Identification of protease substrates by combinatorial profiling on TentaGel beads†

Jacob Kofoed‡ and Jean-Louis Reymond*

Received (in Cambridge, UK) 5th September 2007, Accepted 5th October 2007
First published as an Advance Article on the web 11th October 2007
DOI: 10.1039/b713595e

Reacting a 65 536 member combinatorial library of octapeptides on TentaGel beads with various proteases followed by selective staining of the free amino termini at the reacted bead surface and sequence determination by amino acid analysis allowed the rapid identification of protease substrates.

Protease substrates are used for tracking and studying proteases in relation to their role in biological processes and for screening inhibitors towards drug development.[1–3] Protease substrates may be discovered through profiling experiments, such as the MS-based identification of cleavage sites in reference proteins and peptides,[4,5] the selection of reactive peptides from phage-displayed peptide libraries,[6] activity measurements with synthetic fluorogenic peptide substrates[7] including combinatorial libraries of FRET substrates on a solid support,[8] series of positional scanning mixtures,[9] PNA-tagged libraries,[10] microarray displayed substrates,[11] and active-site directed probes.[12] Although efficient, all of these protease profiling methods involve complex experiments in terms of synthesis and analysis and remain inaccessible to most laboratories. Herein, we report a protease profiling experiment based on the direct reaction of the protease with a combinatorial library of peptides on TentaGel beads.[13] The method is simpler than previously reported profiling experiments and provides rapid access to protease substrates.

We envisioned a simple assay in which protease substrates would be identified by surface proteolysis ('bead shaving')[14,15] of the synthesis beads of a split-and-mix combinatorial peptide library,[16] followed by selective staining of beads carrying a proteolysed peptide with an amine-selective reagent, and bead decoding.[17] Because reactive sequences are often found in even relatively small series of peptides for any given protease,[18,19] a typical assay with 50 mg of solid support, corresponding to approximately 65 000 beads, should be sufficient for such an experiment.

A small combinatorial library **AcL** of octapeptides, a suitable length for peptide substrates of proteases,[1] was prepared by split-and-mix synthesis using four different amino acids per variable position, resulting in 4^8 = 65 536 octapeptides (Fig. 1). Sixteen different amino acids were used including hydrophobic (Val, Leu, Ile, Met, Pro, Hyp), aromatic (Phe, Tyr), positively charged (Arg, His), negatively charged (Asp, Glu), and small and polar amino acids (Gly, Ala, Ser, Thr). Each amino acid appeared at only two variable positions in this library, allowing sequence determination

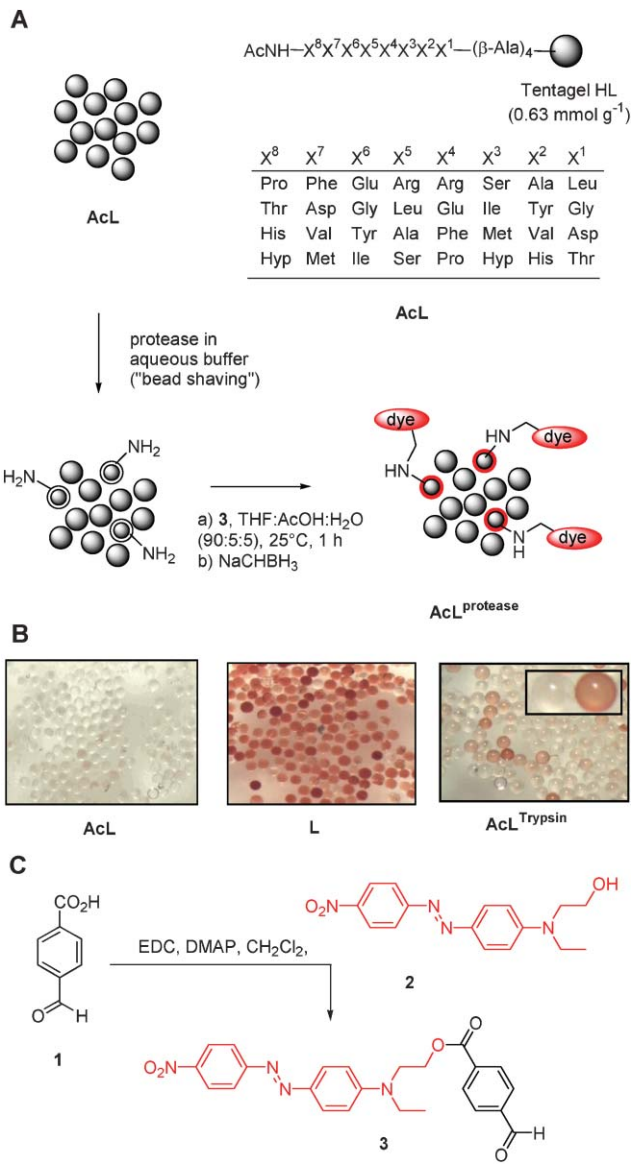

Fig. 1 Protease profiling on TentaGel beads. **A**. Library composition and bead shaving protocol with aldehyde **3**. **B**. Microscope pictures of resin beads after reductive alkylation conditions with **AcL**, **L** (non-acetylated library), and **AcL** after tryptic digest. The zoom in the picture on the right shows the surface stain typical for bead shaving. The beads are 90 μm in diameter. **C**. Synthesis of aldehyde **3**.

Department of Chemistry and Biochemistry, University of Berne, Freiestrasse 3, CH-3012 Berne, Switzerland. E-mail: jean-louis.reymond@ioc.unibe.ch; Fax: +41 31 631 80 57; Tel: +41 31 631 43 25
† Electronic supplementary information (ESI) available: HPLC and MS analyses of protease digests. See DOI: 10.1039/b713595e
‡ *Present address:* Novo Nordisk A/S, Diabetes Protein and Peptide Chemistry, Novo Nordisk Park, DK-2760 Måløv, Denmark.

by amino acid analysis as described previously for dendritic peptides,[20] which allows sequence determination at low cost.[21] Aromatic residues and arginine were placed near the center to offer optimal cleavage for trypsin and chymotrypsin as model proteases. Lysine was omitted and the N-terminus was acetylated, to allow identification of cleaved beads by an amine-selective tagging reagent (see below). The library was synthesized by Fmoc-solid phase split-and-mix peptide synthesis using carbodiimide couplings (2.5 equiv. Fmoc amino acid and 2.5 equiv. DIPCDI–HOBt in DMF) on a 1.0 g batch of TentaGel HL resin (Rapp Polymere, 0.63 mmol g^{-1}). After removal of the last Fmoc protecting group (20% piperidine in DMF) and acetylation (Ac_2O–DCM (1 : 1, v/v)), the side-chain protecting groups were removed by acidic treatment (TFA–TIS–H_2O (95 : 2.5 : 2.5, v/v/v)).

Staining of the free N-terminus liberated by proteolysis was tested by comparing library **AcL** as negative control with the non-acetylated library (**L**) and a sample of the *N*-acetylated library treated with trypsin (**AcL**$^{\mathbf{Trypsin}}$) as positive controls. Standard amine-staining reagents for peptide coupling or a biotinylation procedure with biotin *N*-hydroxysuccinimide ester at neutral pH[22] stained all three test samples equally well, probably due to the presence of reactive unprotected side chains (Thr, Ser, Tyr, His). To our delight, reaction of aldehyde **3**, prepared by coupling of 4-carboxybenzaldehyde (**1**) with Disperse Red 1 (**2**), in the presence of $NaBH_3CN$ produced the desired chemoselective staining of **L** and **AcL**$^{\mathbf{Trypsin}}$ against **AcL** (Fig. 1). Reductive alkylation is known as a preparative procedure in solid-supported synthesis,[23] but has not been used previously for staining as shown here.

The protease profiling experiment was carried out for the serine proteases trypsin, chymotrypsin and subtilisin, and the aspartic protease pepsin. Analysis of stained beads gave octapeptide sequences consistent with the known specificity of the proteases (Table 1).[24] In the case of trypsin, all stained beads contained at least one arginine residue in their sequence, while non-stained beads had arginine only at Arg-Pro and Arg-Hyp dyads. Arginine next to proline is indeed known to be resistant to tryptic digestion. In the case of α-chymotrypsin all sequences contained Phe or Tyr, following the well-known selectivity of this enzyme. For subtilisin, Thr was often found in position X^8 as well as Phe in X^7 and Arg in X^5/X^4. The pepsin assay was rather unspecific, but Asp in X^7 and Leu in X^5 might be important.

The data from the on-bead cleavage assays were confirmed by resynthesis of seven selected octapeptides and determination of their cleavage pattern in solution against each of the four proteases. The cleavage sites were determined using MS analysis of the protease digest mixture. This assay confirmed the results of the on-bead assays with cleavage sites corresponding to the known protease specificities (Fig. 2).[24] Several proteases also hydrolyzed peptides stemming from screening with a different protease, as

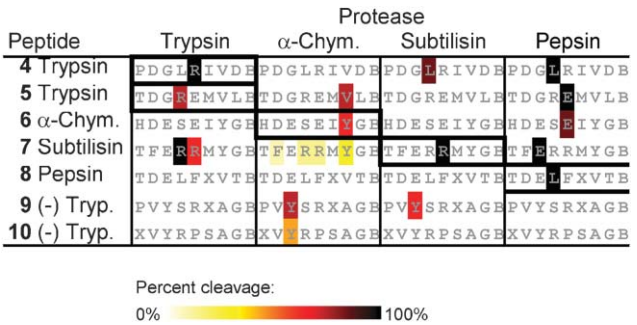

Fig. 2 Color-coded protease activity with peptides **4–10**. The percentage cleavage relative to the parent peptide is shown by coloring the P1 position of each detected fragment according to the scale indicated. Standard one-letter amino acid codes are used. X = hydroxyproline; B = β-alanine. The solution assays were run with 1 mM peptide and 50 μg mL^{-1} protease in the same buffer as the on-bead assay. The reactions were followed by RP-HPLC at 214 nm for 2 h, and the cleavage sites were determined using MS analysis.

Table 1 Protease-reactive sequences with library **AcL** identified by amino acid analysis (AAA) of stained beads[21]

Trypsin[a]	Negative trypsin[a]	α-Chymotrypsin[a]	Subtilisin[b]	Pepsin[c]
AAA giving single sequences:				
PDGLRIVD	PDISPXHS	**HDESEIYG**	**TFERRMYG**	**TDELFXVT**
XDERFSYD	TVEAPSVG	HFGAESHT	TFERRMVG	TFERRMAG
TMGLRSHL		HMGRFIHG	TDILPXAT	TDILEXAT
TVGRFSHT		HFGAESVG	TDYARSAG	TDEAPIYT
XMILRSHL		HFGAEMAT	TFYSEXVS	XDELPMAT
XDERFMVD		TFYRFMYG	TFGRESAD	
TDEARMHD		TFIAPSHT	TDERFSAG	
PFGRRXHG		TFERFMAL	TFGRESVD	
		TFELFXYT	TDERFSVG	
			TVESFXYD	
AAA giving double sequences:[d]				
TDELRMVG	**PVYSRXAG**	TFEARIHG	TDESPIHL	TVGAPIYD
TDGREMVG	**XVYRPSAG**	HFGREIAT	PDILESHT	TDYAPIVG
XDELRIYG	TFYLPSAG		TDILPSYG	TMGLFXVD
XDGREIYL	TFGAPSYL		TDGSPIYL	XDELFMVT
	XVYLGIAG		TFERPIHG	
	XVGAEIYL		PFGREIHT	
	TFYRPXAG		TFYARSHD	
	PFGARXYT		HDYRFSAT	

[a] Conditions: aq. pH 8.0, protease (1 mg mL^{-1}), 25 °C, 24 h, then 10 mM **3** in THF–H_2O–AcOH (90 : 5 : 5, v/v/v), 1 h, then $NaBH_3CN$. [b] Same as *a* using aq. pH 6.5. [c] Same as *a* with aq. pH 4.0. All experiments gave 30–40% stained beads. The surface proteolysis did not significantly alter bead composition for amino acid analysis (< 10% conversion of the material on the bead). [d] See ref. 21 for details of the sequencing method. Sequences in bold were re-synthesized.

could be expected due to the presence of multiple cleavage sites in their sequence.

The experiments above demonstrate that direct proteolysis of combinatorial peptide libraries on a solid support followed by selective tagging of free N-termini and decoding by amino acid analysis provides a rapid entry into protease substrates. The survey of a 65 536-member tags-free library, although only covering a very limited sequence space, is sufficient to provide reactive substrates for each case studied. The library composition can be adapted for specific needs depending on the targeted protease family.

Notes and references

1 (*a*) J. Powers, J. Asgian, O. Ekici and K. James, *Chem. Rev.*, 2002, **102**, 4639–4750; (*b*) M. Meldal, *QSAR Comb. Sci.*, 2005, **24**, 1141–1148; (*c*) S. L. Diamond, *Curr. Opin. Chem. Biol.*, 2007, **11**, 46–51.

2 X. S. Puente, L. M. Sanchez, C. M. Overall and C. Lopez-Otin, *Nat. Rev. Genet.*, 2003, **4**, 544–568.

3 P. L. Richardson, *Curr. Pharm. Des.*, 2002, **8**, 2559–2581.

4 I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.*, 1967, **27**, 157–162.

5 J. E. Sheppeck, II, H. Kar, L. Gosink, J. B. Wheatley, E. Gjerstadt, S. M. Loftus, A. R. Zubirua and J. W. Janc, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 2639–2642.

6 (*a*) D. J. Matthews and J. A. Wells, *Science*, 1993, **260**, 1113–1117; (*b*) J. F. Chaparro-Riggers, R. Breves, K.-H. Maurer and U. Bornscheuer, *ChemBioChem*, 2006, **7**, 965–970.

7 J. L. Harris, in *Enzyme Assays: High-throughput Screening, Genetic Selection and Fingerprinting*, ed. J.-L. Reymond, Wiley-VCH, Weinheim, Germany, 2006, pp. 303–331.

8 M. Meldal, I. Svendsen, K. Breddam and F. Auzanneau, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 3314–3318.

9 (*a*) N. A. Thornberry, T. A. Rano, E. P. Peterson, D. M. Rasper, T. Timkey, M. Garcia-Calvo, V. M. Houtzager, P. A. Nordstrom, S. Roy, J. P. Vaillancourt, K. T. Chapman and D. W. Nicholson, *J. Biol. Chem.*, 1997, **272**, 17907–17911; (*b*) J. H. Harris, B. J. Brackes, F. Leonetti, S. Mahrus, J. A. Ellmann and C. S. Craik, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7754–7759.

10 (*a*) N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick and J. L. Harris, *Chem. Biol.*, 2004, **11**, 1351–1360; (*b*) J. J. Díaz-Mochón, L. Bialy and M. Bradley, *Chem. Commun.*, 2006, 3984–3986.

11 (*a*) C. M. Salisbury, D. J. Maly and J. A. Ellman, *J. Am. Chem. Soc.*, 2002, **124**, 14868–14870; (*b*) D. N. Gosalia, C. M. Salisbury, J. A. Ellman and S. L. Diamond, *Mol. Cell. Proteomics*, 2005, **4**, 626–636.

12 (*a*) M. Bogyo, S. Verhelst, V. Bellingard-Dubouchaud, S. Toba and D. Greenbaum, *Chem. Biol.*, 2000, **7**, 27–38.

13 TentaGel is a low crosslinked polystyrene matrix grafted with polyethylene glycol, and is a standard support for solid-phase synthesis compatible with aqueous conditions for bioassays. See www.rapp-polymere.com.

14 J. Vagner, G. Barany, K. S. Lam, V. Krchnak, N. F. Sepetov, J. A. Ostrem, P. Strop and M. Lebl, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 8194–8199.

15 (*a*) J. Kress, R. Zanaletti, A. Amour, M. Ladlow, J. G. Frey and M. Bradley, *Chem.–Eur. J.*, 2002, **8**, 3769–3772; (*b*) I. Basso, B. Paolo, C. Ebert, L. Gardossi and P. Linda, *J. Chem. Technol. Biotechnol.*, 2006, **81**, 1626–1640.

16 (*a*) A. Furka, F. Sebestyén, M. Asgedom and G. Dibó, *Int. J. Pept. Protein Res.*, 1991, **37**, 487–493; (*b*) K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski and R. J. Knapp, *Nature*, 1991, **354**, 82–84; (*c*) R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo, *Nature*, 1991, **354**, 84–86.

17 (*a*) A. W. Czarnik, *Curr. Opin. Chem. Biol.*, 1997, **1**, 60–66; (*b*) K. S. Lam, M. Lebl and V. Krchnak, *Chem. Rev.*, 1997, **97**, 411–448; (*c*) F. Guillier, D. Orain and M. Bradley, *Chem. Rev.*, 2000, **100**, 2091–2158; (*d*) R. L. Affleck, *Curr. Opin. Chem. Biol.*, 2001, **5**, 257–263.

18 J.-L. Reymond and P. Babiak, *Adv. Biochem. Eng. Biotechnol.*, 2007, **105**, 31–58.

19 Y. Yongzheng and J.-L. Reymond, *Mol. BioSyst.*, 2005, **1**, 57–63.

20 (*a*) A. Clouet, T. Darbre and J.-L. Reymond, *Angew. Chem., Int. Ed.*, 2004, **43**, 4612–4615; (*b*) J. Kofoed, T. Darbre and J.-L. Reymond, *Org. Biomol. Chem.*, 2006, 3268–3281; (*c*) T. Darbre and J.-L. Reymond, *Acc. Chem. Res.*, 2006, **39**, 925–934.

21 The peptides in this library correspond to 54 730 different amino acid analysis results, 43 960 of which match with only a single peptide sequence (67.2% of the beads), 10 752 with two sequences (32.8% of the beads), and 72 with quadruple sequences (< 0.5% of the beads). J. Kofoed and J.-L. Reymond, *J. Comb. Chem.*, 2007, DOI: 10.1021/cc7001155.

22 I. Selo, L. Negroni, C. Creminon, J. Grassi and J. M. Wal, *J. Immunol. Methods*, 1996, **199**, 127–138.

23 E. G. Brown and J. M. Nuss, *Tetrahedron Lett.*, 1997, **38**, 8457–8460.

24 *Handbook of Proteolytic Enzymes*, ed. A. J. Barrett, N. D. Rawlings and J. F. Woessinger, Elsevier Academic Press, London, 2003.