Supporting Information

"Developing novel activity-based fluorescent probes that target different classes of proteases"

Qing Zhu, Aparna Girish, Souvik Chattopadhaya and Shao Q Yao *

Departments of Chemistry & Biological Science, National University of Singapore, 3 Science Drive 3, Singapore 117543.

1. General:

All chemicals were purchased from commercial vendors and used without further purification, unless otherwise noted. ^{1}H NMR spectra were taken on a Bruker 300 MHz NMR spectrometer. Chemical shifts are reported in parts per million referenced with respect to residual solvent ($CHCl_3$ = 7.26 ppm).

The alkaline phosphatase (P-7640), papain (P-4762), α-chymotrypsin (C-4129), lysozyme(L-6876), lipase(L-1754), trypsin (T-8003), thermolysin (T-7902); renin (83553) and bovine serum albumin(B-4287) were purchased from Sigma (St Louis, USA) while recombinant TEV protease (10127-017) was from Invitrogen.

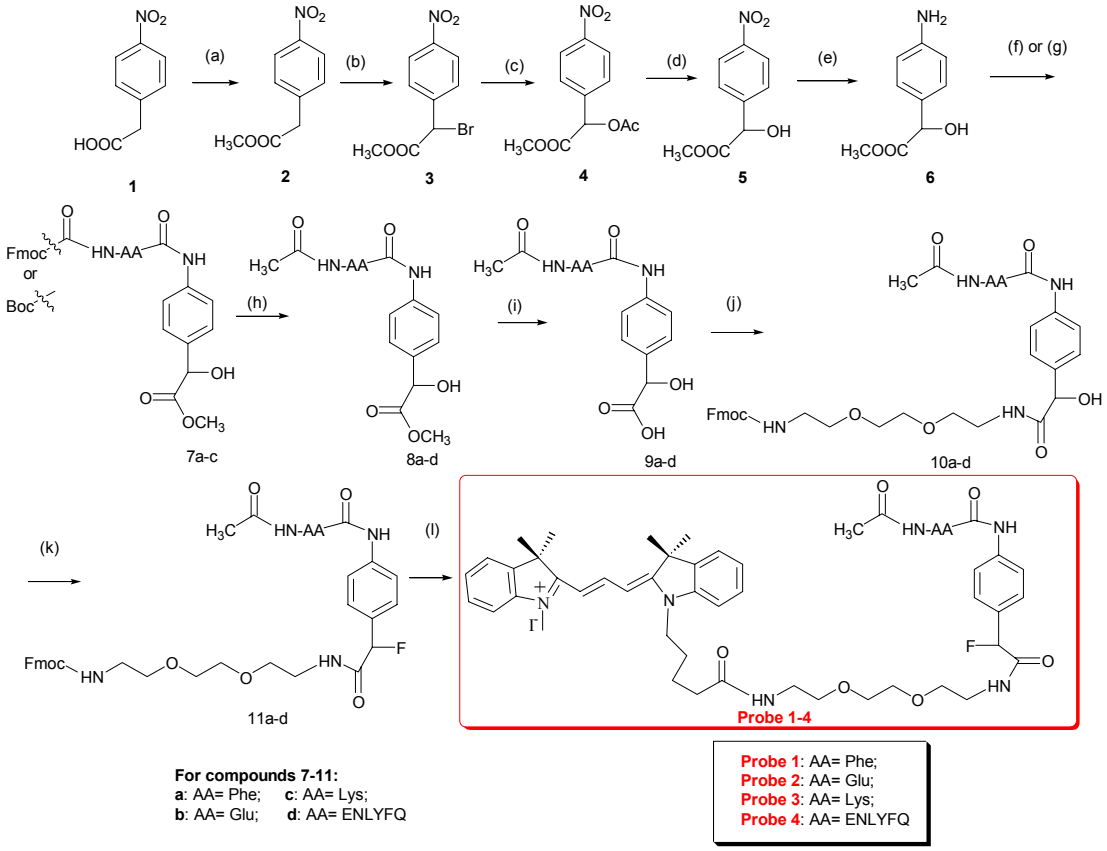
2. Chemical synthesis

4-Nitrophenylacetic acid ethyl ester (**2**):

To a flask containing 4-nitrophenylacetic acid (9.06 g, 50 mmol) dissolved in 100 mL of methanol at 0 °C, thionyl chloride (1.04 g, 5 mmol) was added. The flask was warmed to room temperature and the reaction was allowed to proceed for 24 hours with constant stirring. The solution was then evaporated and the resulting white precipitate was filtered, washed with methanol and dried overnight under reduced pressure to give **2,** a white solid (9.10 g, 93%).[^{1}H-NMR (300 MHz, CDCl3) δ 8.14 (d, *J* = 9.1 Hz, 1H), 7.41 (d, *J* = 9.1 Hz, 1H), 3.72 (s, 2H), 3.71 (s, 3H)]

1'-Bromo-4-nitrophenylacetic acid ethyl ester (**3**):

NBS (1.96 g, 11 mmol) was added to a solution of compound **2** (1.95 g, 10 mmol) in 20 mL of CCl_4 in a reaction flask equipped with a condenser. The flask was irradiated with an 100W tungsten lamp for 5 h resulting in the formation of a white precipitate. Following filtration, the filtrate was concentrated, subjected to silica gel column chromatography and eluted with hexane/ethyl acetate (4/1) to give the ester product **3** (2.33 g, 85%) as a pale yellow solid. ^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.21 (d, *J* = 8.9 Hz, 2H), 7.2 (d, *J* = 8.9 Hz, 2H), 5.40 (s, 1H), 3.81 (s, 3H).



(a) MeOH, $SOCl_2$, 12 h, 93%; (b) NBS, CCl_4, light, 5 h, 85%; (c) CH_3COONa, DMF, 12 h, 73%; (d) HBF_4, 24 h, 74%; (e) H_2, Pd/C,12 h, 92%; (f) Boc-AA-OH or Fmoc-AA-OH, HATU, collidine, DMF, 8 h, 80-85%; (g) Protected hexapeptide ENYLFQ, HATU, collidine, DMF, 8 h, 70%; (h) 50% TFA or 20% pieeridine in DMF, 30 min; then $(CH_3CO)_2O$, 1 h, 45-53%; (i) LiOH, MeOH/H2O=3:1,1 h; (j) EDC, HOBt, FmocNH-(O-O)-NH_2 , 12 h, 80%; (k) DAST, DCM, 1 h, 80%; (k) 20% pieperidin/DMF, 30 min; then Cy_3NHS, DCM, DIEA, 70%.

p-Nitro-mandelic acid methyl ester acetate (ester) (**4**):

Compound **3** (13.7 g, 50 mmol) was dissolved in a mixture of 50 mL DMF and 25 mL H_2O under nitrogen with constant stirring. Following addition of CH_3COONa (8.16 g, 60 mmol), the mixture was stirred at 100 °C. After 12 hours, the solvent was removed by evaporation under reduced pressure, and the residue was partitioned between ethyl acetate (200 mL) and brine (200 mL). The organic layer was then separated, washed with 10% hydrochloric acid (100 mL) and brine (100 mL), dried with Na_2SO_4 and then concentrated under reduced pressure to give a brown oil which was subsequently purified by flash chromatography (silicon gel, ethyl acetate/hexane = 4:1) to afford the pale yellow solid **4** (9.33 g, 73.2%). [^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.25 (d, J = 8.8 Hz, 2H), 7.68 (d, J = 8.8 Hz, 2H), 6.05 (s, 1H), 3.75 (s, 3H), 2.34 (s, 3H)].

p-Nitro-mandelic acid methyl ester (**5**):

To a solution of **4** (0.656 g, 2.57 mmol) in methanol (50 mL), 4.5 mL of 48 % (wt) HBF_4 was added and the reaction was allowed to proceed at room temperature for 24 hours, after which the solvent was removed by evaporation. The crude material was

dissolved in ethyl acetate (50 mL), washed with saturated $NaHCO_3$ (2 x 50 mL), brine (2 x 50 mL), dried over Na_2SO_4 and concentrated under reduced pressure to give a yellow oil. Purification by flash chromatography (silicon gel, ethyl acetate/hexane = 2:1) afforded **5** as a pale yellow solid (0.40 g, 74%). ^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.21 (d, *J* = 8.8 Hz, 2H), 7.64 (d, *J* = 8.8 Hz, 2H), 5.30 (s, 1H), 3.78 (s, 3H).

p-Amino-mandelic acid methyl ester (**6**):

Compound **5** (0.40 g, 1.73 mmol) was dissolved in 50 mL of methanol and Pd/C (40 mg, 10% wt) was added. Hydrogen gas was introduced into the solution and the reaction was allowed to proceed for 8 hours at room temperature with constant stirring. The resulting mixture was then filtered, followed by concentration of the filtrate under reduced pressure to give product **6** as a pale yellow solid (0.29 g, 92.7%). ^{1}H-NMR (300 MHz, CD_3OD) δ 7.17 (d, *J* = 8.4 Hz, 2H), 6.73 (d, *J* = 8.4 Hz, 2H), 5.04 (s, 1H), 3.67 (s, 3H).

*Synthesis of the protected hexapeptide **ENYLFQ***:

The synthesis of the peptide was carried out on trityl chloride resin using standard Fmoc solid phase peptide synthesis. CH_3CO-Glu(OBut)-Asn(Trt)-Leu-Tyr(OBut)-Phe-Gln(Trt)-OH was purified by HPLC (CH_3CN/H_2O/0.1% TFA; 15/85 → 60/40 gradient in 60 min) to give the pure product. ESI-MS: *m/z* 1451.2 $[M + H]^+$.

*General Procedure for coupling reactions between the p-amino-mandelic acid methyl ester(**6**) and different amino acids/peptide*:

The desired protected amino acid/peptide (1 mmol) and HATU (0.456 g, 1.2 mmol) was dissolved in DMF and collidine (158 μL, 1.2 mmol) was added via a syringe. The resulting solution was agitated for 10 min and compound **6** (0.181 g, 1 mmol) was added. The mixture was further agitated for another 12 hrs, after which the DMF was removed *in vacuo* and the residue partitioned between ethyl acetate (50 mL) and 10% HCl (50 mL). The organic layer was washed with brine, dried over $MgSO_4$ and concentrated under reduced pressure to give a yellow oily product. Purification of this compound by flash chromatography (silicon gel, ethyl acetate/hexane = 3:1) afforded **7a-c** and **8d** as a white solid.

[4-(Boc-Phe-amino)-phenyl]-hydroxy-acetic acid methyl ester **7a:**
Yield = 80%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.48-7.12 (m, 9H), 5.12 (s, 1H), 4.95 (m, 1H), 3.77 (s, 3H), 3.18 (m, 2H), 1.41 (s, 9H); ESI-MS: *m/z* 429.2 $[M + H]^+$.

{4-[Fmoc-Glu(OBut-*t*)amino]-phenyl}-hydroxy-acetic acid methyl ester **7b:**
Yield = 85%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.86 (s, 1H), 7.74 (d, *J* = 7.6 Hz, 2H), 7.54 (m, 4H), 7.30 (m, 6H), 6.08 (s, 1H), 5.14 (s, 1H), 4.38 (m, 3H), 4.18 (m, 1H), 3.71 (m, 3H), 2.41 (m, 2H), 2.04 (m, 2H), 1.44 (m, 9H); ESI-MS: *m/z* 611.1 $[M + Na]^+$.

{4-[Fmoc-Lys(Boc)amino]-phenyl}-hydroxy-acetic acid methyl ester **7c:**

Yield = 84%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.91 (s, 1H), 7.72 (d, *J* = 7.6 Hz, 2H), 7.50 (m, 4H), 7.42 (m, 6H), 5.99 (s, 1H), 5.12 (s, 1H), 4.36 (m,2H), 4.13 (m, 2H), 3.69 (m, 3H), 3.05 (m, 2H), 1.90 (m, 1H), 1.70 (m, 1H), 1.41 (m, 11H), 1.26 (m, 2H); ESI-MS: *m/z* 654.3 $[M + Na]^+$.

{[CH_3CO-Glu-Asn(Trt)-Leu-Tyr(OBut)-Phe-Gln(Trt)-amino]-phenyl}-hydroxy-acetic acid methyl ester **8d:**
Yield = 70%; The product was purified by HPLC (CH_3CN/H_2O/0.1% TFA; 15/85 → 60/40 gradient in 60 min). ESI-MS: *m/z* 1614.2 $[M + H]^+$.

*General Procedure for the acetylation of compound **7a-c***:

20% Piperidine or 50% TFA in DMF (5 mL) was added to compound **7** (0.5 mmol) and the resulting solution was stirred for 30 minutes. Following removal of the solvent *in vacuo*, the residue was dissolved in anhydrous DCM and a mixture of acetic anhydride (0.153 g, 1.5 mmol) and DIEA (0.194 g, 1.5 mmol) was added. After 1 hour, the DCM was removed and the residue was purified by flash chromatography (silicon gel, ethyl acetate/hexane = 3:1) to give **8a-c** as a white solid.

[4-(Acetyl -Phe-amino)-phenyl]-hydroxy-acetic acid methyl ester **8a:**
Yield = 45%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.48-7.12 (m, 9H), 5.12 (s, 1H), 4.95 (m, 1H), 3.77 (s, 3H), 3.18 (m, 2H), 1.99 (s, 3H); ESI-MS: *m/z* 370.9 $[M+H]^+$.

{4-[Acetyl -Glu(OBut-*t*)amino]-phenyl}-hydroxy-acetic acid methyl ester **8b:**
Yield = 50%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.49 (d, *J* = 8.43 Hz, 2H), 7.37 (d, *J* = 8.43 Hz, 2H), 7.30 (m, 6H), 5.12 (s, 1H), 4.56 (m, 1H), 3.67 (s, 3H), 2.31 (m, 4H), 2.04 (s, 3H), 1.44 (s, 9H); ESI-MS: *m/z* 409.2 $[M+H]^+$.

{4-[Acetyl -Lys(Boc)amino]-phenyl}-hydroxy-acetic acid methyl ester **8c:**
Yield = 53%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.45 (d, *J* = 8.4 Hz, 2H), 7.29 (d, *J* = 8.43 Hz, 2H), 7.42 (m, 6H), 5.12 (s, 1H), 4.53 (m,1H), 3.67 (s, 3H), 3.00 (m, 2H), 2.03 (s, 3H), 1.90 (m, 2H), 1.41 (m, 11H), 1.26 (m, 2H); ESI-MS: *m/z* 452.1 $[M+H]^+$.

*General Procedure for the synthesis of the compound **11***:

To a solution of compound **8** (0.1 mmol) in methanol (9 mL), 0.1 M LiOH (3 mL) was added and the mixture was stirred at room temperature. After 2 hours, the resulting mixture was concentrated under reduced pressure. Following addition of 1M HCl solution (10 mL), the product was extracted with ethyl acetate (20 mL) and the organic layer was washed with brine (20 mL), dried over $MgSO_4$ and concentrated *in vacuo* to afford **9** as a yellow solid.

To a flask containing compound **9** dissolved in DMF (15 mL), Fmoc protected diamine (44 mg, 0.12 mmol), HOBt (18 mg, 0.12 mmol) and EDC (23 mg, 0.12 mmol) were added and the reaction was allowed to proceed at 0 °C for 30 min with stirring. At the end of this period, the reaction mixture was warmed to room temperature and further stirred for an additional 8 hours, after which aqueous NH_4Cl was added and the product

extracted with ethyl acetate. The combined organic extracts was washed with 1M HCl (50 mL), brine (50 mL), dried with $MgSO_4$ and evaporated to give the crude product **10**. Without further purification, this compound was dissolved in CH_2Cl_2 (12 ml) at 0 °C and (diethylamino)sulfur trifluoride (0.033 ml, 0.3 mmol) was added dropwise. The reaction temperature was then increased to 25 °C. After stirring for 30 minutes, the solvent was removed and the residue partitioned between ethyl acetate (50 ml) and water (50 ml). The organic layer was washed with saturated aqueous NaCl (50 ml), dried over Na_2SO_4 and concentrated under reduced pressure to give the crude product, which was further purified by flash chromatography (silica gel, hexane/ethyl acetate= 1/3) to afford **11**.

11a:
Yield = 53%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.2 Hz, 2H), 7.68 (m, 4H), 7.36 (m, 6H), 5.70 (d, *J* = 48.6 Hz, 1H), 5.35 (m, 1H), 4.54 (m, 1H), 4.39 (m, 2H), 4.20 (m, 2H), 3.53 (m, 10H), 3.38 (m, 2H), 2.46 (m, 2H), 2.10 (m, 2H), 2.01 (s, 3H), 1.45 (m, 9H); ESI-MS: *m/z* 771.3 $[M + Na]^+$.

11b:
Yield = 60%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.2 Hz, 2H), 7.58 (d, *J* = 7.7 Hz, 2H), 7.31 (m, 12H), 7.06 (m, 1H), 5.70 (d, *J* = 48.5 Hz, 1H), 5.43 (m, 1H), 4.41 (m, 2H), 4.20 (m, 1H), 3.63 (m, 10H), 3.37 (m, 2H), 3.07 (m, 2H), 2.04 (s, 3H); ESI-MS: *m/z* 733.3 $[M + Na]^+$.

11c:
Yield = 66%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.75 (d, *J* = 7.2 Hz, 2H), 7.58 (m, 4H), 7.36 (m, 6H), 5.70 (d, *J* = 48.6 Hz, 1H), 5.33 (m, 1H), 4.59 (m, 2H), 4.40 (m, 1H), 3.61 (m, 10H), 3.37 (m, 2H), 3.09 (m, 2H), 2.04 (s, 3H), 1.90 (m, 2H), 1.43 (m, 13H); ESI-MS: *m/z* 814.4 $[M + Na]^+$.

11d:
Yield = 55%; The product was purified by HPLC (CH_3CN/H_2O/0.1% TFA; 15/85 → 60/40 gradient in 60 min). ESI-MS: *m/z* 1977.0 $[M + Na]^+$.

*General Procedure for the synthesis of **Probes 1-4***:

To compound **11** (0.05 mmol), 20% piperidine in DMF (3 mL) was added and the resulting solution was stirred for 30 min. At the end of this period, the solvent was removed *in vacuo* and the residue dissolved in DCM (2 mL). Cy_3-NHS (20.0 mg, 0.06 mmol) and triethylamine (8 μL, 0.06 mmol) were then added and the reaction mixture was stirred at room temperature for 18 h followed by concentration under reduced pressure to give the crude product as syrup. Following de-protection with 50% TFA/DCM, the product was further purified by HPLC (CH_3CN/H_2O/0.1% TFA; 15/85 → 60/40 gradient in 60 min) to give the desired product (**Probes 1-4**).

Probe 1:
Yield = 70%; ^{1}H-NMR (300 MHz, CD_3OD) δ 8.53 (m, 1H), 7.53 (m, 4H), 7.41 (m, 4H), 7.31 (m, 9H), 6.42 (m, 2H), 5.70 (d, *J* = 48.5 Hz, 1H), 3.60 (m, 12H), 3.30 (m, 4H), 2.29 (t, *J* = 6.8 Hz, 2H), 1.99 (s, 3H), 1.76 (s, 19H); ESI-MS: *m/z* 913.5 $[M - I]^+$.

Probe 2:
Yield = 68%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.37 (m, 1H), 8.19 (m, 2H), 7.69 (m, 2H), 7.37 (m, 8H), 7.14 (m, 2H), 6.52 (m, 2H), 5.70 (d, *J* = 48.5 Hz, 1H), 4.62 (m, 1H), 4.02 (m, 2H), 3.62 (m, 14H), 2.21 (m, 2H), 2.03 (s, 3H), 1.71 (m, 10H), 1.25 (s, 6H), 0.89 (m, 3H); ESI-MS: *m/z* 895.6 $[M - I]^+$.

Probe 3:
Yield = 70%; ^{1}H-NMR (300 MHz, $CDCl_3$) δ 8.41 (m, 1H), 8.13 (m, 2H), 7.69 (m, 2H), 7.48 (m, 6H), 7.15 (m, 4H), 6.38 (m, 2H), 5.70 (d, *J* = 48.5 Hz, 1H), 4.60 (m, 1H), 4.00 (m, 2H), 3.57 (m, 14H), 3.35 (m, 2H), 3.11 (m, 2H), 2.04 (m, 7H), 1.71 (m, 10H), 1.25 (s, 6H), 0.89 (m, 3H); ESI-MS: *m/z* 894.6 $[M - I]^+$.

Probe 4:
Yield = 53%; ESI-MS: *m/z* 1560.6$[M - I]^+$.

3. Procedure for Protein labeling with the 4 probes:

Stock solutions of the different proteins, except TEV, were prepared as ~2 mg/ml solutions in distilled water, desalted with a NAP5 column (Amersham, USA) and stored at -20 ^oC until use. TEV protease (10 U/μL) was used directly according to the manufacturer's protocol. Since the different proteins, as obtained from commercial sources, contain different amount of salts, the final absolute weights of the protein in the estimated 2 mg/ml stock solution likely differ substantially. Indeed, Coomassie gels of the 2 mg/ml protein stocks revealed that this was the case: same volumes of the 2 mg/ml proteins, when loaded and separated on SDS-PAGE gel, differ in the band intensity greatly, indicating the amount of each protein differs. Furthermore, the catalytic activity (as measured by unit/μL) of each commercial protein also differs greatly, which further complicate the accurate quantitation of our labeling reactions. Consequently, the degree of fluorescence labeling in each protein should only be used as a qualitative measure to determine the relative activity of each probe against the said protein.

The stock solutions for the four probes was prepared as 200 μM solutions in methanol and stored at -20 ^oC.

Fluorescence imaging was done using the Typhoon 9200 scanner (Amersham, USA)

General procedure for labeling experiments of non-proteases using any of ***Probes 1-4*** *(negative control)*:

2 μl of non-proteases (BSA, alkaline phosphatase, lipase, or lysozyme) were incubated with 0.2 μl of any of these four probes in 18 μl of Tris buffer (50 mM, pH 8). After 30 min, 2 h and 8 h, the samples were boiled at 95 ^{o}C for 5 minutes with 4 μl of 6X

SDS-PAGE loading dye. The samples were then resolved on 12% gels by SDS-PAGE and visualized by fluorescence scanning. SDS-PAGE results show no labeling for any of the proteins with the probes even after 8 h of labeling reaction.

General procedure for labeling experiments of proteases using ***Probes 1-4***:

2 µl of the protein stock solution (~2 mg/ml) was diluted with 18 µl of Tris buffer (50 mM, pH 8). 0.2 µl of the probe (200 µM) was then added and the reaction mix was incubated at room temperature in the dark for 30 min, 2 hours and 8 hours, respectively. The reaction was quenched by boiling the sample at 95 ^{o}C for 5 minutes with 4 µl of 6X SDS-PAGE loading dye. The sample was then resolved on 12% gels by SDS-PAGE and visualized by fluorescence scanning.

For labeling experiments using TEV, the reaction was done in a buffer containing 50 mM Tris, 0.5 mM EDTA and 1 mM DTT with all other conditions remaining the same.

Some representative gel results were shown below:

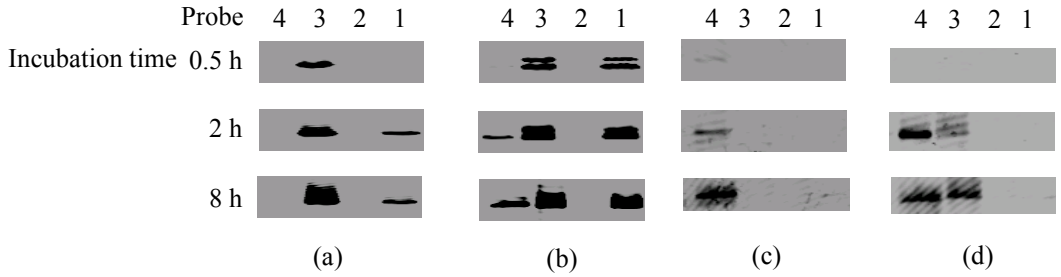


Figure 4. The labeling experiments were carried out with (a) Trypsin, (b) α-Chymotrypsin, (c) TEV NI a protease, (d) Papain, in 50 mM Tri buffer (pH 8) for 0.5 h, 2 h, 8 h respectively at room temperature with **Probe 1-4**.

Above results indicate that a longer incubation time to some extent increases the protein labeling for most probes, with 2 h being the optimal incubation time that accurately assess whether a protein is active against a particular probe.

Labeling experiments using heat-denatured proteins:

2 µl of the protein stock solution (~2 mg/ml) was diluted with 18 µl of Tris buffer (50 mM, pH 8) and heated at 95 ^{o}C for 3 minutes, allowed to cool to room temperature and labeled with the probes. The manner of labeling and detection was the same as the experiments described above.

Labeling experiments in the presence of protease inhibitors:

To 2 µl of the protein stock solution (~2 mg/ml), 0.2 µl of inhibitor, PMSF or IA (~200 µM) and 18 µl of Tris buffer (50 mM, pH 8) was added and the samples were incubated at room temperature for 1 hour. 0.2 µl of the probe was then added and the

reaction was incubated at room temperature in dark for an additional 2 hours. At the end, the reaction was quenched by boiling the samples at 95 ^{o}C for 5 minutes with 4 μl of 6X SDS-PAGE loading dye. The samples were then resolved on 12% gels by SDS-PAGE and visualized by fluorescence scanning.

Other relaed issues regarding our labeling approach:

Previous studies (e.g. reference 6a in maintext) using similar probes to target tyrosine phosphatases (PTP) indicated that, multiple copies of the reactive quinone methide may be added to the same target enzyme, generating a multiple alkylation product. In addition, it has also been observed that quinone methide may also diffuse away from the active site of the target enzyme, and end up nonspecifically labeling other enzymes.

In our approach, we generated a reactive quinolimine methide intermediated, which is likely similar to quinine methide in its chemical reactivity. Therefore, some of the above observations with quinine methide may occur in our approach. Consequently, our current research is directed at a better understanding of the detail mechanism and some potential limitations of our labeling approach: (1) Are our labeling reagents specific for Cys versus Lys nucleophiles in the enzyme active site, or other nucleophiles? (2) How many molecules of our labeling reagent are being added to each labeled enzyme, does it involve a single alkylation or multiple alkylation reaction? (3) The species generated from our labeling reagents are likely extremely reactive. They may diffuse from the active site of the target protein which could potentially provide false positives for activity-based profiling studies.

We will report our findings in due course.

We thank Reviewer 2 for pointing out some of these potential issues.