gly-Gly-Gly-Gly	16
C-HT7	Val-Ser-Leu-Glu-Pro-Thr-Thr-**Glu-Asp-Leu-Tyr-Phe-Gln-Ser**-Asp-Asn-Asp	17

[a] N or C refers to the orientation of the fusion (N, tag on N-terminus; C, tag on C-terminus).
[b] TEV protease recognition site is shown in bold; cleavage occurs between P1 (Gln) and P′ (Ala/Ser).



Linker optimization. Fusions (46 kDa) between HT7 and Id were overexpressed in *E. coli* KRX at 30 °C, and soluble fractions of lysates labeled to completion with the TMR-ligand and incubated with (+) or without (-) TEV protease for 30 min at 30 °C. Reactions were resolved by SDS-PAGE and scanned for fluorescence (E_{ex}/E_{em} = 532/580 nm). Fusions to the N-terminus of Id (panel **A**) contained either linker N-1 or N-HT7, while fusions to the C-terminus of Id (panel **B**) contained either linker C-1 or C-HT7 (Table **S2**). The expected size of the labeled cleavage product, free HT7, is 34 kDa.

Fig. (S5).

Table S3. Expression benefit[a] provided by HT7 (relative to HT2)

Fusion Orientation (Location of Tag)			TNT Cell-free Lysates[c]	
	E. coli[b]	HeLa Cells	Rabbit Reticulocyte	Wheat Germ
N-terminus	38–140 (3)[d]	N.D.[e]	2–175 (9)	2–20 (6)
C-terminus	130 (1)	5–30 (3)	4–140 (3)	N.D.[e]

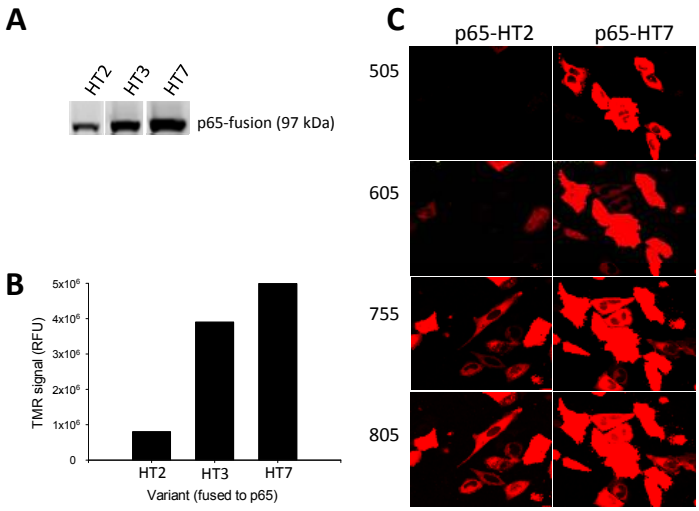
[a] fold improvement over original HT2

[b] expression was induced at 25 °C using autoinduction protocol.

[c] expression was carried out at 30 °C for 90 min (Rabbit reticulocyte) or at 25 °C for 2 h (Wheat germ)

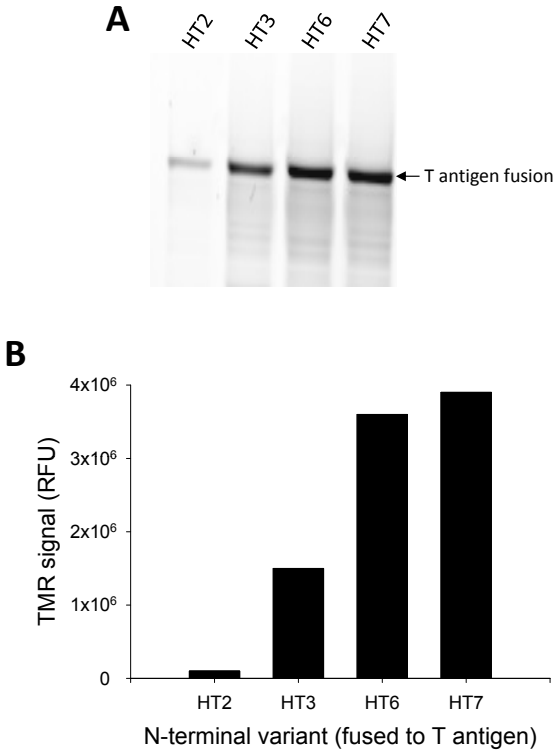
[d] value in parenthesis indicates the number of different partner proteins examined

[e] magnitude of improvement not determined with respect to original HT2



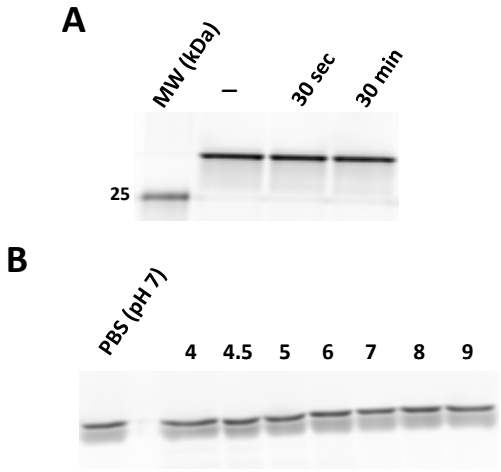
Expression in HeLa cells. Cells expressing p65-HT2, p65-HT3, or p65-HT7 were labeled with 0.2 μM TMR-ligand for 2 h, washed, and then lysed and analyzed by SDS-PAGE. The gel was scanned for fluorescence (E_{ex}/E_{em} = 532/580 nm) and bands representing full length fusion protein (97 kDa) were quantitated. The image is shown in panel **A** and the quantitation in panel **B**. Note the image in panel A shows non-adjacent lanes from the same gel reconfigured for presentation. Cells expressing p65-HT2 or p65-HT7 (panel **C**) were labeled with 5 μM TMR ligand for 15 min, washed, and imaged using inverted confocal microscopy (543 nm Ar; 570 nm emission filter). HT2 and HT7 paired images acquired using the same PMT settings (505, 605, 755, or 805) are shown for comparison.

Fig. (S6).



Expression in cell-free lysates. Wheat germ lysates (TNT, SP6) were used to generate T antigen fusions to the C-terminus of HT2, HT3, HT6, or HT7. Lysates were labeled with 1 μM TMR-ligand for 1 h and analyzed by SDS-PAGE. The gel was scanned for fluorescence (E_{ex}/E_{em} = 532/580 nm) and bands representing full length fusion quantitated. The fluorescence image is shown in panel **A** and the quantitation in panel **B**.

Fig. (S7).



Stability of the attachment between HT7 and ligand. HT7 was labeled to completion with the TMR-ligand and then exposed to either elevated temperature (95 °C, 30 sec or 30 min; panel **A**) or a range of pH values (pH 4–9, 30 min; panel **B**). For the temperature experiment, samples were denatured in SDS-gel loading buffer prior to exposure to 95 °C. The control (-) sample was incubated at 25 °C in the presence of SDS. For the pH experiment, a sample in PBS (pH 7.0) was used as a control for neutral pH. Samples were resolved by SDS-PAGE, scanned for fluorescence, and quantitated for total functional protein (TMR fluorescence; E_{ex}/E_{em} = 532/580 nm).

Fig. (S8).

Table S4. Expression Benefit[a] Provided by HT7

Fusion Orientation (location of HT7)	*E. coli*	HEK-293 cells	TNT High-Yield (Wheat Germ)
N-terminus (upstream)	2–22	2–12	2–8
C-terminus (downstream)	Not measured	2–8	2–12

[a] fold improvement over non-tagged protein; values represent the range of fold-improvement observed for Src, Luc, p65, and p53 when fused to HT7.

Table S5.

Effect of various detergents on HT7 ligand binding kinetics (TMR-ligand)

Detergent	CMC[a]	Concentration dependent effect on binding kinetics[b]		
		<CMC	at CMC	>CMC
Non-ionic				
Triton X-100	**0.2–0.9 (0.0125–0.0563%)**	**+**	**+**	**+**
BigCHAP	**3.4 (0.299%)**	**+**	**NE**	**NE**
NP-40	**0.05–0.3 (0.003–0.018%)**	**NE**	**+**	**– (>0.05%)**
TWEEN 20	**0.06 (0.0074%)**	**–**	**–**	**–**
IGEPAL	**0.08 (0.0048%**	**NE**	**NE**	**–**
Digitonin	**<0.5 (<0.0614%)**	**+**	**NE**	**–**
Cationic				
DTAB[c]	**15 (0.462%)**	**–**	**–**	**–**
BTAB[d]	*	**Slight inhibition (10%) when used at >0.05%**		
Anionic				
Sodium deoxycholate	**2–6 (0.083–0.249%)**	**+**	**NE**	**–**
SDS	**7–10 (0.202–0.289%)**	**–**	**–**	**–**
Zwitterionic				
CHAPS	**6 (0.369%)**	**+**	**NE**	**NE**

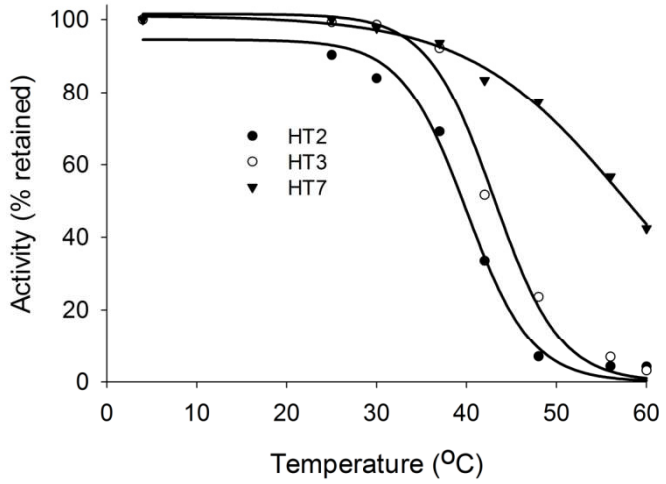
[a] critical micelle concentration; mM (source: Sigma or EMD Biosciences; except for DTAB (see reference 13)
[b] (+), enhanced kinetics; (–) reduced kinetics; NE, no effect
[c] dodecyl-trimethylammonium bromide
[d] (5-bromopentyl)trimethylammonium bromide
*data not available

Table S6.

Effect of various additives on HT7 ligand binding (TMR-ligand)

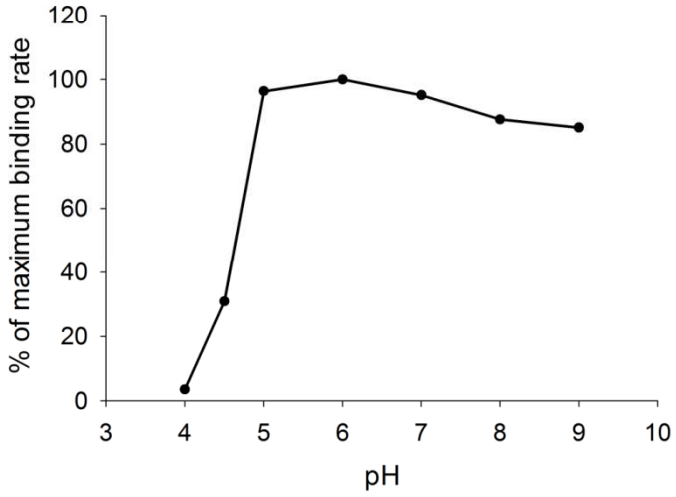
Additive	Observations	Acceptable concentration[a]
Divalent cations		
$MnCl_2$	Accelerated binding from 5–50 mM	10 mM
$CaCl_2$	No effect at 1 mM; buffer precipitated when >1 mM	1 mM
$MgCl_2$	Accelerated binding from 1–50 mM	25 mM
$ZnCl_2$	Accelerated binding with 5 mM; buffer precipitated when >5 mM	5 mM
DTT	40% inhibition with 5 mM	1 mM
Formaldehyde	80% activity retained after 2 min exposure to 0.5%[b]	0.5%
Paraformaldehyde	80% activity retained after 2 min exposure to 2%[c]	2%

[a]concentrations that can be tolerated by HT7. Note in some cases using higher amounts of an additive may be beneficial.
[b]compared to 60% activity retained for original HT2
[c]compared to 35% activity retained for original HT2



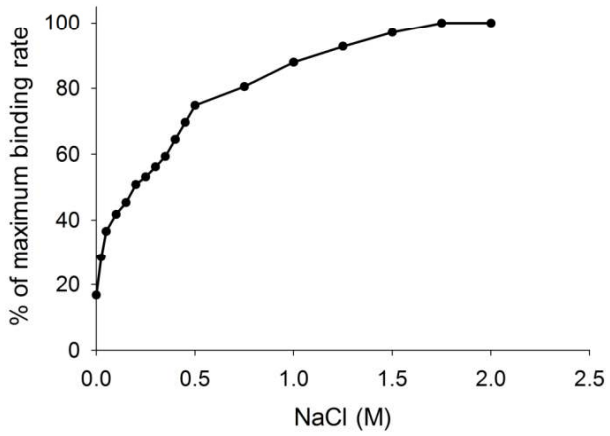
Thermal stability of HT7. HT2, HT3, and HT7 were exposed to elevated temperature for 30 min and the remaining activity (relative to 4 °C) was calculated based on initial rate determinations (FP, FAM-ligand).

Fig. (S9).



Effect of pH on HT7 activity. HT7 activity (as determined by the initial binding rate to the FAM-ligand using FP) was measured at different pH.

Fig. (S10).



Effect of NaCl on HT7 activity. The activity of HT7 in the presence of NaCl as determined by the initial binding rate to the FAM-ligand using FP.

Fig. (S11).

Table S7. Mass Spectrometry

			Peptides	Peptides
Protein(s) Identified	**Accession#**	**MW**	**RPS9-HaloTag**	**HaloTag Ctrl**
RPS3 40S ribosomal protein S3	IPI00011253	27 kDa	42	11
GNB2L1 38 kDa protein	IPI00641950 (+1)	38 kDa	42	8
PABPC1 Isoform 1 of Polyadenylate-binding protein 1	IPI00008524 (+2)	71 kDa	39	0
RPL4 60S ribosomal protein L4	IPI00003918	48 kDa	36	0
RPL5 60S ribosomal protein L5	IPI00000494	34 kDa	34	3
RPS4X 40S ribosomal protein S4, X isoform	IPI00217030	30 kDa	33	8
RPS3A 40S ribosomal protein S3a	IPI00419880	30 kDa	32	2
RPL6 60S ribosomal protein L6	IPI00329389 (+2)	33 kDa	30	7
RPSAP15;RPSA 33 kDa protein	IPI00413108 (+2)	33 kDa	30	4
RPL3 60S ribosomal protein L3	IPI00550021	46 kDa	30	3
RPS2 40S ribosomal protein S2	IPI00013485 (+2)	31 kDa	29	2
RPLP0 60S acidic ribosomal protein P0	IPI00008530	34 kDa	27	5
RPL23 60S ribosomal protein L23	IPI00010153 (+1)	15 kDa	26	2
RPL7A 60S ribosomal protein L7a	IPI00299573	30 kDa	24	6
RPL10A 60S ribosomal protein L10a	IPI00412579 (+1)	25 kDa	23	0
PABPC4 Isoform 1 of Polyadenylate-binding protein 4	IPI00012726 (+3)	71 kDa	23	0
RPS12 40S ribosomal protein S12	IPI00013917	15 kDa	22	0
RPL17;LOC100133931 60S ribosomal protein L17	IPI00413324	21 kDa	21	2
RPS9 40S ribosomal protein S9	IPI00221088	23 kDa	21	0
RPS18;LOC100130553 40S ribosomal protein S18	IPI00013296	18 kDa	20	6
RPS16 40S ribosomal protein S16	IPI00221092	16 kDa	20	2

Table S7. Contd…..

Protein(s) Identified	Accession#	MW	Peptides RPS9-HaloTag	Peptides HaloTag Ctrl
RPL26 60S ribosomal protein L26	IPI00027270	17 kDa	20	0
RPL7;RPL7P32 60S ribosomal protein L7	IPI00030179 (+2)	29 kDa	19	8
RPS6 40S ribosomal protein S6	IPI00021840	29 kDa	19	2
RPS15A 40S ribosomal protein S15a	IPI00221091	15 kDa	19	2
RPL11 Isoform 1 of 60S ribosomal protein L11	IPI00376798	20 kDa	19	0
HSPA1A;HSPA1B Heat shock 70 kDa protein 1	IPI00304925 (+1)	70 kDa	19	0
RPS8 40S ribosomal protein S8	IPI00216587 (+1)	24 kDa	18	6
RPL18 60S ribosomal protein L18	IPI00215719	22 kDa	18	4
RPL13 60S ribosomal protein L13	IPI00465361	24 kDa	18	3
RPS14 40S ribosomal protein S14	IPI00026271	16 kDa	18	2
RPS5 40S ribosomal protein S5	IPI00008433	23 kDa	18	0
HSPA8 Isoform 1 of Heat shock cognate 71 kDa protein	IPI00003865	71 kDa	18	0
RPLP2 60S acidic ribosomal protein P2	IPI00008529	12 kDa	17	0
RPL9 60S ribosomal protein L9	IPI00031691	22 kDa	17	0
RPL30 60S ribosomal protein L30	IPI00219156 (+1)	13 kDa	16	0
RPL14 Ribosomal protein L14 variant	IPI00555744	24 kDa	15	3
RPL12 Isoform 1 of 60S ribosomal protein L12	IPI00024933	18 kDa	15	2
RPL18A 60S ribosomal protein L18a	IPI00026202	21 kDa	15	2
RPL38 60S ribosomal protein L38	IPI00215790	8 kDa	15	0
RPL13A 60S ribosomal protein L13a	IPI00304612	24 kDa	14	3
RPS20 40S ribosomal protein S20	IPI00012493	13 kDa	14	0
RPS10 40S ribosomal protein S10	IPI00008438	19 kDa	14	0
RPS7 40S ribosomal protein S7	IPI00013415	22 kDa	14	0
NCL cDNA FLJ45706 fis, clone FEBRA2028457, highly similar to Nucleolin	IPI00444262 (+2)	66 kDa	14	0
UBC;RPS27A;UBB ubiquitin and ribosomal protein S27a precursor	IPI00179330	18 kDa	13	0
RPL10 60S ribosomal protein L10	IPI00554723 (+2)	25 kDa	13	0
PA2G4 Proliferation-associated protein 2G4	IPI00299000 (+2)	44 kDa	13	0
RPL8 60S ribosomal protein L8	IPI00012772	28 kDa	12	6
RPS13 40S ribosomal protein S13	IPI00221089	17 kDa	12	2
RPS19 40S ribosomal protein S19	IPI00215780	16 kDa	12	2
RPS21 40S ribosomal protein S21	IPI00017448	9 kDa	12	0
RPL37A 60S ribosomal protein L37a	IPI00414860	10 kDa	12	0
SNORA7A;RPL32 60S ribosomal protein L32	IPI00395998 (+4)	16 kDa	12	0
RPS17 40S ribosomal protein S17	IPI00221093	16 kDa	12	0
RPL23A 60S ribosomal protein L23a	IPI00021266 (+1)	18 kDa	12	0

Table S7. Contd…..

Protein(s) Identified	Accession#	MW	Peptides RPS9-HaloTag	Peptides HaloTag Ctrl
HNRNPC Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	IPI00216592 (+2)	32 kDa	12	0
RPS15 40S ribosomal protein S15	IPI00479058	17 kDa	12	0
RPL15 60S ribosomal protein L15	IPI00470528 (+1)	24 kDa	11	3
RPL27 60S ribosomal protein L27	IPI00219155	16 kDa	11	2
RPL28 60S ribosomal protein L28	IPI00182533	16 kDa	11	2
HIST1H1C Histone H1.2	IPI00217465	21 kDa	11	0
RPL35 60S ribosomal protein L35	IPI00412607	15 kDa	11	0
SERBP1 Isoform 1 of Plasminogen activator inhibitor 1 RNA-binding protein	IPI00410693 (+1)	45 kDa	11	0
TUBA1C Tubulin alpha-1C chain	IPI00218343 (+2)	50 kDa	10	2
RPL27A 60S ribosomal protein L27a	IPI00456758	17 kDa	10	0
RPS11 40S ribosomal protein S11	IPI00025091	18 kDa	10	0
EIF4A3 Eukaryotic initiation factor 4A-III	IPI00009328	47 kDa	10	0
RPL36 60S ribosomal protein L36	IPI00216237	12 kDa	10	0
RPS24 Isoform 1 of 40S ribosomal protein S24	IPI00029750 (+4)	15 kDa	10	0
EEF1A1 Elongation factor 1-alpha 1	IPI00396485 (+1)	50 kDa	9	3
RPS25 40S ribosomal protein S25	IPI00012750	14 kDa	9	2
RPL21P19;RPL21;RPL21P16 60S ribosomal protein L21	IPI00247583	19 kDa	9	0
TUBB Tubulin beta chain	IPI00011654 (+1)	50 kDa	9	0
RPL19 60S ribosomal protein L19	IPI00025329	23 kDa	9	0
UPF1 Isoform 1 of Regulator of nonsense transcripts 1	IPI00034049 (+1)	124 kDa	9	0
RPS23 40S ribosomal protein S23	IPI00218606	16 kDa	8	0
RPL24 60S ribosomal protein L24	IPI00306332 (+2)	18 kDa	7	2
RPL22 60S ribosomal protein L22	IPI00219153	15 kDa	7	2
RPS26;RPS26P25 40S ribosomal protein S26	IPI00655650 (+1)	13 kDa	7	0
YBX1 Nuclease-sensitive element-binding protein 1	IPI00031812	36 kDa	7	0
RPL31 60S ribosomal protein L31	IPI00026302 (+6)	14 kDa	7	0
RPS27 40S ribosomal protein S27	IPI00513971 (+1)	9 kDa	7	0
EIF6 Eukaryotic translation initiation factor 6	IPI00010105	27 kDa	6	0
RPL35A 60S ribosomal protein L35a	IPI00029731	13 kDa	6	0
RPL34 60S ribosomal protein L34	IPI00219160	13 kDa	5	2
ZCCHC3 Zinc finger CCHC domain-containing protein 3	IPI00011550	44 kDa	5	0
EEF2 Elongation factor 2	IPI00186290	95 kDa	5	0
RPLP1 60S acidic ribosomal protein P1	IPI00008527	12 kDa	5	0
HNRNPA1 Isoform A1-B of Heterogeneous nuclear ribonucleoprotein A1	IPI00215965 (+2)	39 kDa	5	0
RPS29 40S ribosomal protein S29	IPI00182289	7 kDa	5	0
RPL36AL 60S ribosomal protein L36a-like	IPI00056494 (+1)	12 kDa	4	0

Table S7. Contd…..

Protein(s) Identified	Accession#	MW	Peptides RPS9-HaloTag	Peptides HaloTag Ctrl
H1FX Histone H1x	IPI00021924	22 kDa	4	0
RPS28 40S ribosomal protein S28	IPI00719622 (+2)	8 kDa	4	0
RALY RNA binding protein, autoantigenic (HnRNP-associated with lethal yellow homolog (Mouse)), isoform CRA_a (Fragment)	IPI00011268 (+3)	33 kDa	4	0
RPL22L1 60S ribosomal protein L22-like 1	IPI00856049	15 kDa	4	0
SLC25A5 ADP/ATP translocase 2	IPI00007188	33 kDa	3	2
RPL29P31 similar to ribosomal protein L29	IPI00173589 (+2)	17 kDa	3	0
HNRNPK Isoform 1 of Heterogeneous nuclear ribonucleoprotein K	IPI00216049 (+4)	51 kDa	3	0
LARP1 Isoform 1 of La-related protein 1	IPI00185919 (+2)	124 kDa	3	0
SNRPD2 Small nuclear ribonucleoprotein Sm D2	IPI00017963	14 kDa	3	0
ACTB Actin, cytoplasmic 1	IPI00021439 (+5)	42 kDa	3	0
HNRNPA2B1 Isoform B1 of Heterogeneous nuclear ribonucleoproteins A2/B1	IPI00396378 (+3)	37 kDa	3	0
HNRNPH1 Heterogeneous nuclear ribonucleoprotein H	IPI00013881 (+1)	49 kDa	3	0
RSL1D1 Ribosomal L1 domain-containing protein 1	IPI00008708 (+1)	55 kDa	3	0
SFRS1 Isoform ASF-1 of Splicing factor, arginine/serine-rich 1	IPI00215884	28 kDa	3	0
HNRNPD Isoform 1 of Heterogeneous nuclear ribonucleoprotein D0	IPI00028888 (+1)	38 kDa	2	0
HNRNPM Isoform 1 of Heterogeneous nuclear ribonucleoprotein M	IPI00171903 (+1)	78 kDa	2	0
SFRS6 Isoform SRP55-1 of Splicing factor, arginine/serine-rich 6	IPI00012345 (+2)	40 kDa	2	0
NHP2L1 NHP2-like protein 1	IPI00026167 (+1)	14 kDa	2	0
NACA Nascent polypeptide-associated complex subunit alpha	IPI00023748 (+3)	23 kDa	2	0
CNBP Isoform 1 of Cellular nucleic acid-binding protein	IPI00430812 (+6)	19 kDa	2	0
HSP90AB1 Heat shock protein HSP 90-beta	IPI00414676	83 kDa	2	0
BAT1 Isoform 2 of Spliceosome RNA helicase BAT1	IPI00641829 (+3)	51 kDa	2	0
IGF2BP3 Isoform 2 of Insulin-like growth factor 2 mRNA-binding protein 3	IPI00165467 (+1)	22 kDa	2	0
RBMX Heterogeneous nuclear ribonucleoprotein G	IPI00304692 (+1)	42 kDa	2	0
FAU ubiquitin-like protein fubi and ribosomal protein S30 precursor	IPI00019770 (+1)	14 kDa	2	0

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

See main text.

REFERENCES

[1] Schindler JF, Naranjo PA, Honaberger DA, *et al.* Haloalkane dehalogenases: steady-state kinetics and halide inhibition. Biochemistry 1999; 38(18): 5772-8.

[2] Newman J, Peat TS, Richard R, *et al.* Haloalkane dehalogenases: structure of a Rhodococcus enzyme. Biochemistry 1999; 38(49): 16105-14.

[3] Los GV, Encell LP, McDougall MG, *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 2008; 3(6): 373-82.

[4] Schagat T, Friedman-Ohana R, Otto P, Hartnett J, Slater M. KRX Autoinduction protocol: a convenient method for protein expression. Promega Notes 2008; 98: 16-8.

[5] Dube DK, Black ME, Munir KM, Loeb LA. Selection of new biologically active molecules from random nucleotide sequences. Gene 1993; 137(1): 41-7.

[6] Williams JW, Morrison JF. Chemical mechanism of the reaction catalyzed by dihydrofolatereductase from Streptococcus faecium: pH studies and chemical modification. Biochemistry 1981; 20(21): 6024-9.

[7] Janssen DB. Evolving haloalkane dehalogenases. Curr Opin Chem Biol 2004; 8(2): 150-9.

[8] Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng 2001; 14(8): 529-32.

[9] Argos P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol 1990; 211(4): 943-58.

[10] Kurz M, Cowieson NP, Robin G, *et al.* Incorporating a TEV cleavage site reduces the solubility of nine recombinant mouse proteins. Protein Expr Purif 2006; 50(1): 68-73.

[11] Robinson CR, Sauer RT. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proc Natl Acad Sci USA 1998; 95(11): 5929-34.

[12] Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ. Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. Virology 1989; 172(1): 302-10.

[13] Neugebauer JM. Detergents: An overview. Meth enzymol 1990; 182: 239-52.

Received: March 02, 2012 Revised: April 04, 2012 Accepted: April 16, 2012