ELSEVIER

Methods for mapping protease specificity

Scott L Diamond

The study of protease specificity provides information on active-site structure and function, protein–protein interaction, regulation of intracellular and extracellular pathways, and evolution of protease and substrate genes. Peptide libraries that include fluorogenic and binding tags are often generated by solid-phase synthesis. Even larger explorations of cleavage site preferences employ positional scanning libraries or phage display. Microarrays enable presentation of individual peptides to proteases, DNA sequences for capture of peptide nucleic acid encoded peptides, or nanodroplets containing soluble peptide sequences. These new methods continue to inform the design of chemical inhibitors and the identification of substrates of proteases.

Addresses
Penn Center for Molecular Discovery, Institute for Medicine and Engineering, Department of Chemical and Biomolecular Engineering, 1024 Vagelos Research Laboratories, University of Pennsylvania, Philadelphia PA 19104, USA

Corresponding author: Diamond, Scott L (sld@seas.upenn.edu)

Current Opinion in Chemical Biology 2007, **11**:46–51

This review comes from a themed issue on
Proteomics and genomics
Edited by Matthew Bogyo and Benjamin F Cravatt

Available online 6th December 2006

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2006.11.021

Introduction

Protease substrate profiling begins by assignment of non-prime and prime amino acid preferences at the site of substrate cleavage, following the standard nomenclature [1] shown in Figure 1a. The subsites (S_1, S_2, and so on) correspond to locations of the active site that functionally interact with the residues (P_1, P_2, and so on) of the substrate. Peptide substrate libraries that map residue preferences of a given protease enable prediction of candidate proteins that might serve as substrates. For example, the genome of *Plasmodium falciparum* (the protozoan parasite that causes malaria in humans) contains over 80 poorly characterized proteases, the protein substrates of which are unknown and might be of protozoan, mosquito, avian or mammalian origin. Moreover, protease substrate mapping has proven crucial to drug design and can be applied with or without ligand–protease structures. This review covers the chemistry approaches, experimental platforms and diversity display strategies for mapping protease specificity.

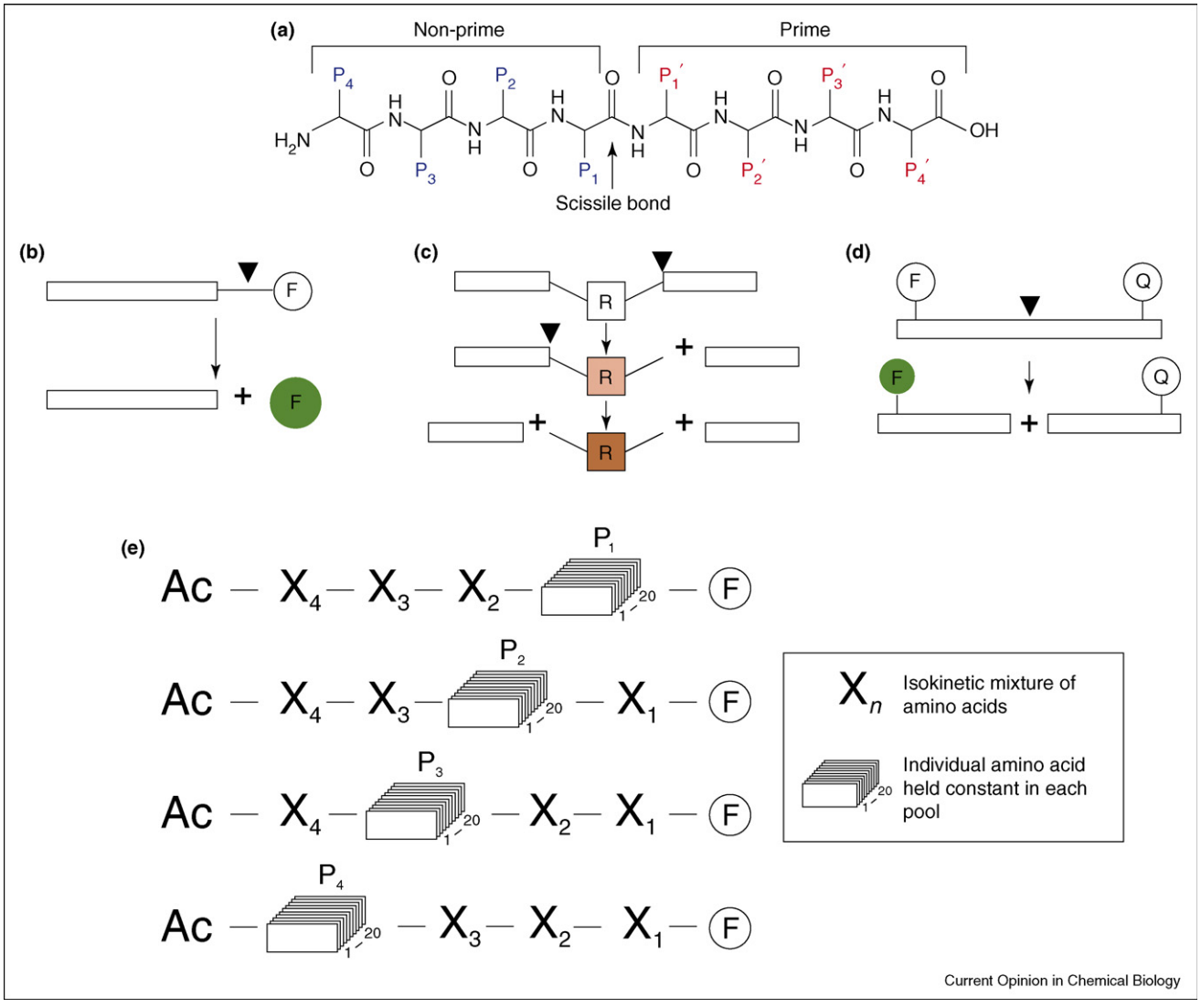
Peptide substrates and libraries

Non-prime residues (P_1, P_2, and so on) are studied using fluorogenic leaving groups where the prime residues (P_1', P_2', and so on) are all replaced by the leaving group (Figure 1b). The most common fluorogenic leaving group 7-amino-4-methylcoumarin (AMC) is easily detected in plate readers at nanomolar levels in $\leq$10 μL reactions. One limitation of coumarin or 6-amino-1-napthalenesulfonamide (ANSN) leaving groups [2] is their low excitation wavelength (λ_{EX} = 350–390 nm), which prevents their use with typical lasers found in microarray scanners. Mono- and di-substituted rhodamine 110 (R110) substrates [3] are easily excited by lasers; however, $[X_n]_2$–R110 substrates produce a highly nonlinear signal in time as the first and then the second peptide is cleaved (Figure 1c). To study protease recognition of non-prime and prime residues, quenched substrate libraries are required (Figure 1d).

Fluorogenic libraries

Positional scanning libraries were developed in the early 1990s where a single position of the peptide was held constant in each sublibrary, while other positions were diversified with all possible combinations of natural amino acids, typically excluding cysteine [4,5] (Figure 1e). An early study with a positional scanning library Ac–X–X–X–Asp–AMC (separate P_4, P_3, P_2 sublibraries of 8000 compounds each with 20 wells of 20^2 compounds) led to the development of a potent aldehyde inhibitor, Ac–WEHD–CHO (K_i = 56 pM) of caspase 1 (IL-1β-converting enzyme [ICE]) [6]. Harris *et al.* [7] reported the solid-phase synthesis of fluorogenic 7-amino-4-carbamoylmethylcoumarin (ACC) peptides. Because ACC has a quantum yield threefold greater than AMC, mixtures of greater numbers of peptides could be pooled. Using isokinetic mixtures for synthesis, the P_1-diverse library comprised 20 pools, with P_1 fixed in each pool and each pool containing 6859 compounds (19^3 to randomize P_4–P_3–P_2 with all natural amino acids excluding cysteine and norleucine substituted for methionine). Measuring signal from only 20 wells per protease provided P_1 preferences: basic amino acids (arginine and lysine) for trypsin, thrombin and plasmin; large hydrophobic amino acids (tyrosine, phenylalanine and tryptophan) for chymotrypsin; aspartic acid (but not glutamic acid) for granzyme B; alanine and valine for human neutrophil elastase; and broad specificity for papain and cruzain. Also, the full P_2, P_3 and P_4 preferences were mapped by positional scanning for plasmin, thrombin, urokinase (uPA), tissue plasminogen activator, factor Xa, papain and cruzain.

Figure 1



Peptide-based chemistries for mapping protease specificity. **(a)** Peptide numbering nomenclature for individual residues of the substrate for proteolytic cleavage between P_1 and P_1'. **(b)** Release of a fluorogenic leaving group (F) such as AMC owing to hydrolysis of the amide bond distal of P_1. **(c)** Use of rhodamine 110 (R) fluorogenic leaving group for evaluation of non-prime residues. **(d)** Quenched substrates (Q) for evaluation of non-prime and prime residues. **(e)** Design of positional scanning synthetic combinatorial libraries (PS-SCL) where P_i libraries consist of 20 sublibraries with the position P_i held constant in each sublibrary whereas all other positions $X_{j\neq i}$ are randomized with mixtures of amino acids. Arrow heads denote scissile bonds. Abbreviation: Ac, acetyl.

Positional scanning libraries were subsequently used to profile β-tryptases I and II [8] and seven different human tissue kallikreins (hK3-7, -10 and -11) [9•]. Using a scoring algorithm based on positional scanning probabilities, more than 70 proteins were predicted to contain potential hK4 (prostase, KLK4) cleavage sites [10].

A full P_4–P_3–P_2–P_1 positional scanning fluorogenic library (160 000 sequences) was recently synthesized by Choe *et al.* [11] to scan the non-prime site preferences of cathepsins L, V, K, S, F and B, papain and bromelain. Positional scanning revealed subtle functional differences between cathepsin S and K, especially the acceptance of proline in the P_2 position by cathepsin K, enabling design of a selective cathepsin K inhibitor.

Quenched peptide libraries

Quenched libraries to study prime preferences tend to be smaller and more focused to a particular enzyme family. Stennicke *et al.* [12] developed quenched substrates of Abz–GDEVD–SVY(NO_2)D for caspase analysis, where the P_4 or P_1 or P_1' positions (shown underlined) were each individually varied (Abz, *O*-aminobenzoyl; Y(NO_2), nitrotyrosine). As expected, a strong preference for P_1 = Asp

was confirmed. These studies also confirmed the preference of small residues of glycine, serine and alanine in P_1' by caspases 1, 3, 6, 7 and 8, as expected from cleavage sites found in natural substrates.

Using a consensus Factor Xa cleavage site of Abz–VQFR–SLGDQ(EDDnp) for P_1 = arginine and P_1' = serine, Bianchini *et al.* [13] individually varied the P_3, P_2, P_1', P_2' or P_3' position (shown underlined) in five sublibraries of 19 compounds each. They found that the Factor Xa active site had minor peptide substrate preferences that were unaffected by its cofactor, Factor Va — thus indicating that exosite interactions with prothrombin have a major role in specificity in coagulation. Similarly, following evaluation of the non-prime P_4–P_1 specificity, Petrassi [14] developed positional scanning libraries of methoxy-coumarin–Leu–Thr–Pro–Arg–X–X–X–X–Lys(DNP)–Arg library (DNP, 2,4-dinitrophenol) to explore thrombin function and methoxy-coumarin–Asp–Glu–Val–Asp–X–X–X–X–Lys(DNP)–Arg library to explore caspase 3 function.

Exosite studies

Distinct from the cleavage site, other sequences within the protein substrate (exosites) can dictate recognition. As discussed above for Factor Xa, exosites provide proteases of the coagulation pathway with their exquisite protein substrate specificity [15••]. In this multistep pathway of recognition, exosite binding of the substrate to the protease helps deliver the cleavage site to an active site that might actually be fairly tolerant of substantial variations in a cleavage sequence. Fewer generalized methods for making large exosite libraries are available.

Microarrays

Microarrays provide advantages over well plates when library size multiplied by sample number exceeds 2000–5000, depending on the availability of the library, protease, and robotic liquid handling. Whereas positional scanning with mixtures provides information on average preferences, microarrays enable direct assessment of individual sequences to rank substrates and to discover subsite cooperativity.

Solid-phase presentation of peptides

Recently, Salisbury *et al.* [16] generated fluorogenic substrate arrays. The 361-member substrates of the form Ac–Ala–P_3–P_2–Lys–ACC–linker employed an alkoxyamine linker for covalent oxime-forming reaction to aldehyde-derivatized surfaces (Figure 2a). Direct attachment of the peptides to aldehyde-functionalized glass slides were poorly cleaved by trypsin, probably owing to steric hindrance. Microarraying of the library to BSA–aldehyde slides created ~100 μm features that could be protease activated (15 μL of 250 nM thrombin for 60 min) and fluorescently read (λ_{EX} 390 nm/λ_{EM}460 nm) using a charge-coupled device (CCD)-based slide imager. Fluorogenic peptide arrays confirmed the strong preference of thrombin for proline in the P_2 position and broad specificity in the P_3 position.

Solution-phase presentation of peptides

Gosalia and Diamond [17] developed a nanodroplet microarray where substrates are arrayed in glycerol/DMSO whereby the DMSO evaporates out of each nanoliter droplet within minutes (Figure 2b). The arrays, when activated with aerosol deposition of enzyme and incubated under humidity (final reaction condition of 10% glycerol in water dictated by liquid vapor equilibrium), produce fluorogenic signals from homogenous reactions. Nanodroplet microarrays confirmed the strict proline requirement in the P_2 position for human, bovine and salmon thrombin [18], despite 400 million years of divergent evolution. This approach also revealed subsite cooperation where use of lysine in the P_1 position (instead of arginine) changed the glycine preference to phenylalanine in the P_2 position. Such subsite cooperativity would be difficult to discern with positional scanning libraries.

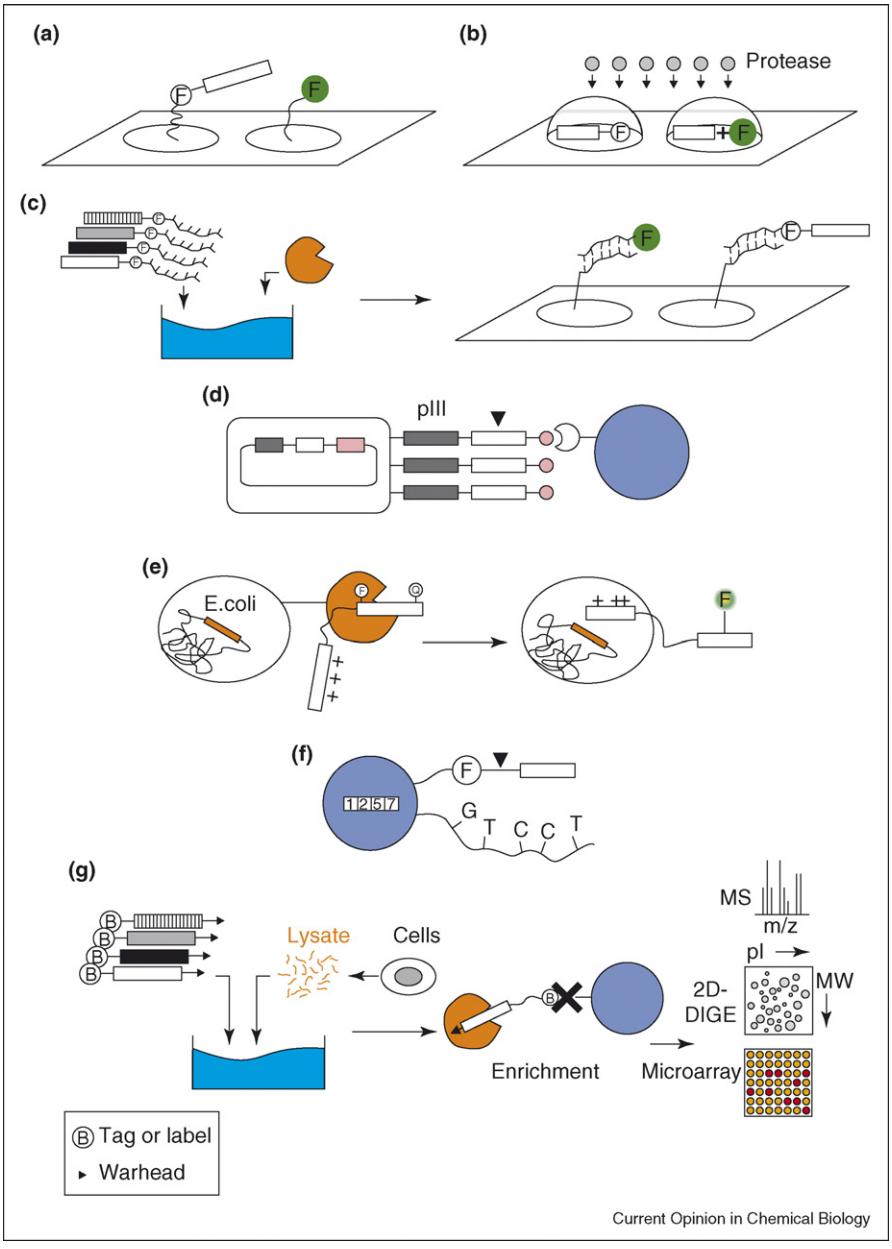
The method was also used with Ac–Ala–P_3–P_2–(Arg/Lys)–ACC libraries for 2×19^2 individual, spatially separated fluorogenic reactions with 13 different serine proteases and 11 cysteine proteases [19•] including rhodesain from *Trypanasoma brucei* rhodesiense, which displayed a strong P_2 preference for leucine, valine, phenyalanine and tyrosine. With the individual coagulation serine proteases fully mapped, the method was further extended by mathematical analysis of the signal pattern generated with protease mixtures found in citrated and recalcified plasma and kaolin or uPA-treated plasma [20].

Other routes to obtain protease microarrays include microarrayed compound screening (microARCS), which is used primarily for compound high-throughput screening; however, substrate profiling is also feasible with the overlay of a protease-laden agarose gel on an array of fluorogenic substrates [21]. Alternatively, multicomponent nanoliter-scale reactions for substrate mapping can be assembled on microarrays by re-spotting locations with pin-contact microarrayers [22].

Detection microarrays

DNA microarrays offer a powerful tool for sequence-specific capture of molecules through base pairing to achieve spatial deconvolution of a library mixture. For protease mapping, peptide–peptide nucleic acid (PNA) libraries can be subjected to proteases in tube reactions followed by spatial deconvolution on the microarray (Figure 2c). Winssinger *et al.* [23••] generated a library of 192 quenched PNA-encoded peptide substrates and demonstrated the sensing nature of the chip surface with thrombin, plasmin, caspase 3, an apoptotic lysate, and human plasma (normal or warfarin-treated).

Figure 2



Platforms for evaluation of protease specificity. **(a)** Covalent linkage of fluorogenic peptides (F) to microarrays enables direct activation of the microarray with a protease whereby fluorescent spot intensities indicate sequence preferences. **(b)** Use of aerosol deposition of proteases to nanodroplets containing fluorogenic substrates, allowing for solution phase reactions with no need for covalent attachment of substrates. **(c)** Spatial deconvolution of PNA-encoded fluorogenic peptide sequences through the use of DNA microarrays. **(d)** Phage display of peptide sequences proximal to an affinity capture tag, facilitating protease elution of phage with optimal sequences. **(e)** Enzyme engineering using biological display of protease variants that cleave a cationic quenched peptide sequence whose fluorogenic product remains electrostatically bound to the bacteria surface. **(f)** Bead display of fluorogenic peptides with beads employing radiofrequency or DNA or PNA encoding to record split-pool synthesis history. **(g)** Use of warheads to covalently label protease active sites within complex protein mixtures for subsequent proteomic analysis. Various tags (e.g. biotin or PNA) or labels (isotopic or fluorescent) facilitate enrichment and detection. Abbreviations: MW, molecular weight.

Chemical–protease interaction microarrays

Microarray presentation of covalently attached compounds or pharmacophore fragments enables evaluation of chemical–protease interactions. Dickopf *et al.* [24] used surface plasmon resonance detection of arrays presenting over 1500 compound fragments (~200 Da: amines and anilines, carboxylic acids, aldehydes and ketones) for binding of Factor VIIa. A total of 60 ligands produced detectable signals, ultimately leading to an X-ray structure of dichloroaminophenol in a hydrogen bonding

network with polar residues ^{189}Asp and ^{190}Ser of the S_1 pocket of factor VIIa.

Biological display

Phage display generates a larger number ($>10^7$) and longer length of peptide sequences via expression of a given peptide sequence with an affinity tag on the minor coat protein pIII of bacteriophage M13. Protease-mediated release of phage, multiple rounds of panning, and sequencing of selected phage allowed Matthews and Wells [25] to find optimal proteases substrates (Figure 2d). Phage display has been applied to numerous proteases, including caspases, factor Xa, HIV protease, furin, metal metalloproteases, subtilisin and rat α-chymase.

A histidine-tagged phage library of X_4–Arg–X_4 was used to profile substrate specificity of human C1s protease [26]. Following repeated rounds of C1s-mediated elution of phage from nickel-chelated sepharose beads, multiple phage plaques were randomly selected and sequenced, 40% of which yielded the sequence YLGR–SYKV. This sequence was cleaved fivefold faster than sequences taken from the natural substrates C2 and C4. Frequency mapping of phage revealed that leucine or valine were preferred at P_2, with little evidence for specificity at prime sites other than leucine at P_2'.

Phage display also enables disulfide-constrained sequences. Hansen *et al.* [27•] incubated uPA with 10^{11} plaque-forming units of each library of X_7, Cys–X_n–Cys for n = 7 or 10, and Cys–X_3–Cys–X_3–Cys–X_3–Cys with uPA followed by capture on anti-uPA-coated tubes. A theoretical diversity of 10^9 produced 19 of 28 individual phages with sequences of CSWRGLENHRMC (K_d ~0.5 μM when expressed with a fusion protein). Subsequent alanine scanning of this sequence and the uPA active site defined peptide–protease interactions on both sides of the binding interface. The cyclic structure and P_2 = tryptophan prevented hydrolysis of the sequence. A further modification of phage display includes the use of the BirA sequence for production of biotinylated phage peptide substrates [28].

Display methods have also been used to evolve the enzyme active site to recognize a particular substrate (Figure 2e). A cationic-quenched OmpT substrate was used to label *Escherichia coli* and screen via flow cytometry a total of 6×10^5 random OmpT variants [29]. In an example of active-site engineering to alter substrate specificity, OmpT was mutated to a variant that accepted Ala–Arg cleavage instead of Arg–Arg with a three-million-fold selectivity [30•].

Other methods

Beads

Rosse *et al.* [31] described the use of bead-linked Lys(dabsyl)–Thr–Ser–Arg–(Pro/Ala)–X_4–Glu(lucifer yellow)–Gly–bead, with arginine in the P_1 position. Individual beads became fluorescent when activated with napcin A. Fluorescent beads were then isolated and the peptides sequenced by Edman degradation. This approach results in the diversity scale of positional scanning combined with the detection of actual individual sequences, such as that provided by microarrays, individual well reactions, or phage display. Beads can be fluorescently encoded, DNA encoded, or radiofrequency encoded (Figure 2f), but encoded bead-based fluorogenic peptide libraries have yet to be reported.

Reactive probes

Peptide libraries presenting reactive warheads are useful for evaluation of proteases in complex mixtures (Figure 2g). Greenbaum *et al.* [32] created peptidyl epoxide P_2, P_3 and P_4 positional scanning libraries for covalent labeling of proteases. On the basis of an isotope-competition assay, a cluster analysis of 12 papain family proteases revealed distinct subsite specificities. Also, a P_2 diversity library included 19 natural amino acids and 41 nonnatural hydrophobic amino acids to facilitate inhibitor design. PNA-encoded peptide libraries presenting an acrylate reactive group inhibited cysteine proteases in complex mixtures followed by spatial deconvolution on microarrays [33].

Conclusions

The number of peptide libraries, formats and methodologies for evaluation of protease substrate interactions has grown considerably. Data from these types of studies can be found in the MEROPS database (http://www.merops.sanger.ac.uk), a searchable repository of information on proteases, substrates, inhibitors and structures [34••]. Future trends will continue with the development of new synthetic routes, fluorogenic chemistries, exosite scanning and whole protease proteome methods to take large numbers of proteases against large substrate libraries.

Acknowledgements

This work was supported by the National Institutes of Health (grants NIH R33-HL87317, R01-HL56621 and U54-HG3915).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Schechter I, Berger A: **On the size of the active site in proteases. I. Papain**. *Biochem Biophys Res Commun* 1967, **27**:157-162.

2. Butenas S, Orfeo T, Lawson JH, Mann KG: **Aminonaphthalenesulfonamides, a new class of modifiable fluorescent detecting groups and their use in substrates for serine protease enzymes**. *Biochemistry* 1992, **31**:5399-5411.

3. Cai SX, Zhang HZ, Guastella J, Drewe J, Yang W, Weber E: **Design and synthesis of rhodamine 110 derivative and caspase-3 substrate for enzyme and cell-based fluorescent assay**. *Bioorg Med Chem Lett* 2001, **11**:39-42.

4. Pinilla C, Appel JR, Blanc P, Houghten RA: **Rapid identification of high affinity peptide ligands using positional scanning**

synthetic peptide combinatorial libraries. *Biotechniques* 1992, **13**:901-905.

5. Dooley CT, Houghten RA: **The use of positional scanning synthetic peptide combinatorial libraries for the rapid determination of opioid receptor ligands**. *Life Sci* 1993, **52**:1509-1517.

6. Rano TA, Timkey T, Peterson EP, Rotonda J, Nicholson DW, Becker JW, Chapman KT, Thornberry NA: **A combinatorial approach for determining protease specificities: application to interleukin-1beta converting enzyme (ICE)**. *Chem Biol* 1997, **4**:149-155.

7. Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, Craik CS: **Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries**. *Proc Natl Acad Sci USA* 2000, **97**:7754-7759.

8. Harris JL, Niles A, Burdick K, Maffitt M, Backes BJ, Ellman JA, Kuntz I, Haak-Frendscho M, Craik CS: **Definition of the extended substrate specificity determinants for beta-tryptases I and II**. *J Biol Chem* 2001, **276**:34941-34947.

9. • Debela M, Magdolen V, Schechter N, Valachova M, Lottspeich F, Craik CS, Choe Y, Bode W, Goettig P: **Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences**. *J Biol Chem* 2006, **281**:25678-25688.
This paper describes complete P_1 to P_4 positional profiling of seven members of a 15-member family of human tissue kallikrein. Tissue kallikreins have been implicated in malignancies and inflammation.

10. Matsumura M, Bhatt AS, Andress D, Clegg N, Takayama TK, Craik CS, Nelson PS: **Substrates of the prostate-specific serine protease prostase/KLK4 defined by positional-scanning peptide libraries**. *Prostate* 2005, **62**:1-13.

11. Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Bromme D, Ellman JA, Craik CS: **Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities**. *J Biol Chem* 2006, **281**:12824-12832.

12. Stennicke HR, Renatus M, Meldal M, Salvesen GS: **Internally quenched fluorescent peptide substrates disclose the subsite preferences of human caspases 1, 3, 6, 7 and 8**. *Biochem J* 2000, **350**:563-568.

13. Bianchini EP, Louvain VB, Marque PE, Juliano MA, Juliano L, Le Bonniec BF: **Mapping of the catalytic groove preferences of factor Xa reveals an inadequate selectivity for its macromolecule substrates**. *J Biol Chem* 2002, **277**:20527-20534.

14. Petrassi HM, Williams JA, Li J, Tumanut C, Ek J, Nakai T, Masick B, Backes BJ, Harris JL: **A strategy to profile prime and non-prime proteolytic substrate specificity**. *Bioorg Med Chem Lett* 2005, **15**:3162-3166.

15. •• Krishnaswamy S: **Exosite-driven substrate specificity and function in coagulation**. *J Thromb Haemost* 2005, **3**:54-67.
In-depth analysis of exosites found to enforce enzyme specificity among serine proteases of the human coagulation cascade with particular focus on prothrombinase (Factor Xa/Va) recognition and cleavage of prothrombin.

16. Salisbury CM, Maly DJ, Ellman JA: **Peptide microarrays for the determination of protease substrate specificity**. *J Am Chem Soc* 2002, **124**:14868-14870.

17. Gosalia DN, Diamond SL: **Printing chemical libraries on microarrays for fluid phase nanoliter reactions**. *Proc Natl Acad Sci USA* 2003, **100**:8721-8726.

18. Gosalia DN, Salisbury CM, Maly DJ, Ellman JA, Diamond SL: **Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays**. *Proteomics* 2005, **5**:1292-1298.

19. • Gosalia DN, Salisbury CM, Ellman JA, Diamond SL: **High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays**. *Mol Cell Proteomics* 2005, **4**:626-636.
The authors present a 722-member Ala–X–X–(Arg/Lys)–ACC library on microarrays for mapping of specificity of 13 serine proteases and 11 cysteine proteases. This is the first report of rhodesain substrate specificity.

20. Gosalia DN, Denney WS, Salisbury CM, Ellman JA, Diamond SL: **Functional phenotyping of human plasma using a 361-fluorogenic substrate biosensing microarray**. *Biotechnol Bioeng* 2006, **94**:1099-1110.

21. Gopalakrishnan SM, Karvinen J, Kofron JL, Burns DJ, Warrior U: **Application of micro arrayed compound screening (microARCS) to identify inhibitors of caspase-3**. *J Biomol Screen* 2002, **7**:317-323.

22. Angenendt P, Lehrach H, Kreutzberger J, Glokler J: **Subnanoliter enzymatic assays on microarrays**. *Proteomics* 2005, **5**:420-425.

23. •• Winssinger N, Damoiseaux R, Tully DC, Geierstanger BH, Burdick K, Harris JL: **PNA-encoded protease substrate microarrays**. *Chem Biol* 2004, **11**:1351-1360.
First report of encoded fluorogenic peptides using peptide nucleic acid for spatial deconvolution on microarrays. This approach is useful for substrate mapping and biomarker detection.

24. Dickopf S, Frank M, Junker HD, Maier S, Metz G, Ottleben H, Rau H, Schellhaas N, Schmidt K, Sekul R *et al.*: **Custom chemical microarray production and affinity fingerprinting for the S1 pocket of factor VIIa**. *Anal Biochem* 2004, **335**:50-57.

25. Matthews DJ, Wells JA: **Substrate phage: selection of protease substrates by monovalent phage display**. *Science* 1993, **260**:1113-1117.

26. Kerr FK, O'Brien G, Quinsey NS, Whisstock JC, Boyd S, de la Banda MG, Kaiserman D, Matthews AY, Bird PI, Pike RN: **Elucidation of the substrate specificity of the C1s protease of the classical complement pathway**. *J Biol Chem* 2005, **280**:39510-39514.

27. • Hansen M, Wind T, Blouse GE, Christensen A, Petersen HH, Kjelgaard S, Mathiasen L, Holtet TL, Andreasen PA: **A urokinase-type plasminogen activator-inhibiting cyclic peptide with an unusual P2 residue and an extended protease binding surface demonstrates new modalities for enzyme inhibition**. *J Biol Chem* 2005, **280**:38424-38437.
The authors used phage display to present disulfide-constrained peptide sequences to proteases in a potentially more protein-like structure. Generation of a noncleavable peptide inhibitor confirms the role of local structural context that might be difficult to achieve with synthetic libraries.

28. Scholle MD, Kriplani U, Pabon A, Sishtla K, Glucksman MJ, Kay BK: **Mapping protease substrates by using a biotinylated phage substrate library**. *Chem BioChem* 2006, **7**:834-838.

29. Olsen MJ, Stephens D, Griffiths D, Daugherty P, Georgiou G, Iverson BL: **Function-based isolation of novel enzymes from a large library**. *Nat Biotechnol* 2000, **18**:1071-1074.

30. • Varadarajan N, Gam J, Olsen MJ, Georgiou G, Iverson BL: **Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity**. *Proc Natl Acad Sci USA* 2005, **102**:6855-6860.
A novel selection method is described to reprogram protease activity towards new substrates. The method exploits cationic capture of quenched substrate to bacterial surface to achieve physical linkage of genotype and proteolytic phenotype.

31. Rosse G, Kueng E, Page MG, Schauer-Vukasinovic V, Giller T, Lahm HW, Hunziker P, Schlatter D: **Rapid identification of substrates for novel proteases using a combinatorial peptide library**. *J Comb Chem* 2000, **2**:461-466.

32. Greenbaum DC, Arnold WD, Lu F, Hayrapetian L, Baruch A, Krumrine J, Toba S, Chehade K, Bromme D, Kuntz ID *et al.*: **Small molecule affinity fingerprinting. A tool for enzyme family subclassification, target identification, and inhibitor design**. *Chem Biol* 2002, **9**:1085-1094.

33. Winssinger N, Ficarro S, Schultz PG, Harris JL: **Profiling protein function with small molecule microarrays**. *Proc Natl Acad Sci USA* 2002, **99**:11139-11144.

34. •• Rawlings ND, Morton FR, Barrett AJ: **MEROPS: the peptidase database**. *Nucleic Acids Res* 2006, **34**:D270-D272.
Overview of MEROPS, a comprehensive web-based and searchable database that uses a hierarchical, structure-based classification of peptidases. Information on protease gene and protein structure, substrates and inhibitors can be found for 180 families of proteases covering >30 000 sequences.