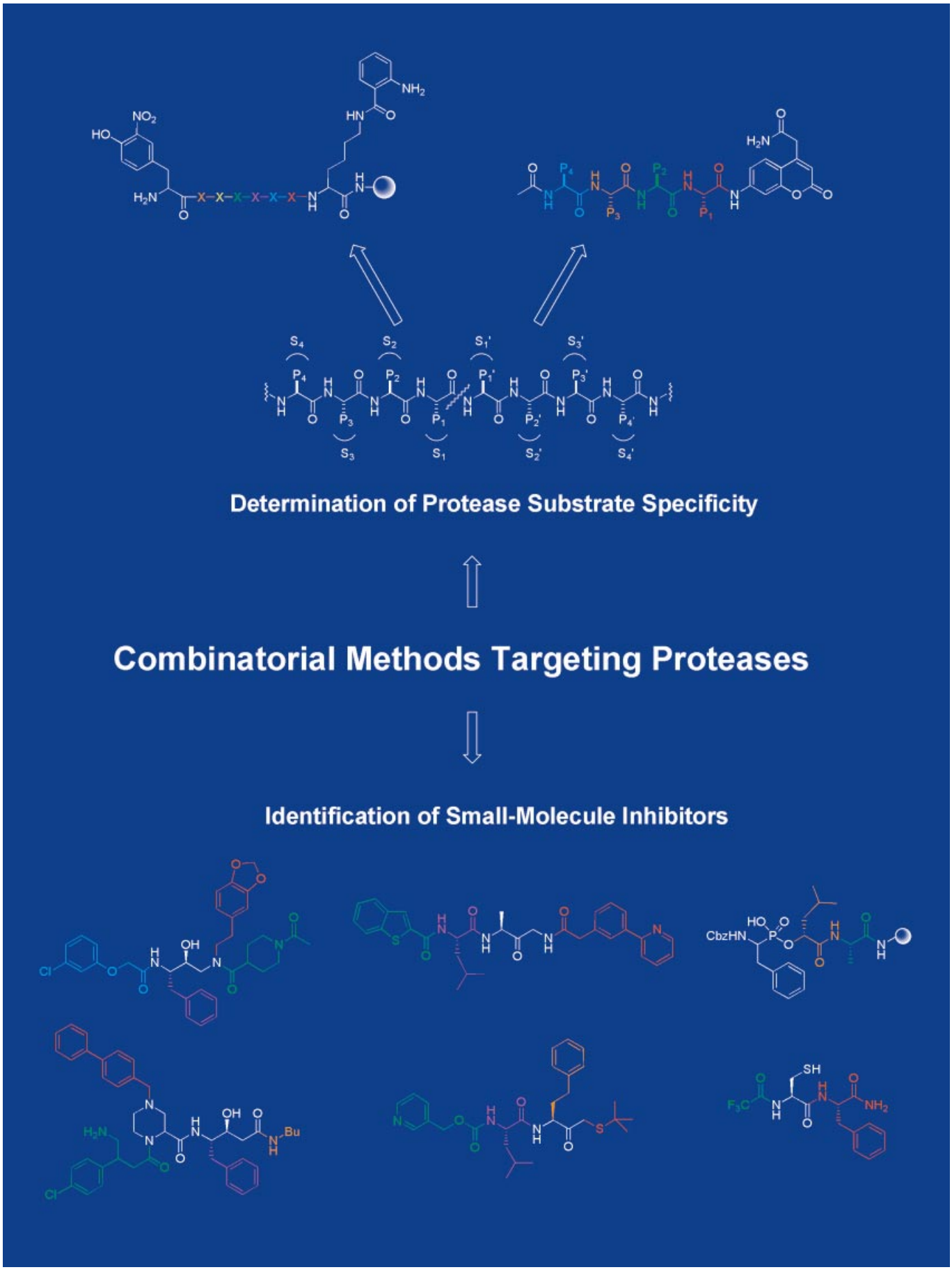


1439-4227/02/03/01 $ 17.50+.50/0 CHEMBIOCHEM **2002**, 3, 16 – 37

Combinatorial Strategies for Targeting Protein Families: Application to the Proteases**

Dustin J. Maly, Lily Huang, and Jonathan A. Ellman*[a]

Tens of thousands of proteins have been identified as a result of recent large scale genomic and proteomic efforts. With this large influx of new proteins, the formidable task of elucidating their function begins. However, this task becomes more manageable if proteins are divided into families based upon sequence homology, thereby allowing tools for their systematic study to be developed based upon their common structural and mechanistic characteristics. Combinatorial chemistry is ideally suited for the systematic study of protein families because a large amount of diversity can be readily displayed about a common scaffold designed to target a given protein family. Targeted combinatorial libraries have been particularly effective for the study of a ubiquitous family of proteins, the proteases. Substrate-specificity profiles of many proteases have been determined by using combinatorial libraries of appropriately labeled peptides. This specificity information has been utilized to identify the physiological protein substrates of these enzymes and has facilitated inhibitor design efforts. Furthermore, combinatorial libraries of small molecules prepared with mechanism-based scaffolds have resulted in the identification of potent, small-molecule inhibitors of numerous proteases. Cell-permeable small-molecule inhibitors identified by these methods have served as powerful chemical tools to study protease function in vitro and in vivo and have served as leads for the development of therapeutic agents.

KEYWORDS:

combinatorial chemistry · inhibitors · proteases · solid-phase synthesis · substrate specificity

1. Introduction

The initial sequencing of the human genome suggests the presence of 30000 – 40000 genes.[1, 2] The number of proteins encoded by these genes is much greater, when splice variants and post-translational modifications are taken into consideration. Furthermore, tens of thousands of additional encoded proteins have been identified from genomic and proteomic analysis of other organisms. For years to come, establishing the function of these proteins will be one of the most important goals of research in the biological sciences.

Characterization of this huge number of proteins is a seemingly insurmountable challenge. However, this task becomes less daunting when we consider that a majority of the proteins may be categorized into a much smaller collection of protein families, based upon sequence homology and, consequently, protein structure and mechanism.[1, 2] Combinatorial chemistry methods designed to exploit common structural or mechanistic features of protein families will play a critical role in establishing protein function, since a well-designed library synthesis method can potentially be applied to any member of a protein family. Targeted library methods may be designed to aid in the identification of the physiological substrates or ligands of members of a protein family. Targeted libraries may also be designed to enable rapid identification of cell-permeable ligands that selectively activate or inactivate any member of a protein family. These ligands can serve as powerful tools to elucidate the role of a protein in cells and in model organisms, in addition to serving as leads for drug discovery efforts.

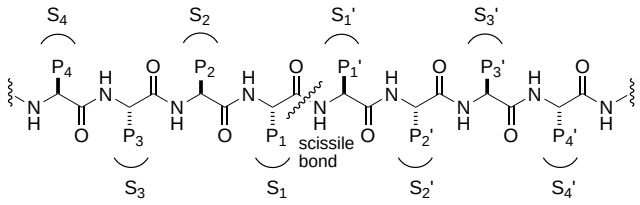
This review will summarize current combinatorial strategies designed to target one of the most important classes of proteins, the proteases. Libraries have been designed to both establish substrate specificity and to identify potent and selective small-molecule inhibitors.

2. Role of Proteases in Biological Processes

A majority of biological processes are regulated by proteases, which catalyze the hydrolysis of amide bonds in peptides and proteins (Scheme 1) to cause activation or inactivation of key protein targets. For example, in apoptosis, certain caspases inactivate proteins responsible for cell maintenance and DNA repair. In contrast, proteolytic cleavage of prohormones generates active peptide hormones that regulate many physiological processes. The homeostasis of numerous essential functions is tightly regulated by protease cascades. Two such processes, blood coagulation and fibrinolysis, are required for wound healing but have severe consequences when the body's machinery malfunctions, resulting in stroke and heart attacks.

[a] *Prof. J. A. Ellman, D. J. Maly, L. Huang*
Department of Chemistry
University of California
Berkeley, CA 94720-1460 (USA)
Fax: (+1) 510-642-8369
E-mail: jellman@uclink.berkeley.edu

[**] A list of abbreviations can be found at the end of the text.



Scheme 1. *Standard nomenclature for protease substrate cleavage. P_n, P_3, P_2, P_1, P_1', P_2', P_3', P_n', etc. designate amino acid side chains of a peptide substrate. Cleavage occurs between the P_1 and P_1' residues. The corresponding binding sites in the protease active site are designated as the S_n, S_3, S_2, S_1, S_1', S_2', S_3', S_n', etc. subsites.*[3]

Proteases are also essential to bacterial, viral, and parasitic pathogens where they can play a key role in replication, nutrition, and host invasion.

Due to the essential biological roles of proteases, a number of these enzymes are important therapeutic targets for a wide variety of illnesses, which range from inflammation, neurodegenerative disease, cardiovascular disease, and cancer to viral, parasitic, and bacterial infections. The therapeutic utility of protease inhibitors has been demonstrated by the treatment of AIDS with HIV protease inhibitors and the treatment of hypertension with angiotensin-converting enzyme inhibitors.

Jonathan Ellman received his S.B. degree from the Massachusetts Institute of Technology in 1984, where he worked in the laboratory of K. B. Sharpless. He completed his Ph.D. degree with D. A. Evans at Harvard University in 1989. In 1992, after an National Science Foundation postdoctoral fellowship at the University of California at Berkeley with P. G. Schultz, he joined the faculty at the University of California at Berkeley where he is currently Professor of Chemistry. He holds a joint appointment at University of California at San Francisco in the Department of Cellular and Molecular Pharmacology. His laboratory is currently engaged in the development of new design strategies for the preparation of small-molecule libraries and in the application of small-molecule libraries to different research problems in chemistry and biology. In addition, his group places a major emphasis on the development of practical and general synthetic methods that are applicable to both library synthesis and more traditional synthesis efforts. Professor Ellman has received a number of awards, including the American Chemical Society Cope Scholar Award, the University of California at Berkeley Department of Chemistry Teaching Award, the Burroughs Wellcome Foundation George Herbert Hitchings Award for Drug Design and Discovery, a Beckman Foundation Award, an Eli Lilly Grantee award, and an Alfred P. Sloan Fellowship.

3. Protease Substrate Specificity

The ability of proteases to selectively act upon a small number of targets in a myriad of potential physiological substrates is necessary for maintaining the fidelity of most biological functions. While substrate selection by proteases is governed by many factors, a principal determinant is the substrate specificity of the enzyme's active site (Scheme 1).[3] Knowledge of a protease's substrate specificity can greatly facilitate the elucidation of the physiological substrates of the protease, which is essential for defining its role in complex biological pathways. Determination of substrate specificity also can provide the basis for the design of potent and selective substrates and inhibitors.

Traditionally, panels of peptides were analyzed individually as potential substrates for a specific protease. While this process provides valuable information, it is tedious and impractical for the screening of a large number of potential substrates. Over the last few years a number of elegant combinatorial approaches have been developed to determine the substrate specificity of proteases. All of these combinatorial methods involve the generation of libraries of potential substrates, proteolysis of favorable substrates, and identification of the cleaved substrate sequences. Combinatorial substrate libraries can be divided into two categories depending on their method of generation: biological and synthetic. Biological library methods include the use of substrate phage display,[4, 5] the randomization of amino acids at physiological cleavage sites,[6] and the use of in vitro expression cloning of complementary DNA (cDNA) libraries to determine physiological substrates.[7] While biological combinatorial methods are very powerful, they are complicated by the lack of homogeneity in their presentation of potential substrates and the fact that these methods are often time consuming and technically demanding.

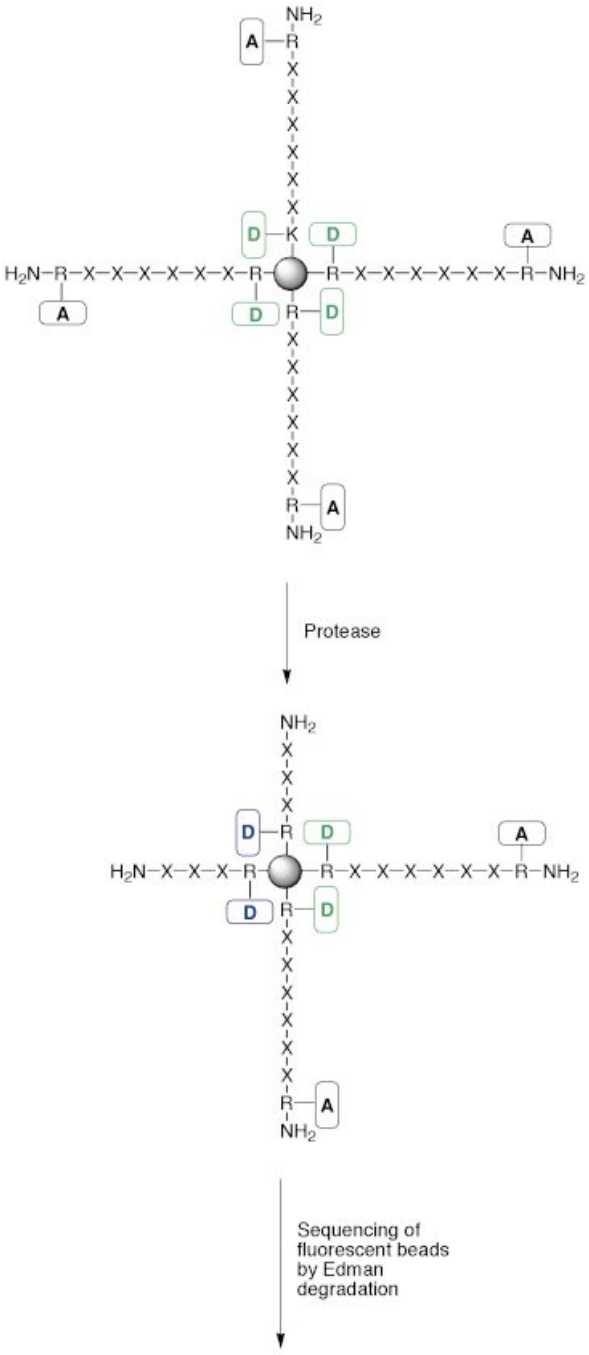
3.1. Synthetic protease substrates

Solid-phase peptide synthesis methods allow for the reproducible production of libraries of millions of potential substrates in a defined manner. These libraries can be synthesized in parallel arrays or by split and mix synthesis to provide discrete compounds or defined mixtures. Once these libraries are prepared they can then be screened against the desired protease directly on solid support or in solution. Finally, the analysis of protease cleavage can be monitored in a continuous or discontinuous manner.

3.2. Synthesis and screening of fluorescence-quenched peptide substrate libraries

Internally quenched fluorescent peptide substrates have proven to be of great utility for the assay of protease activity.[8] These substrates demonstrate long-range energy transfer between a donor fluorophore on one end of the peptide and a suitable acceptor on the other end. Upon cleavage of the substrate, the efficiency of this energy transfer is diminished and an increase in donor fluorescence of 10- to 50-fold is observed.[8]

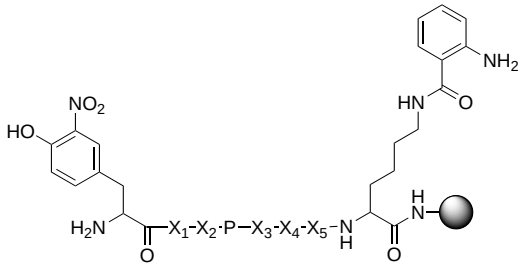
Internally quenched fluorescent peptide substrates have been utilized to develop a powerful combinatorial method for determining protease substrate specificity (Scheme 2).[9] An essential component of this methodology is the utilization of a solid support (for example, Tentagel, polyethylene glycol – poly(*N*,*N*-dimethylacrylamide) copolymer (PEGA), or superpermeable organic combinatorial chemistry resin (SPOCC)) that allows direct enzymatic screening of libraries of support-bound molecules. An amino acid that is conjugated with a suitable fluorescence donor is attached to the solid support. A peptide library is then synthesized by split and mix synthesis,[10] such that a single peptide sequence is synthesized on each resin bead. Finally, an amino acid labeled with a fluorescence acceptor is coupled to the N terminus of each peptide. This library of internally quenched fluorescent peptide substrates is then partially proteolyzed by the protease of interest and the most fluorescent beads are chosen for sequence determination. Sequencing of the free N-terminal amino group with Edman degradation allows determination of the substrate specificity on the prime and nonprime sides (Schemes 1 and 2).



Scheme 2. *Overview of substrate-specificity profiling with fluorescence-quenched peptide substrate libraries. D = fluorescence donor, A = fluorescence acceptor.*

3.2.1. The o-aminobenzamide – 3-nitrotyrosine donor – acceptor pair

The utility of combinatorial libraries of internally quenched fluorescent substrates was first demonstrated by determination of the substrate specificity of the serine protease subtilisin Carlsberg.[9] Two resins, PEGA and Kieselguhr-supported poly(*N*,*N*′-dimethylacrylamide), were utilized to construct the libraries due to their excellent swelling properties in both organic and aqueous media, which allowed library synthesis, as well as direct enzymatic screening, on solid support. The donor – acceptor amino acid pair *o*-aminobenzamide (Abz) and 3-nitrotyrosine ($Y(NO_2)$) was utilized to create a split and mix library of general structure H_2N-$Y(NO_2)$-X_1-X_2-P-X_3-X_4-X_5-K(Abz)-spacer-PEGA (Scheme 3), where X is all proteinogenic amino acids.



Scheme 3. *General structure of the o-aminobenzamide and 3-nitrotyrosine fluorescence-quenched substrate library used to determine the substrate specificity of subtilisin Carlsberg.*

Placement of a fixed proline residue was used to orient the library for cleavage between the X_3 and X_4 residues due to the absolute specificity of subtilisin Carlsberg for proline at the P_2 position. The screening results demonstrated that subtilisin Carlsberg has a distinct preference for glutamic acid at the P_2' position and a preference for leucine or phenylalanine at the P_1 site. Other subsites were found to be much less stringent in substrate specificity, with the P_4 subsite showing a distinct pH dependence.

This strategy has been utilized to determine the substrate specificities of several other proteases. In one study, a library of the general structure H_2N-$Y(NO_2)$-X_1-X_2-X_3-X_4-(R/K)-X_5-X_6-K(Abz)-spacer-$PEGA_{1600}$ was constructed and used to gain information about the substrate specificity of the cysteine protease cruzain.[11] The substrate specificity of the cysteine protease CPB2.8

from the protozoan parasite *Leishmania mexicana* was also evaluated by synthesizing and screening a library of general structure H_2N-Y(NO_2)-X_1-X_2-X_3-X_4-X_5-X_6-X_7-K(Abz)-spacer-$PEGA_{4000}$.[12] It was determined that the substrate specificity of CPB2.8 is very similar to that of cruzain with a few subtle differences. In a different study, a library of structure H_2N-X_1-X_2-Y(NO_2)-X_3-X_4-X_5-X_6-K(Abz)-spacer-$PEGA_{4000}$ was constructed and screened against the matrix metalloprotease, MMP-9.[13] Several good substrates for MMP-9 were identified by this method. Unfortunately, for this enzyme, there was not a good correlation between the efficiency of cleavage of support-bound and solution-phase substrates.

This methodology has been further utilized to develop a series of proteolytically cleavable peptide linkers for bioconjugates. These bioconjugates were designed to allow selective release of radiolabeled reagents from a targeting moiety (for example, an antibody) upon entering the lysosome. A library of the general structure H_2N-Y(NO_2)-X_1-X_2-X_3-X_4-F-K(Abz)-spacer-PEGA was constructed, where X is any one of nine amino acids known to be compatible with simple bioconjugate construction. This support-bound library was then screened against the abundant, lysosomal proteases cathepsin B and cathepsin D.[14] This methodology allowed for the rapid discovery of efficiently cleaved peptide linkers for these proteases, even with the small subset of amino acids used.

A thorough study of the subsite specificity of the cysteine protease papain has recently been published.[15] This study demonstrates the utility of fluorescence-quenched combinatorial libraries for determining protease substrate specificity, while also revealing some of the potential limitations in applying this methodology. A library of the general structure H_2N-Y(NO_2)-X_1-X_2-X_3-X_4-X_5-X_6-X_7-K(Abz)-spacer-$PEGA_{4000}$ was constructed and screened against papain. This library selected preferred substrates with a large percentage of aspartic and glutamic acid residues at all sites. However, these amino acids were selected due to the ability of these acidic residues to increase the effective local concentration of basic papain (pI = 8.75) in the solid support rather than because of the substrate preference of papain. The fluorescence quencher also influenced the site of cleavage of the peptide substrate, with the aromatic *o*-nitrotyrosine residue preferentially binding in the enzyme's S_2 site. To overcome these limitations, a second library was constructed of the general structure H_2N-Y(NO_2)-P-X_1-X_2-X_3-X_4-X_5-X_6-X_7-K(Abz)-spacer-$PEGA_{4000}$, with a proline residue introduced to prevent binding of *o*-nitrotyrosine in the S_2 subsite. To prevent the uneven distribution of papain in the solvated resin beads, acidic amino acids were not included in this library. Screening this second library successfully provided the $P_4 - P_4'$ substrate specificity of papain (Figure 1). The kinetics of the best substrates determined by the solid-phase assay were then compared to solution-phase donor – quencher substrates. For a number of the

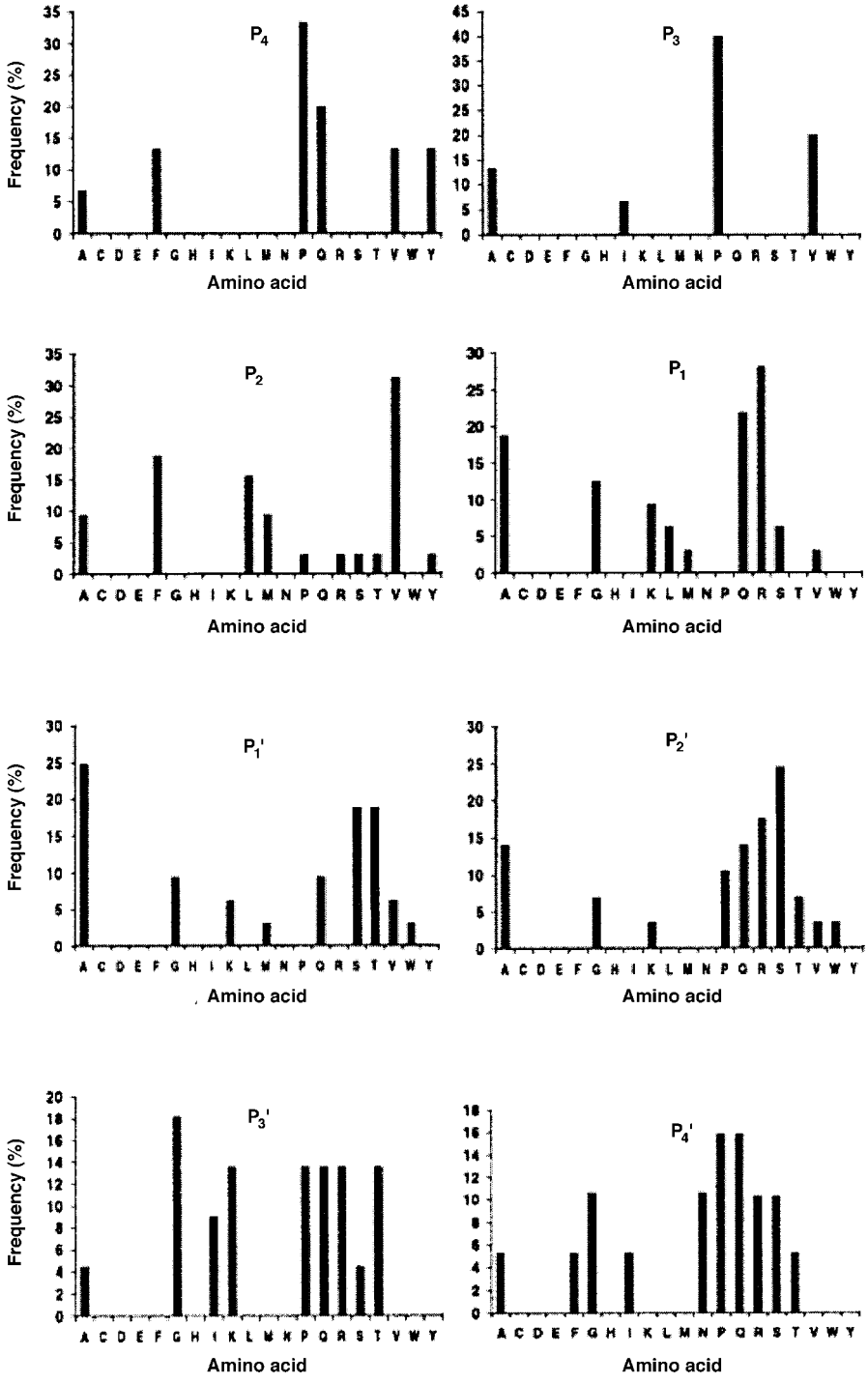
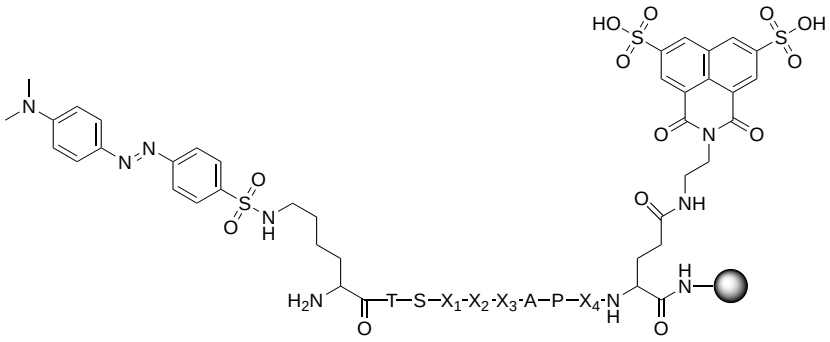


Figure 1. *Substrate specificity of papain. Reproduced with permission from ref. [15]. Copyright (1999) American Chemical Society.*

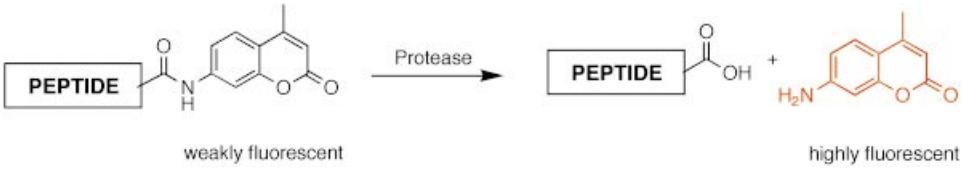
best substrates there was not a good relative correlation of efficiency for cleavage in solution compared to that on solid support. However, a number of efficiently cleaved substrates were still identified by this method.

3.2.2. The Lucifer yellow – Dabsyl donor – acceptor amino acid pair

Recently, fluorescence-quenched solid-phase combinatorial libraries with an alternative *N*-(2-aminoethyl)-4-amino-3,6-disulfo-1,8-naphthalimide (Lucifer yellow) and 4-dimethylaminoazobenzene-4′-sulfonyl (Dabsyl) fluorescence donor – acceptor amino acid pair (Scheme 4) have been used to determine the substrate specificities of *Escherichia coli* leader peptidase and the human aspartyl protease, napsin A.[16] A small percentage of the combinatorial library was utilized to demonstrate that the substrate specificity for *E. coli* leader peptidase was consistent with previous studies, thereby validating the use of the Lucifer yellow – Dabsyl donor – acceptor pair. The library was then used to determine the substrate specificity of the previously uncharacterized aspartyl protease napsin A. These studies not only provided insight for inhibitor design but also yielded a good synthetic substrate for high-throughput screening of this enzyme.



Scheme 4. *General structure of the Dabsyl and Lucifer yellow fluorescence-quenched peptide substrate library used to determine the substrate specificity of napsin A.*



Scheme 5. *Fluorogenic AMC peptide substrates.*

3.3. Cleavage of support-bound fluorophores

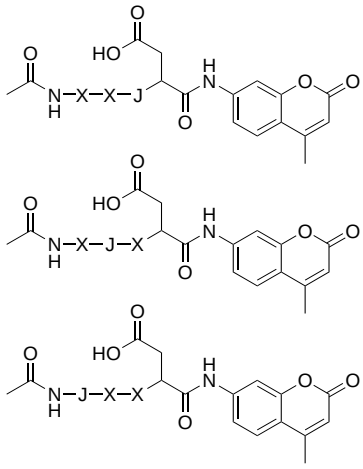
An alternative strategy for the construction and screening of combinatorial libraries on solid support has also been reported. This method relies on capping the N terminus of the peptide with a fluorophore, and observing the bead's decrease in fluorescence after the support-bound peptide is cleaved by the protease. This methodology has been utilized to evaluate the substrate specificities of elastase,[17] chymotrypsin, and papain.[18]

3.4. 7-Aminocoumarin fluorogenic substrates

The use of 7-amino-4-methylcoumarin (AMC) peptide substrates is a well-established method for the assay of protease activity. Cleavage of the anilide bond liberates the fluorogenic AMC leaving group (Scheme 5) and allows the simple and continuous determination of cleavage rates for individual substrates. Libraries of fluorogenic AMC peptide substrates can be used to define the nonprime substrate specificity of proteases. In particular, positional-scanning, synthetic combinatorial libraries (PS-SCLs) can establish protease substrate-specificity profiles extremely rapidly, since few assays (<100) need to be performed and information is provided for all side chains at every position. Finally, in contrast to the aforementioned combinatorial methods, this library format provides information on the substrate specificity of a protease in solution rather than in a solid-phase format.[19]

3.4.1. PS-SCLs of 7-amino-4-methylcoumarin peptides

The positional-scanning method was first demonstrated in determining the substrate-specificity profiles for members of the caspase family of proteases. The design of the library Ac-X-X-X-D-AMC (Scheme 6) was based on the absolute specificity of caspases for an aspartic acid residue at the P_1 position (Ac = acetyl). With this PS-SCL of fluorogenic AMC substrates, the



Scheme 6. *AMC tetrapeptide positional-scanning synthetic combinatorial library. The PS-SCL consists of three separate sublibraries of 8000 compounds each (20 wells of 400 compounds). "X" represents an equimolar mixture of proteinogenic amino acids (excluding cysteine and methionine) and "J" represents a fixed amino acid residue.*

$P_4 - P_2$ substrate specificities of *Caenorhabditis Elegans* CED-3, nine human caspases, and granzyme B were readily determined.[20]

Analysis of the complete substrate profile of the caspases (Figure 2) provided a great deal of insight into the physiological roles of these essential enzymes. First, the caspases could be clustered into three distinct groups based upon similarities in their sequence-specificity profiles. Second, key proteins involved in cell maintenance and DNA repair were found to be substrates for Group II caspases by comparing the determined specificity fingerprint to the cleavage sites of proteins known to be proteolysed during apoptosis. Further, it was determined that the preferred substrate specificity of the Group III caspases was very similar to the activation sites of several Group II proenzymes. This result was found to be consistent with the hypothesized role of the Group III caspases as upstream components in a proteolytic cascade that amplifies the death signal. Finally, based upon the substrate specificity of the Group I caspases, it was determined that this group of enzymes was unlikely to be involved in the apoptotic process. The information provided by these libraries greatly enhanced the understanding of the role of caspases in apoptosis and provided an essential conceptual framework for their further study.

The aforementioned Ac-X-X-X-D-AMC PS-SCL (Scheme 6) was also utilized in conjunction with substrate phage display to determine the extended substrate specificity of the serine protease granzyme B.[21] The PS-SCL was used to determine the $P_4 - P_2$ substrate spe-

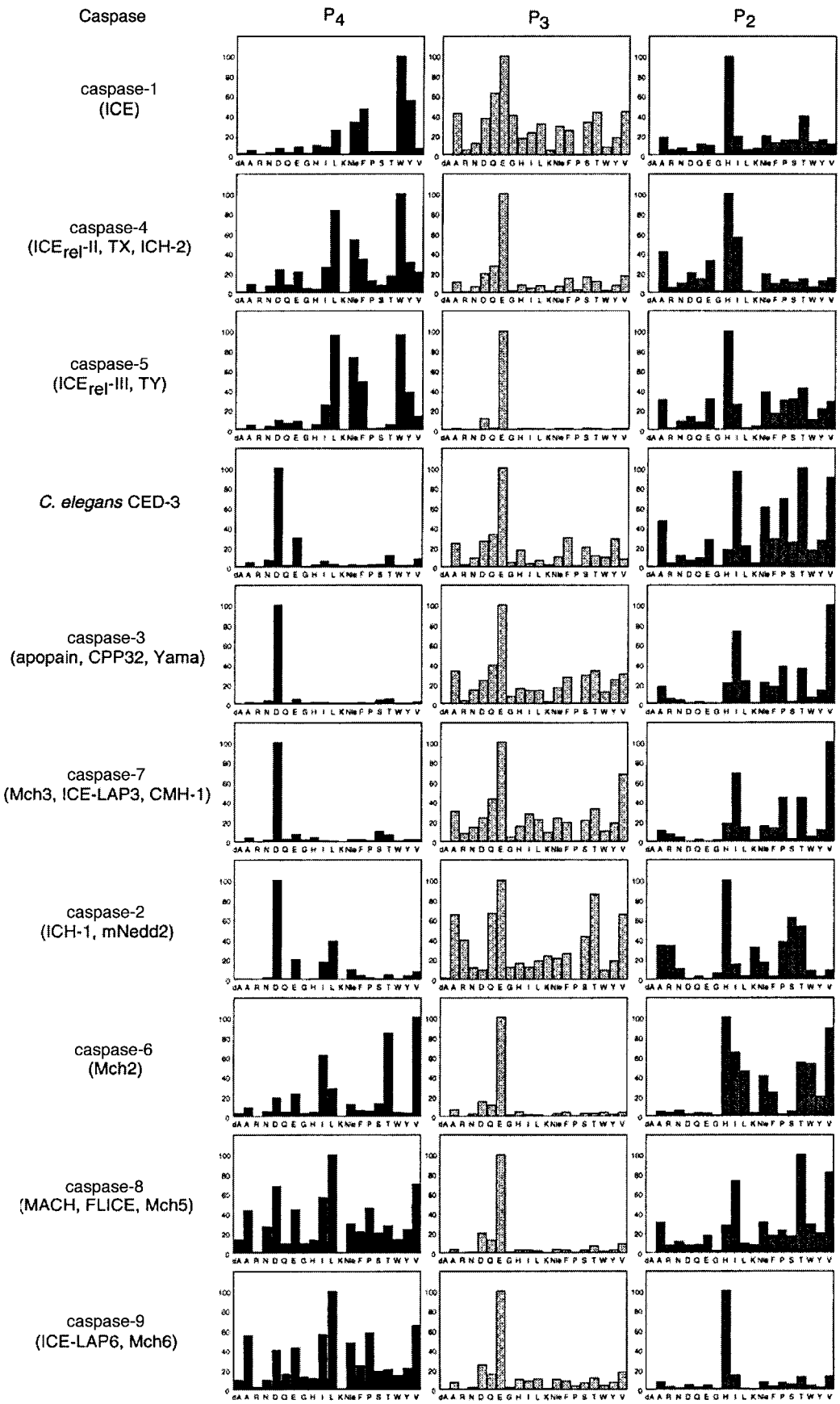


Figure 2. *Substrate specificities of the caspases. Reproduced with permission from ref. [20]. Copyright (1997) The American Society for Biochemistry and Molecular Biology.*

cificity of granzyme B and to direct the cleavage site for the determination of P_3, P_1', and P_2' specificity by substrate phage display. The use of these two methods allowed the determination of the preferred $P_4 - P_2'$ substrate sequence of I-E-X-D-X-G. This preferred substrate sequence was found to closely resemble the activation sites of caspase 3 and caspase 7, which is consistent with the role of granzyme B in the activation of these proteases during apoptosis. Further, many caspase substrates were found to contain potential granzyme B cleavage sites, which suggests a redundancy in function.

The AMC substrate libraries with aspartic acid at the P_1 position (Scheme 6) were conveniently synthesized by linking the carboxylic acid side chain of *O*-allyl-aspartic acid protected at the N terminus with 9-fluorenylmethyloxycarbonyl (Fmoc) to the insoluble support; this method allowed the library synthesis by standard solid-phase methods.[19, 22] Unfortunately, the method used to synthesize the library is specific for an aspartic acid at the P_1 site and is of limited utility for most proteases, which either have a different or unknown P_1 substrate specificity. Although it is possible to develop strategies to link through other amino acid side chains (for example, the δ-guanido group of arginine),[23] these strategies are cumbersome and not feasible for all 20 proteinogenic amino acids. General chemical strategies to incorporate all the proteinogenic amino acids at the P_1 position of a PS-SCL have therefore been developed to provide the extended $P_4 - P_1$ specificity.

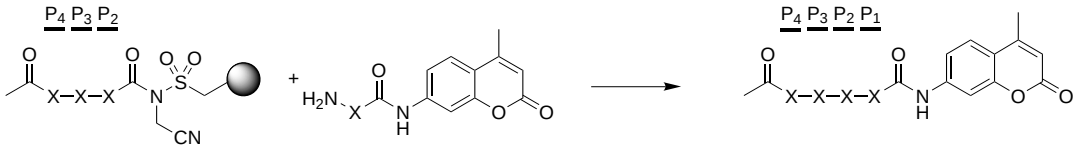
In one general method, fluorogenic AMC peptide substrates with any residue at the P_1 position can be prepared by using amino acid nucleophiles with AMC attached to displace activated tripeptides from solid supports (Scheme 7).[24, 25] Any positional-scanning library can be prepared since the AMC nucleophiles can be synthesized from any amino acid. The extended substrate specificities of the serine proteases plasmin and thrombin were determined by using a PS-SCL with a lysine at the P_1 position that was prepared by this method. Plasmin, a protease traditionally characterized as having a broad substrate specificity, was shown to have a distinct preference for aromatic amino acids at the P_2 position and a moderate preference for lysine and hydrophobic amino acids at the P_4 site. Thrombin was found to have a complete preference for proline at the P_2 position and a moderate preference for aliphatic amino acids at the P_4 site. The substrate specificities of plasmin and thrombin were found to be consistent with the known macromolecular substrates of these two enzymes and provided further insight into their roles in the blood coagulation pathway. This synthesis method has also recently been used to construct an 8800-member array of individual tripeptides that was utilized to determine the substrate specificity of the cysteine protease cathepsin B.[26]

The extended substrate specificity of membrane-type serine protease 1 (MT-SP1) was determined utilizing a PS-SCL with lysine at the P_1 position in combination with substrate phage display.[27] The two methods were found to be complementary, as the specificity information determined from the PS-SCL proved critical for the design of a biased substrate phage library with a preferred cleavage site. This biased substrate phage library was then used to determine the preferred P_1' residues, while it also detected an interdependency between the P_3 and P_4 sites that the PS-SCL did not. The preferred $P_4 - P_1'$ substrate sequences of R/K-X-S-R-A and X-R/K-S-R-A were determined by merging the data from these two methods. The determined substrate consensus sequence was utilized to search for potential macromolecular targets for MT-SP1. Protease activated receptor 2 (PAR2) and single-chain urokinase-type plasminogen activator (sc-uPA), two proteins that are colocalized with MT-SP1, were identified as efficient macromolecular substrates for MT-SP1.

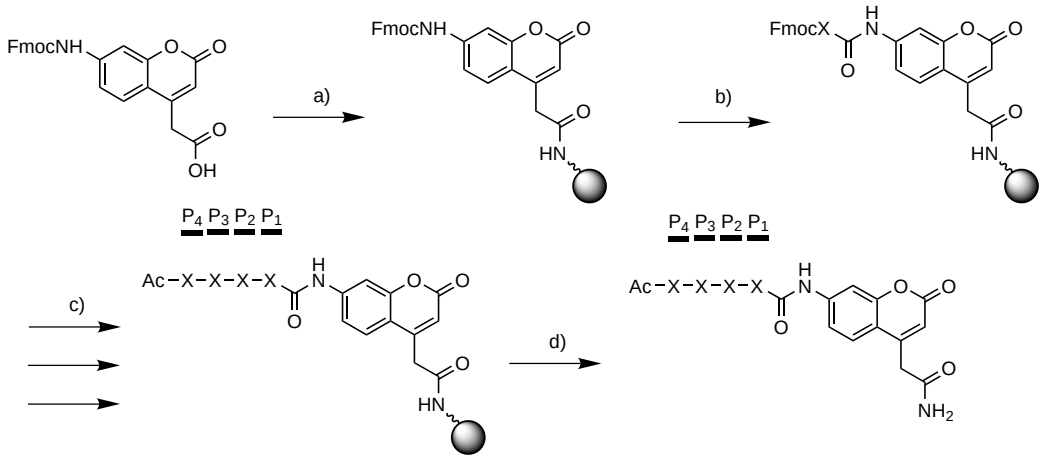
3.4.2. PS-SCLs of 7-amino-4-carbamoylmethylcoumarin peptides

Most recently, an even more efficient method has been developed to prepare positional-scanning libraries and arrays of fluorogenic substrates. The direct incorporation of any amino acid at all possible positions in a peptide substrate, including the P_1 position, can be accomplished by solid-phase peptide synthesis with a bifunctional fluorophore, 7-amino-4-carboxymethylcoumarin, which allows straightforward attachment to a solid support (Scheme 8).[28] Highly activated coupling conditions are used to couple the first amino acid residue to the poorly nucleophilic aniline. Any unreacted aniline is then capped with an acetyl group. Standard solid-phase synthesis methods then provide support-bound fluorogenic substrates that are released as peptidyl 7-amino-4-carbamoylmethylcoumarins (ACC) upon treatment of the support with trifluoroacetic acid (TFA).

To assess the performance of the modified fluorogenic ACC leaving group, an ACC PS-SCL with lysine at the P_1 position was compared to an AMC PS-SCL which also had lysine at that site. The serine proteases plasmin and thrombin showed comparable kinetic profiles and substrate specificities, which demonstrates the equivalency of ACC to AMC as a fluorogenic leaving group. To further validate ACC as a comparable leaving group to AMC, single fluorogenic peptide substrates were prepared for thrombin and the steady-state kinetics constants were shown to be comparable. The only major difference between the ACC and AMC fluorogenic leaving groups is the increased fluorescence of the ACC leaving group. This increased fluorescence allows the



Scheme 7. *General method for the preparation of AMC PS-SCLs. "X" denotes any natural or nonnatural amino acid.*



Scheme 8. *General method for the preparation of ACC PS-SCLs. "X" denotes any natural or nonnatural amino acid. a) Rink amine resin, DIC, HOBt; b) 1. 20% piperidine in DMF; 2. Fmoc-X-OH, HATU, 2,4,6-collidine; 3. AcOH, nitrotriazole, DIC; c) standard Fmoc SPPS; d) TFA/triisopropylsilane/H_2O (95:2.5:2.5).*

assay of proteases at lower enzyme and peptide-substrate concentrations.

To demonstrate the utility of the ACC substrates a P_1-diverse tetrapeptide library was created. With this library, the P_1 substrate specificities of a large number of different proteases, including the serine proteases thrombin, plasmin, factor Xa, urokinase-type plasminogen activator, tissue plasminogen activator, granzyme B, trypsin, chymotrypsin, human neutrophil elastase, and the cysteine proteases papain and cruzain were determined (Figure 3). The determined P_1 substrate specificities correlated well with the known P_1 preferences for these enzymes. The extended P_4-P_2 substrate specificities of the proteases listed above were also determined with this methodology. As observed with the P_1-diverse tetrapeptide library, the specificity profiles determined with the PS-SCLs were consistent with the limited substrate preference information that had been previously determined for these proteases.

In a separate effort an ACC PS-SCL was utilized to determine the substrate specificities of the serine proteases βI and βII tryptase.[29] This fingerprint information demonstrated the identical substrate preference for these two enzymes, which differ by only a single glycosylation site. The consensus sequence of P_4 = proline, P_3 = arginine/lysine, P_2 = X, and P_1 = lysine/arginine was used to search protein databases (SwissPROT) to reveal potential physiological substrates. Furthermore, the substrate specificity was used to generate a potent and selective chloromethyl ketone peptide inhibitor of βI and βII tryptase.

3.5. Use of irreversible inhibitors to determine substrate specificity

Recently, PS-SCLs of peptide vinyl sulfones, with an invariant asparagine at the P_1 position, have been utilized to determine the substrate specificity of serine proteases (Scheme 9).[30] These peptides serve as substrate mimics that covalently label the enzyme active site through Michael addition of the active-site nucleophile onto the vinyl sulfone. Thus, the side chains of the peptide inhibitor interact with the protease binding pockets that are responsible for defining substrate specificity. Analysis of inhibitor binding is performed with a competition assay in which mixtures or single compounds are first incubated with a protease. The effectiveness of the peptide vinyl sulfone as a substrate mimic is visualized by the addition of a general radiolabeled inhibitor that covalently modifies any remaining active site nucleophiles of the protease. The degree of inhibition is then quantified by separation of the protease with sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and determinination of the decrease in radioactivity compared to a standard.

By using PS-SCLs of peptide vinyl sulfones, the substrate specificities of individual catalytic subunits of the proteasome were determined under a variety of physiological conditions. It was found that the substrate specificities of the individual subunits varied from the specificity determined for the proteasome as a whole. Furthermore, the specificity information determined allowed for the design of inhibitors that selectively inactivate a single active site in the complex.

3.6. Edman degradation of peptide pools

Other combinatorial methods have been developed to analyze pools of synthetic peptides (Scheme 10). In these strategies, a combinatorial mixture of terminal *N*-acetylated peptides is incubated with a protease.[31, 32] The peptides compete as substrates for the enzyme and those with the highest specificity constants are cleaved. After a percentage of the peptide mixture has been cleaved, the mixture is analyzed. Only peptides that have been cleaved present a free α-amino group and, thus, are susceptible to N-terminal Edman degradation. This method of analysis provides information about the prime side of the cleaved peptide, where the first round of sequencing provides information about the P_1' position, the second round of sequencing provides information about the P_2' position, and so forth.

In a recent demonstration of this approach, proteolytic cleavage of acetyl-capped peptide pools followed by Edman

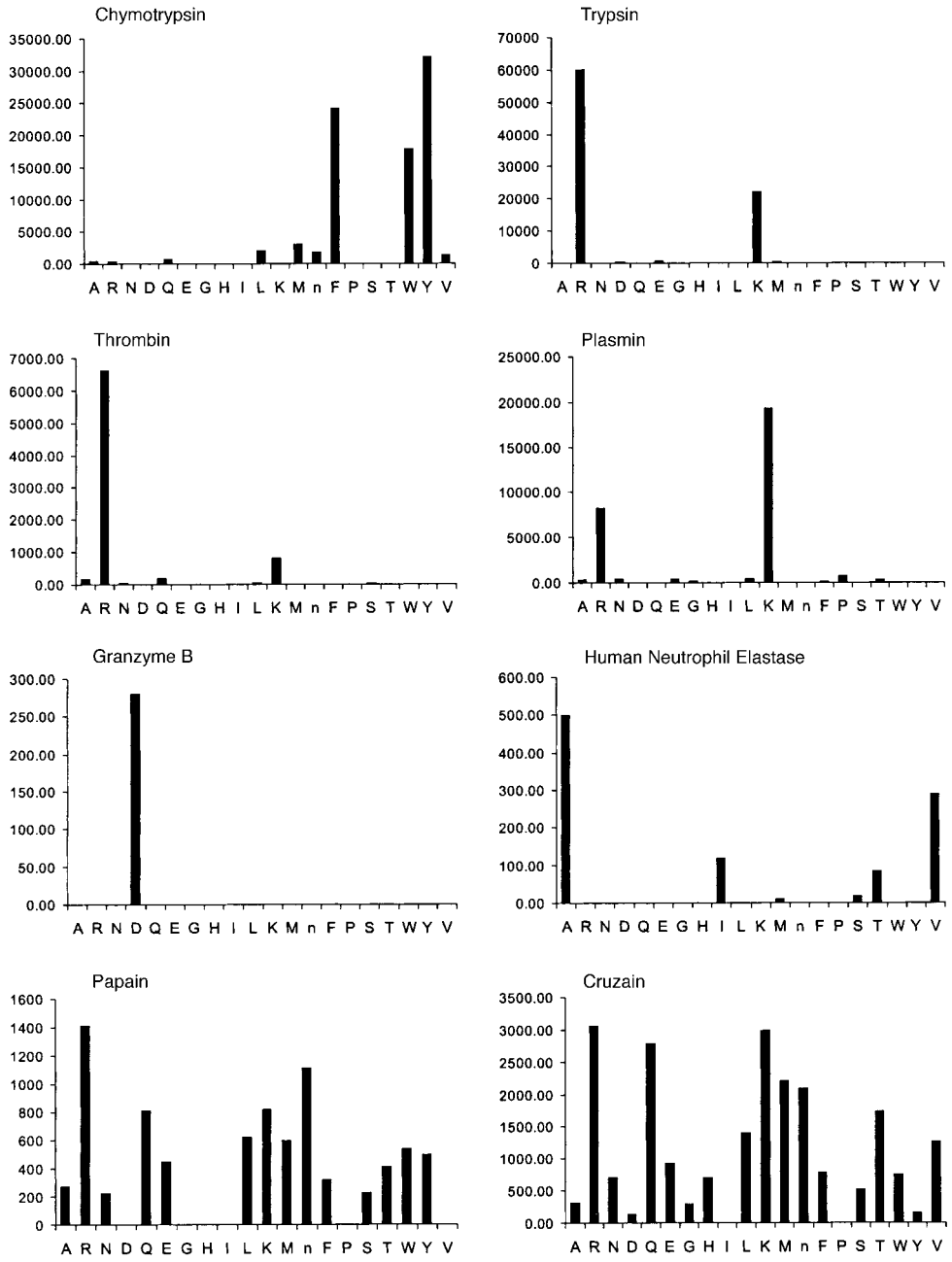
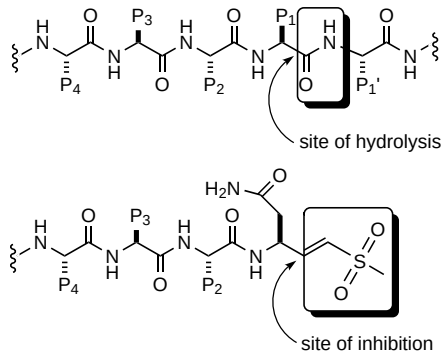


Figure 3. *P_1 substrate specificity of eight serine and cysteine proteases.*



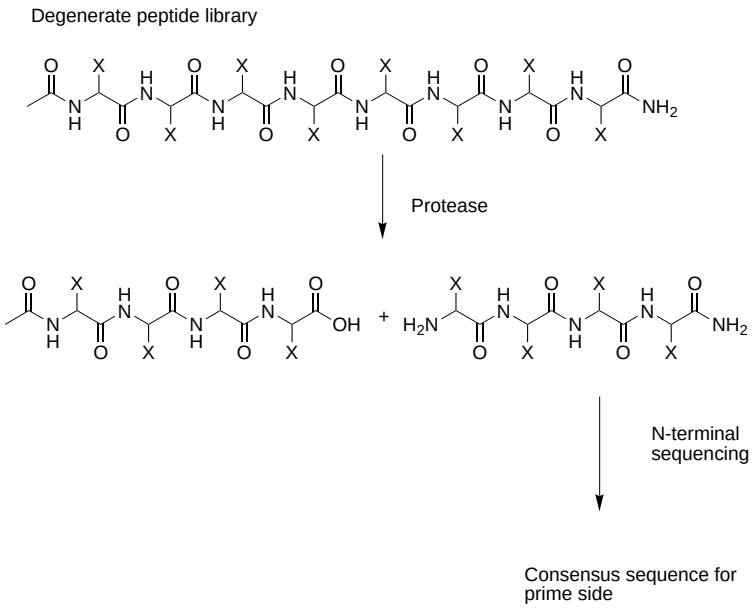
Scheme 9. *Peptide vinyl sulfones.*

degradation was applied to the aspartyl protease cathepsin E.[33] A pool of 2×10^{14} peptides with the general structure Ac-G-A-X_1-X_2-X_3-X_4-X_5-X_6-X_7-X_8-X_9-X_{10}-A-G-NH_2, where X represents any one of 20 naturally occurring amino acids, was constructed and digested. Edman degradation analysis of the resultant cleaved pools revealed some specificity determinants for the prime side of cathepsin E. The identified determinants correlated with the limited consensus sequence information provided by the digestion of a known protein substrate of cathepsin E. The substrate specificity of cathepsin E for the $P_5 - P_1$ sites was evaluated by C-terminal peptide sequencing. However, analysis is complicated by the low efficiency of C-terminal sequencing.[34]

A similar strategy was utilized to determine the prime-side substrate specificities of the zinc metallopeptidases meprin A and B.[35] A peptide library of the general structure Ac-X_1-X_2-X_3-X_4-X_5-X_6-X_7-X_8-X_9-X_{10}-NH_2, where X represents any one of 19 amino acids, was constructed. This pool of 19^{10} acetylated dodecamers was incubated with meprin A and B until 5 – 10% of the pool was digested, and the mixture was subjected to N-terminal sequencing. This work suggested that meprin B prefers acidic residues at the P_1' position, whereas meprin A does not (Figure 4). Furthermore, these two highly related proteases were shown to have different substrate specificities at the P_2' site. Very recently, this strategy was used to determine the prime-side substrate specificities of six enzymes in the MMP family.[36] By utilizing the determined prime-side specificity, information was obtained about the nonprime-side specificity of these six proteases through the use of an oriented peptide library.

4. Protease Inhibitor Design

The dominant strategy for the design of protease inhibitors is to utilize a mechanism-based pharmacophore as a key binding element. A variety of pharmacophores have been developed to target each of the four major protease classes, which are defined



Scheme 10. *Edman degradation of peptide pools.*

by their mechanism of protease-catalyzed amide bond hydrolysis. A secondary alcohol has been utilized as a pharmacophore when targeting aspartic acid proteases. Ketone and aldehyde pharmacophores have been used when targeting cysteine and serine proteases. Finally, zinc-binding groups, such as hydroxamic acids, thiols, and carboxylic acids, have been used when targeting matrix metalloproteases.

The identification of cell-permeable inhibitors is a key challenge in library design. Several factors should be considered to achieve this goal, including low molecular weights and the minimization of the peptidic nature of the inhibitor structure.[*] In this review we will emphasize libraries focused on these goals.

4.1. Aspartic acid proteases

Aspartyl proteases are a ubiquitous class of enzymes that play an important role in mammals, plants, fungi, parasites, and retroviruses.[38] The pharmaceutical industry has shown an intense interest in this class of enzymes since they are targets for therapeutic intervention in many important diseases. For example, HIV protease is involved in the development of AIDS, renin modulates hypertension, plasmepsin I and II are implicated in malaria, and cathepsin D and β-secretase have been identified as targets for the treatment of Alzheimer's disease (AD).

The aspartyl proteases are endopeptidases that use two aspartic acid residues to catalyze the hydrolysis of amide bonds (Scheme 11). Potent inhibitors of the aspartyl proteases have been developed that utilize a secondary alcohol, which is a stable mimetic of the tetrahedral intermediate and serves as the minimal pharmacophore. Indeed, clinically approved HIV-1 protease inhibitors are based upon hydroxyethylamine and statine-based inhibitors (Scheme 12).[39]

4.1.1. Libraries based upon the hydroxyethylamine inhibitors

Libraries of potential hydroxyethylamine-based inhibitors **1a** and **1b** have been synthesized by displaying functionality

[*] *Many properties are essential for a useful drug. These properties have been the subject of a number of excellent reviews.*[37a,b]

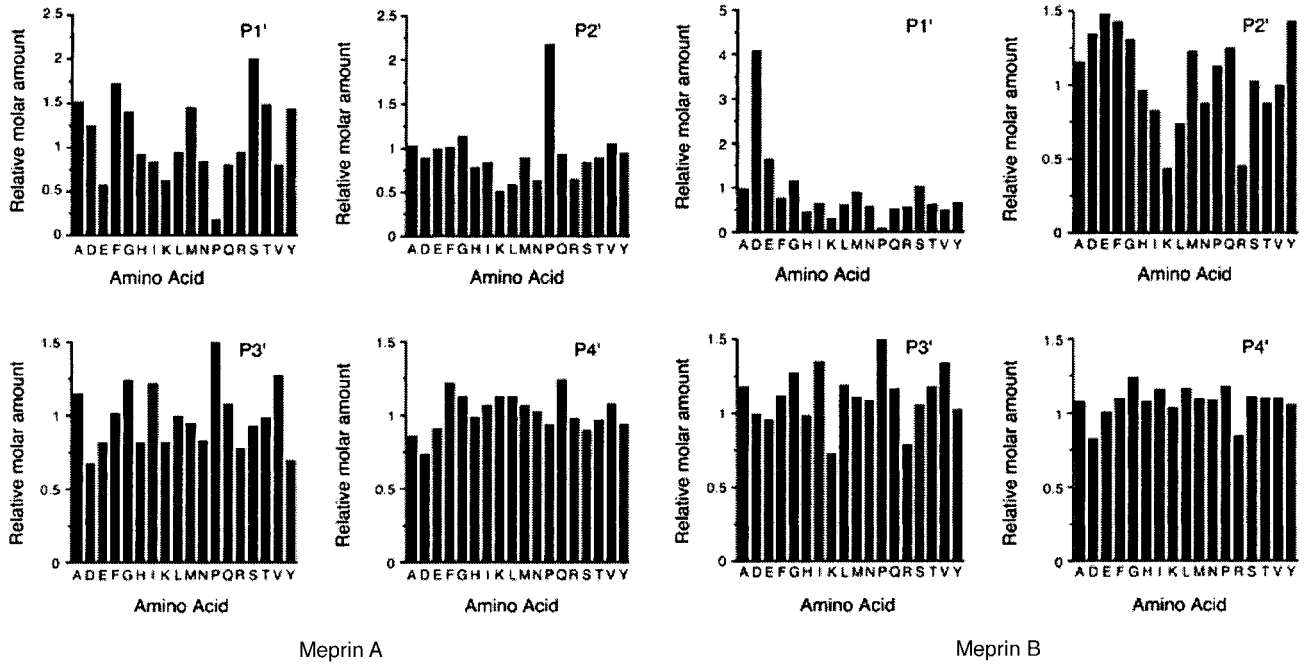
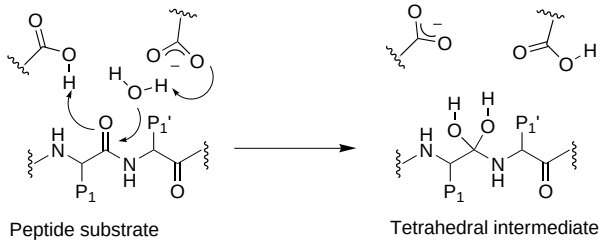
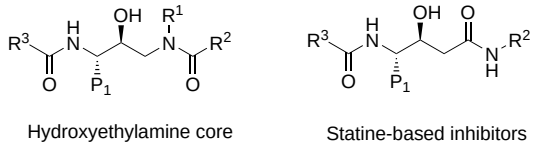


Figure 4. *Substrate specificities of Meprin A and B. Reproduced with permission from ref. [35]. Copyright (2001) The American Society for Biochemistry and Molecular Biology.*



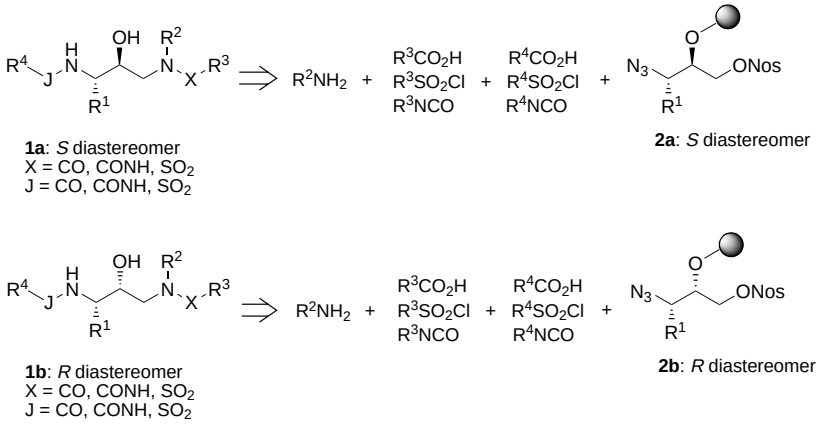
Scheme 11. *Peptide hydrolysis catalyzed by aspartyl proteases.*



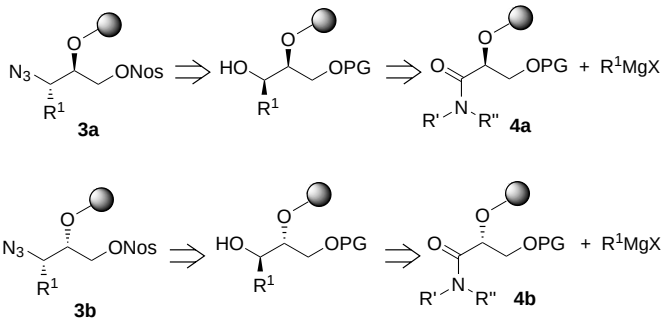
Scheme 12. *Peptide isosteres found in many potent aspartyl protease inhibitors.*

around a minimal scaffold **2a** and **2b** (Scheme 13).[40] Diverse functionality can be introduced at three sites on a preformed core that is linked to solid-support through the invariant secondary alcohol pharmacophore. Attachment at this site allows diversity to be placed on both sides of the inhibitor structure and protects the alcohol functionality throughout the synthesis sequence. Both secondary alcohol diastereomers **1a** and **1b** can be accessed by using either diastereomer **2a** or **2b** in this sequence. In an important extension of this synthesis method,[41] diverse R^1 substituents (P_1 side chains) are introduced by using Grignard reagents (Scheme 14). Either diastereomeric intermediate **3a** or **3b** may be accessed from the readily accessible support-bound amides of glyceric acid **4a** or **4b**, respectively.

These solid-phase synthesis methods have been used to prepare libraries that have resulted in the identification of potent inhibitors to a number of different aspartyl protease targets. Cathepsin D, which functions in protein metabolism and antigen processing, was the first protease targeted with this synthesis method. It has been implicated in a number of diseases, including cancer,[42] AD,[43a,b] and muscular dystrophy.[44] Initial



Scheme 13. *Retrosynthesis of hydroxyethylamine-based inhibitors.*

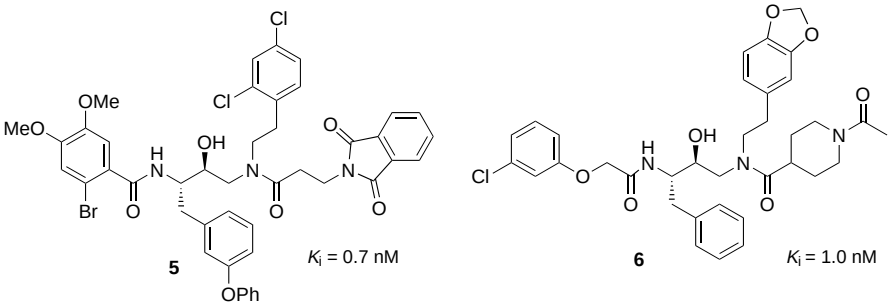


Scheme 14. *Introduction of R^1 diversity.*

inhibitors were identified by screening two libraries, each containing 1000 discreet compounds. For one of the libraries, building blocks were selected by using the modeling program DOCK; this provides one of the first examples of integrating combinatorial chemistry and structure-based methods.[45] Building blocks for the second library were selected by using diversity-based methods. Low nanomolar inhibitors of cathepsin D were identified from both of the libraries. The synthesis and screening of small, focused libraries then resulted in extremely potent, small-molecule inhibitors of cathepsin D with single-digit nanomolar to high picomolar K_i values (Scheme 15). Inhibitors were also optimized for desirable characteristics in addition to potency, including minimal binding to serum proteins and good calculated logP values.

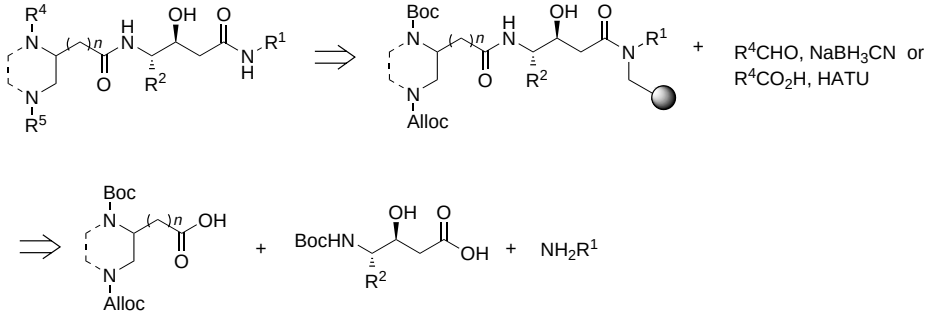
The identified cathepsin D inhibitors have served as important pharmacological tools for probing the enzyme's effect in human disease. For example, researchers have studied the relationship between cathepsin D and the presence of neurofibrillary tangles, one of the defining features of AD. Due to lysosomal dysfunction, elevated levels of cathepsin D in AD-vulnerable neurons is observed in advance of pathology.[46] Furthermore, in vitro, cathepsin D cleaves the tau protein at neutral pH values, which results in fragments corresponding in mass to those found in tangles.[47] This implicates cathepsin D as the protease responsible for tau protein processing. To test this hypothesis, the cathepsin D inhibitors were evaluated in a human hippocampal cell model of neurofibrillary tangle formation.[48a,b] At 1 μM inhibitor concentration, tau proteolysis was inhibited by 50%, with complete suppression at higher concentrations. These results support the hypothesis that cathepsin D links lysosomal dysfunction to the etiology of AD and suggests a new treatment approach.

By using the same library-synthesis methods, single-digit nanomolar inhibitors have also been identified against the malarial proteases plasmepsin I and II. The plasmepsins have been implicated as potential targets for the treatment of malaria, a parasitic disease that afflicts 300 – 500 million people worldwide, killing 1 – 2 million annually.[49] These inhibitors were optimized for drug-like characteristics, such as molecular weights below 600 amu, good calculated logP values, minimal binding to serum



Scheme 15. *Representative cathepsin D inhibitors from combinatorial libraries.*

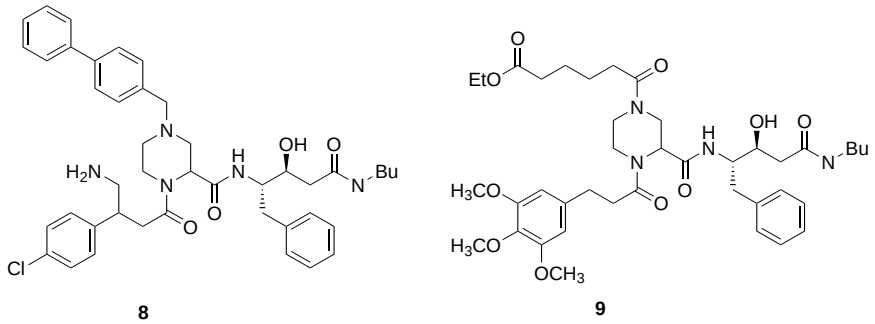
proteins, and greater than 15-fold selectivity over the most closely related human aspartyl protease, cathepsin D.[50] The most promising inhibitors have single-digit micromolar to submicromolar IC_{50} values against the parasite in cell culture (Scheme 16; IC_{50} = the concentration at which activity is inhibited by 50%). By using the same library methods, inhibitors have been identified towards β-secretase[51] and Yapsin A.[52]



Scheme 17. *Retrosynthesis of statine-based inhibitors.*

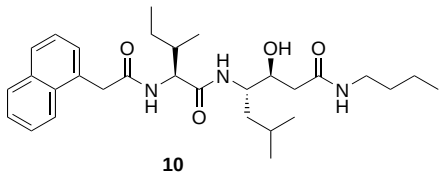
4.1.2. Libraries based upon the statine pharmacophore

In a series of reports, small-molecule inhibitors with low micromolar K_i values to plasmepsin II were identified by several library design, synthesis, and screening iterations.[53] By using a binary-encoded tagging strategy, a statine library of 18900 compounds containing cyclic diamino acids was prepared through the synthesis method outlined in Scheme 17. This scaffold was based on a docking exercise with the X-ray crystal structure of a cathepsin D-pepstatin complex. Screening of this library led to the identification of selective inhibitors for both plasmepsin II and cathepsin D (Scheme 18). One example of a selective compound is **8**, which has a K_i of 490 nM against plasmepsin II and a K_i of 45 μM against cathepsin D. Another potent compound, **9**, shows the opposite selectivity, displaying a K_i of ~100 μM against plasmepsin II and a K_i of 1.1 μM against cathepsin D.

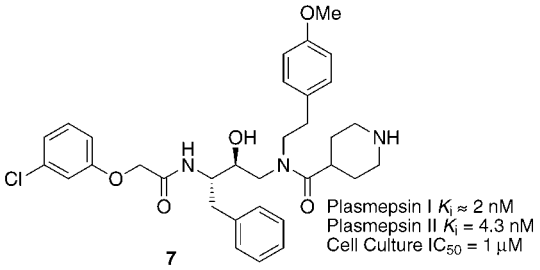


Scheme 18. *Selective inhibitors from the statine-based library.*

In a related report, a combinatorial library with 13020 compounds was synthesized by using the above synthesis scheme.[54] Potent low-nanomolar small-molecule inhibitors of plasmepsin I and II were identified, such as compound **10** (Scheme 19), which has a K_i of 50 nM against plasmepsin II and a K_i of 320 nM against cathepsin D.



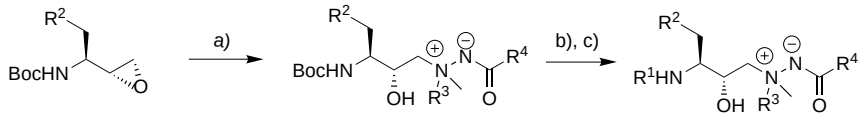
Scheme 19. *Selective plasmepsin II inhibitor.*



Scheme 16. *Potent plasmepsin inhibitor.*

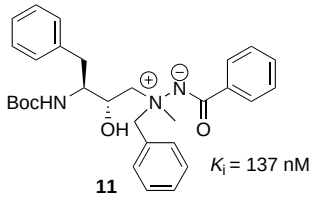
4.1.3. Libraries based upon aminimide peptide isosteres

A novel aminimide transition-state analogue has been identified and utilized as a general aspartyl protease pharmacophore.[55] This isostere was designed to mimic the placement of hydrogen-bonding atoms in the aspartyl protease active site, while being resistant to proteolysis. A library targeting HIV-1 protease was prepared by solution-phase methods in a one-step synthesis by



Scheme 20. *Synthesis of aminimide inhibitors. a) $R^3CH_3NH_2$, R^4CO_2Me; b) TFA; c) R^1X.*

using an ester, a hydrazine, and an epoxide (Scheme 20). Compound **11** (Scheme 21) was identified as a moderate inhibitor of HIV-1 protease, with a K_i value of 137 nM. The general applicability of the aminimide pharmacophore to the identification of other aspartyl protease inhibitors remains to be demonstrated.

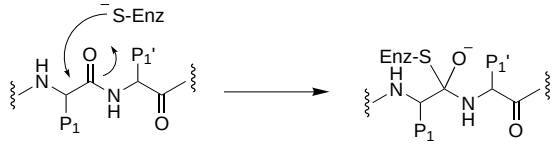


Scheme 21. *HIV-1 protease inhibitor.*

4.2. Cysteine proteases

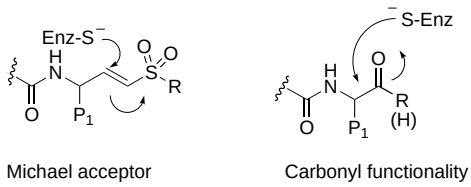
Cysteine proteases are essential to many biological processes. These proteases are characterized by an active-site cysteine thiol that attacks the carbonyl of an amide bond (Scheme 22).[56] This class of proteases includes the calpains, which have been implicated in neurodegenerative disorders, cathepsin K, which has been linked to osteoporosis,[57] and the caspase family of proteases, which is involved in programed cell death.[58]

A common feature of virtually all cysteine protease inhibitors is an electrophilic functionality, such as a carbonyl or Michael acceptor, which is attacked by the cysteine thiol (Scheme 23).[59]



Scheme 22. *Peptide hydrolysis catalyzed by cysteine proteases.*

The first class of reversible inhibitors to be reported was peptidyl aldehydes. However, the inherent reactivity of the aldehyde pharmacophore to nucleophilic attack and oxidation are considerable liabilities for achieving good pharmacokinetics. In addition, aldehyde-based inhibitors only allow display of functionality on one side of the

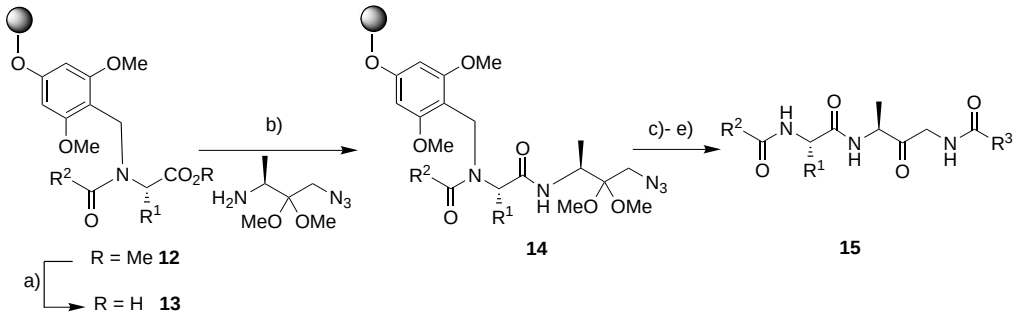


Scheme 23. *Cysteine protease pharmacophores.*

carbonyl. In contrast, ketone-based pharmacophores are chemically more stable, and enable the display of functionality on both sides of the carbonyl to potentially achieve enhanced specificity through multiple interactions with the active site.

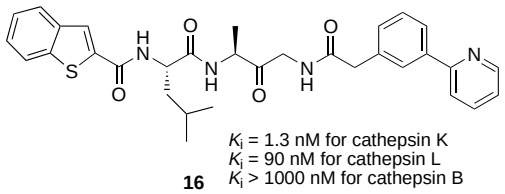
4.2.1. Amidomethyl ketone based libraries

Extensive efforts to develop inhibitors to cathepsin K have resulted in the identification of potent amidomethyl ketone based inhibitors that are active in animal models.[60] Recently, a solid-phase strategy for rapidly accessing this inhibitor class was reported (Scheme 24).[61] Amino acid methyl esters are loaded onto resin derivatized with the backbone amide linker (BAL) linker by reductive amination. A carboxylic acid is then coupled to the resulting secondary amine to provide a second diversity input. The methyl esters **12** are then hydrolyzed with potassium trimethylsilyloxide and a preformed masked ketone is then coupled to form an amide bond with the carboxylic acid. Reduction of the azide **14** is followed by acylation to introduce the third diversity input. Acid-mediated cleavage from support and concomitant hydrolysis of the ketal protecting group affords the desired ketone inhibitors **15**. Unfortunately, the ketone products **15** were prepared as a mixture of diastereomers due to approximately 20% epimerization upon cleavage. A small library was synthesized and screened against cathepsins K, L, and B.



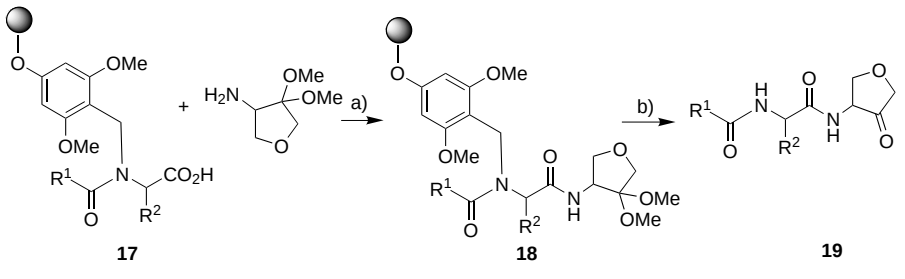
Scheme 24. *Solid-phase synthesis of amidomethyl ketones. a) $KOSiMe_3$, THF; b) EDC, NMP; c) $SnCl_2$, PhSH, Et_3N, THF; d) R_3CO_2H; e) TFA, Me_2S, H_2O.*

A representative potent inhibitor **16** towards cathepsin K that was also selective against cathepsins B and L is shown in Scheme 25.

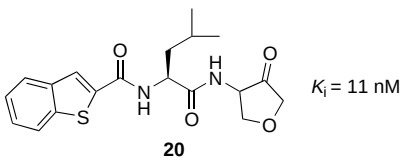


Scheme 25. *Potent and selective amidomethyl ketone inhibitor of cathepsin K.*

A related synthesis method has recently been reported to access cyclic alkoxy ketones structures. This class of inhibitors has also been identified to inhibit cathepsin K, as well as other cysteine proteases.[62] In this method, an α-amino-substituted cyclic ketal is coupled to the common carboxylic acid intermediate **17** (Scheme 26). Acid treatment releases the compound from support with concomitant hydrolysis of the ketal to afford the desired ketone product **19** as a mixture of two diastereomers in good yield and purity. By using this sequence, a library was prepared and screened against cathepsin K; this resulted in the identification of a number of potent inhibitors, as exemplified by **20** (Scheme 27).

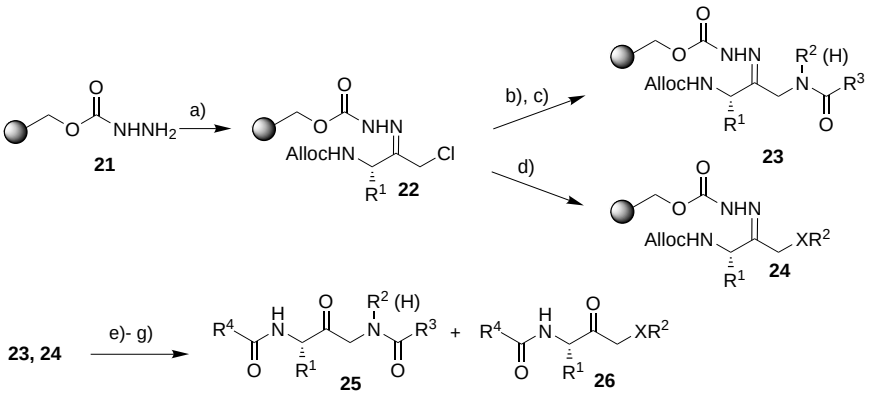


Scheme 26. *Synthesis of cyclic alkoxyketone inhibitors. a) EDC, DMF; b) TFA/CH_2Cl_2/H_2O (7:2:1).*



Scheme 27. *Potent cyclic alkoxyketone inhibitor of cathepsin K.*

tionality at several sites about a ketone carbonyl moiety has also been reported.[63] A chloromethyl ketone (Scheme 28) introduces the P_1 side chain and provides sites for further functionalization on both sides of the ketone carbonyl group. Linking to support through the ketone carbonyl moiety is ideal since it is the only invariant part of a ketone-based inhibitor regardless of the cysteine protease that is targeted. Attaching the ketone to the support as a carbazone further allows nucleophilic substitution at the α position while simultaneously preventing nucleophilic attack at the carbonyl group. Another advantage of the carbazone linkage is that it prevents racemization, which is problematic for the corresponding enolizable α-acylamino-substituted chiral ketone and the dimethoxy ketal strategy

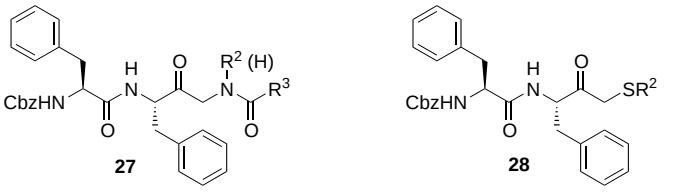


Scheme 28. *General method to prepare diverse ketone-based inhibitors. a) Chloromethyl ketone derivative of* N*-Alloc amino acid, THF; b) NH_2R^2, DMF or (1) NaN_3, MeOH; (2) $SnCl_2$, PhSH, iPr_2EtN, THF; c) R^3CO_2H, PyBOP, HOAt, iPr_2EtN, DMF; d) R^2CO_2H or R^2SH, iPr_2EtN, DMF; e) $Pd(PPh_3)_4$, CH_2Cl_2, $TMSN_3$, TBAF · 3 H_2O; f) R^4CO_2H, PyBOP, HOAt, iPr_2EtN, DMF; g) TFA/H_2O/TFE (1:4:15).*

4.2.2. Libraries of amidomethyl, mercaptomethyl, and acyloxymethyl ketones

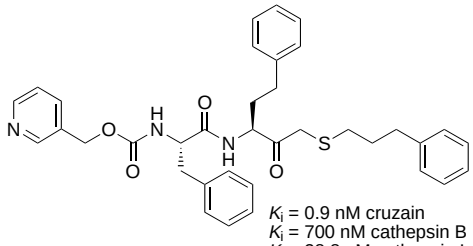
A general method for the solid-phase preparation of small-molecule cysteine protease inhibitors displaying diverse functionality at several sites about a ketone carbonyl moiety has also been reported.[63] A chloromethyl ketone (Scheme 28) introduces the P_1 side chain and provides sites for further functionalization on both sides of the ketone carbonyl group. Linking to support through the ketone carbonyl moiety is ideal since it is the only invariant part of a ketone-based inhibitor regardless of the cysteine protease that is targeted. Attaching the ketone to the support as a carbazone further allows nucleophilic substitution at the α position while simultaneously preventing nucleophilic attack at the carbonyl group. Another advantage of the carbazone linkage is that it prevents racemization, which is problematic for the corresponding enolizable α-acylamino-substituted chiral ketone and the dimethoxy ketal strategy mentioned above (Scheme 26). Nucleophilic displacement of the support-bound α-chlorohydrazones with amines followed by acylation provides access to the amidomethyl ketone class **25** of reversible cysteine protease inhibitors. Displacement with carboxylates or thiolates provides entry into the acyloxymethyl ketone **26** (X = O_2C) and mercaptomethyl ketone **26** (X = S) cysteine-protease inhibitors. Removal of the protecting group, acylation, and subsequent cleavage off of support results in the fully functionalized ketone products **25** and **26** in 40 – 100% yields, without racemization.

Initial library efforts were designed to identify potent inhibitors of cruzain, the essential cysteine protease in the parasite *Trypanosoma cruzi*. This parasite has been shown to be the causative agent in Chagas' disease, which is estimated to afflict over 60 million people in Central and South America and is the leading cause of heart disease in South America. The first generation libraries were based in part on Cbz-F-F (or homophenylalanine (hPhe)) fluoromethyl ketone derivatives, which have previously been identified as irreversible inhibitors of cruzain (Cbz = benzyloxycarbonyl). Libraries of both amidomethyl ketones **27** and mercaptomethyl ketones **28** (Scheme 29) were



Scheme 29. *Representative cysteine protease pharmacophores.*

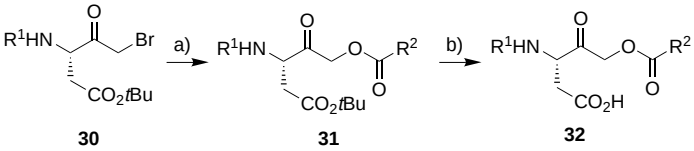
prepared and screened, which resulted in the identification of mercaptomethyl ketones with low nanomolar K_i values. Additional optimization libraries have further resulted in even more potent inhibitors with good aqueous solubilities and low molecular weights (477 – 610 amu). For example, compound **29** was identified as a very potent inhibitor of cruzain, with a K_i of 0.9 nM (Scheme 30). The compounds showed greater than 700-fold and 30-fold selectivity relative to two of the most closely homologous human proteases, cathepsins B and L, respectively.



Scheme 30. *Potent and selective cruzain inhibitor.*

An alternate, general, solid-phase route to amidomethyl ketones has also recently been reported. This strategy allows the display of a diverse set of functionality about the amidomethyl ketone core, although the method is limited in that P_1 side-chain diversity cannot be incorporated into the synthesis scheme.[64] The diamino propanol core, which is orthogonally protected with two different amine protecting groups, is linked to a tetrahydropyran (THP) linker in excellent yield (Scheme 31). Selective modification is possible through the removal of each orthogonal amine protecting group, followed by acylation with a variety of carboxylic acids and sulfonyl chlorides. TFA-mediated cleavage followed by oxidation of the secondary alcohol with Dess – Martin periodinane affords the desired ketones in 20 – 60% yields and > 90% purity.

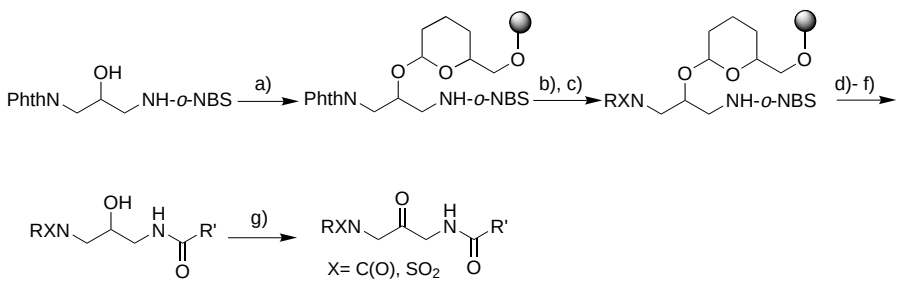
A solution-phase strategy to generate a library of 590 acyloxymethyl ketones has also been reported (Scheme 32).[65] In this synthetic scheme, the bromomethyl ketone **30** is coupled with 50 carboxylic acids, deprotected by HCl, and purified by using a support-bound thiourea resin to remove unreacted bromomethyl ketone. This library was screened against *N*-His (D381E) interleukin-1β converting enzyme, which resulted in a number of inhibitors with a greater than tenfold improvement in activity compared with a previously identified lead.



Scheme 32. *Alternative solution-phase synthesis of acyloxymethyl ketones. a) R^2CO_2H, Et_3N, DMF; b) HCl, THF; c) thiourea scavenger.*

4.3. Serine proteases

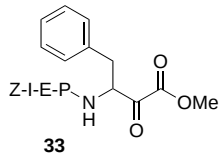
Serine proteases, which rely upon the attack of an active-site serine upon the scissile amide carbonyl, also play a large role in many biological processes.[38] Important pharmaceutical targets include thrombin and factor Xa, which have been implicated in homeostasis of blood coagulation, and human leukocyte elastase, which is important in lung diseases. Many peptidic libraries have been designed towards serine proteases.[66a–c] However, since peptide-based inhibitors generally have poor pharmacokinetics and bioavailability, more recent inhibitor-design efforts toward serine proteases have focused on the display of nonpeptidic functionality about an activated carbonyl group. The identification of inhibitors that show high selectivity for a single serine protease is a particularly difficult challenge because more than one hundred serine proteases are encoded by the human genome.[1] Furthermore, these enzymes tend to interact with only a small portion (3 – 5 amino acid residues) of the peptide and protein substrates.



Scheme 31. *Alternative solid-phase synthesis of amidomethyl ketones. a) THP linker, PPTS, 1,2-dichloroethane; b) NH_2NH_2, THF, DMF; c) RCO_2H, DIC, HOBt, DMF or RSO_2Cl, iPr_2EtN, 1,2-dichloroethane; d) β-mercaptoethanol, DBU, DMF; e) $R'CO_2H$, DIC, HOBt; f) TFA/H_2O (6:1); g) Dess – Martin periodinane.*

4.3.1. α-Ketoacid-derivative libraries

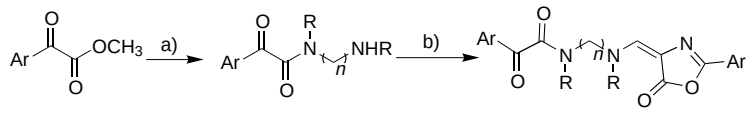
α-Ketoacid derivatives have served as an important pharmacophore for targeting serine proteases. Early reports described a library of peptidyl α-ketoamides that was targeted against human heart chymase and chymotrypsin.[67] Human heart chymase is important because of its role in hypertension. These inhibitors were synthesized in a spatially separate, parallel array and were screened against two targets by



Scheme 33. *Potent chymase inhibitor.*

using an on-bead assay. From this library, optimal amino acid sequences were identified. The corresponding α-ketoesters were potent inhibitors, such as **33** (Scheme 33), which has a K_i value of 1 nM against chymase.

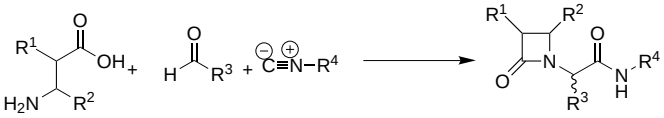
A convergent, parallel, solution-phase synthesis method to prepare nonpeptidic α-ketoamide derivatives has also been reported.[68] Reaction of α-ketoesters with symmetrical diamines afforded α-ketoamide products, which are then reacted with ethoxymethyleneoxazolones (Scheme 34). An array of 1600 compounds was prepared and screened against the serine proteases α-thrombin, factor Xa, trypsin, and plasmin. Several inhibitors with low micromolar K_i values were reported, but the structures of the active compounds were not disclosed. Further synthesis and biological evaluation will be needed to evaluate the versatility of this inhibitor synthesis strategy.



Scheme 34. *Solution-phase synthesis of α-ketoamide inhibitors. a) RHN(CH$_2$)$_n$NHR; b) ethoxymethyleneoxazolone.*

4.3.2. β-Lactam inhibitor libraries

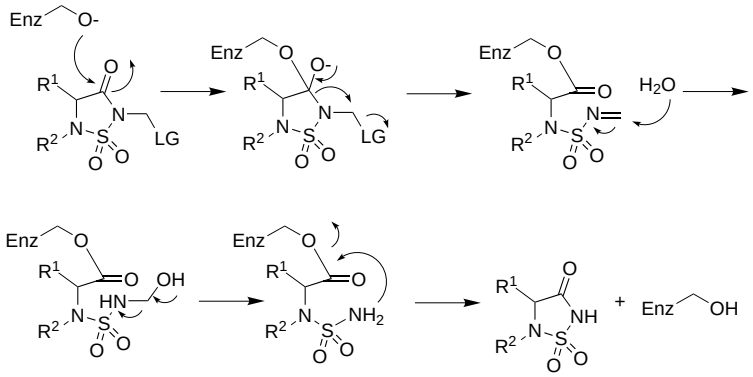
Another general serine protease pharmacophore is the β-lactam ring. A library of 126 β-lactam dipeptides was synthesized by employing an Ugi condensation strategy.[69] A wide variety of functionality was introduced with β-amino acids, aldehydes, and isonitriles (Scheme 35). However, an inherent limitation of this strategy was the formation of products as a mixture of diastereomers.



Scheme 35. *Synthesis of β-lactam inhibitors through an Ugi condensation.*

4.3.3. Libraries based on a 1,2,5-thiadiazolidin-3-one 1,1-dioxide scaffold

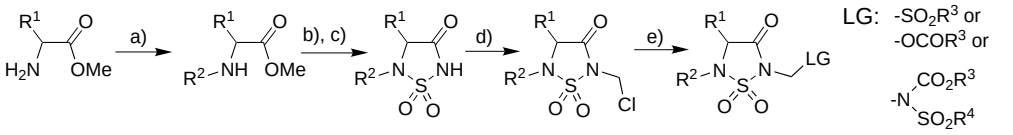
The structure-based design of a general class of heterocyclic mechanism-based inhibitors has been reported.[70] These inhibitors contain a sulfamide functionality and have been shown to be potent, time-dependent, highly selective inhibitors of human leukocyte elastase, cathepsin G, and proteinase 3. Selectivity can be predicted based upon the nature of the P_1 substitutent. The unique mechanism (Scheme 36) of interaction is postulated to occur through the nucleophilic addition of the serine alcohol to the activated amide carbonyl group, which results in the ring opening and formation of a covalent enzyme – inhibitor complex. Upon addition of water, formaldehyde is liberated and ring closure releases the enzyme and a heterocyclic product. From the commercially available amino acid methyl ester, these compounds can be synthesized in a five-step, solution-phase, parallel synthesis (Scheme 37). To demonstrate the generality of the inhibitors, a collection of 27 compounds was made and assayed against the enzymes. Second-order rate constants established the potency and selectivity of these inhibitors. For example, one compound had a K_{inact}/K_i value of 186 000 $M^{-1}s^{-1}$ against chymase, but was inactive against chymotrypsin.



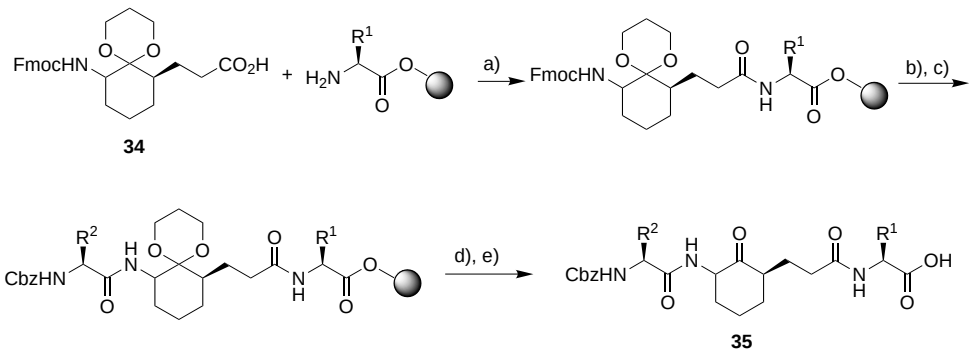
Scheme 36. *Mechanism of action.*

4.3.4. Cyclohexanone-based inhibitor libraries

In a recent report, cyclohexanone inhibitors **35** were prepared to target both serine and cysteine proteases in order to probe for specificity in the S and S′ binding sites.[71] Cyclohexanone ketal

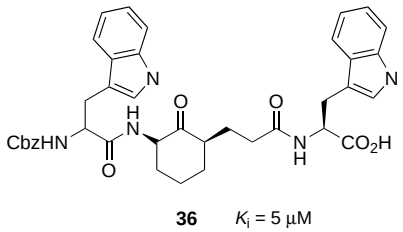


Scheme 37. *Synthesis of 1,2,5-thiadiazolidin-3-one 1,1-dioxide inhibitors. a) R^{2}CHO, NaBH(OAc)$_3$, 1,2-dichloroethane, AcOH; b) H$_2$NSO$_2$Cl, Et$_3$N; c) NaH, THF; d) HOCH$_2$SO$_3$Na, SOCl$_2$; e) HSR^3, DBU, then mCPBA; or NaI, acetone, then R^{3}CO$_2$H, DBU; or NaI, acetone, then R^{3}OCONHSO$_2$R^4, DBU.*



Scheme 38. *Synthesis of cyclohexanone inhibitors. a) HBTU, iPr_2EtN; b) piperidine, DMF; c) Cbz-aa-OH, HBTU, iPr_2EtN; d) TFA; e) TFA, H_2O.*

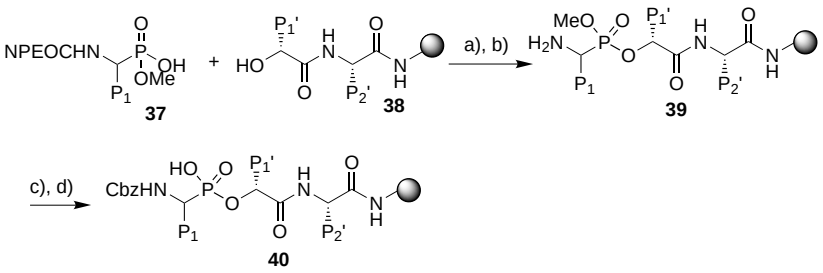
34 is utilized as the key pharmacophore building block in standard peptide synthesis (Scheme 38). A library of 400 compounds was prepared and screened against plasmin, cathepsin B, and papain. One of the most potent inhibitors identified in this library is **36** (Scheme 39), which has a K_i value of 5 μM against plasmin.



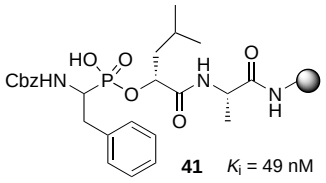
Scheme 39. *Cyclohexanone-based inhibitor of plasmin.*

phonic acid is followed by a TFA-mediated cleavage, which also simultaneously removes side-chain protecting groups. A 540-member library was constructed by using split and mix synthesis and was screened against thermolysin through an on-bead assay. Iterative deconvolution was used to determine the identity of active compounds such as **41** ($K_i = 49$ nM; Scheme 41).

Potent and selective phosphinic peptide inhibitors have also been



Scheme 40. *Synthesis of phosphonate inhibitors. a) Tris(4-chlorophenyl)phosphine, DIAD, iPr_2EtN, THF; b) 5% DBU/NMP; c) CbzCl, iPr_2EtN, dioxane; d) 1. PhSH:Et_3N:dioxane (1:2:2); 2. triethylsilane, TFA.*



Scheme 41. *Potent phosphonic acid based inhibitor of thermolysin.*

4.4. Metalloproteases

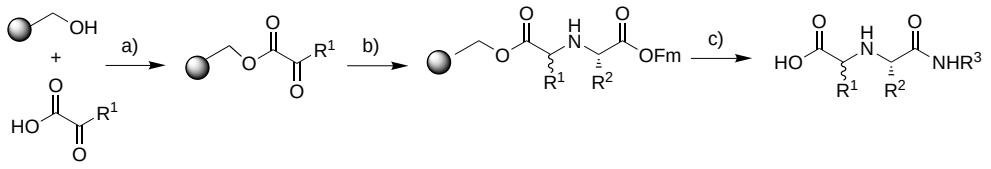
The fourth class of proteases catalyzes peptide-bond hydrolysis by an active-site, zinc-mediated activation of the substrate amide carbonyl group to attack by water.[72] Improper regulation of metalloproteases contributes to disorders such as rheumatoid arthritis, osteoporosis, and cancer. Almost all inhibitors of metalloproteases contain a zinc-binding group (ZBG). A number of different ZBGs have been discovered, such as thiols, carboxylic acids, and hydroxamic acids.

4.4.1. Phosphonate- and phosphinic-based libraries

One of the first solid-phase strategies for synthesizing inhibitors of metalloproteases was based upon the display of peptidic functionality about a phosphonate pharmacophore.[73] Instead of binding to the active-site zinc center, the phosphonate group is an effective transition-state analogue that mimics the scissile amide bond. The phosphonate inhibitors are prepared by coupling a 2-(4-nitrophenyl)ethoxycarbonyl (NPEOC) protected alkylphosphonic acid **37** to a support-bound α-hydroxy acid **38** (Scheme 40). Selective demethylation of the protected phosphonic acid is followed by a TFA-mediated cleavage, which also simultaneously removes side-chain protecting groups. A 540-member library was constructed by using split and mix synthesis and was screened against thermolysin through an on-bead assay. Iterative deconvolution was used to determine the identity of active compounds such as **41** ($K_i = 49$ nM; Scheme 41).

Potent and selective phosphinic peptide inhibitors have also been reported.[74] A library of several hundred peptides having the general formula Z-(L,D)Fψ(PO_2CH_2)(L,D)X-X-X were synthesized as a mixture of four diastereomers and screened against rat brain zinc endopeptidases neurolysin (24 – 16) and thimet oligopeptidase (24 – 15). From this library, potent inhibitors were identified, such as Z-(L,D)Fψ(PO_2CH_2)(L,D)A-R-M, which has a K_i value of 70 pM for endopeptidase 24 – 15. In a related synthesis strategy, the same authors reported a series of phosphinic peptides,[75] which varied according to peptide size, position of the phosphinic bond, and capping of the N-terminal peptide. The most potent inhibitor in this library, P-L-Fψ(PO_2CH_2)G-P, has a K_i value of 4 nM for endopeptidase 24 – 16 and is 2000 times less potent for endopeptidase 24 – 15.

Another recent phosphinic acid strategy described a library of 165 000 peptide inhibitors, which were screened against MMP-12.[76] The library was designed as a one-bead/two-compounds library, with every bead containing the same quenched fluorogenic substrate and a different inhibitor. Potent inhibitors were identified, such as L-M-Y-Gψ(PO_2HCH_2)L-Y-M-P-G, which had a K_i value of 6 nM for MMP-12.

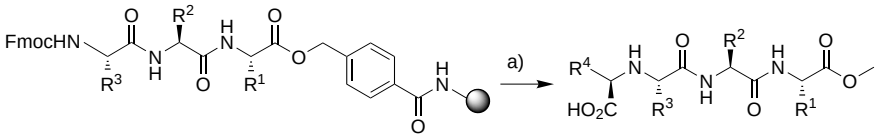


Scheme 42. *Synthesis of* N*-carboxyalkyl amino acid derivative libraries. a) HBTU, iPr_2EtN, DMF; b)* O*-fluorenylmethyl* α*-amino ester, $NaCNBH_3$, iPr_2EtN, AcOH, DMF; c) 1. piperidine, DMF; 2. HSpfp, DIC, DMF; 3. NH_2R^3, iPr_2EtN, HOBt, DMF; 4. HCl, dioxane.*

4.4.2. N-*Carboxyalkyl amino acid derivative libraries*

A targeted library of *N*-carboxyalkyl amino acid derivatives that utilizes a carboxylic acid ZBG has also been described.[77] Acylation of a support-bound hydroxy group with substituted pyruvic acids introduces the first diversity input (Scheme 42). Reductive amination with an *O*-fluorenylmethyl α-amino ester then provides the second diversity input. The final diversity input is introduced by removal of the *O*-fluorenylmethyl protecting group followed by coupling with a substituted amine. Acidic cleavage from the support affords the *N*-carboxyalkyl amino acid derivatives as a 1:1 mixture of epimers. By using this synthesis method, iterative libraries were designed through structure-based design, synthesized, and screened against stromelysin.

A related library of *N*-carboxyalkyl tripeptides was synthesized in order to generate new leads against metalloproteases (Scheme 43).[78] First, a support-bound tripeptide library was

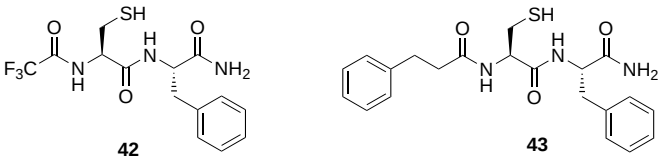


Scheme 43. *Synthesis of* N*-carboxyalkyl tripeptides. a) 1. Piperidine, DMF; 2.* α*-ketoacid, BH_3, pyridine; 3. TFA; 4. neutralize; 5. 10% Et_3N, MeOH.*

prepared by using standard solid-phase synthesis methods, and subsequently the carboxyalkyl functionality was introduced by reductive alkylation. This 20000-member library was prepared through both split and mix and indexed combinatorial techniques and was tested against a number of MMPs. In a representative example, these inhibitors were screened against MMP-3 as 100 mixtures each of 200 compounds, resulting in the identification of a 400 nM inhibitor of this enzyme.

4.4.3. *Thiol-based libraries*

Targeted libraries utilizing a thiol pharmacophore have been prepared by standard solid-phase peptide synthesis methods.[79] Screening of several iterative libraries resulted in the identification of selective inhibitors of both gelatinase A and gelatinase B (Scheme 44). Examples of selective compounds include **42**,



42

Gelatinase A IC_{50} = 40 nM
Gelatinase B IC_{50} = >1000 nM

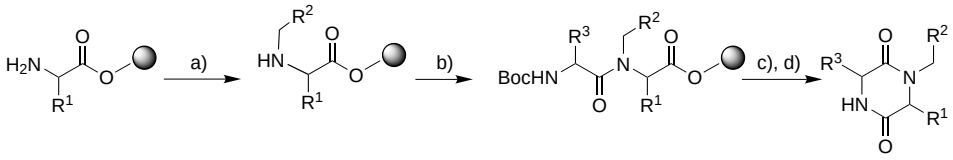
43

Gelatinase A IC_{50} = 3498 nM
Gelatinase B IC_{50} = 38 nM

Scheme 44. *Potent thiol-based MMP inhibitors.*

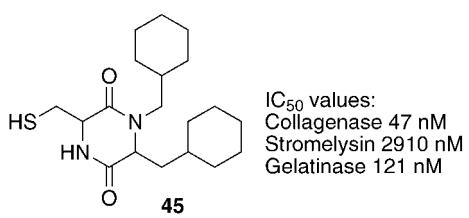
which has an IC_{50} value of 40 nM against gelatinase A and of >1000 nM against gelatinase B. Another potent compound **43** shows the opposite selectivity, displaying an IC_{50} value of 38 nM against gelatinase B and of ~3500 nM against gelatinase A.

In another published report, two libraries of diketopiperazine (DKP) inhibitors **44** with a thiol pharmacophore were synthesized and screened against collagenase-1, stromelysin 1, and gelatinase B.[80] Two 684-membered libraries were prepared using the synthesis sequence outlined in Scheme 45. These libraries were screened as mixtures and iterative deconvolution was used to determine the identity of potent and selective inhibitors, such as **45** (Scheme 46).



44

Scheme 45. *Synthesis of diketopiperazine inhibitors. a) R_2CHO, $NaBH_3CN$, MeOH; b) Boc-X-OH, HATU, iPr_2EtN, CH_2Cl_2; c) TFA; d) toluene.*



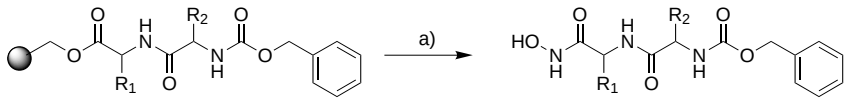
45

IC_{50} values:
Collagenase 47 nM
Stromelysin 2910 nM
Gelatinase 121 nM

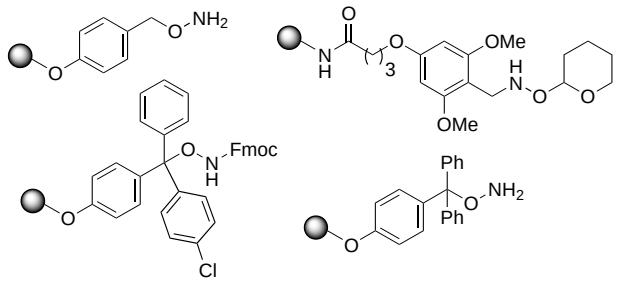
Scheme 46. *Potent and selective MMP inhibitor based upon the diketopiperazine scaffold.*

4.4.4. Hydroxamic acid based libraries

Finally, the hydroxamic acid functionality is one of the most extensively studied pharmacophores of metalloproteases and several solid-phase synthesis strategies for preparing hydroxamic acids have been described.[81a–f] Solid-phase synthesis is typically performed by using a hydroxylamine that is attached to the solid support through an acid-labile linker, such as those shown in Scheme 47.



Scheme 48. *Formation of hydroxamic acid from solid-support. a) 50% aq. NH_2OH (25 equiv), THF.*



Scheme 47. *Representative hydroxamic acid linkers.*

An alternative method for preparing hydroxamic acid inhibitors is to release the inhibitor from a support by nucleophilic attack of a hydroxylamine derivative upon the final support-bound intermediate (Scheme 48).[82] By using this strategy, several iterative libraries of hydroxamic acid dipeptides and tripeptides were prepared and screened to identify potent inhibitors of Procollagen C proteinase, a matrix metalloprotease that is responsible for procollagen processing.

5. Summary and Outlook

We have summarized the most current combinatorial techniques for the systematic study of proteases. First, combinatorial techniques have enabled the study of protease substrate specificity, which provides valuable insight into the elucidation of physiological substrates as well as in the design of good synthetic substrates and potent inhibitors. Second, small-molecule library synthesis methods have been developed to exploit a common pharmacophore to a protease class, such that they can potentially be applied to any member of that protease class. Small-molecule inhibitors that have been identified with these library methods have been applied to the in vitro and in vivo study of protease function and have served as leads for drug development.

Large-scale genomic and proteomic efforts over the last ten years have resulted in an enormous number of new biological targets that must be studied and understood. To tackle this huge number of targets the development of new techniques for the systematic analysis of proteins is necessary. Combinatorial methods are highly effective systematic tools because a large number of relevant chemical tools towards a protein family may rapidly be prepared by using common scaffolds or synthesis sequences designed to target that protein family. Targeted combinatorial approaches have been most widely utilized for the study of proteases and kinases,[83] but should be of general utility for most protein families.

Abbreviations

aa	amino acid
Abz	*o*-aminobenzamide
Ac	acetyl
ACC	7-amino-4-carbamoylmethylcoumarin
AD	Alzheimer's disease
All	allyl
Alloc	allyloxycarbonyl
AMC	7-amino-4-methylcoumarin
Bal	backbone amide linker
Boc	*tert*-butyloxycarbonyl
Cbz	benzyloxycarbonyl
cDNA	complementary DNA
Dabsyl	4-dimethylaminoazobenzene-4′-sulfonyl
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DKP	diketopiperazine
DIAD	diisopropyl azodicarboxylate
DIC	diisopropylcarbodiimde
DMF	*N*,*N*-dimethylformamide
EDC	1-(3-dimethylaminopropyl)-3-ethylcarbodiimide
Fm	fluorenylmethyl
Fmoc	9-fluorenylmethyloxycarbonyl
HATU	*O*-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate
HBTU	*O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxy-1*H*-benzotriazole
hPhe	homophenylalanine
HSpfp	pentafluorothiophenol
IC_{50}	the concentration at which activity is inhibited by 50%
LG	leaving group
mCPBA	*m*-chloroperoxybenzoic acid
MMP	matrix metalloproteinase
MT-SP1	membrane-type serine protease 1
o-NBS	*o*-nitrobenzenesulfonyl
NMM	*N*-methylmorpholine
NMP	*N*-methylpyrrolidone
Nos	*o*-nitrobenzenesulfonyl
NPEOC	2-(4-nitrophenyl)ethoxycarbonyl
PAR2	protease-activated receptor 2
PEGA	polyethylene glycol – poly(*N*,*N*-dimethylacrylamide) copolymer
PG	protecting group

Phth	phthaloyl
PPTS	pyridinium *p*-toluenesulfonate
PS-SCL	positional-scanning, synthetic combinatorial library
PyBOP	benzotriazol-1-yloxytris(pyrrolidino)phosphonium hexafluorophosphate
sc-uPA	single-chain urokinase-type plaminogen activator
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SPPS	solid-phase peptide synthesis
TBAF	tetrabutylammonium fluoride
TFA	trifluoroacetic acid
TFE	trifluoroethanol
THF	tetrahydrofuran
THP	tetrahydropyran
TMS	trimethylsilyl
$Y(NO_2)$	3-nitrotyrosine
ZBG	zinc-binding group

Protease work in this lab has been supported by the National Institutes of Health (GM54051).

[1] International Human Genome Sequencing Consortium, *Nature* **2001**, *409*, 860 – 921.
[2] J. C. Venter et al., *Science* **2001**, *291*, 1304 – 1351.
[3] I. Schechter, A. Berger, *Biochem. Biophys. Chem. Commun.* **1968**, *27*, 157 – 162.
[4] D. J. Matthews, J. A. Wells, *Science* **1993**, *260*, 1113 – 1117.
[5] M. M. Smith, L. Shi, M. Navre, *J. Biol. Chem.* **1995**, *270*, 6440 – 6449.
[6] A. Bevan, C. Brenner, R. S. Fuller, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10384 – 10389.
[7] S. Kothakota, T. Azuma, C. Reinhard, A. Klippel, J. Tang, K. Chu, T. J. McGarry, M. W. Kirschner, K. Koths, D. J. Kwiatkowski, *Science* **1997**, *278*, 294 – 298.
[8] C. G. Knight, *Methods Enzymol.* **1995**, *248*, 18 – 34.
[9] M. Meldal, I. Svendsen, K. Breddam, F. I. Auzanneau, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3314 – 3318.
[10] K. S. Lam, M. Lebl, *Methods Mol. Biol.* **1998**, *87*, 1 – 6.
[11] E. Del Nery, M. A. Juliano, M. Meldal, I. Svendsen, J. Scharfstein, A. Walmsley, L. Juliano, *Biochem. J.* **1997**, *323*, 427 – 433.
[12] P. M. St. Hilaire, L. C. Alves, S. J. Sanderson, J. C. Mottram, M. A. Juliano, L. Juliano, G. H. Coombs, M. Meldal, *ChemBioChem* **2000**, *1*, 115 – 122.
[13] M. Renil, M. Ferreras, J. M. Delaisse, N. T. Foged, M. Meldal, *J. Pept. Sci.* **1998**, *4*, 195 – 210.
[14] J. J. Peterson, C. F. Meares, *Bioconjugate Chem.* **1998**, *9*, 618 – 626.
[15] P. M. St. Hilaire, M. Willert, A. M. Juliano, L. Juliano, M. Meldal, *J. Comb. Chem.* **1999**, *1*, 509 – 523.
[16] G. Rosse, E. Kueng, M. G. P. Page, V. Schauer-Vukasinovic, T. Giller, H.-W. Lam, P. Hunziker, D. Schlatter, *J. Comb. Chem.* **2000**, *2*, 461 – 466.
[17] J. Singh, M. P. Allen, M. A. Ator, J. A. Gainor, D. A. Whipple, J. E. Solowiei, A. M. Treasurywala, B. A. Morgan, T. D. Gordon, D. A. Upson, *J. Med. Chem.* **1995**, *38*, 217 – 219.
[18] S. Leon, G. Lowe, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2997 – 3002.
[19] T. A. Rano, T. Timkey, E. P. Peterson, J. Rotonda, D. W. Nicholson, J. W. Becker, K. T. Chapman, N. A. Thornberry, *Chem. Biol.* **1997**, *4*, 149 – 155.
[20] 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, D. W. Nicholson, *J. Biol. Chem.* **1997**, *272*, 17907 – 17911.
[21] J. L. Harris, E. P. Peterson, D. Hudig, N. A. Thornberry, C. S. Craik, *J. Biol. Chem.* **1998**, *273*, 27364 – 27373.
[22] D. Lee, J. L. Adams, M. Brandt, W. E. DeWolf, Jr., P. M. Keller, M. A. Levy, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1667 – 1672.
[23] P. D. Edwards, R. C. Mauger, K. M. Cottrell, F. X. Morris, K. K. Pine, M. A. Sylvester, C. W. Scott, S. T. Furlong, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2291 – 2294.
[24] B. J. Backes, J. A. Ellman, *J. Org. Chem.* **1999**, *64*, 2322 – 2330.
[25] B. J. Backes, J. L. Harris, F. Leonetti, C. S. Craik, J. A. Ellman, *Nat. Biotechnol.* **2000**, *18*, 197 – 193.
[26] J. E. Sheppeck, H. Kar. J. B. Wheatley, E. Gjerstad, S. M. Loftus, A. R. Zubiria, J. W. Janc, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2639 – 2642.
[27] T. Takeuchi, J. L. Harris, W. Huang, K. W. Yan, S. R. Coughlin, C. S. Craik, *J. Biol. Chem.* **2000**, *275*, 26333 – 26342.
[28] J. L. Harris, B. J. Backes, F. Leonetti, S. Mahrus, J. A. Ellman, C. S. Craik, *Proc. Natl. Acad. Sci.* **2000**, *97*, 7754 – 7759.
[29] J. L. Harris, A. Niles, K. Burdick, M. Maffitt, B. J. Backes, J. A. Ellman, I. Kutz, M. Haak-Frendscho, C. S. Craik, *J. Biol. Chem.* **2001**, in press.
[30] T. Nazif, M. Bogyo, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2967 – 2972.
[31] J. R. Petithory, F. R. Masiarz, J. F. Kirsch, D. V. Santi, B. A. Malcolm, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 11510 – 11514.
[32] A. J. Birkett, D. F. Socer, R. L. Wolz, J. S. Bond, J. Wiseman, J. Berman, R. B. Harris, *Anal. Biochem.* **1991**, *196*, 137 – 143.
[33] D. Arnold, W. Keilholz, H. Schild, T. Dumrese, S. Stevanovic, H.-G. Rammensee, *Eur. J. Biochem.* **1997**, *249*, 171 – 179.
[34] T. Bergman, E. Cederlund, H. Jornvall, *Anal. Biochem.* **2001**, *290*, 74 – 82.
[35] G. P. Bertenshaw, B. E. Turk, S. J. Hubbard, G. L. Matters, J. B. Bylander, J. M. Crisman, L. C. Cantley, J. S. Bond, *J. Biol. Chem.* **2001**, *276*, 13248 – 13255.
[36] B. E. Turk, L. L. Huang, E. T. Piro, L. C. Cantley, *Nat. Biotechnol.* **2001**, *19*, 661 – 666.
[37] a) W. P. Walters, A. Murcko, M. A. Murcko, *Curr. Opin. Chem. Biol.* **1999**, *3*, 384 – 387; b) D. E. Clark, S. D. Pickett, *Drug Discovery Today* **2000**, *5*, 49 – 58.
[38] D. Leung, G. Abbenante, D. P. Fairlie, *J. Med. Chem.* **2000**, *43*, 305 – 341.
[39] J. P. Vacca, J. H. Condra, *Drug Discovery Today* **1997**, *2*, 261 – 272.
[40] E. K. Kick, J. A. Ellman, *J. Med. Chem.* **1995**, *38*, 1427 – 1430.
[41] C. E. Lee, E. K. Kick, J. A. Ellman, *J. Am. Chem. Soc.* **1998**, *120*, 9735 – 9748.
[42] B. R. Westley, F. E. B. May, *Eur. J. Cancer* **1996**, *32*, 15 – 24.
[43] a) U. S. Ladror, S. W. Snyder, G. T. Wang, T. F. Holzman, G. A. Krafft, *J. Biol. Chem.* **1994**, *269*, 18422 – 18428; b) A. M. Cataldo, D. J. Hamilton, J. L. Barnett, P. A. Paskevisch, R. A. Nixon, *J. Neurosci.* **1996**, *16*, 186 – 199.
[44] P. Gopalan, M. J. Dufresne, A. H. Warner, *Can. J. Physiol. Pharmacol.* **1987**, *65*, 124 – 129.
[45] E. K. Kick, D. C. Roe, A. G. Skillman, G. Liu, T. J. A. Ewing, Y. Sun, I. D. Kuntz, J. A. Ellman, *Chem. Biol.* **1997**, *4*, 297 – 307.
[46] J. C. Troncoso, A. M. Cataldo, R. A. Nixon, J. L. Barnett, M. K. Lee, F. Cheeler, D. R. Fowler, J. E. Smialek, B. Crain, L. J. Martin, *Ann. Neurol.* **1998**, *43*, 673 – 676.
[47] E. Bednarski, G. J. Lynch, *Neurochemistry* **1996**, *67*, 1846 – 1855.
[48] a) E. Bednarski, C. E. Ribak, G. J. Lynch, *Neurosci. Lett.* **1997**, *17*, 4006 – 4021; b) X. Bi, B. Lin, T. Haque, C. E. Lee, A. G. Skillman, I. D. Kuntz, J. A. Ellman, G. Lynch, *J. Neurochem.* **1999**, *74*, 1469 – 1477.
[49] D. Wyler, *J. Clin. Infect. Dis.* **1993**, *16*, 449 – 458.
[50] T. S. Haque, A. G. Skillman, C. E. Lee, H. Habashita, I. Y. Gluzman, T. J. A. Ewing, D. E. Goldberg, I. D. Kuntz, J. A. Ellman, *J. Med. Chem.* **1999**, *42*, 1428 – 1440.
[51] J. A. Ellman, X. Lin, unpublished results.
[52] J. A. Ellman, N. Cawley, unpublished results.
[53] C. D. Carroll, T. O. Johnson, S. Tao, G. Lauri, M. Orlowski, I. Y. Gluzman, D. E. Goldberg, R. E. Dolle, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3203 – 3206.
[54] C. D. Carroll, H. Patel, T. O. Johnson, T. Guo, M. Orlowski, Z.-M. Hie, C. L. Cavallaro, J. Guo, A. Oksman, I. Y. Gluzman, J. Connelly, D. Chelsky, D. E. Goldberg, R. E. Dolle, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2315 – 2320.
[55] E. E. Rutenber, F. McPhee, A. P. Kaplan, S. Gallion, J. C.Hogan, Jr., C. S. Craik, R. M. Stroud, *Bioorg. Med. Chem.* **1996**, *4*, 1545 – 1558.
[56] For a review, see: H.-H. Otto, T. Schirmeister, *Chem. Rev.* **1997**, *97*, 133 – 171.
[57] M. J. Bossard, T. A. Tomaszek, S. K. Thompson, B. Y. Amegadzie, C. R. Hanning, C. Jones, J. T. Kurdyla, D. E. McNulty, F. H. Drake, M. Gowen, M. A. Levy, *J. Biol. Chem.* **1996**, *271*, 12517 – 12524.
[58] N. A. Thornberry, *Chem. Biol.* **1998**, *5*, R97-R103.
[59] See also the symposium in print on cysteine protease inhibitors featured in: *Bioorg. Med. Chem.* **1999**, *7*, 571 – 644.
[60] R. W. Marquis, Y. Ru, S. M. LoCastro, J. Zeng, D. S. Yamashita, H.-J. Oh, K. F. Erhard, L. D. Davis, T. A. Tomaszek, D. Tew, K. Salyers, J. Proksch, K. Ward, B. Smith, M. Levy, M. D. Cummings, R. C. Haltiwanger, G. Trescher, B. Wang, M. E. Hemling, C. J. Quinn, H.-Y. Cheng, F. Lin, W. W. Smith, C. A. Janson, B. Zhao, M. S. McQueney, K. D'Alessio, C.-P. Lee, A. Marzulli, R. A. Dodds, S. Blake, S.-M. Hwang, I. E. James, C. J. Gress, B. R. Bradley, M. W. Lark, M. Gowen, D. F. Veber, *J. Med. Chem.* **2001**, *44*, 1380 – 1395.

[61] D. S. Yamashita, X. Dong, H.-J. Oh, C. S. Brook, T. A. Tomaszek, L. Szewczuk, D. G. Tew, D. F. Veber, *J. Comb. Chem.* **1999**, *1*, 207 – 215.
[62] A. E. Fenwick, B. Garnier, A. D. Gribble, R. J. Ife, A. D. Rawlings, J. Witherington, *Bioorg. Med. Chem. Lett*. **2001**, *11*, 195 – 198.
[63] A. Lee, L. Huang, J. A. Ellman, *J. Am. Chem. Soc*. **1999**, *121*, 9907 – 9914.
[64] C. Subramanyam, S.-P. Chang, *Tetrahedron Lett*. **2000**, *41*, 7145 – 7149.
[65] J. S. Warmus, T. R. Ryder, J. C. Hodges, R. M. Kennedy, K. D. Brady, *Bioorg. Med. Chem. Lett*. **1998**, *8*, 2309 – 2314.
[66] a) J. Eichler, R. A. Houghten, *Biochemistry* **1993**, *32*, 11 035 – 11 041; b) J. Eichler, A. W. Lucka, R. A. Houghten, *Pept. Res*. **1994**, *7*, 300 – 307; c) T. Carell, E. A. Wintner, A. J. Sutherland, J. Rebek, Jr., Y. M. Dunayevskiy, P. Vouros, *Chem. Biol.* **1995**, *2*, 171 – 183.
[67] M. Bastos, N. J. Maeji, R. H. Abeles, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6738 – 6742.
[68] C. M. Baldino, D. S. Casebier, J. Caserta, G. Slobodkin, C. Tu, D. L. Coffen, *Synlett* **1997**, 488 – 490.
[69] J. Pitlik, C. A. Townsend, *Bioorg. Med. Chem. Lett*. **1997**, *24*, 3129 – 3134.
[70] R. Kuang, J. B. Epp, S. Ruan, H. Yu, P. Huang, S. He, J. Tu, N. M. Schechter, J. Turbov, C. J. Froelich, W. C. Groutas, *J. Am. Chem. Soc*. **1999**, *121*, 8128 – 8129.
[71] P. Abato, J. L. Conroy, C. T. Seto, *J. Med. Chem*. **1999**, *42*, 4001 – 4009.
[72] For a review, see: M. Whittaker, C. D. Floyd, P. Brown, A. J. H. Gearing, *Chem. Rev.* **1999**, *99*, 2735 – 2776.
[73] D. A. Campbell, J. C. Bermak, T. S. Burkoth, D. V. Patel, *J. Am. Chem. Soc*. **1995**, *117*, 5381 – 5382.
[74] J. Jirácek, A. Yiotakis, B. Vincent, A. Lecoq, A. Nicolaou, F. Checler, V. Dive, *J. Biol. Chem*. **1995**, *270*, 21 701 – 21 706.
[75] J. Jirácek, A. Yiotakis, B. Vincent, F. Checler, V. Dive, *J. Biol. Chem*. **1996**, *271*, 19 606 – 19 611.
[76] J. Buchardt, C. B. Schiødt, C. Krog-Jensen, J.-M. Delaissé, N. T. Foged, M. Meldal, *J. Comb. Chem.* **2000**, *2*, 624 – 638.
[77] A. Rockwell, M. Melden, R. A. Copeland, K. Hardman, C. P. Decicco, W. F. DeGrado, *J. Am. Chem. Soc*. **1996**, *118*, 10 337 – 10 338.
[78] C. K. Esser, N. J. Kevin, N. A. Yates, K. T. Chapman, *Bioorg. Med. Chem. Lett*. **1997**, *7*, 2639 – 2644.
[79] M. A. Foley, A. S. Hassman, D. H. Drewry, D. G. Greer, C. D. Wagner, P. L. Feldman, J. Berman, D. M. Bickett, G. M. McGeehan, M. H. Lambert, M. Green, *Bioorg. Med. Chem. Lett*. **1996**, *6*, 1905 – 1910.
[80] A. K. Szardenings, D. Harris, S. Lam, L. Shi, D. Tien, Y. Wang, D. V. Patel, M. Navre, D. A. Campbell, *J. Med. Chem*. **1998**, *41*, 2194 – 2200.
[81] a) U. Bauer, W.-B. Ho, A. M. P. Koskinen, *Tetrahedron Lett*. **1997**, *41*, 7233 – 7236; b) L. S. Richter, M. C. Desai, *Tetrahedron Lett*. **1997**, *38*, 321 – 322; c) S. L. Mellor, C. McGuire, W. C. Chan, *Tetrahedron Lett*. **1997**, *38*, 3311 – 3314; d) K. Ngu, D. V. Patel, *J. Org. Chem*. **1997**, *62*, 7088 – 7089; e) J. J. Chen, A. F. Spatola, *Tetrahedron Lett*. **1997**, *38*, 1511 – 1514; f) C. D. Floyd, C. N. Lewis, S. R. Patel, M. Whittaker, *Tetrahedron Lett*. **1996**, *37*, 8045 – 8048.
[82] S. M. Dankwardt, R. J. Billedeau, L. K. Lawley, S. C. Abbot, R. L. Martin, C. S. Chan, H. E. Van Wart, K. A. M. Walker, *Bioorg. Med. Chem. Lett*. **2000**, *10*, 2513 – 2516.
[83] N. S. Gray, L. Wodicka, A. M. W. H. Thunnissen, T. C. Norman, S. J. Kwon, F. H. Espinoza, D. O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S. H. Kim, D. J. Lockhart, P. G. Schultz, *Science* **1998**, *281*, 533 – 538.

Received: June 18, 2001 [A 255]