Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 3162–3166
Bioorganic & Medicinal Chemistry Letters

A strategy to profile prime and non-prime proteolytic substrate specificity

H. Michael Petrassi, Jennifer A. Williams, Jun Li, Christine Tumanut, Jared Ek, Takashi Nakai, Brian Masick, Bradley J. Backes*,† and Jennifer L. Harris*

Genomics Institute of the Novartis Research Foundation, Drug Discovery, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA

Received 4 January 2005; revised 3 April 2005; accepted 5 April 2005
Available online 4 May 2005

Abstract—A strategy was developed to determine the prime and non-prime substrate specificity of serine, threonine and cysteine proteases. ACC positional scanning technology was employed to determine the P4–P1 non-prime site substrate specificity. The data was used to synthesize biased donor–quencher positional scanning libraries to profile the P1′–P4′ prime site substrate specificity. Directed sorting using the Irori Nanokan system allowed for the archiving of multiple P1′–P4′ positional scanning libraries. From these libraries focused donor–quencher libraries incorporating P4–P1 data for each protease under study could be rapidly prepared. The profiling of thrombin and caspase-3 P4–P4′ substrate specificity, comparison of the library specificity data to single substrates, and the analysis of physiological cleavage sites are described.

Proteases are an important class of proteins constituting a relatively large percentage (about 2%) of the gene products in the human genome.[1,2] Proteases play critical roles in the initiation and regulation of biological processes including fertilization, development, homeostasis, immunity, and cell death. From a therapeutic perspective, modulation of protease activity offers considerable promise for the treatment of human diseases. Indeed, it has been estimated that in addition to proteases targeted from infectious diseases, over 14% of human proteases are under investigation as drug targets.[3]

A characteristic function of proteases is the ability to discriminate among many potential substrates, termed the substrate specificity of a protease. Substrate specificity is a critical factor that maintains the fidelity of the biological processes in which a protease acts. For researchers, substrate specificity can serve as a handle by which proteases can be discriminated from others in its class, even in cases in which a large degree of structural homology exists.

Several methods have been developed to obtain substrate specificity data for proteases.[4] Traditionally, defining the substrate specificity of a protease consisted of iteratively testing the activity of a protease against many individual protein and peptide substrates.[5,6] More recently, combinatorial synthesis and genetic methods have allowed for a more systematic identification of optimal substrate sequences.[7,8] Many of these are selection methods that identify the optimal substrates but often miss the identification of negative substrate determinants. In contrast, the screening of positional scanning synthetic combinatorial libraries (PSSCLs) of 7-amino peptide coumarin substrates represent a rapid method to screen for positive and negative substrate specificity determinants of proteases.[9–11] One major disadvantage of coumarin substrates is that the latent fluorescent leaving group occupies the prime site preventing examination of prime site substrate interactions. This limits the utility of these libraries to examining only the non-prime specificity of proteases. Donor–quencher libraries that span the cleavage site on both the prime and the non-prime sides of the peptide have been effective in identifying substrate specificity determinants of proteases, however several limitations exist. The number of positions that can be evaluated at one time is

Keywords: Protease substrate specificity; Positional scanning library; Donor–quencher substrate; Directed sorting.
* Corresponding authors. Tel.: +1 858 812 1500; fax: +1 858 812 1648 (J.L.H.); e-mail: jharris@gnf.org
† Present address: Metabolic Disease Research, Global Pharmaceutical Research and Development, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064-6098, USA.

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.04.019

restricted since each additional site investigated causes an exponential increase in the substrate number. There is a finite limit to substrate pool sizes due to solubility concerns and decreased assay sensitivity given the reduction in concentration of individual substrates. In addition, data can be difficult to interpret since cleavage at any position along the peptide results in a fluorescence signal. Accordingly, prior knowledge of a protease's substrate specificity can be used to bias the library construction to limit the number of sites investigated and to fix the catalytic register.

A strategy for evaluating prime and non-prime substrate specificity data of serine, cysteine and threonine proteases was devised (Fig. 1). The two-stage method provides screening information for both optimal and suboptimal substrates, and does not require prior specificity information. Proteases were profiled in two discrete stages. In the first step, a 7-amino coumarin-based PSSCL[9–11] was employed to query the P4–P1 specificity using previously published methods (Fig. 2).[12–14] Since cleavage of the coumarin anilide bond of the substrates is the only manner in which a fluorogenic signal can be generated, the data was straightforward to interpret. The optimal residues determined for the P4–P1 positions and the donor were attached to the N-terminus of pre-made positional scanning libraries bearing a quencher that were designed to query the P1′–P4′ positions. Since the method relies on archived PSSCL, directed sorting automation and microtiter plate-based fluorescence scanning instrumentation, many proteases can profiled in a short period of time.

The synthesis of the PSSCL was performed using standard Fmoc peptide synthesis, split and pool directed sorting, and Irori Nanokan technology.[15–17] These technologies allowed for the synthesis of the PSSCL employing only 20 of reaction flasks in four rounds of synthesis and sorting. The synthesis was initiated with the attachment of Fmoc–Arg–(Pbf)–OH, to increase the water solubility of the final donor–quencher constructs, to Rink Resin (Fig. 1). Commercially available Fmoc–Lys–(DNP)–OH was attached as the quencher, and the resin was loaded into Nanokans. To synthesize the P4′ fixed positions 19 Nanokan collections were placed in individual flasks and a single amino acid was attached to fix the position (O). In order to obtain a statistical mixture of amino acids (X), the 20th collection, comprising all other Nanokans, were treated with an isokinetic mixture of 19 amino acids (for fixed positions and isokinetic mixtures, Cys was left out and norleucine (n, Nle) was substituted for Met).[9] Upon completion of the first PSSCL coupling all Nanokans were combined and prepared for the second sort. To complete the P1′–P4′ PSSCL three more rounds of sorting and synthesis followed. For each protease a set of 76 Nanokan PSSCL was placed in single vessel and the optimal non-prime site P4–P1 tetrapeptide, determined from the coumarin PSSCL in stage 1, for the protease was attached.

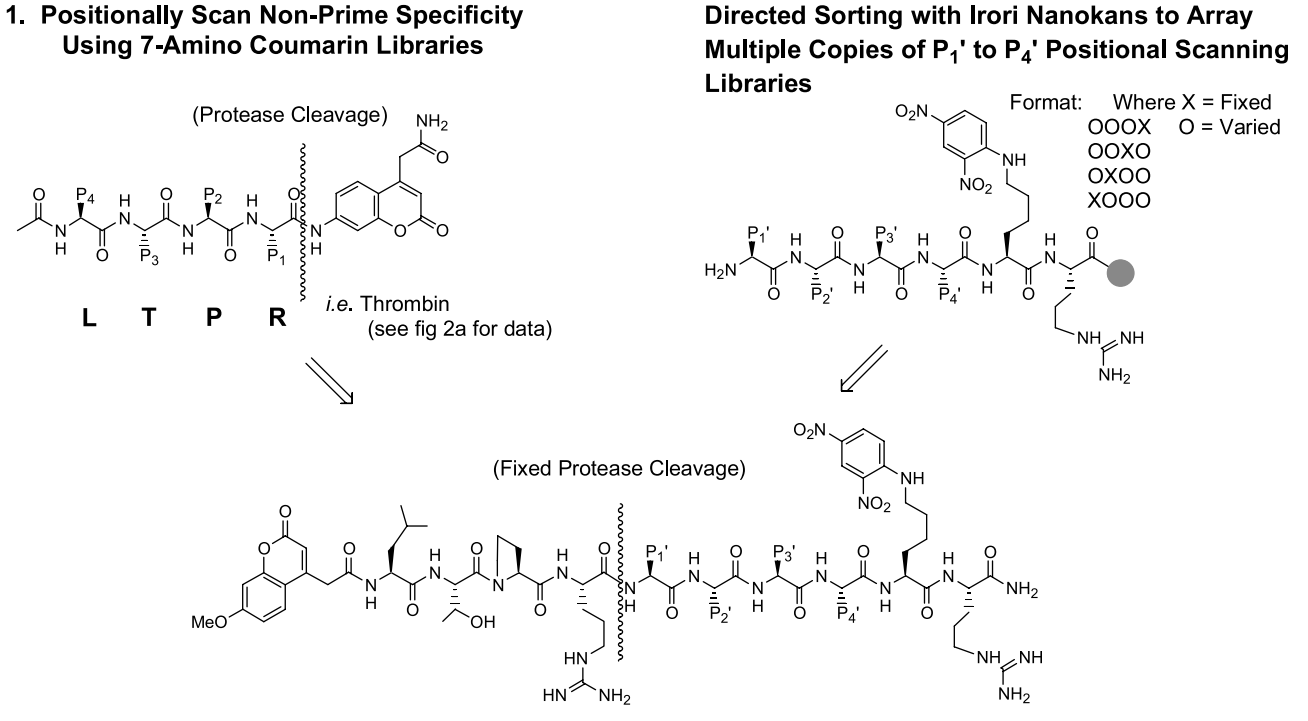

Figure 1. Profiling of prime and non-prime substrate specificity. Step 1: Functional characterization of non-prime substrate specificity using 7-amino coumarin libraries.[11–13] Optimal residues at each non-prime side position of the substrate are identified. Step 2: Biased donor–quencher library prepared from a single copy of archived positional scanning library by attaching the optimal non-prime sequence.

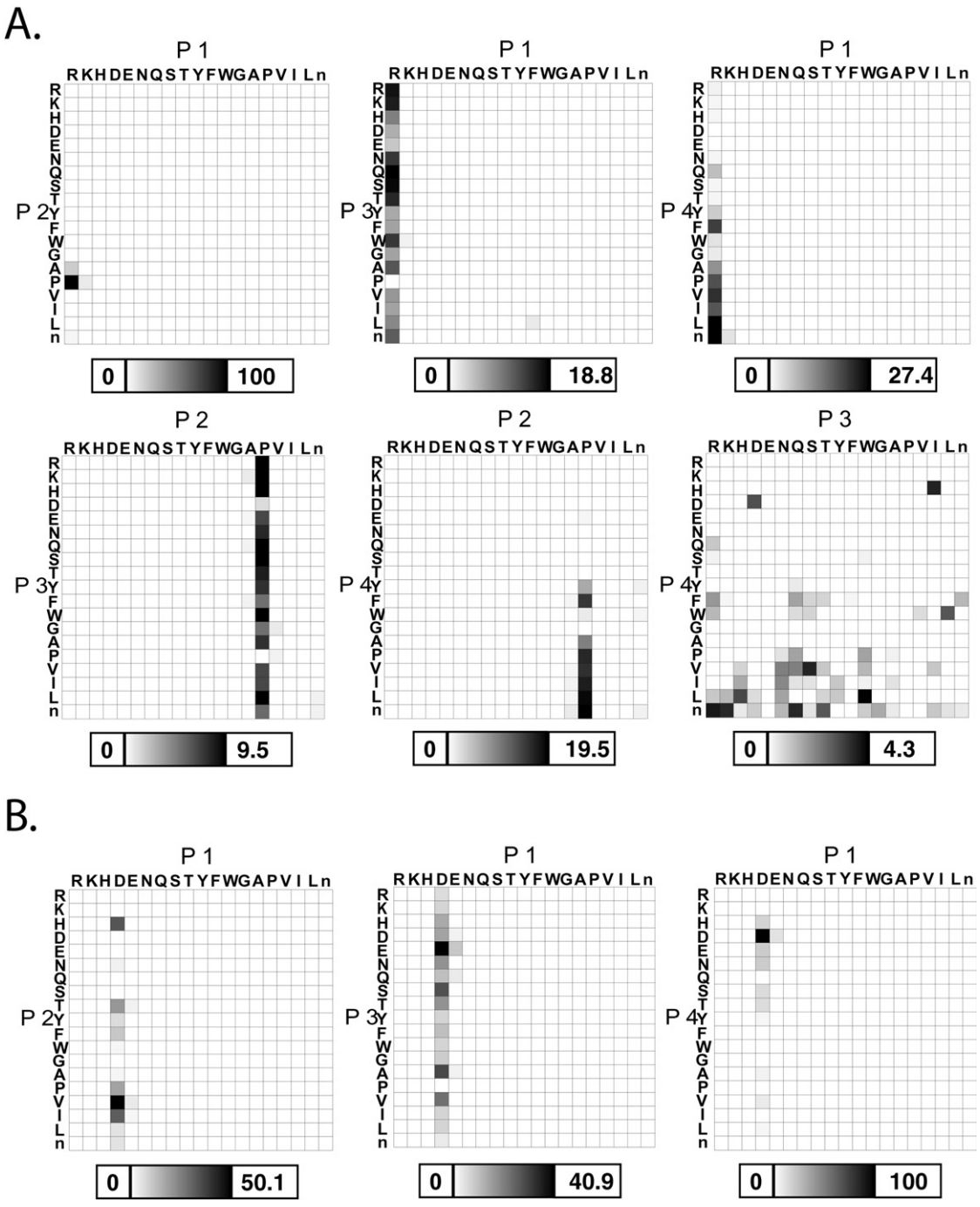

Figure 2. Results from the two-position fixed tetrapeptide coumarin libraries for (A) thrombin and (B) caspase-3. The *x*- and *y*-axes represent the fixed amino acid, and the other two non-fixed positions contain an equimolar mixture of 19 amino acids (cysteine was excluded and norleucine (n) was substituted for methionine) to give a total of 361 substrates assayed in each well (represented by one square on the 2D array). Each 2D array depicts, by relative intensity, normalized reaction velocities in RFU/s. Less than 2% maximal activity was observed in libraries P2 × P3, P2 × P4, and P3 × P4 for caspase-3, therefore the data is not shown. The shaded bars below each graph represent the normalized velocities with the strongest determinant being set to 100 (i.e., the P1—Arg and P2—Pro for thrombin, first graph of A). The shaded bars allow for a more thorough representation of the weight of each subsite specificity determinant in the context or the full specificity profile.

To complete the synthesis the methoxy coumarin acetic acid (MCA) was coupled to the N-terminus of the P4 residue to serve as the donor.[18] The Nanokans were then sorted to a single 96 well plate and cleaved employing a standard TFA protocol to provide the PSSCL that was assayed upon dissolution to a stock solution in DMSO. In our efforts, 50 copies of the 76 Nanokan PSSCL (3800 Nanokans) were prepared and stored for this and future studies.

After identifying the non-prime side substrate specificity in the 7-amino coumarin libraries for thrombin as P4—(Leu/Nle), P3—(Gln/Ser/Thr), P2–Pro, and P1—Arg and caspase-3 as P4—Asp, P3—Glu, P2—Val, and

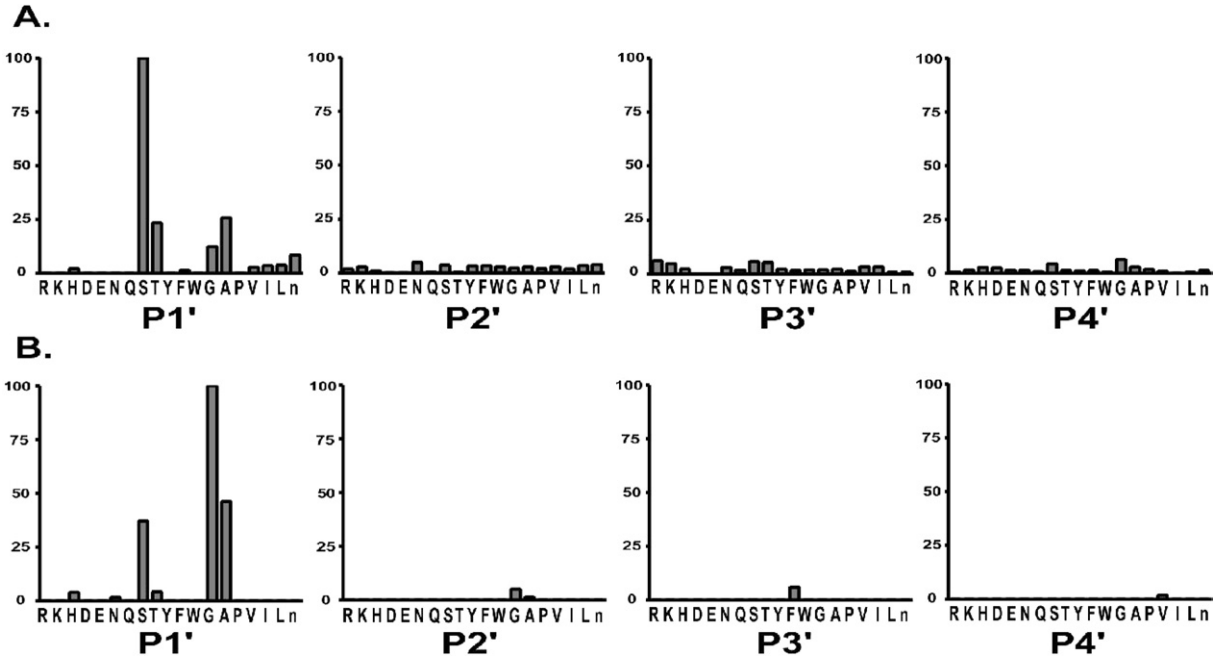

Figure 3. Results of (A) thrombin and (B) caspase-3 activity in the one-position fixed P1′–P4′ donor–quencher substrate libraries. The identity of the amino acid in the fixed position is indicated on the *x*-axis. The remaining three positions contain a mixture of 19 amino acids (cysteine was excluded and norleucine (n) was substituted for methionine) to give a total of 6859 substrates assayed in each well. The *y*-axis is the normalized rate of hydrolysis reaction in RFU/s. The octa-peptide sequence in the thrombin library (A) consisted of Leu-Thr-Pro-Arg▼Xaa-Xaa-Xaa-Xaa and in the caspase-3 library (B) consisted of Asp-Glu-Val-Asp▼Xaa-Xaa-Xaa-Xaa.

P1—Asp (Fig. 2), prime site libraries were synthesized where the P1′–P4′ positions were randomized in a PSSCL format and the P1–P4 positions were held constant (Fig. 1). Construction of a one-position fixed PSSCL in this format addressed the issue of catalytic register of the substrate by satisfying the prime side positions and biasing for cleavage after the internal P1—Arg or P1—Asp amino acid (Fig. 3). The results of thrombin in the Leu-Thr-Pro-Arg-Xaa-Xaa-Xaa-Xaa library showed the major determinant was P1′ Ser with a very little specificity seen in the other prime positions with the exception of an absence of activity for acidic amino acids in the P1′–P3′ positions. This finding is in agreement with previously published results.[19] The results of caspase-3 in the Asp-Glu-Val-Asp-Xaa-Xaa-Xaa-Xaa library showed that the major determinant for P1′ was for small amino acids such as Gly, Ala, and Ser. Like thrombin, a very little activity is observed in the P2′–P4′ positions of the substrate for caspase-3.

Validation of the screening method was obtained by the synthesis and kinetic evaluation of four thrombin substrates that differed in the prime site amino acids (Table 1). Changing the P1′ Ser to an Arg dramatically reduced the k_{cat}/K_m by over 75-fold while the more conservative change to a Gly only resulted in an approximate 4-fold reduction. Strict dependence on the P2′–P4′ was not observed for thrombin in the library, however there was a prominent absence of activity for acidic amino acids in P1′–P3′. Accordingly, changing P2′ from Asn to Asp resulted in a 6-fold reduction in k_{cat}/K_m. LCMS analysis of the hydrolysis products of the first two substrates of Table 1 by Thrombin was performed. In both cases, only the products from the expected cleavage site were observed (Supplementary data). Hydrolysis of the second substrate was not observable after 1 h but an overnight incubation demonstrated hydrolysis at the expected cleavage site (data not shown).

Evaluation of cleavage sites of physiological substrates of thrombin showed a correspondence between the results from the libraries and the sequences found in natural substrates (Table 2).

Table 1. Steady state kinetic analysis of single substrates cleavage by thrombin

P4–P4′		k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu M^{-1}\ s^{-1}$)
Substrate	Sequence			
LTPR	SNRG	80	11	1.55
LTPR	**R**NRG[a]			0.02
LTPR	S**D**RG	53	213	0.25
LTPR	**G**NRG	189	460	0.41

[a] Saturation kinetics not observed for this substrate.

Table 2. Alignment of thrombin cleavage sites in some macromolecular substrates

Substrate	P4	P3	P2	P1	▼	P1′	P2′	P3′	P4′
PAR1	L	D	P	R		S	F	L	L
PAR2	L	F	F	R		T	R	S	I
Factor V	L	G	I	R		S	F	R	N
Factor V	L	S	P	R		T	F	H	P
Factor VIII	I	Q	I	R		S	V	A	K
Factor VIII	I	E	P	R		S	F	S	Q
Factor VIII	Q	S	P	R		S	F	Q	K
Factor XI	I	K	P	R		I	V	G	G
Factor XIII	V	V	P	R		G	V	N	L

The iterative strategy presented here represents a new method for rapidly screening both non-prime and prime subsites of protease peptide substrates. The substrate specificity information obtained will be useful for the elucidation of optimal substrates for assay development and the design of highly selective biological probes.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2005.04.019.

References and notes

1. *Handbook of Proteolytic Enzymes*; Barrett, A., Rawlings, N., Woessner, J., Eds., 2nd ed.; Elsevier Academic Press: London, 2004.
2. Puente, X. S.; Sanchez, L. M.; Overall, C. M.; Lopez-Otin, C. *Nat. Rev. Genet.* **2003**, *4*, 544.
3. Southan, C. *Drug Discovery Today* **2001**, *6*, 681.
4. Richardson, P. L. *Curr. Pharm. Des.* **2002**, *8*, 2559.
5. Tsu, C. A.; Craik, C. S. *J. Biol. Chem.* **1996**, *271*, 11563.
6. Frederiks, W. M.; Mook, O. R. F. *J. Histochem. Cytochem.* **2004**, *52*, 711.
7. Matthews, D. J.; Wells, J. A. *Science (Washington, DC, United States)* **1993**, *260*, 1113.
8. St. Hilaire, P. M.; Willert, M.; Juliano, M. A.; Juliano, L.; Meldal, M. *J. Comb. Chem.* **1999**, *1*, 509.
9. Houghten, R. A.; Appel, J. R.; Blondelle, S. E.; Cuervo, J. H.; Dooley, C. T.; Pinilla, C. *BioTechniques* **1992**, *13*, 412.
10. Rano, T. A.; Timkey, T.; Peterson, E. P.; Rotonda, J.; Nicholson, D. W.; Beckeer, J. W.; Chapman, K. T.; Thornberry, N. A. *Chem. Biol.* **1997**, *4*, 149.
11. Harris, J. L.; Backes, B. J.; Leonetti, F.; Mahrus, S.; Ellman, J. A.; Craik, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7754.
12. Harris, J. L.; Alper, P. B.; Li, J.; Rechsteiner, M.; Backes, B. J. *Chem. Biol.* **2001**, *8*, 1131.
13. Maly, D. J.; Leonetti, F.; Backes, B. J.; Dauber, D. S.; Harris, J. L.; Craik, C. S.; Ellman, J. A. *J. Org. Chem.* **2002**, *67*, 910.
14. Shipway, A.; Danahay, H.; Williams, J. A.; Tully, D. C.; Backes, B. J.; Harris, J. L. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 953.
15. Nicolaou, K. C.; Pfefferkorn, J. A.; Mitchell, H. J.; Roecker, A. J.; Barluenga, S.; Cao, G. Q.; Affleck, R. L.; Lillig, J. E. *J. Am. Chem. Soc.* **2000**, *122*, 9954.
16. Donor–quencher PSSCLs to determine prime site specificity of thrombin: Nanokans were loaded with RINK amide AM resin that had been functionalized first with Arginine(Pbf) and then with *N*-α-Fmoc-*N*-ε-2,4-dinitrophenyl-L-lysine (Lys (DNP)). The resin loading was 0.37 mmol/g and each Nanokan was loaded with 8 mg of resin to give a theoretical yield of 0.3 μmol/Nanokan. A total of 3800 Nanokans were employed for the planned synthesis of 50 copies of the P1′–P4′ PSSCLs. The first four positions from the Lys (DNP) were varied by creating sub-libraries in which a single position was fixed with one of 19 amino acids (cysteine was excluded and norleucine was substituted for methionine) and the other three were addressed with an isokinetic mixture of 19 amino acids.[9] Four rounds of synthesis with library planning and sorting performed by the IRORI automation yielded multiple copies of 76 Nanokan PSSCL libraries. A single copy of the PSSCL library was placed in a round bottom and the four optimal amino acids defined through ACC profiling were coupled in sequence. The non-prime preferred thrombin cleavage sequence identified was: P1—Arg, P2—Pro, P3—Ser, P4—Leu. Synthesis was performed using standard solid phase peptide techniques. FMOC deprotections were performed by agitating Nanokans in a 20% piperidine NMP solution for 20 min, followed by a wash with NMP and a further 20 min subjection to 20% piperidine in NMP. Amino acid couplings were performed using a 0.3 M solution of the FMOC acid, HOBt, and DICI in NMP. The reaction was allowed to agitate for 5 h or overnight then the Nanokans washed with NMP (×3) and DCM (×3). For Nanokans requiring a mixture, an isokinetic mixture was employed.[9] Following peptide synthesis the methoxy coumarin acetic acid (MCA)[18] was added to the N-terminus of the peptides using a 0.3 M solution of MCA, HOBt, and DICI in NMP. Cleavage from the resin was performed in the IRORI cleavap using a 95% TFA, 2.5% water, 2.5% TIS solution to afford solid PSSCL upon removal of cleavage cocktail in IRORI cleavap.
17. Donor–quencher PSSCL to determine prime site specificity of caspase-3 was created in exactly the same manner as the thrombin library with the non-prime site specificity determined as P4—Asp, P3—Val, P2—Glu, P1—Asp.
18. Knight, C. G.; Willenbrock, F.; Murphy, G. *FEBS Lett.* **1992**, *296*, 263.
19. Le Bonniec, B. F.; Myles, T.; Johnson, T.; Knight, C. G.; Tapparelli, C.; Stone, S. R. *Biochemistry* **1996**, *35*, 7114.