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

Qing Zhu,[a] Aparna Girish,[b] Souvik Chattopadhaya[b] and Shao Q. Yao*[ab]

[a] Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543.
E-mail: chmyaosq@nus.edu.sg; Fax: +65 6779 1691; Tel: +65 6874 1683
[b] Department of Biological Sciences, National University of Singapore, 3 Science Drive 3, Singapore
117543. E-mail: chmyaosq@nus.edu.sg; Fax: +65 6779 1691; Tel: +65 6874 1683

In this article, we report the design and synthesis of a group of novel activity-based probes that target different protease sub-classes based on their substrate specificities, rather than their enzymatic mechanisms. The feasibility of our approach has been demonstrated by using representative members of the different protease sub-classes.

Proteases are enzymes that catalyze the cleavage of peptide bonds.[1] As one of the largest and most important groups of enzymes, they play significant roles in mediating different cellular processes such as DNA replication, cell-cycle progression, cell proliferation, differentiation, migration *etc*.[2] No wonder that malfunctioning of the cellular proteolysis system has been associated with many diseases such as emphysema, stroke, viral infections, cancer and Alzheimer's disease. It is therefore critical to develop strategies which, in a diseased state of the cell, would allow for the rapid profiling of different classes of proteases based on their activity. Such strategies, in turn, would greatly aid in the identification of potential pharmaceutical targets.

In the post-genomic era, proteomics, with its ability to bridge the gap between the genome sequence and cellular functionalities of proteins, has emerged as a major tool for the high-throughput identification and characterization of proteins.[3] In particular, activity-based proteomic approaches have been successfully applied to study enzymes present in different proteomes.[4] By taking advantage of mechanism-based chemical probes which covalently modify different enzyme classes in an activity-dependent fashion, it is now possible to simultaneously monitor not only the expression levels of different enzymes but also their localization, regulation and inhibition. Given the importance of proteases, it is not surprising that the majority of activity-based probes developed so far are tailored toward different classes of proteases.[5] For example, fluorophosphonate/fluorophosphate derivatives have been developed to selectively profile serine hydrolases, including serine proteases.[5a,b] For cysteine proteases, different classes of chemical probes have been reported, including probes containing α-halo or (acyloxy)methyl ketone substituents,[5c,d] epoxy- and vinyl sulfone-derivatized peptides.[5e–g] However, till now, there has been no report on activity-based probes capable of profiling aspartic proteases or metalloproteases, which constitute the other two major classes of proteases. This is due to the lack of known mechanism-based inhibitors that can form covalent adducts with these enzymes, a prerequisite of all existing activity-based chemical profiling approaches.[4] Other known activity-based probes include sulfonate ester-containing probes that target a few different classes of enzymes,[6a] as well as probes conjugated to *p*-hydroxymandelic acid which specifically label protein phosphatases.[6b,c] Herein, we report the design and synthesis of a group of activity-based chemical probes, which can target and profile different classes of proteases (*i.e.* serine/cysteine/aspartic/metallo proteases) by virtue of their ability to be recognized as potential enzyme substrates (Fig. 1 and Fig. 2). This profiling approach allows for the covalent labeling of target enzymes purely based on their intrinsic catalytic

activities. In future, the strategy may also be extended to profile other classes of hydrolytic enzymes, besides proteases.

Each probe was designed to mimic an enzyme substrate which would then be recognized and hydrolyzed by the specific enzyme. Hydrolysis would result in the release of a reactive intermediate which then covalently labels the enzyme, making it easily identifiable. Since the targets in our study encompass different classes of proteases, we designed probes which contain amino acids/peptides linked, *via* the scissile bond (*e.g.* the amide bond which the enzyme hydrolyzes), to a *p*-aminomandelic acid moiety, which, upon cleavage by the protease would rearrange to generate the reactive quinolimine methide (Fig. 1A). Previous work from our group and others had shown that *p*-hydroxymandelic acid-containing probes can be used to specifically label protein tyrosine phosphatases (PTP) in an activity-based fashion.[6b,c] Mechanistically, these probes work by the formation of a reactive quinone methide intermediate within the active site of PTP (Fig. 1B), which then readily alkylates nucleophilic groups of the enzyme to form a covalent enzyme–probe adduct. We therefore hypothesized that quinolimine methide, by virtue of its structural similarity to quinone methide, should possess similar chemical reactivity in alkylating neighboring nucleophilic groups on the enzymes. Accordingly, we designed a series of protease probes, which contain recognition heads made of either amino acids or peptides, corresponding to the P-site residues in protease substrates (Fig. 2A). The recognition head was linked to the *p*-aminomandelic acid moiety *via* an amide bond which imitates the scissile bond in a protease substrate. A fluorescent reporter group, Cy3, was attached to the other end of *p*-aminomandelic acid with 2,2-(ethylenedioxy)bis(ethylamine) as the linker. We envisaged that, if the probe is recognized as a potential substrate by a target protease, it will be recruited to the enzyme active site, where proteolytic cleavage will
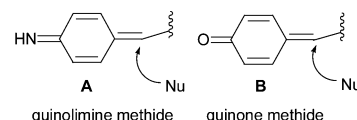

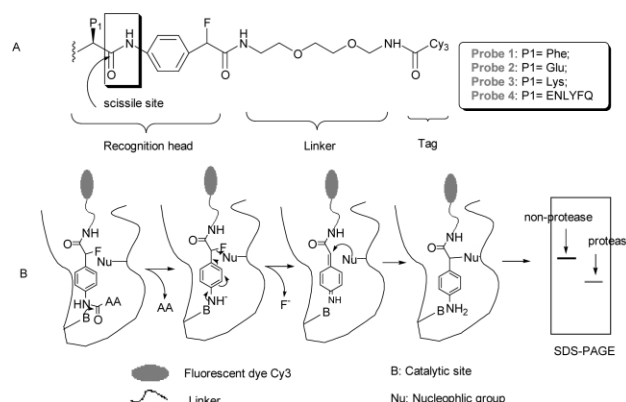**Fig. 1** Structures of the two reactive intermediates.

Fig. 2 A. The structure of probes, with their scissile cleavage sites highlighted; B. The activity-based labeling mechanism of the probes.

† Electronic supplementary information (ESI) available: experimental details. See http://www.rsc.org/suppdata/cc/b4/b404471a/

occur within the recognition head of the probe (Fig. 2B). However, only the cleavage of the scissile site, and not any other site within the probe, will trigger the release of the reactive quinolimine methide, which would subsequently react with nearby nucleophilic groups within the active site of the enzyme, leading to an irreversible fluorescent labeling of the enzyme which can then be easily identified after resolving the protein mixture by denaturing SDS polyacrylamide gel electrophoresis.

To test the feasibility of our strategy, three probes (**Probes 1** to **3**) were chemically synthesized (see ESI†), each containing only a single amino acid-phenylalanine, glutamic acid or lysine. These amino acids correspond to the P_1 position of a protease substrate. The three amino acids were so chosen that they may target different classes of proteases conferring different substrate specificities. The choice of the P_1 position amino acid was based on the assumption that P_1 position is the most critical residue in a protease substrate for defining protease specificity. This is true to some extent, as many proteases are known to accept substrates which contain only single amino acids (i.e. P_1 residues). Other proteases such as the Tobacco Etch Virus (TEV) NIa protease, are known to be highly specific and recognize only well-defined peptide sequences. In order to validate that our approach is equally amenable for activity-based profiling of these type of proteases, we synthesized **Probe 4**, which contains an ENLYFQ hexapeptide sequence (Fig. 2A), corresponding to the TEV protease recognition sequence.

To demonstrate the labeling characteristics, the four probes were tested against a panel of commercially available proteins, both proteases and non-proteases. No labeling, even after prolonged hours of incubation, was seen with any of the proteins (e.g. BSA, alkaline phosphatases, lysozyme and lipase) which do not belong to the protease family (see ESI†). Having shown that the labeling is restricted only to proteases, we then used six different proteases covering all the four protease classes for the labeling experiments. Trypsin and α-chymotrypsin are serine proteases; TEV NIa protease and papain are cysteine proteases; thermolysin is a metalloprotease while renin is an aspartic protease. As can be seen from Fig. 3A, each representative member of a protease class was labeled only by the probe bearing the P_1 amino acid corresponding to its known substrate specificity. For instance, trypsin, a protease strongly preferring basic residues at the P_1 position, was preferentially labeled by the Lys-bearing **Probe 3**; α-chymotrypsin, known to prefer hydrophobic P_1 residues, was labeled by the Phe-bearing **Probe 1** while the highly specific TEV protease was labeled exclusively by **Probe 4**, which has the ENLYFQ TEV recognition

sequence. Thermolysin and renin both have broad substrate specificities and hence were labeled by all 4 probes. Papain is known to prefer substrates containing a hydrophobic residues at the P_2 rather than the P_1 position. It is therefore not surprising that it was preferentially labeled only by **Probe 4**, which has Phe at the P_2 position in the hexapeptide sequence. We also observed some labeling with the Phe-bearing **Probe 1** and Lys-bearing **Probe 3** to trypsin and α-chymotrypsin, respectively, indicating that amino acid-containing probes (i.e. **Probes 1** to **3**), by virtue of their minimal enzyme recognition sequence, may also be useful for broad-based protease profiling experiments.
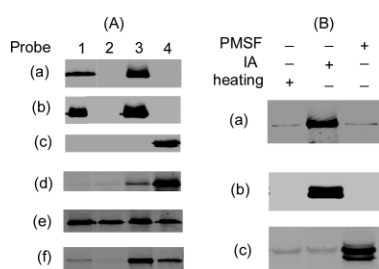
Having successfully shown the efficacy of our labeling strategy, we next set to confirm that the observed labeling was dependent on the active state of the enzymes. Prior to labeling with the probes, the different enzymes trypsin, α-chymotrypsin and TEV NIa protease were treated either with phenylmethylsulfonyl fluoride (**PMSF**) or iodoacetamide (**IA**), which are known serine and cysteine protease inhibitors, respectively, or heat inactivated. (Fig. 3B). As expected, the labeling of trypsin and α-chymotrypsin was inhibited by **PMSF** but not **IA** while TEV NIa protease was inhibited by **IA**, but not **PMSF**. No labeling was seen with any of the enzymes upon heating (the first column in Fig. 3B). Taken together, the results validate the activity-based design strategy of the probes.

In conclusion, a new class of activity-based probes has been successfully designed, synthesized and tested. Since the labeling by these probes is dependent on the cleavage of the amide bond in the recognition head of the probes, they can be used to selectively profile different classes of proteases in a broad-based or specific manner, depending on the choice of the amino acids/peptides used as recognition elements. We believe this substrate-based, activity-dependent protein profiling approach will find wide ranging applications in the field of proteomics. We are currently investigating the detailed mechanism and potential limitations of our labeling approach.

Fig. 3 (A) The selectivity of probes towards different proteases: the labeling experiments were carried out with (a) trypsin, (b) α-chymotrypsin, (c) TEV NIa protease, (d) papain, (e) thermolysin, (f) renin for 2 h. Noted the relative catalytic activity of each enzyme differs, which may cause differences in labeling intensity. (B) Inhibition experiments with (a) trypsin labeled with **Probe 3**; (b) a-chymotrypsin labeled with **Probe 3**; (c) TEV protease labeled with **Probe 4**.

Notes and references

1 A. J. Barrett, *Handbook of Proteolytic Enzymes*, Academic Press, London, 1998.
2 M. D. Sternlicht and Z. Werb, *Annu. Rev. Cell. Dev. Biol.*, 2001, **17**, 463.
3 (*a*) M. Tyers and M. Mann, *Nature*, 2003, **422**, 193; (*b*) S. Fields, *Science*, 2001, **291**, 1221; (*c*) G. Y. J. Chen, M. Uttamchandani, R. Y. P. Lue, M. L. Lesaicherre and S. Q. Yao, *Curr. Top. Med. Chem.*, 2003, **3**, 705.
4 (*a*) B. F. Cravatt and E. J. Sorensen, *Curr. Opin. Chem. Biol.*, 2000, **4**, 663; (*b*) D. A. Jeffery and M. Bogyo, *Curr. Opin. Biotechnol.*, 2003, **14**, 87; (*c*) X. Huang, E. L. P. Tan, G. Y. J. Chen and S. Q. Yao, *Appl. Genomics Proteomics.*, 2003, **2**, 225.
5 (*a*) Y. S. Liu, M. P. Patricelli and B. F. Cravatt, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 14694; (*b*) D. Kidd, Y. S. Liu and B. F. Cravatt, *Biochemistry*, 2001, **40**, 4005; (*c*) L. Faleiro, R. Kobayashi, H. Fearnhead and Y. Lazebnik, *EMBO J.*, 1997, **16**, 2271; (*d*) M. L. Liau, R. C. Panicker and S. Q. Yao, *Tetrahedron Lett.*, 2003, **44**, 1043; (*e*) D. Greenbaun, K. F. Medzihradszky, A. Burlingame and M. Bogyo, *Chem. Biol.*, 2000, **7**, 569; (*f*) T. Nizif and M. Bogyo, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 2967; (*g*) G. Wang, M. Uttamchandani, G. Y. J. Chen and S. Q. Yao, *Org. Lett.*, 2003, **5**, 737.
6 (*a*) G. C. Adam, E. J. Sorensen and B. F. Cravatt, *Nat. Biotechnol.*, 2002, **20**, 805; (*b*) L. C. Lo, T. L. Pang, C. H. Kuo, Y. L. Chiang, H. Y. Wang and J. J. Lin, *J. Proteome Res.*, 2002, **1**, 35; (*c*) Q. Zhu, X. Huang, G. Y. J. Chen and S. Q. Yao, *Tetrahedron Lett.*, 2003, **44**, 2669.