Pergamon

Bioorganic & Medicinal Chemistry Letters 10 (2000) 2639–2642

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Synthesis of a Statistically Exhaustive Fluorescent Peptide Substrate Library for Profiling Protease Specificity

James E. Sheppeck II,[a,*] Heidi Kar,[a] Luke Gosink,[a] Jeffrey B. Wheatley,[a] Erik Gjerstad,[b] Siobhan M. Loftus,[b] Alexi R. Zubiria[b] and James W. Janc[b]

[a]*ChemRx Advanced Technologies, 385 Oyster Point, South San Francisco, CA 94080, USA*
[b]*Axys Pharmaceuticals, Enzymology Department, 180 Kimball Way, South San Francisco, CA 94080, USA*

Received 17 August 2000; accepted 14 September 2000

Abstract—A statistically exhaustive, 8800 compound tripeptidal amidomethylcoumarin library was synthesized as discreet compounds using solid-phase combinatorial chemistry. A subset of the compounds was purified by HPLC and tested in a high-throughput fluorometric assay against several known serine and cysteine proteases to demonstrate the utility of this library for profiling protease substrate specificity.

Introduction

Proteases constitute one of the largest superfamilies of enzymes, many of which embody some of today's most exciting drug targets.[1] A first step in understanding the biological function of a particular protease is knowing its substrate specificity. This information can be used to synthesize fluorogenic substrates for use in high-throughput screening and to design peptide-based inhibitors for accelerating medicinal chemistry programs. Fluorescent substrates for assaying protease activity have traditionally come from transposing the in vivo substrate sequence onto a fluorophore, screening the limited set of commercially available substrates (presently about 200), or using the consensus sequence specificity of homologous proteases as a starting point. Although sometimes successful, these methods are no substitute for assaying a protease using a stastistically complete strategy whereby all possible amino acid sequences are examined for efficient proteolytic cleavage and selectivity.

Combinatorial chemistry has made a significant impact in profiling protease substrate specificity. Use of pooled compound strategies, such as phage display libraries[2] and combinatorial peptide libraries assayed on and off solid support,[3] have afforded a rapid and exhaustive means of elucidating the preferred consensus sequence of many proteases. However, non-fluorometric assays only yield indirect information about the specific sequence under study and require post-proteolytic analysis such as Edman degradation or mass spectrometry. Although on-bead enzymology is a powerful method, it suffers from a disparity between enzyme kinetics on and off solid support and fails altogether for some enzymes.

Many of these shortcomings were overcome by the Merck group who made a positional scanning synthetic combinatorial library (PS-SCL) of peptidal 7-amino-4-methylcoumarin (AMC) substrates based on the structure Ac-aa_4-aa_3-aa_2-Asp-AMC. Pools of these compounds were then screened to assess the optimal P_4–P_2 sequence for several members of the caspase family—proteases that have a strict requirement for aspartic acid in the P_1 position.[4] The SmithKline Beecham group synthesized a similar but smaller caspase-directed substrate library as a combinatorial array (CA) of single compounds per well rather than the pooled, PS approach.[5] More recently, Ellman et al. published two new synthetic routes which they used to make pooled, positional scanning (PS) libraries of the general structure Ac-aa_4-aa_3-aa_2-aa_1-AMC (or ACC) for determining the specificity of several proteases.[6]

Our approach was to make a statistically exhaustive tripeptidal AMC library of the general structure Ac-aa_3-aa_2-aa_1-AMC in which all three amino acid positions were independently randomized using an expanded set of the 20 naturally occurring amino acids (Table 1). The library was synthesized as a combinatorial array of

*Corresponding author. Tel.: +1-650-829-1142; +1-fax: 650-829-1266; e-mail: jim_sheppeck@axystech.com

0960-894X/00/$ - see front matter
PII: S0960-894X(00)00545-X

single compounds per well followed by high-throughput HPLC purification. Although a CA cannot practically access the numbers of compounds synthesized in a PS library, it does have distinct advantages, among them: the ability to obtain and directly compare kinetic data for single substrates without resynthesis, higher quality control, and actual yields of the final compounds for rigorous molarity determination. Because the PS approach involves making simultaneous pairwise changes to the peptide sequence, the implicit assumption is that there are no context effects (such as conformational changes) and that such changes are always additive when sometimes, they are not.

Since our CA contained 22 P_1 amino acids, high-throughput screening of the library against any protease should rapidly establish its specificity without the need for prior knowledge of the given protease's in vivo target, structure, or homology to other proteases. The best substrate sequences could then serve as a starting point for a follow-on library to improve their kinetic properties and specificity between proteases.

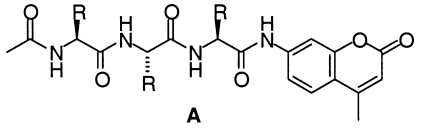
Chemistry

The Merck group's caspase-directed substrate library had the advantage of an aspartic acid in the P_1 position for attachment to a resin through a side chain ester linkage. Since we wished to randomize the P_1 position as well, synthesis of our library required the use of a traceless linker strategy that would take advantage of either Fmoc or Boc solid-phase peptide synthesis. Because the peptide side chains of each structure represent the 'diversity' of the library and we elected not to structurally perturb the well-entrenched AMC fluorophore, the potential attachment points of the solid support were limited to one of the backbone amides. Two synthetic methods are precedented in the literature that are amenable to traceless solid-phase peptide synthesis: the 'safety-catch' approach originally described by Kenner[7] and significantly improved by Ellman and co-workers[8] and, the backbone amide linker (BAL) approach pioneered by Jenson et al.[9] We investigated both approaches but found the safety catch route to be more general and tolerant of difficult peptide couplings.

The library was synthesized as a CA with the dimensions: 20 P_3 × 20 P_2 × 22 P_1 (8800 members) using the starting materials listed in Table 1. The preparation of the library began with the synthesis of 22 individual aminoacyl-AMC starting materials **5a**–**v** (Scheme 1). Boc-protected amino acids coupled cleanly to AMC using the original mixed anhydride conditions reported by Zimmerman[10] followed by standard TFA-mediated Boc deprotection. Fmoc-protected amino acids were coupled to AMC using the mixed anhydride conditions reported by Alves and co-workers.[11] Selective removal of the Fmoc group in solution was accomplished using a catalytic DBU/thiol reagent system that we have already described.[12]

As depicted in Scheme 2, synthesis of the library began by coupling the 20 P_2 amino acids **6** listed in Table 1 to modified 'safety catch' resin using the low temperature

Table 1. Structure of the library and the diversity elements used

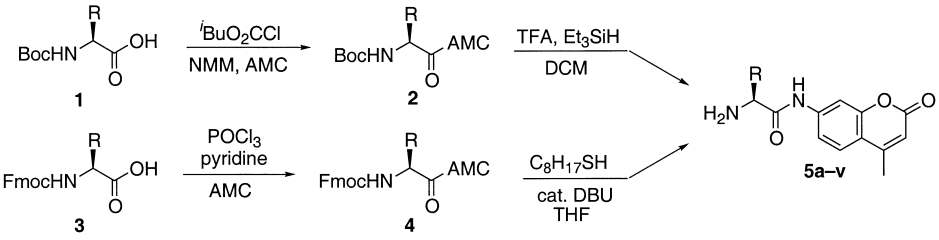


A

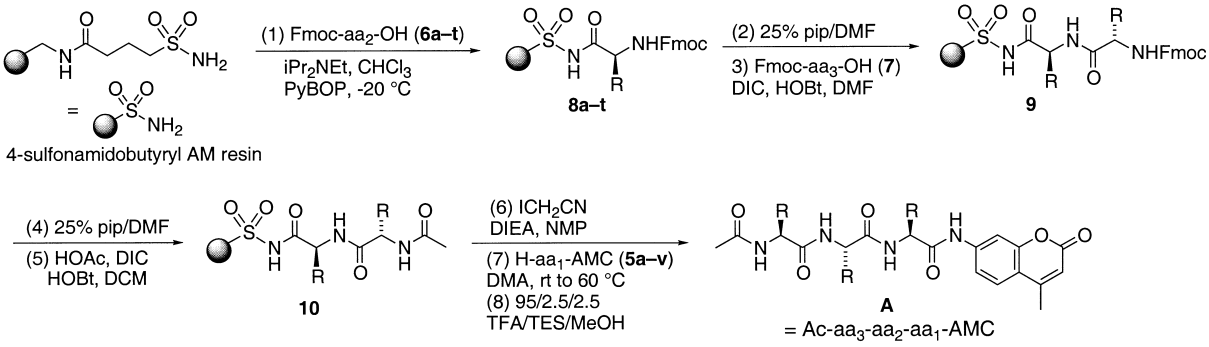
Entry	Fmoc-aa$_3$-OH, **7** (P_3 amino acid)	Fmoc-aa$_2$-OH, **6** (P_2 amino acid)	H-aa$_1$-AMC, **5** (P_1 amino acid)
a	Fmoc-Arg(Pbf)-OH	Fmoc-Arg(Pbf)-OH	H-Arg(Pbf)-AMC
b	Fmoc-Asn(Trt)-OH	Fmoc-Asn(Trt)-OH	H-Asn(Trt)-AMC
c	Fmoc-Asp(OtBu)-OH	Fmoc-Asp(OtBu)-OH	H-Asp(OtBu)-AMC
d	Fmoc-Glu(O*t*Bu)-OH	Fmoc-Glu(O*t*Bu)-OH	H-Cys(Trt)-AMC
e	Fmoc-Gln(Trt)-OH	Fmoc-Gln(Trt)-OH	H-Glu(O*t*Bu)-AMC
f	Fmoc-His(Boc)-OH	Fmoc-His(Boc)-OH	H-Gln(Trt)-AMC
g	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	H-His-AMC
h	Fmoc-Orn(Boc)-OH	Fmoc-Orn(Boc)-OH	H-Lys(Boc)-AMC
i	Fmoc-Ser(*t*Bu)-OH	Fmoc-Ser(*t*Bu)-OH	H-Orn(Boc)-AMC
j	Fmoc-Thr(*t*Bu)-OH	Fmoc-Thr(*t*Bu)-OH	H-Cit-AMC
k	Fmoc-Trp(Boc)-OH	Fmoc-Trp(Boc)-OH	H-Ser(*t*Bu)-AMC
l	Fmoc-Tyr(*t*Bu)-OH	Fmoc-Tyr(*t*Bu)-OH	H-Thr(*t*Bu)-AMC
m	Fmoc-Ala-OH	Fmoc-Ala-OH	H-Trp(Boc)-AMC
n	Fmoc-Gly-OH	Fmoc-Gly-OH	H-Tyr(*t*Bu)-AMC
o	Fmoc-Ile-OH	Fmoc-Ile-OH	H-Ala-AMC
p	Fmoc-Leu-OH	Fmoc-Leu-OH	H-Gly-AMC
q	Fmoc-Met-OH	Fmoc-Met-OH	H-Ile-AMC
r	Fmoc-Phe-OH	Fmoc-Phe-OH	H-Leu-AMC
s	Fmoc-Pro-OH	Fmoc-Pro-OH	H-Met-AMC
t	Fmoc-Val-OH	Fmoc-Val-OH	H-Phe-AMC
u			H-Pro-AMC
v			H-Val-AMC

conditions optimized by Backes and Ellman (PyBOP/$CHCl_3$/iPr_2NEt/−20 °C).[8a]

The 20 resulting resins **8a**–**t**, were dispensed into 96-well plates. Standard solid-phase Fmoc amino acid synthesis was used to prepare the solid-supported dipeptides which were *N*-acetyl capped (HOAc, HOBt, DIC) to give resins of structure **10**. Optimization of the resin cleavage reaction identified dimethylacetamide (DMA) as the optimal solvent that provided both good solubility of **5a**–**v** and reaction rates. The linker was activated with iodoacetonitrile (20 equiv), washed, and cleaved with the 22 P_1 elements **5a**–**v** listed in Table 1 to give the library compounds in solution. Cleavage from the activated resin with sterically hindered amino acids such as H-Val-AMC and H-Ile-AMC required elevated temperatures and were accompanied by partial racemization of the P_2 amino acid. Cleavages using



Scheme 1. Synthesis of the aminoacyl-AMC starting amines **5a**–**v** (H-aa_1-AMC).



Scheme 2. Synthesis of the tripeptidal AMC library.

Table 2. Results from screening Ac-Phe-aa_2-aa_1-AMC[a]

Protease	Best AMC substrate sequences	% of control substrate
Cathepsin B[b]	Ac-Phe-Arg-Arg-AMC	46
	Ac-Phe-Lys-Arg-AMC	43
	Ac-Phe-Arg-Lys-AMC	37
	Ac-Phe-Val-Lys-AMC	23
	Ac-Phe-Orn-Lys-AMC	21
	Ac-Phe-Orn-Arg-AMC	18
	Ac-Phe-Met-Lys-AMC	18
	Boc-Leu-Lys-Arg-AMC[c]	100
Elastase	Ac-Phe-Pro-Val-AMC	11
	Ac-Phe-Thr-Val-AMC	1.6
	Ac-Phe-Glu-Val-AMC	1.0
	Ac-Phe-Pro-Ala-AMC	0.8
	MeOSuc-Ala-Ala-Pro-Val-AMC[c]	100
Factor Xa	Ac-Phe-Tyr-Arg-AMC	1.1
	$MeSO_2$-D-Cha-Gly-Arg-AMC[c]	100
Urokinase	Ac-Phe-Ser-Arg-AMC	4.2
	Ac-Phe-Gly-Arg-AMC	3.9
	Ac-Phe-Ala-Arg-AMC	2.0
	Glt-Gly-Arg-AMC[c]	100
Thrombin	Ac-Phe-Pro-Arg-AMC	5.4
	Z-Gly-Pro-Arg-AMC[c]	100

[a]Values are relative fluorescence units (RFU)/min determined at 100 μM concentration of substrate diluted from DMSO stock (1 mM). Final DMSO concentration was < 10% v/v.
[b]Additional cathepsin B data is shown in the Fig. 1 histogram.
[c]Positive control substrates.

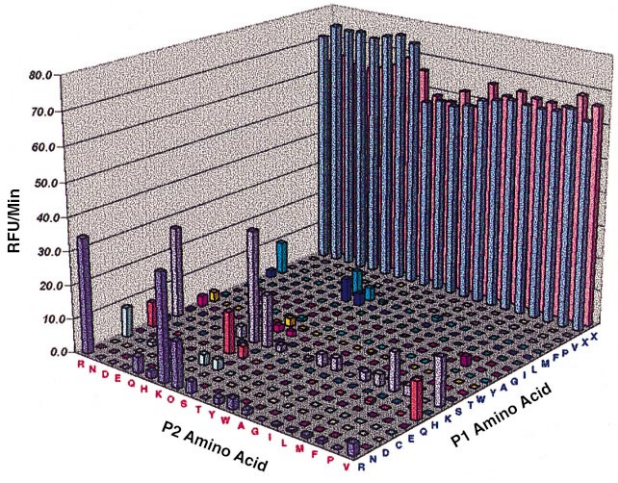


Figure 1. 2D Histogram of the complete cathepsin B data. 'X' represents Boc-Leu-Lys-Arg-AMC positive control substrate (the two sets of data are taken from different assay plates).'O' is ornithine.

H-Pro-AMC, **5u**, were kinetically the slowest reactions observed. The compounds were concentrated in vacuo, deprotected (95:2.5:2.5, TFA:TES:MeOH), and concentrated again to give the final library compounds of general structure **A**. Cysteine containing sequences were treated with DTT to reduce cystine disulfide bonds. Each substrate was subjected to high-throughput reverse-phase HPLC purification to remove unreacted **5**, protecting group impurities, and minor diastereomeric impurities. The identity of each compound was confirmed by direct injection MS.

Biology

Of the 8800 compounds, we sampled 400 of them where the P_1 and P_2 elements were randomized with the precursors listed in Table 1 and P_3 was fixed with *N*-acetyl phenylalanine. Each substrate was assayed against five typical serine and cysteine proteases: cathepsin B, elastase, factor Xa, urokinase, and thrombin using the known and preferred substrates for each protease as positive controls (Table 2).

The assay data for the cysteine protease cathepsin B (Table 2 and Fig. 1) clearly illustrate that it prefers basic amino acids (Arg, Lys, and Orn) in both the P_1 and P_2 positions. The optimal sequence for elastase is Ac-Phe-Pro-Val with a clear preference for Val in P_1. The optimal sequences for the trypsin-like serine proteases factor Xa, urokinase, and thrombin were Ac-Phe-Tyr-Arg, Ac-Phe-Ser-Arg, and Ac-Phe-Pro-Arg, respectively, confirming their preference for Arg in P_1 and in the case of fXa and thrombin, a high selectivity for the P_2 amino acids. Although none of the substrates was superior to the control substrates, all of the data are consistent with the well-established substrate specificity for all five enzymes[13] and there appear to be no false positives. Since the 400 compounds that were screened had P_3 fixed as Phe, screening of the entire library should elucidate the preferred P_3 amino acids as well.

Conclusion

We have completed a convergent synthesis of an 8800 compound tripeptidal AMC library using the Ellman-modified Kenner safety catch linker. Assay of a subset of these compounds against several known representative serine and cysteine proteases gave results consistent with the previously determined substrate specificity of these enzymes. Results of screening novel proteases against the entire 8800 membered library and use of this and other synthetic methods to make improved and less peptidic substrates will be reported in due course.

Acknowledgements

We wish to thank Brad Backes and Jon Ellman for helpful discussions about the use of their 4-sulfonamidobutyryl AM resin and synthetic method. We would also like to acknowledge Todd Somers and Nathan Collins for their insightful discussions and Bing Yan for analytical support.

References and Notes

1. Leung, D.; Abbenante, G.; Fairlie, D. P. *J. Med. Chem.* **2000**, *43*, 305.
2. Matthews, D. J.; Wells, J. A. *Science* **1993**, *260*, 1113.
3. For a review, see: (a) Lien, S.; Francis, G. L.; Graham, L. D. *Comb. Chem. High Throughput Screening* **1999**, *2*, 73. (b) St. Hilaire, P. M.; Willert, M.; Juliano, M. A.; Juliano, L.; Meldal, M. *J. Comb. Chem.* **1999**, *1*, 509. (c) Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G. *J. Org. Chem.* **1999**, *64*, 8761.
4. (a) Rano, T. A.; Timkey, T.; Peterson, E. P.; Rontonda, J.; Nichoson, D. W.; Becker, J. W.; Chapman, K. T.; Thornberry, N. A. *Chem. Biol.* **1997**, *4*, 149. (b) Thornberry, N. A.; Rano, T. A.; Peterson, E. P.; Rasper, D. M.; Timkey, T.; Garcia-Calvo, M.; Houtzager, V. M.; Nordstrom, P. A.; Roy, S.; Vaillancourt, J. P.; Chapman, K. T.; Nicholson, D. W. *J. Biol. Chem.* **1997**, *272*, 17907.
5. Lee, D.; Adams, J. L.; Brandt, M.; DeWolf, W. E., Jr.; Keller, P. M.; Levy, M. A. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1667.
6. (a) Backes, B. J.; Harris, J. L.; Leonetti, R.; Craik, C. S.; Ellman, J. A. *Nat. Biotechnol.* **2000**, *18*, 187. (b) 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.
7. Kenner, G. W.; McDermott, J. R.; Sheppard, R. C. *J. Chem. Soc., Chem. Commun.* **1971**, *636*.
8. (a) Backes, B. J.; Ellman, J. A. *J. Org. Chem.* **1999**, *64*, 2322. (b) Backes, B. J.; Virgilio, A. A.; Ellman, J. A. *J. Am. Chem. Soc.* **1996**, *118*, 3055.
9. Jensen, K. J.; Alsina, J.; Songster, M. F.; Vagner, J.; Albericio, F.; Barany, G. *J. Am. Chem. Soc.* **1998**, *120*, 5441.
10. (a) Zimmerman, M.; Yurewicz, E.; Patel, G. *Anal. Biochem.* **1976**, *70*, 258. (b) Zimmerman, M.; Ashe, B.; Yurewicz, E.; Patel, G. *Anal. Biochem.* **1977**, *78*, 47.
11. Alves, L. C.; Almeida, P. C.; Franzoni, L. J.; Juliano, M. A. *Pept. Res.* **1996**, *9*, 92.
12. Sheppeck, J. E., II; Kar, H.; Hong, H. *Tetrahedron Lett.* **2000**, *41*, 5329.
13. *Handbook of Proteolytic Enzymes*; Barrett, A. J. Rawlings, N. D., Woessner, J. F., Eds.; Academic: San Diego, 1998.