Supporting Information

A facile and efficient L-Method for incorporation of multiple unnatural amino acids into single protein

Niraikulam Ayyadurai[a], Kanagavel Deepankumar[a], Nadarajan Saravanan Prabhu[a], Sungu Lee[b], and Hyungdon Yun[a]*

[a]Dr. N. Ayyadurai, K. Deepankumar, N. Saravanan Prabhu, Prof. H. Yun*
School of Biotechnology
Yeungnam University
Gyeongsan, South Korea. E-mail: hyungdon@ynu.ac.kr

[b]Prof. S. G. Lee
Dept of Chemical Engineering
Pusan National University
South Korea

[*] Corresponding author

Materials

PCR reagents, T4 DNA ligase and restriction endonucleases were purchased from Promega (Madison, WI, USA). The isopropyl-D-thiogalactopyranoside (IPTG) was purchased from sigma chemicals (St. Louis, MO, USA). The host bacterium *Escherichia coli* (*E.coli*) strain XL1-blue (Stratagene, CA, USA) was used for plasmid DNA preparation, in this study. *E. coli* cells with plasmids were grown aerobically in Luria-Bertani (LB) broth (Difco Laboratories, Detroit, Michigan, USA) or on LB agar plate, supplemented with appropriate antibiotics for the selection of transformants. *E.coli* strain B834(DE3) Methionine auxotroph was obtained from Prof. William Studier (Brookhaven National Laboratory, USA). Natural amino acids, unnatural amino acid 3, 4-dihydroxy-L-phenylalanine (L-DOPA) and chitosan were purchased from Sigma (St. Louis, MO, USA). The L-Methionine surrogate L-homopropargylglycine (Hpg) and L-azidohomoalanine (AHA) was purchased from Chiralix (Nijmegen, The Netherland). The pQE 80L plasmid and nickelnitrilotriacetic acid (Ni-NTA) affinity column were purchased from Qiagen (Valencia, CA, USA). The vector pBAD and azide bearing Alexa fluor 594 were obtained from Invitrogen (Invitrogen. Life Technologies, Carlsbad, CA, USA)

Construction of plasmids and strains

The DNA manipulations were performed according to the procedures described by Sambrook and Russel[1]. The PCR reaction (50 µl) contains 10 pM of each primer, 50 ng of template DNA, 1X Taq DNA polymerase buffer, 1U of *Taq* DNA polymerase (Promega, Madison, WI, USA), 0.2 mM of each deoxyribonucleotide triphosphates and 1.5 mM $MgCl_2$. Amplification was performed in a DNA thermal cycler (Master Gradient thermal cycler, Eppendorf, Hamburg, Germany) programmed for an initial denaturation (94 ºC for 1 min) followed by 30 cycles of 1 min at 94ºC, 1 min at 60ºC and 0.5 min at 72ºC with an extension at 72ºC for 10 min. Coding sequences of GFP (GFPhs2 was used in this study) was previously cloned into expression vector pQE-80 L (Qiagen,USA) to generate the plasmid pQE80GFP[2]. The constructed pQE80GFP plasmid was transformed into *E. coli* B834(DE3) Methionine auxotroph for the production of recombinant unnatural proteins. All these constructs were sequenced and confirmed for their target protein sequence.

Optimization of L-Methionine concentration for L-Hpg incorporation

The *E. coli* B834(DE3) Methionine auxotroph bearing pQE80GFP plasmid were grown in the minimal medium supplemented with different concentration of L-Methionine (0.02-0.06 mM). Samples were collected for every one hour and values were plotted using Origin software version: 6 (Microcal software, MA, USA).

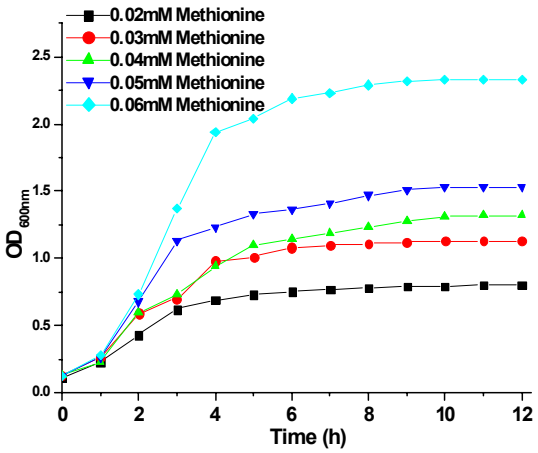


Figure S1. Growth curve analysis of B834 (DE3) Methionine auxotroph using different concentration of L-Methionine.

Expression and purification of pQE80-GFPHpg

Incorporation of L-Hpg was carried out in the optimized condition by as described earlier [1]. Briefly, the limiting concentration of L-Met (0.03 mM) allowed the cells to attain an OD_{600} 0.6-0.8 and the target proteins were induced with 1 mM IPTG followed by simultaneous addition of L-Met analogues (0.5 mM **2**) and allowed the expression for 7 h. The harvested cells were subjected to centrifugation and stored at -70°C until further use. Lysis of the cells was carried out by using BugBuster protein extraction kit (Novagen) followed by sonication. Briefly, collected cell pellet corresponding to 1 mL of culture was resuspended in 100 µl of lysis buffer, incubated at room temperature for 10 min, and centrifuged at 9000 g, 4ºC for 20 min. The supernatant was saved as a soluble protein fraction, and analyzed by SDS-PAGE (12% acrylamide gel). The remaining soluble protein fractions were purified by Ni-NTA column chromatography (GE Healthcare Bio-Sciences, Sweden) by standard protocol. Elution fractions

were analyzed by SDS-PAGE, and those that were enriched in the desired GFP variants were pooled and dialyzed against 1X phosphate buffered saline (PBS).

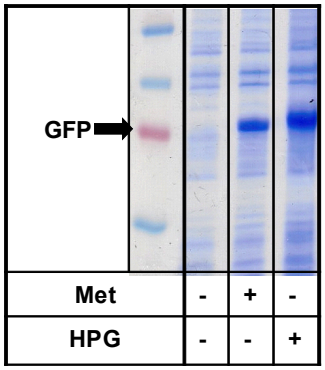


Figure S2. Protein expression profiles of residue specific incorporation of L-HPG in pQE-GFP. Lane 1: GFP induced in the presence of 19 amino acids (except L-Methionine); Lane 2: GFP induced in the presence of 20 amino acids; Lane 3: GFP induced in the presence of 19 amino acids with L-HPG (except L-Methionine).

Site directed mutagenesis in pQE80GFP

Coding sequence of GFP was previously cloned into the expression vector pQE80L used to generate the plasmid pQE80GFP$_{Y66am}$. The following mutant vectors carrying amber stop codon (TAG) in the chromophore tyrosine (GCG) was generated using Quick-change site directed mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's description in the manual. The mutagenesis was confirmed by DNA sequencing analysis at Cosmo Genetech, Dajeon, South Korea.

Site specific incorporation of L-DOPA

In a typical procedure of incorporation, *E.coli* B834 (DE3) containing pQE80-GFP$_{Y66am}$ and pAC-DOPA-6TRN which contain mutant tRNA/synthetase for site specific incorporation were grown in LB broth supplemented with ampicillin (50 mg/mL), tetracycline (7.5 mg/mL), and grown overnight at 37°C. The cells in the small starting cultures were collected, resuspended in phosphate buffer saline, and inoculated into glycerol minimal media

supplemented with ampicillin (50 mg/mL) and tetracycline (7.5 mg/mL). Cultures were grown at 37°C until OD_{600} reach approximately 0.7 and 1 mM of L-DOPA (Sigma Aldrich, USA) was added before induction. Cultures were further induced with 1 mM IPTG for 7 h at 37°C and the cells were harvested and frozen at -80°C for purification.

Incorporation of multi-unnatural amino acids into recombinant protein using pQE80GFP system

The above mentioned constructs were grown in M9 minimal medium supplemented with 1% (w/v) of glucose, 1% glycerol, 0.1 mM of $CaCl_2$, 1.0 mM of $MgSO_4$, 35 µg/mL of thiamine, 19 amino acids (40 mg/L) without L-Met, 0.03 mM of L-Met and appropriate antibiotics. When an optical density of the culture reached 0.8 to 1.0 at OD_{600}, the cultures were reached the saturated level. The cells were added to different test tubes containing L-L-Methionine (40 mg/ml) and L-tyrosine (40 mg/ml), L-DOPA (1 mM), L-HPG (0.5 mM) and combination of L-Methionine + L-DOPA, L-tyrosine +L-HPG, and L-HPG + L-DOPA. After incubating in 37^0C for 10 min, 1 mM IPTG was added to induce the protein expression. The cells were harvested finally after 7 h of induction and OD_{600} was measured. The harvested cells were subjected to centrifugation and stored at -80°C until further use. Lysis of the cells was carried out by using BugBuster protein extraction kit (Novagen) followed by sonication. Briefly, collected cell pellet corresponding to 1 mL of culture was resuspended in 100 µl of lysis buffer, incubated at room temperature for 10 min, and centrifuged at 9000 g, 4ºC for 20 min. The supernatant was saved as a soluble protein fraction, and analyzed by SDS-PAGE (12% acrylamide gel).

Whole cell fluorescence assay

The cells were harvested finally after induction and OD_{600} was measured. The GFP variant's whole cell fluorescence were performed by measuring fluorescence intensity by exciting at 485 nm and emission at 520 nm with excitation/emission slits of 5.0 nm and was recorded on Perkin Elmer/Wallac Victor 2 Multilabel Counter (1420-011).

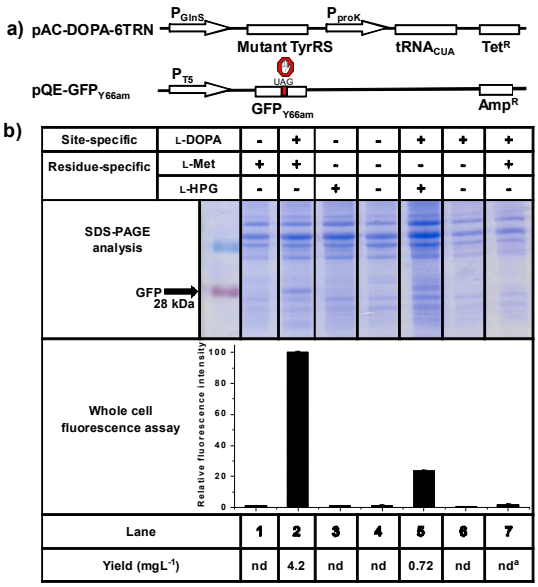


Figure S3. a) Expression system used for MUAA incorporation. b)The expression level and relative fluorescence analysis of fullength GFP with chromophore amber mutation at position Y66 for site specific incorporation of L-DOPA and residue specific incorporation of L-Hpg at L-Met position, (SDS-PAGE (Lane 1-7) and fluorescence analysis (lane 1-7) results uptained in ideal experimental conditions). nd-not determined. a-without addtion of IPTG.

Construction of pBAD expression system for MUAA incorporation

The DNA manipulations were performed as described earlier. The gene specific primers were designed and amplified as mentioned in the earlier section. Further GFP sequences were amplified from pQE80GFP and sucessfully cloned into pBAD-HisA (Invitrogen, Carlsbad, CA, USA) to generate the plasmid pBADGFP. The construct of pBADGFP plasmid was used to introduce the amber stop codon (TAG) in the chromophore tyrosine (GCG) to generate pBADGFP$_{Y66am}$ using Quick-change site directed mutagenesis kit (Stratagene, La Jolla, CA). The constructed pBADGFP$_{Y66am}$ plasmid was transformed into *E. coli* B834(DE3) Methionine auxotroph bearing pAC-DOPA-6TRN for the purpose of MUAA incorporation. All these constructs were sequenced and confirmed for their target protein sequence.

Incorporation and optimization of multi-unnatural amino acids into recombinant protein using pBADGFP system

The multi-unnatural amino acid incorporation experiment was carried out by similar to pQE80GFP system using $pBADGFP_{am}$ in Methionine auxotroph B834 (DE3) system. It is well known that pBAD expression system is consistent with the autocatalytic mechanism of induction which affects the optimum level of expression. To further to optimize the inducer concentration, *E. coli* B834 (DE3) bearing pBAD-GFP_{Y66am} and pAC-DOPA-6TRN cells were grown in MMGG medium supplemented with 0.03mM Methionine to reach OD_{600} for 0.8-1 and the induction was carried out by the addition different concentration of L-arabinose (0.02%, 0.02% and 0.2%). The protein expression level was monitored by SDS-PAGE and whole cell fluorescence analysis.

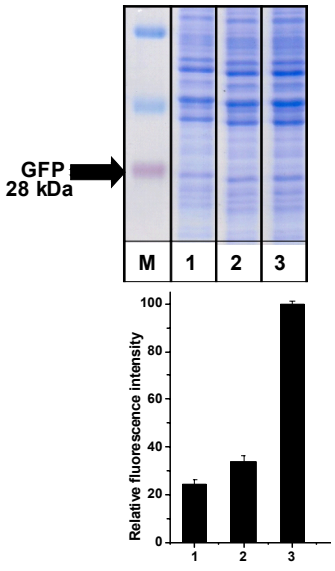


Figure S4. Protein expression profiles and whole cell fluorescence assay of GFP (Y66Dopa)[Hpg] in the optimized condition. Lane 1: GFP(Y66Dopa)[Hpg] induced with 0.002% L-arabinose; Lane 2: GFP(Y66Dopa)[Hpg] induced with 0.02% L-arabinose; Lane 3: GFP(Y66Dopa)[Hpg] induced with 0.2% L-arabinose.

Incorporation and Optimization of carbon source for the production MUAA incorporated protein in pBAD expression system

To evaluate the effect of carbon source in culture media, carbon sources such as glucose and glycerol were used in different concentrations such (0.5%, 1%, 2.5%, 5%, 7.5% and 10%) and growth curve was moniter with addition 0.03 mM L-Methionine in the medium. Simultaneously, cells were grown to reach OD_{600} for 0.8-1 and the induction was carried out by the addition of L-arabinose (0.2%) and along with 1mM **1** and 0.5 mM **2** were added. After 7 hrs induction, cell were harvested and subjected to SDS-PAGE and whole cell fluorescence analysis.

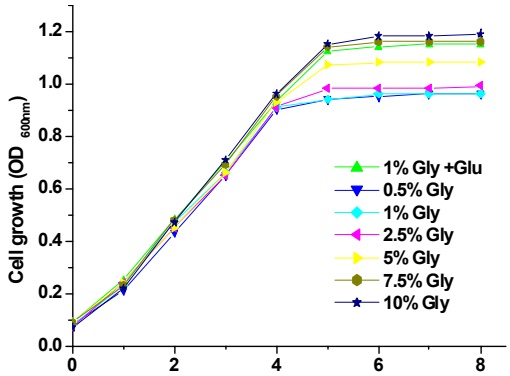


Figure S5. Growth curve analysis of B834 (DE3) Methionine auxotroph bearing pBADGFPam and pAC-DOPA-6TRN plasmid at difference cocentrations of carbon source.

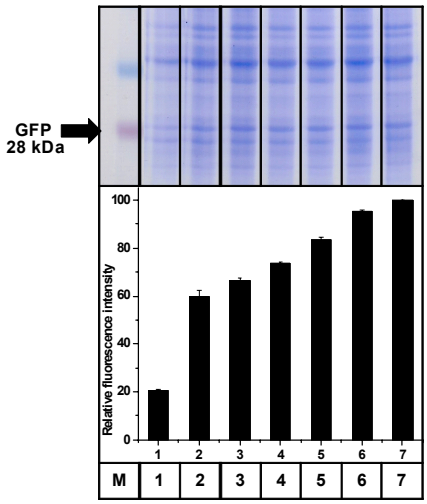


Figure S6. Protein expression profiles and whole cell fluorescence assay of mutant congener GFP(Y66Dopa)[Hpg] at different concentration of carbon source. Lane 1: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 1% glycerol and glucose respectively. Lane 2: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 0.5% glycerol; Lane 3: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 1% glycerol. Lane 4: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 2.5% glycerol. Lane 5: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 5% glycerol. Lane 6: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 7.5% glycerol. Lane 7: mutant congener GFP(Y66Dopa)[Hpg] incorporated in the presence of 10% glycerol.

Protein purification and quantification

Purification of recombinant proteins were performed by using Ni-NTA HisBind resin (Novagen, USA). Further Histag purified samples were subjected to gel permeation chromatography (GPC) purification by AKTA explorer FPLC system containing superdex 75 HR column at 4 °C. Absorption spectrum and quantification of proteins were measured as describe elsewhere.

In-gel digestion

Samples were resolved by SDS-PAGE and stained with Colloidal Coomassie blue R250 solution (ProteoL-Metech Inc. Korea). Protein in-gel digestion was performed with modifications as described previously[4]. Gel bands were sliced into 1 mm^3 cubes and washed twice with 50 mM ammonium bicarbonate containing 50% acetonitrile. Protein reduction and alkylation were performed consecutively with 10 mM DTT in 50 mM ammonium bicarbonate for 1 h at 56 °C and then with 10 mM acrylamide in 50 mM ammonium bicarbonate for 45 min at room temperature, respectively. Gel pieces were washed twice with 50 mM ammonium bicarbonate, 50% acetonitrile, dehydrated with 100% acetobitrile, and dried in a speed vac. The dried gel pieces were rehydrated with 20 ng μL-1 sequencing grade trypsin (Promega, CA, USA) in 50 mM ammonium bicarbonate and incubated overnight at 37 °C. Supernatants were transferred to fresh tubes, and the remaining peptides in gel slices were extracted by sonicating with 50% acetonitrile in 5% TFA in water bath for 40 min. The extracts were combined, dried about 2 μL and desalted using a GeLoader tip column packed with POROSR2/R3 (Applied Biosystems, CA, USA) and the eluted peptides were subjected to mass spectroL-Metric analysis.

Nano LC-MS/MS

Tryptic digests were separated by an Ultimate nanoLC systems including the FAMOS autosampler and Switchos column switching valve (LC-Packings, Amsterdam, The Netherlands) and on-line analyzed by a QSTAR Mass spectroL-Meter (Applied Biosystems, CA, USA) with a nanospray interface as described[5]. Peptide binding was occurred in a 2 cm 200 μm ID self-packed column Zorbax 300SB-C18, 5μm (Agilent) with 5% acetonitrile in 0.1% formic acid with a flow rate 4 μL /min for 10 min. The bound peptides were separated in a 15 cm fused silica emitter (75 μm ID) in-house packed with the same resin, eluted with 5 – 50 % (v/v) acetonitrile in 0.1 % formic acid over 45 min at a flow rate of 0.2 μL /min. MS/MS spectra were acquired in an automated switching mode tandem mass spectroL-Meter with m/z-dependent set of collision offset values.

Peptide identification via MASCOT database search

The mass spectral data were processed into peak lists with Analyst QS (version 1.1; Applied Biosystems) and searched against GFP protein sequences using in-house Mascot server (version 2.1, matrix Sciences, UK). The initial mass tolerance in MS mode was set to 0.3 Da and MS/MS mass tolerance was 1 Da. Two tryptic missed cleavages were allowed. Propionamide (C), Deamindation (NQ), Oxidation (M), L-Met-HPG (M^{*}), L-DOPA-chromophor ($Y^{\#}$), L-DOPA (Y^{*}) were searched as variable modifications. The candidate peptides whose probability MOWSE scores are significant ($p<0.05$) were considered and the peptides with greater than 11 ion scores were selected. The MS/MS spectra of selected peptides were manually validated as well.

Protein sequence

>GFP

MRGSHHHHHHGSMSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKITLKLICTTGKLP

VPWPTLVTTCGYGVQCFARYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTFKTRAEVKFEGDTIV

NRIKLKGIDFKEDGNILGHKLEYNFNSHKVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPI

GDGPVRLPDNHYLSTQSVILEDPNEKRDHMVLHEFVTAAGITHGMDELYK

The protein sequence starts from 13th position of L-Methionine will be considered as number 1 in the manuscript but in the case of ESI-MS/MS analysis the sequence starts from 1st position of L-Methionine.

Table 1. List of peptide amino acid sequences of GFP observed by ESI-MS/MS.

amino acids	obseved mass (Da)	Calculated mass (Da)	Ions score[¥]	Sequence
GFP				
16 - 38	1219.0861	2436.2535	85	K.GEELFTGVVPILVELDGDVNGHK.F
16 - 42	732.3969	2925.5235	63	K.GEELFTGVVPILVELDGDVNGHKFSVR.G
39 - 53	762.3423	1522.7011	79	K.FSVRGEGEGDATNGK.I
43 - 53	517.7182	1033.4312	63	R.GEGEGDATNGK.I
58 - 64	403.7094	805.4368	40	K.LICTTGK.L
65 - 85	1190.5206	2379.1755	30	K.LPVPWPTLVTTCGYGVQCFAR.Y
86 - 91	403.6722	805.3429	36	R.YPDHMK.R
86 - 92	473.7204	945.4491	53	R.YPDHMKR.H
92 - 97	425.2151	848.4293	40	K.RHDFFK.S
98 - 108	633.7415	1265.571	66	K.SAMPEGYVQER.T
109 - 119	629.7613	1257.6241	65	R.TISFKDDGTFK.T
114 - 125	683.831	1365.6888	17	K.DDGTFKTRAEVK.F
122 - 134	493.2492	1476.7572	74	R.AEVKFEGDTIVNR.I
126 - 134	525.7321	1049.5141	56	K.FEGDTIVNR.I
135 - 143	531.3202	1060.6644	14	R.IKLKGIDFK.E
137 - 152	446.7412	1782.9628	73	K.LKGIDFKEDGNILGHK.L
139 - 161	892.7442	2675.2979	79	K.GIDFKEDGNILGHKLEYNFNSHK.V
144 - 161	529.5066	2114.018	20	K.EDGNILGHKLEYNFNSHK.V
153 - 161	576.2627	1150.5407	31	K.LEYNFNSHK.V
162 - 168	405.2161	808.433	49	K.VYITADK.Q
162 - 170	533.2857	1064.5866	32	K.VYITADKQK.N
162 - 174	739.9032	1477.814	30	K.VYITADKQKNGIK.A
179 - 206	1039.4752	3115.5071	86	K.IRHNVEDGSVQLADHYQQNTPIGDGPVR.L
181 - 206	949.4257	2845.3379	143	R.HNVEDGSVQLADHYQQNTPIGDGPVR.L
207 - 226	1156.5345	2311.1332	122	R.LPDNHYLSTQSVILEDPNEK.R
207 - 227	1234.5823	2467.2343	109	R.LPDNHYLSTQSVILEDPNEKR.D
228 - 250	872.0356	2613.2355	126	R.DHMVLHEFVTAAGITHGMDELYK.-

[¥]Peptide ion scores greater than 11 are selected.

Table 2. List of peptide amino acid sequences of GFP[Hpg] observed by ESI-MS/MS.

amino acids	obseved mass (Da)	Calculated mass (Da)	Ions score[¥]	Sequence
GFP[Hpg]				
3 - 15	485.2119	1452.651	12	R.GSHHHHHHGSM*SK.G
16 - 38	1219.083	2436.254	84	K.GEELFTGVVPILVELDGDVNGHK.F
16 - 42	976.5	2926.508	21	K.GEELFTGVVPILVELDGDVNGHKFSVR.G
43 - 53	517.692	1033.431	64	R.GEGEGDATNGK.I
98 - 108	622.7754	1243.583	69	K.SAM*PEGYVQER.T
109 - 119	629.7916	1257.624	34	R.TISFKDDGTFK.T
122 - 134	739.3592	1476.757	74	R.AEVKFEGDTIVNR.I
126 - 134	525.746	1049.514	56	K.FEGDTIVNR.I
137 - 143	410.7413	819.4854	36	K.LKGIDFK.E
139 - 152	771.8775	1541.784	71	K.GIDFKEDGNILGHK.L
144 - 152	491.7396	981.4879	58	K.EDGNILGHK.L
153 - 161	576.2472	1150.541	51	K.LEYNFNSHK.V
162 - 168	405.2107	808.433	45	K.VYITADK.Q
162 - 170	533.2809	1064.587	33	K.VYITADKQK.N
171 - 178	446.2198	890.4973	12	K.NGIKANFK.I
181 - 206	949.4184	2845.338	125	R.HNVEDGSVQLADHYQQNTPIGDGPVR.L
207 - 226	1156.525	2311.133	120	R.LPDNHYLSTQSVILEDPNEK.R
227 - 250	909.4325	2725.361	123	K.RDHM*VLHEFVTAAGITHGM*DELYK.-
228 - 250	857.4089	2569.26	132	R.DHM*VLHEFVTAAGITHGM*DELYK.-

[¥]Peptide ion scores greater than 11 are selected.

M* represents L-Met ⟶ HPG substitution in GFP[Hpg] protein which was prepared by the residue specific incorporation of L- Hpg

Table 3. List of peptide amino acid sequences of GFP(Y66Dopa) observed by ESI-MS/MS.

amino acids	obseved mass (Da)	Calculated mass (Da)	Ions score[¥]	Sequence
GFP(Y66Dopa)				
16 - 38	813.3794	2437.2375	67	K.GEELFTGVVPILVELDGDVNGHK.F
16 - 42	976.4849	2926.5075	58	K.GEELFTGVVPILVELDGDVNGHKFSVR.G
43 - 53	517.728	1033.4312	64	R.GEGEGDATNGK.I
43 - 57	745.8608	1489.7259	61	R.GEGEGDATNGKITLK.L
86 - 91	403.6665	805.3429	28	R.YPDHMK.R
92 - 97	425.2117	848.4293	40	K.RHDFFK.S
98 - 108	633.7594	1265.571	69	K.SAMPEGYVQER.T
98 - 108	641.7634	1281.5659	51	K.SAMPEGYVQER.T
109 - 119	420.2042	1257.6241	39	R.TISFKDDGTFK.T
109 - 121	505.9195	1514.7729	38	R.TISFKDDGTFKTR.A
122 - 134	493.2441	1476.7572	70	R.AEVKFEGDTIVNR.I
126 - 134	526.2329	1050.4981	15	K.FEGDTIVNR.I
137 - 143	410.7377	819.4854	29	K.LKGIDFK.E
137 - 152	595.3063	1782.9628	48	K.LKGIDFKEDGNILGHK.L
139 - 152	771.8748	1541.7837	76	K.GIDFKEDGNILGHK.L
144 - 152	491.7375	981.4879	70	K.EDGNILGHK.L
144 - 161	705.6641	2114.018	17	K.EDGNILGHKLEYNFNSHK.V
153 - 161	576.2536	1150.5407	51	K.LEYNFNSHK.V
162 - 168	405.2146	808.433	48	K.VYITADK.Q
162 - 170	533.2873	1064.5866	14	K.VYITADKQK.N
181 - 206	1423.6402	2845.3379	128	R.HNVEDGSVQLADHYQQNTPIGDGPVR.L
207 - 226	1156.5163	2311.1332	116	R.LPDNHYLSTQSVILEDPNEK.R
207 - 227	1234.5814	2467.2343	103	R.LPDNHYLSTQSVILEDPNEKR.D
227 - 250	929.4218	2785.3315	62	K.RDHMVLHEFVTAAGITHGMDELYK.-
228 - 250	1307.5731	2613.2355	110	R.DHMVLHEFVTAAGITHGMDELYK.-
65 – 85[a]	1113.5151	3337.6498	2	K.LPVPWPTLVTTCY[#]VQCFARYPDHMKR.H

[¥]Peptide ion scores greater than 11 are selected.
Y[#] (GY*G) - represents the chromophore formation. The altered chromophore was formed with site specifically incorporated L-DOPA of chromphore region. Y* indicates the site specific replacement of L-DOPA inside of L-tyrosine in the GFP(Y66D) protein
[a] The low ion score containing peptide sequence to show successfull site specific incorporation of L-DOPA in the GFP containing amber mutation in the chromophore.

Table 4. List of peptide amino acid sequences of GFP(Y66Dopa)[Hpg] observed by ESI-MS/MS.

amino acids	obseved mass (Da)	Calculated mass (Da)	Ions score[¥]	Sequence
GFP(Y66Dopa)[Hpg]				
16 - 38	1219.0886	2436.2535	96	K.GEELFTGVVPILVELDGDVNGHK.F
43 - 53	518.2007	1034.4152	40	R.GEGEGDATNGK.I
43 - 57	745.8392	1489.7259	116	R.GEGEGDATNGKITLK.L
65 - 85	1152.5201	2303.123	56	K.LPVPWPTLVTTCY[#]VQCFAR.Y
90 - 95	425.2089	848.4293	35	K.RHDFFK.S
96 - 106	622.7722	1243.5833	56	K.SAM*PEGYVQER.T
107 - 117	420.2018	1257.6241	54	R.TISFKDDGTFK.T
107 - 119	505.9219	1514.7729	33	R.TISFKDDGTFKTR.A
120 - 132	739.3649	1476.7572	69	R.AEVKFEGDTIVNR.I
124 - 132	525.751	1049.5141	52	K.FEGDTIVNR.I
135 - 141	410.7389	819.4854	23	K.LKGIDFK.E
137 - 150	771.8724	1541.7837	71	K.GIDFKEDGNILGHK.L
142 - 150	491.7357	981.4879	55	K.EDGNILGHK.L
142 - 159	529.7497	2115.002	25	K.EDGNILGHKLEYNFNSHK.V
151 - 159	576.2361	1150.5407	55	K.LEYNFNSHK.V
151 - 166	647.9862	1940.9632	77	K.LEYNFNSHKVYITADK.Q
160 - 166	405.2124	808.433	45	K.VYITADK.Q
179 - 204	949.4054	2845.3379	142	R.HNVEDGSVQLADHYQQNTPIGDGPVR.L
205 - 224	1156.5253	2311.1332	129	R.LPDNHYLSTQSVILEDPNEK.R
228 - 250	857.388	2569.2601	126	R.DHM*VLHEFVTAAGITHGM*DELYK.-

[¥]Peptide ion scores greater than 11 are selected.
M* represents L-Met⟶► HPG substitution in GFP(Y66Dopa)[Hpg] protein
Y[#] (GY*G) - represents the chromophore formation. The altered chromophore was formed with site specifically incorporated L-DOPA of chromphore region. Y* indicates the site specific replacement of L-DOPA inside of L-tyrosine in the GFP(Y66Dopa) Protein

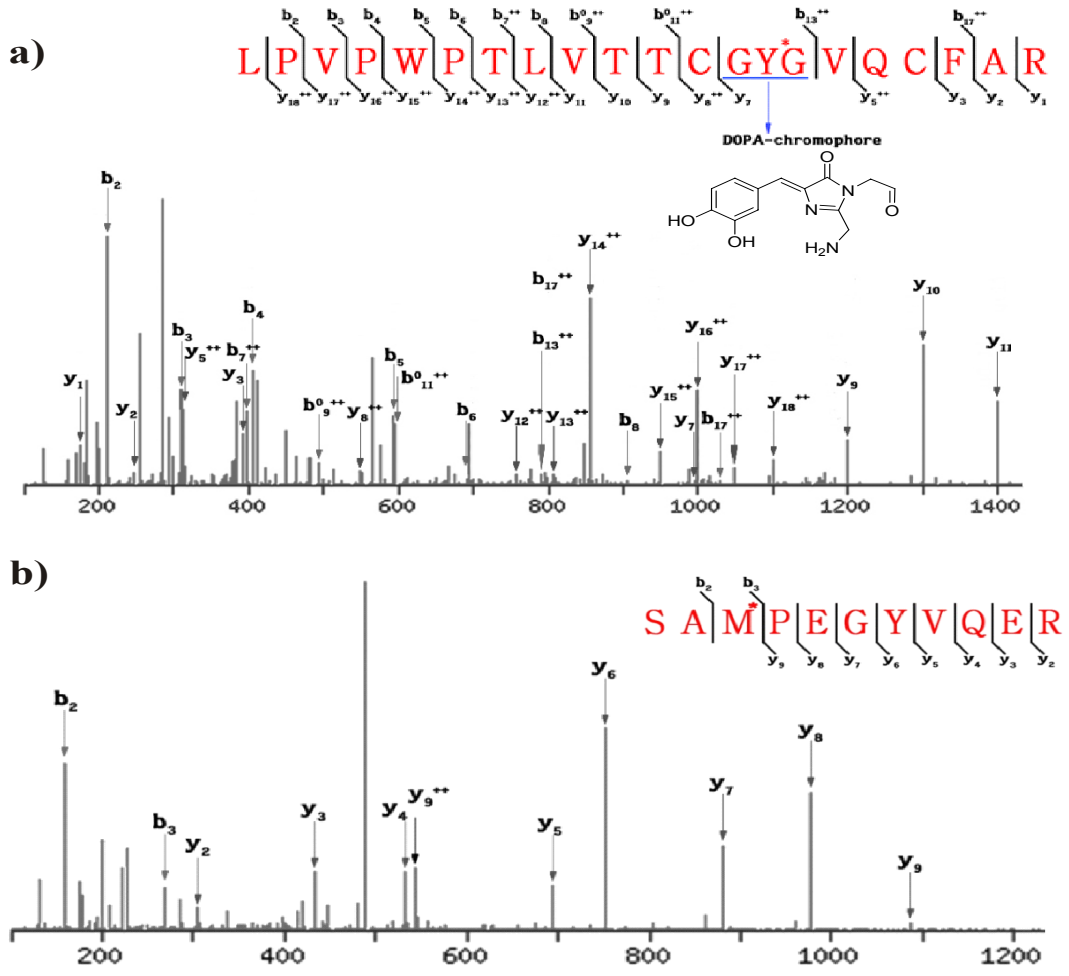


Figure S7. Representative spectra of tryptic digested peptides of GFPdpHpg obtained from ESI–MS/MS analysis. a) The tryptic peptide fingerprint containing GFP chromophore (Y* represents the L-DOPA). b) The tryptic peptide fragment containing L-HPG (M*) and tyrosine (Y).

Cycloaddition reaction

The reaction mixture consisted of azide bearing Alexa fluor 594 (Invitrogen, USA) were prepared with GFP, GFP(Y66Dopa), GFP[Hpg], GFP(Y66Dopa)[Hpg] protein as described earlier[3]. The reaction mixture was incubated in shaking incubator for 24 h at 4°C. As a negative control, azide bearing Alexa fluor 594 was incubated under the same conditions but without addition of $CuSO_4$/L-ascorbic acid. Afterwards all samples were washed three times with 1X PBS at 4°C and analysed by the SDS-PAGE. The concentrated samples were diluted and fluorescence emission were recorded on a JASCO FP-777 spectrofluoriL-Meter equipped with digital software.

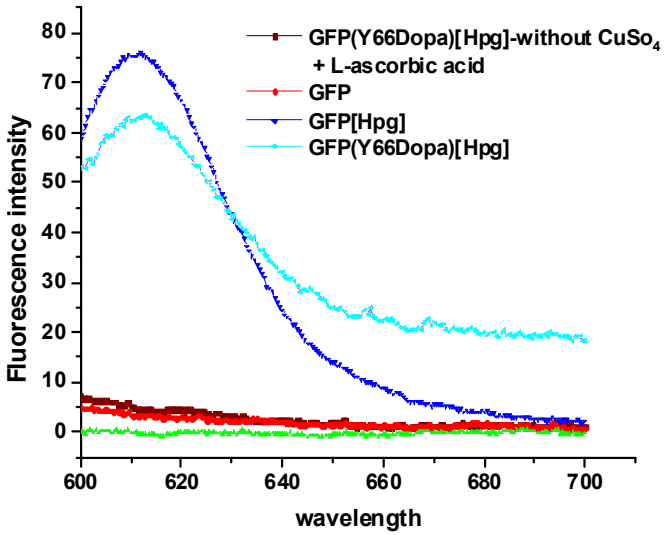


Figure S8. The ‘‘click’’ reaction with Alexa Fluor594 bearing azide group results in strong red fluorescence was emitted only with mutant congener GFP(Y66Dopa)[Hpg] and congener GFP[Hpg] containing alkyne group.

Redox-cycling staining for the proteins expressed with L- DOPA

The GFP, GFP(Y66Dopa), GFP[Hpg], GFP(Y66Dopa)[Hpg] proteins were blotted electrophoretically onto poly vinylidene difluoride membrance (PVDF) and visualized by staining the membrane with nitroblue tetrazolium (NBT). The membrane was incubated in the NBT staining solution (2 M sodium glycinate, 0.24 mM NBT, pH 10.0) in dark room for 3 h.

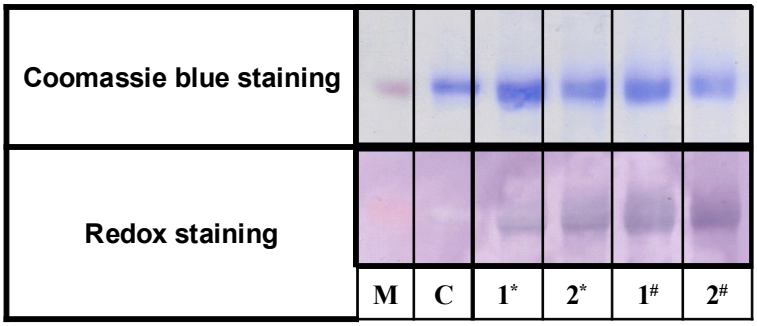


Figure S9. Redox-cycling staining for the L-DOPA containing proteins. All samples were oxidized, purified and resolved by SDS-PAGE, then transferred to a PVDF membrane. DOPA containing proteins were visualized by NBT staining solution. C: Control reaction - GFP protein prepared with 20 amino acids. *pQE expression system; #pBAD expression system; 1: GFP(Y66Dopa) protein; 2: GFP(Y66Dopa)[Hpg].

Fluorescence spectroscopy analysis

Excitaion spectra was recorded between 430 to 550. Emission spectrum of the protein sample was recorded between 500 and 700nm by exciting the sample at specific wavelength using a spectrofluoroL-Meter (Jasco FP-777). Prior to the experiment, the concentration of the protein was quantified. Normalized emission spectrum of the protein was drawn from the base corrected data using Origin software version: 6 (Microcal software, MA, USA).

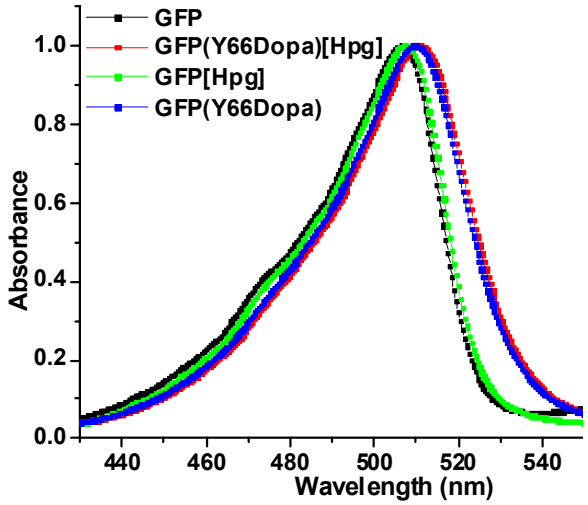


Figure S10. Fluorescence exicitation spectra of GFP, GFP(Y66Dopa), GFP[Hpg] and GFP(Y66Dopa)[Hpg].

Multiunnatural amino acid incorporation in the optimized condition

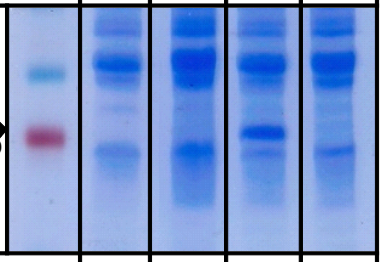
Site-specific	L-DOPA	-	-	+	+
Residue-specific	L-Met	+	-	-	-
	L-AHA	-	-	+	+
SDS-PAGE Gel 					
Lane		1	2	3	4
Yield (mgL^{-1})		nd	nd	3.91	nd[a]

Figure S11. The expression level of full length GFP with chromophore amber mutation at position Y66 for site specific incorporation of L-DOPA and residue specific incorporation of L-AHA at L-Met position, (Lane 1: Medium supplemented with 20 amino acids; Lane 2: Medium supplemented with 18 amino acids; Lane 3 and 4 : Medium supplemented with L-DOPA and L-AHA). nd- not determined; a-without of L-arabinose.

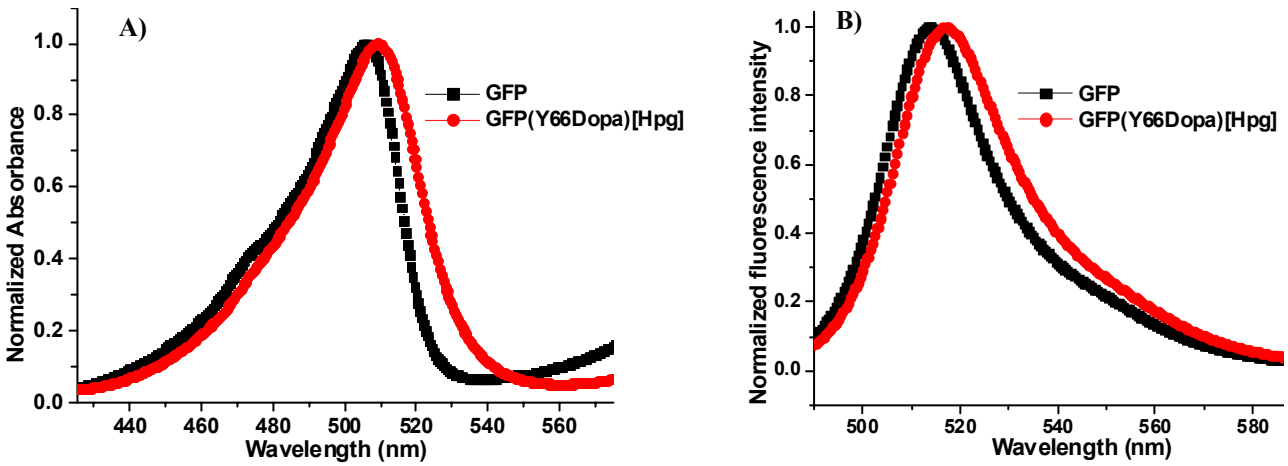


Figure S12. Fluorescence exicitation (A) and emission (B) spectra of GFP and GFP(Y66Dopa)[Aha].

Reference

1. J. Sambrook, E. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual.* 2nd ed., Cold spring harbor laboratory press, Cold spring harbor, NY, USA, **1989.**

2. N. Soundrarajan, N. Ayyadurai, N. Budisa, L. Merkel, G. Raghunathan, S. Sriram,E. Song, B.G. Kim, H. Yun, S. G. Lee, *ChemBioChem* **2010,** 11, 2521.

3. S. Lepthien, L. Merkel, N. Budisa. *Angew. Chem. Int. Ed*. **2010**, *49*, 5446.

4. L. Merkel, M. Schauer, G. Antranikian, N. Budisa, *Chembiochem*. **2010**, *26*, 1505.

5. H. W. Nam, G. Y. Lee, Y. S. Kim, *J Proteome Res.* **2006,** *5*, 1398.