

Identification of protease exosite-interacting peptides that enhance substrate cleavage kinetics

Abeer M. Jabaiah¹, Jennifer A. Getz¹,
Witold A. Witkowski², Jeanne A. Hardy²
and Patrick S. Daugherty^{1,*}

¹Department of Chemical Engineering, University of California, Santa Barbara, CA 93106, USA

²Department of Chemistry, University of Massachusetts, Amherst, 710 North Pleasant Street, Amherst, MA 01003-9336, USA

*Corresponding author.

e-mail: psd@engineering.ucsb.edu

Abstract

Many peptidases are thought to require non-active site interaction surfaces, or exosites, to recognize and cleave physiological substrates with high specificity and catalytic efficiency. However, the existence and function of protease exosites remain obscure owing to a lack of effective methods to identify and characterize exosite-interacting substrates. To address this need, we modified the cellular libraries of peptide substrates (CLiPS) methodology to enable the discovery of exosite-interacting peptide ligands. Invariant cleavage motifs recognized by the active sites of thrombin and caspase-7 were displayed on the outer surface of bacteria adjacent to a candidate exosite-interacting peptide. Exosite peptide libraries were then screened for ligands that accelerate cleavage of the active site recognition motif using two-color flow cytometry. Exosite CLiPS (eCLiPS) identified exosite-binding peptides for thrombin that were highly similar to a critical exosite interaction motif in the thrombin substrate, protease-activated receptor 1. Protease activity probes incorporating exosite-binding peptides were cleaved ten-fold faster than substrates without exosite ligands, increasing their sensitivity to thrombin activity *in vitro*. For comparison, screening with caspase-7 yielded peptides that modestly enhanced (two-fold) substrate cleavage rates. The eCLiPS method provides a new tool to profile the ligand specificity of protease exosites and to develop improved substrates.

Keywords: CLiPS; exosite; protease; substrate; thrombin.

Introduction

The ability of proteases to discriminate between many potential substrates and acquire high selectivity towards specific proteins is critical for regulating and localizing protein activity in diverse physiological processes. Misregulation of the activity of proteases has been implicated

in oncologic, inflammatory, neurodegenerative, and infectious diseases (Stamenkovic, 2000; Borgeño et al., 2004; Ala-aho and Kahari, 2005; Yasuda et al., 2005; Burrage et al., 2006; Lutgens et al., 2007; Tenore and Ferreira, 2009). Therefore, characterization of both substrate specificity and exosite interactions can provide valuable insights to understand biological pathways in health and disease, and to develop therapeutic inhibitors, protease-activated therapeutics, and diagnostic imaging probes.

As protease function is largely defined by the physiological substrate repertoire, our limited understanding of protease substrate recognition mechanisms represents a major obstacle to understanding the biological functions of proteases. With few exceptions, the characterization of protease substrate recognition has been limited to defining the specificity determinants of the active site. However, many proteases utilize non-active site recognition surfaces to influence substrate selectivity and activity, including allosteric sites and exosites (Walker and Royston, 2002; Hardy et al., 2004). This view is supported by the typically large discrepancy in cleavage kinetics between active site recognition sequences, typically fewer than eight amino acids, and bona fide physiological substrates that are often cleaved orders of magnitude faster than these short peptide substrates (Cornille et al., 1997; Bock et al., 2007). Consequently, there remains a need for experimental strategies to identify protease exosite specificity determinants.

Methods to characterize protease substrate specificity typically involve the identification of short peptide substrates that interact with protease active sites. These methods include substrate phage display libraries, positional scanning peptide libraries, combinatorial fluorogenic substrate libraries, mixture-based oriented peptide libraries, and cellular libraries of peptide substrates (CLiPS) (Harris et al., 2000; Turk et al., 2001; Deperthes, 2002; Boulware and Daugherty, 2006; Schneider and Craik, 2009; Boulware et al., 2010). Using these methods, peptide substrates of five to ten amino acids have been identified for numerous proteases. However, typically these active site substrates are not highly selective and are cleaved by other proteases (Talanian et al., 1997; Turk et al., 2001). Thus far, the substrates identified using peptide library methods have not exploited non-active site surfaces, such as exosites, which could dramatically enhance substrate activity and selectivity.

Given the absence of techniques to profile protease exosites, a method was developed to identify exosite-interacting peptide ligands that enhance the cleavage kinetics of a known substrate. The CLiPS method was modified to enable identification of peptide ligands that increase the cleavage rate of a fixed active site recognition motif. In this system, a defined cleavable substrate is expressed on the *N*-terminus

of an *Escherichia coli* outer membrane display scaffold (eCPX) (Rice and Daugherty, 2008) and a random peptide library is expressed as a fusion to the scaffold's C-terminus. To demonstrate the utility of the exosite CLiPS (eCLiPS) approach, libraries were screened using human thrombin, which exhibits two distinct exosites separated from the active site, and caspase-7, which is not known to utilize a peptide-binding exosite. Thrombin is a serine protease that plays a central role in blood coagulation by modulating pro-coagulant and anticoagulant processes, including hydrolysis of fibrinogen into fibrin to promote clot formation. In addition to the active site, the specificity of thrombin toward physiological substrates is typically regulated by interactions with two surface epitopes termed exosite I and exosite II (Bode and Stubbs, 1993; Stubbs and Bode, 1993). Exosite I binds to fibrinogen and protease-activated receptor 1 (PAR1) (Verhamme et al., 2002; Huntington, 2005) and is targeted by the therapeutic inhibitors hirudin and bivalirudin. Exosite II binds to the heparin-antithrombin complex as well as to glycoprotein I α , a protein in the platelet receptor complex and a cofactor for PAR1 cleavage (Huntington, 2005). Interactions with exosites are usually required for efficient cleavage of thrombin substrates by accelerating or stabilizing the formation of the initial thrombin-substrate complex (Huntington, 2005).

In contrast to thrombin, the manner by which most proteases increase their activity and specificity towards short peptide substrates is unknown. For example, the executioner proteases caspase-7 and caspase-3 of the cell apoptotic program were originally thought to have redundant function because they recognize identical substrate motifs (DEVD \uparrow G). However, further investigation suggests that caspase-7 is more specific and the two proteases cleave distinct natural proteins (Walsh et al., 2008). Small molecule ligands that bind an allosteric site in a deep cavity at the dimer interface of caspase-7 can modulate the protease's activity (Hardy et al., 2004), but peptide-binding motifs that enhance the activity of caspase-7 have not been reported. Although a candidate exosite cavity has been identified on caspase-7, approximately 30 Å from the two active sites, it has not been shown to interact with any physiological substrates (Agniswamy et al., 2007). In the present study, we applied eCLiPS to both thrombin and caspase-7 to investigate their exosite ligand preferences. Our results demonstrate that the eCLiPS method can be applied to determine the specificity of protease exosites and may be useful in the development of more active and selective substrates for applications in basic research, therapeutics, and diagnostics.

Results

Development of an exosite library screening method using thrombin and PAR1

To validate the eCLiPS exosite library approach, thrombin was selected as a model protease. Linear exosite-binding sequences

within thrombin's preferred physiological substrate PAR1 have been shown to accelerate the rate of substrate hydrolysis, k_{cat}/K_m , by 325-fold over the active site recognition motif alone. This effect was mediated by a 90-fold reduced K_m value and an increase in k_{cat} by 3.7-fold, with the improved substrate binding contributing the most to the increased cleavage rate (Vu et al., 1991b). Given this substantial exosite-mediated enhancement, only picomolar thrombin concentrations are required for platelet activation (Vu et al., 1991a). The active site recognition motif (LDPR \uparrow SFL) and the exosite-binding region (LRNPNDKYEPFWEDEEK) of PAR1 were displayed on the outer surface of *E. coli* at the N- and C-terminus of the eCPX display scaffold, respectively (Rice and Daugherty, 2008). The extent of substrate conversion could be determined by labeling bacterial cells with fluorescent probes recognizing N- and C-terminal affinity tags. Specifically, N-terminal substrate cleavage was monitored with the red fluorescent probe streptavidin R-phycoerythrin (SA-PE) that binds to the intact N-terminus of the eCPX scaffold (Figure 1A). To ensure that the exosite ligand expressed on the C-terminus is accelerating substrate cleavage and not itself being cleaved, a distinct green fluorescent probe, *Alaj*GFP-Mona SH3, was used to verify that the C-terminal peptide remained intact. Using this labeling scheme, bacterial cells with cleaved substrates will exhibit primarily green fluorescence, while clones with intact peptide substrates will exhibit both red and green fluorescence.

Bacterial cells displaying the PAR1 substrate along with either the corresponding native exosite-binding sequence or a non-binding glycine-serine repeat peptide of equal length (GlySer control) were incubated with different concentrations of thrombin and the rate of substrate cleavage was measured using flow cytometry (Figure 1B). Following incubation with 1 nM thrombin for 60 min, cells displaying both the PAR1 substrate and exosite ligand exhibited a large decrease in red fluorescence, while significantly higher thrombin concentrations were necessary to achieve the same level of conversion with the GlySer control peptide. With 2 nM thrombin, the cells with the active site motif and PAR1 exosite ligand exhibited 90% substrate conversion. By contrast, cells displaying the active site recognition motif along with the GlySer peptide exhibited a slight decrease in red fluorescence with 2 nM thrombin but required 20 nM thrombin for significant substrate cleavage, a 10-fold higher concentration. To test the time evolution of substrate cleavage, cells displaying the active site recognition motif with either the PAR1 exosite ligand or the GlySer control were incubated with 250 pM thrombin and the extent of substrate cleavage after 30, 45, and 60 min was measured by flow cytometry (Supplementary Figure 1). Cells displaying the GlySer control peptide did not exhibit significant substrate cleavage at any of the time points. By contrast, cells co-displaying the PAR1 substrate and exosite ligand exhibited 70% substrate conversion after 60 min, demonstrating that the exosite ligand again enhanced the rate of substrate cleavage. These results demonstrate that the PAR1 exosite ligand substantially enhanced the cleavage kinetics of the substrate when compared with a GlySer control exosite peptide.

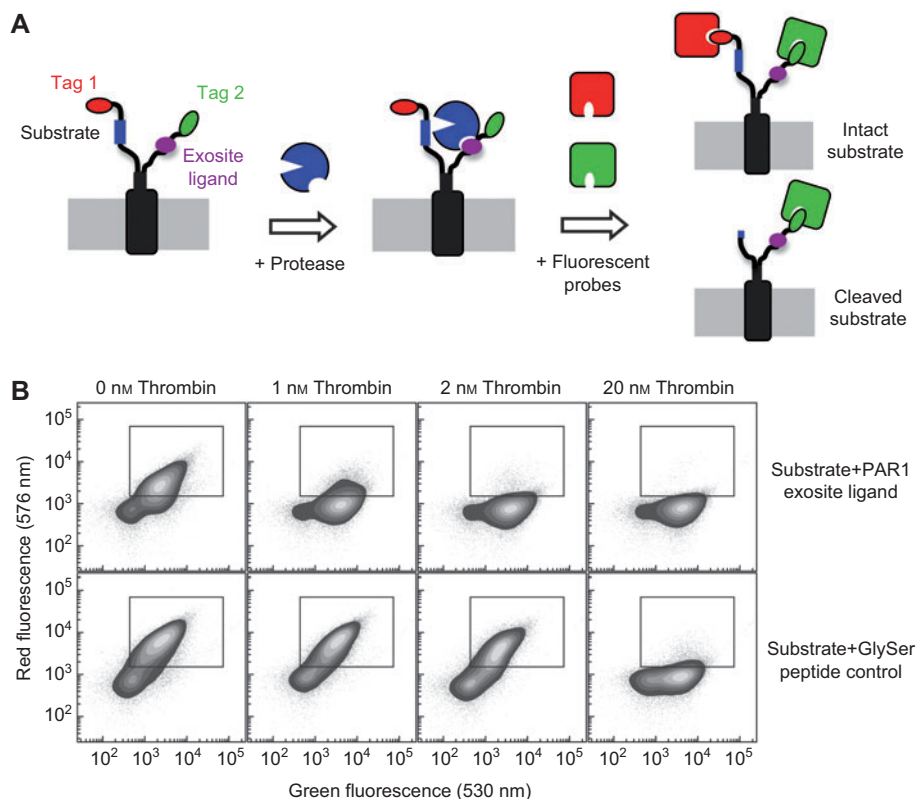


Figure 1 Method for the discovery of exosite ligands using eCLiPS.

(A) Using a fixed substrate recombinantly fused to the *N*-terminus of the bacterial display scaffold, a library of candidate exosite-binding peptides is constructed on the *C*-terminus and quantitatively screened for accelerated substrate cleavage using FACS based on a two-color labeling scheme. (B) Flow cytometric analysis of cell populations displaying the PAR1 substrate and either the PAR1 exosite ligand or the control GlySer peptide, following incubation with three different thrombin concentrations.

Screening exosite peptide libraries with thrombin and caspase-7

To identify exosite-interacting ligands for thrombin that enhance substrate hydrolysis rates of the PAR1-derived active site recognition sequence, an exosite ligand library displayed on the surface of *E. coli* was constructed. To co-display an active site binding sequence and a library of exosite recognition sequences, an invariant PAR1 substrate (LDPR \uparrow SFL) was fused to the *N*-terminus of eCPX, along with a library of 5×10^7 random 12-mer peptides on the *C*-terminus of eCPX. The *E. coli*-displayed exosite peptide library was screened for exosite ligands that accelerate cleavage of the invariant active site sequence using FACS. A two-step screening process was applied to first enrich library members with cleaved substrates after incubation with thrombin (500 μ M for 60 min) and labeling with the fluorescent probes SA-PE and *Alaj*GFP-Mona SH3. In a second screening step without the thrombin incubation, cells exhibiting both green and red fluorescence were sorted to ensure that full-length proteins were enriched. To increase the stringency, sorting was repeated using 125 μ M thrombin to isolate cells displaying ligands that enhance the hydrolysis rate of the thrombin substrate. The enriched library population was analyzed by DNA sequencing to determine the identities of exosite ligands that accelerated substrate

cleavage. Several peptides in the enriched pool exhibited a consensus $^Y/_F/_E/_D$ P[F/I/L]P that is highly similar to the core YEPF exosite-binding motif in PAR1 (Gandhi et al., 2010) (Table 1). Individual identified peptides with the highest cleavage rates also included an anionic residue downstream of the core motif. Interestingly, none of the peptides identified by library screening included a string of anionic residues similar to the PAR1 exosite ligand (EDEE).

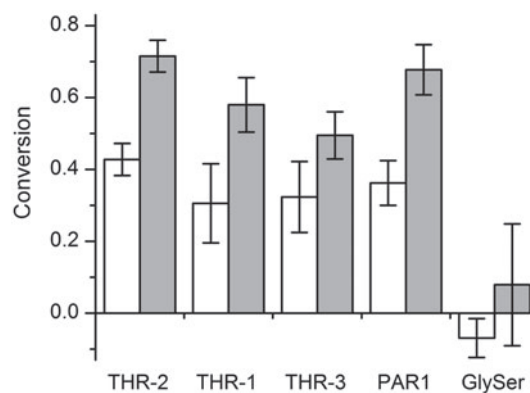
In an effort to identify caspase-7 exosite-interacting peptides, a library was similarly constructed wherein the active site recognition motif DEVD \uparrow G was fixed on the *N*-terminus of eCPX, and a library of candidate exosite ligands was displayed on the *C*-terminus upstream of the affinity tag. Unlike the case of the thrombin-PAR1 interaction, a known exosite ligand was not available for comparison with the enriched clones. The caspase-7 library was sorted using 1 μ M caspase-7 for the initial round followed by 100 nM caspase-7 in the later rounds, using alternating sorts for cleavage and substrate display. Individual isolated clones from the final rounds of screening encoded peptides enriched with basic amino acids including histidine, arginine, and lysine (Table 1), and several peptides exhibited a motif of the form HPxH. In addition, multiple exosite peptides had cysteines spaced by five to seven residues, and the two cysteines would likely form a disulfide bridge.

Table 1 Peptide sequences of potential exosite ligands from the bacterial display selections.

Thrombin exosite ligand	Sequence
PAR1	LRNPNDK YEPF WEDEEK
THR-1	AYERIPYMDSRG
THR-2	SQFEPILLHPN
THR-3	RFDPLPPDLTAH
THR-4	RFFPPTPSSALY
THR-5	NYSFSSHNHSLN
THR-6	PYYCQSFISHLM
THR-7	IFMHPITAPQLYT
THR-8	FRIPNPPRPLSV
THR-9	FLTRAPYRPLMN
THR-10	SPTRLYPHTNPP
Caspase-7 exosite ligand	Sequence
CSP-1	RIHTIHLKYPYG
CSP-2	GLLKKPHHCFMY
CSP-3	CLPKKPHHCFMY
CSP-4	NFYSHTHHVPHA
CSP-5	HPPASCLNYHPC
CSP-6	PLYWPSQLHPQH
CSP-7	LHSVHSILHTPL
CSP-8	NLLHHPSHPLPAFHHS
CSP-9	IVHPPHNSVYWN
CSP-10	HPPHPSDCVLRV
CSP-11	YRVWPTLIHPSH
CSP-12	PSHLLSPVHYLF
CSP-13	LLHHEPDHLILT
CSP-14	STMRTATPHGI
CSP-15	RHLCOHPLPFC
CSP-16	TLTRHHPMSTIY
CSP-17	LHPLYMPDLHRH

Measurement of substrate cleavage kinetics

To determine whether peptides present in the enriched libraries accelerated substrate cleavage, the rate of cleavage was measured first for cell-displayed substrates and then in the context of fluorescent protein FRET substrates. For thrombin, the rates of cleavage for three individual cell-displayed substrates, whose exosite sequences exhibited similarity with PAR1, were measured using flow cytometry. Similarly to the PAR1 exosite ligand, each of the exosite ligand-displaying clones accelerated cleavage of the active site recognition motif by six- to nine-fold over the GlySer control sequence with 250 μ M thrombin (Figure 2). Thrombin at a concentration of 125 μ M resulted in a cell-surface substrate conversion of 30–40%, when the exosite ligands were present. We next tested whether substrate-exosite ligand fusions, when reconstructed as contiguous linear FRET probes, would exhibit similar enhancement of hydrolysis rates. The soluble protease activity probes were constructed using fluorescent proteins optimized for FRET, CyPet and YPet (Nguyen and Daugherty, 2005). The FRET probes were constructed by fusing the PAR1 thrombin substrate to the 12-mer exosite ligands with an intervening 9-mer GGS repeat linker. Two positive control FRET probes were constructed wherein the PAR1 active site motif was directly fused to the PAR1 exosite ligand, or separated by a three-residue GGS linker from the PAR1 exosite ligand (Figure 3). As a non-functional exosite ligand control, the PAR1 active

**Figure 2** Measurement of the extent of conversion of bacterial clones with different exosite ligands.

The cells were incubated with 125 μ M (white bars) and 250 μ M thrombin (gray bars), and cellular red fluorescence was measured by flow cytometry in order to determine the extent of substrate conversion. The error bars are the standard deviation calculated from triplicate samples.

site motif was fused to the GlySer control peptide. The FRET probes were then used to measure second-order rate constants when cleaved by thrombin (Figure 3A). The exosite ligand THR-2 mediated the largest enhancement in hydrolysis rate (k_{cat}/K_m), which was roughly ten-fold greater than that of the substrate with the GlySer control peptide. The two other FRET substrates were cleaved five- to eight-fold faster than the control. Interestingly, adding the GGS linker between the PAR1 substrate and exosite ligand improved the hydrolysis rate of the substrate by 40%. These results demonstrate that thrombin exosite-binding peptides identified using eCLiPS substantially increase the rate of substrate hydrolysis.

In contrast with the thrombin exosite ligands, the caspase-7 exosite ligands only modestly enhanced cleavage rates in the context of the FRET probes (Figure 3B). The best performing exosite ligand, CSP-1, conferred a two-fold increase in hydrolysis rate over the GlySer control peptide. Overall, the second-order rate constants for the caspase-7 FRET probes were approximately 100-fold lower than those observed for the thrombin substrates. Caspase-7 is a cysteine protease that exhibits higher activity in the presence of the reducing agent dithiothreitol (DTT). Because the streptavidin-binding peptide in the eCLiPS system contains a disulfide bond, DTT was not included during the screening procedure with caspase-7. Adding 10 mM DTT along with the FRET probe in caspase-7 reactions increased the protease's activity and promoted substrate cleavage at a lower caspase-7 concentration (0.5 nM vs. 8 nM). Given the small enhancement mediated by caspase-7 selected exosite ligands, we then investigated whether adjusting the linker length between the substrate and the best exosite ligand, CSP-1, could enhance cleavage rates. The original glycine-serine linker (three GGS repeats) was shortened to a single GGS repeat and also lengthened to five or seven GGS repeats. Neither increasing nor decreasing the linker length significantly increased the cleavage kinetics for this substrate-exosite ligand pair with an approximately two- to three-fold

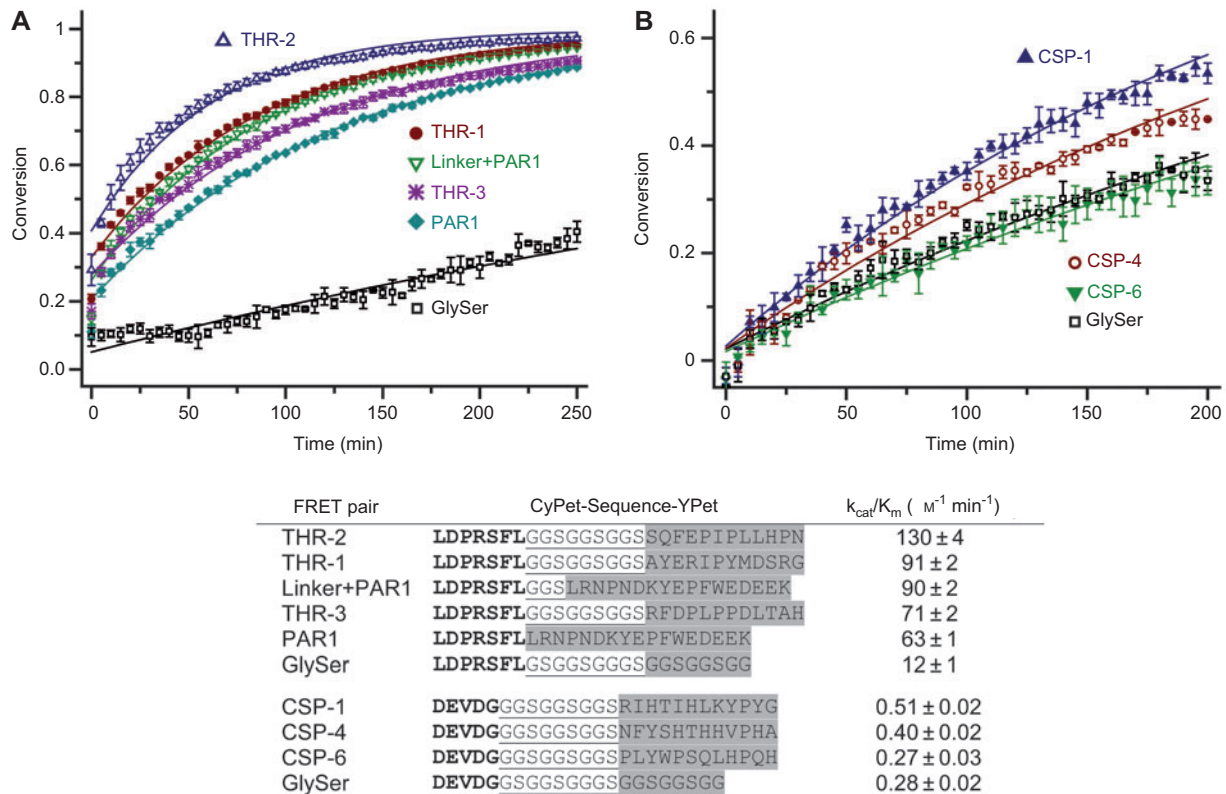


Figure 3 Determining the cleavage kinetics using FRET probes.

(A) The FRET probes with the PAR1 substrate attached to the exosite ligands were incubated with 125 μ M thrombin, and the conversion vs. time was fit (solid lines) to calculate the second-order rate constants (k_{cat}/K_m). The error bars are the standard deviation from triplicate reactions. (B) Similarly to the thrombin probes, the FRET probes with the caspase-7 substrate and exosite binders were incubated with 8 nM caspase-7. In the table, the sequences for the CyPet-YPet FRET pairs are listed with the substrate (bold), glycine-serine linker (underlined), and exosite or control ligand (highlighted in gray). The k_{cat}/K_m values are listed for each probe with the standard deviation of the value from triplicate reactions.

difference between the FRET probes with the CSP-1 peptide vs. the corresponding GlySer control. Unlike thrombin, screening with caspase-7 did not yield peptides that substantially increase substrate cleavage kinetics. These results suggest that caspase-7 may not use an exosite to efficiently cleave extended linear peptides in a manner analogous to that of thrombin.

Discussion

Here we have developed a bacterial surface display strategy to identify ligands for protease exosites and to develop substrates that exhibit higher catalytic turnover and are more selective for the target protease. Although peptide libraries are increasingly used to identify protease substrates, their application has been limited to characterizing the specificity of protease active sites (Harris et al., 2000; Turk et al., 2001; Deperthes, 2002; Boulware and Daugherty, 2006; Schneider and Craik, 2009). CLiPS and its predecessor, substrate phage display, are distinct from widely used positional scanning solution-phase libraries in that they enable screening or selection, respectively, for kinetically optimal substrates directly from very large libraries containing 10^8 – 10^9 members. These methodologies enable the identification of combinations of

amino acids that can exhibit cooperativity in protease binding. CLiPS differs from the substrate phage method in that many copies of the substrate ($\sim 10^4$) are displayed on the surface of *E. coli*, in contrast to the four to five copies displayed on phage when using the gpIII coat protein, such that screening is based upon average substrate conversion per cell. The use of cell-surface display in conjunction with flow cytometry also enables real-time optimization and fine tuning of the library screening stringency. In the present study, the identification of exosite ligands required the use of flow cytometry to determine both effective protease concentrations and reaction times, along with comparison of the cleavage kinetics of the library population vs. the control clones.

An important subtlety is that the eCLiPS system utilizes display on both the *N*- and *C*-termini of the display scaffold. Since the distance between the protease active site and a potential exosite is typically unknown, displaying the substrate and the candidate exosite ligand on different termini reduces spatial constraints to enable identification of ligands that bind exosites either close to, or far from, the active site. Furthermore, if both the substrate and potential exosite ligand were displayed on the *N*-terminus of eCPX, it would be difficult to differentiate candidate peptides that are cleaved from those that act as true exosite ligands. By displaying the

substrate on the *N*-terminus of the display scaffold and the potential exosite binder on the *C*-terminus, both termini can be labeled with different color fluorophores to enable sorting for bacterial clones with a cleaved substrate but intact *C*-terminal exosite binder. Also, display of the substrate and exosite ligand on separate termini may improve the display of certain substrates thereby simplifying library screening. The optimal strategy can be determined empirically by measuring display levels of the substrate along with a glycine-serine control sequence both on the *N*-terminus, both on the *C*-terminus, and on separate termini of the eCPX display scaffold.

While many experimental methods can identify preferred protease active site interaction motifs (e.g., P6-P5-P4-P3-P2-P1- \uparrow -P1'-P2'), it has become increasingly clear that physiological substrate recognition frequently involves additional important protease-substrate interactions. There are now multiple examples of proteases that cleave physiological protein substrates orders of magnitude faster than the best known active site interacting peptide substrate. Urokinase-type plasminogen activator, for example, can cleave the physiological substrate plasminogen more than 100-fold faster than the best known peptide substrates selected using substrate phage display (e.g., PFGR \uparrow SA) (Bergstrom et al., 2003). Although the precise mechanisms used by proteases to cleave physiological substrates so efficiently remain unclear, a subset of proteases can recognize an extended linear peptide substrate using multiple interacting exosites in a manner similar to thrombin. For example, the protease domain of botulinum neurotoxin A selectively and efficiently cleaves the substrate synaptosome-associated protein-25 kDa utilizing a remarkably extensive set of peptide-interacting exosites (Breidenbach and Brunger, 2004). In addition, a potent peptide inhibitor of the tissue factor-factor VIIa complex was discovered by constructing a *C*-terminal extension library on a known exosite-binding peptide and utilizing phage display to select for peptides that simultaneously bind the active site and exosite (Maun et al., 2003). Similarly, many other proteases may utilize one or

more exosites to achieve high selectivity for physiological substrates, and such mechanisms could dramatically restrict the repertoire of substrates that are efficiently cleaved. The characterization of a larger number of protease exosites using eCLiPS could help identify the defining features of peptide-interacting exosites to enable their prediction from structural data. Further application of eCLiPS will provide an improved understanding of substrate recognition specificity to better elucidate the functions of proteases.

The substrate specificity of thrombin has been investigated previously using several methods including substrate phage libraries and positional scanning fluorogenic substrate libraries that target the active site (Backes et al., 2000; Ohkubo et al., 2001). In addition to the active site, thrombin exosites I and II are necessary for enhancement of the specificity of thrombin for physiological substrates as shown for PAR1 (Figure 4A) (Vu et al., 1991b). Using eCLiPS, thrombin-binding peptides were identified that enhanced cleavage of the PAR1 thrombin substrate to a rate equivalent to, or greater than, the physiological thrombin exosite-binding peptide in the PAR1 protein. The commercially available peptide inhibitors of thrombin, bivalirudin, hirudin, and lepirudin (Warkentin, 2004), also utilize both the active site and exosite I to achieve high affinity and specificity. Using the exosite ligands identified here, we were able to construct highly sensitive thrombin-activity detection probes using fluorescent protein FRET probes (Boulware and Daugherty, 2006). The addition of exosite-binding peptides enhanced thrombin detection sensitivity by roughly ten-fold when compared with a conventional active site interacting substrate.

Caspase-7 was selected for exosite profiling because it has been shown previously to possess an allosteric site, a deep pocket adjacent to the active site, that can bind to small molecules to modulate activity (Figure 4B) (Hardy et al., 2004). Although screening via eCLiPS enriched peptides that accelerated catalysis, the effect was modest; substrates with the identified exosite ligands were cleaved two- to three-fold faster than

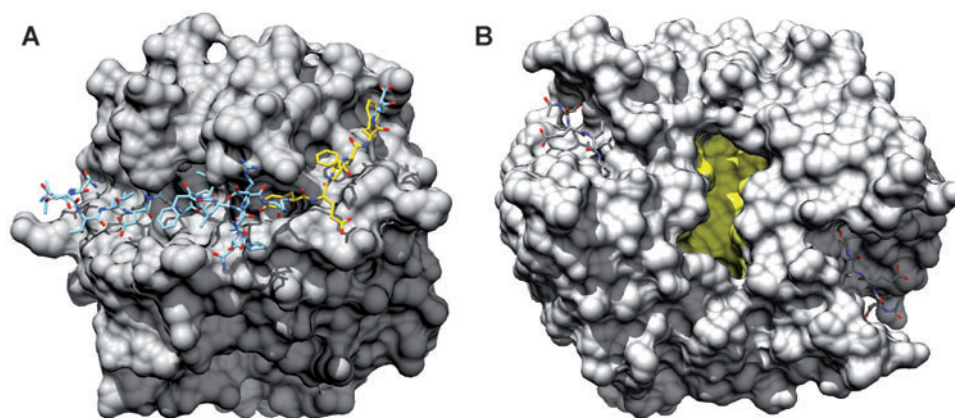


Figure 4 Location of the exosite on thrombin and allosteric site on caspase-7.

(A) An extended exosite groove in human thrombin binds to the soluble, extracellular fragment of PAR1 (YEPFW is highlighted in yellow). (B) The structure of the caspase-7 dimer has a deep pocket (highlighted in yellow) forming an allosteric site located within 10 Å of the active sites, which are bound to the inhibitor DEVD. The images shown were created using PDB:3LU9 (thrombin-PAR1) and PDB:1F1J (caspase-7) with the UCSF Chimera software package.

the control substrate for the most effective linker length. One potential explanation for this small effect could be a non-optimal spacing between or orientation of the active site motif and exosite ligand. Also, with the large diversity remaining in the caspase-7 final population, additional sorting with lower protease concentrations may have improved exosite ligand affinity and substrate cleavage kinetics. Alternatively, caspase-7 may utilize a larger protein interaction surface to recognize a folded polypeptide domain, rather than a linear peptide.

A drawback to the biterminal display strategy is that the length of linker between the substrate and exosite binder needs to be determined subsequently within the context of a contiguous linear substrate sequence. With the thrombin ligands, the sequence of the native PAR1 substrate was used to estimate the needed linker length between the two entities. For caspase-7, although different linker lengths were tested, neither shorter nor longer linkers substantially improved the cleavage kinetics. We cannot rule out the possibility that further manipulation of the linker length and composition, as well as the exosite ligand position (*N*- vs. *C*-terminal to the active site recognition motif) could improve the substrate cleavage. Nevertheless, the eCLiPS approach presented here can help to characterize the functions of protease exosites and identify peptide motifs that are critical to enhancement of protease specificity for physiological protein substrates. Finally, protease-interacting peptides, such as those discovered here, could provide dramatically increased sensitivity for protease-activity detection *in vitro* and *in vivo*.

Materials and methods

Reagents

Bacterial experiments were performed with *E. coli* strain MC1061 and grown in LB medium supplemented with 34 µg/ml chloramphenicol. Human α -thrombin was obtained from Haematologic Technologies, Inc. (Essex Junction, VT, USA). Streptavidin-R-phycoerythrin (SA-PE) was from Invitrogen (Carlsbad, CA, USA). Oligonucleotides were from Operon Biotechnologies (Huntsville, AL, USA), and restriction enzymes were from New England Biolabs (Ipswich, MA, USA).

Caspase-7 mutant generation, expression, and purification

Wild-type human caspase-7 was expressed from a two-plasmid expression system. A construct encoding residues 50–198 in the pRSET(Amp^R) plasmid expresses the large subunit. A second construct encoding residues 199–303 plus one codon for Q plus a six-Histidine tag in the plasmid pBB75 (Kan^R) plasmid (Batchelor et al. 1998) expresses the small subunit. The recombinant large and small subunits were co-expressed in the BL21(DE3) strain of *E. coli* in 2×YT media grown for 18 h after induction with 1 mM Isopropyl β -D-1-thiogalactopyranoside at an OD₆₀₀ of 0.6. Caspase-7 was purified using Ni-affinity liquid chromatography (HiTrap Chelating HP, GE Healthcare, Waukesha, WI, USA). After binding protein to the affinity column, the protein was eluted with a step gradient from 2 mM to 250 mM imidazole. Protein was diluted to 50 mM NaCl and then purified using a Macro-Prep High Q ion exchange column (Bio-Scale Mini 5 ml, Bio-Rad, Hercules, CA, USA) with a

linear gradient from 50 mM to 500 mM NaCl in 20 mM Tris buffer pH 8.0, with 2 mM DTT. Protein eluted in 120 mM NaCl and 20 mM Tris pH 8.0 was assessed for purity by SDS-PAGE to be 98+% pure and stored in elution buffer at -80°C.

Clone and library construction

For the eCLiPS system, the pBAD33 plasmid consisted of a streptavidin-binding peptide (WVCHPMWEVMCLR), linker (GGSGQSQGG), thrombin substrate from PAR1 (LDPRSFL) or a caspase substrate (DEVVDG), and another linker (GSSGGSGGSGG) on the *N*-terminus of eCPX; on the *C*-terminus of eCPX, a linker (GGSGGSGGSGG) was followed by the exosite-interacting ligand, an additional linker (GQQGGGSGGSGG), and a Mona-SH3 domain binding peptide (HISQWKPKVPNREDKYKK, P2X). For a positive and negative control, either a thrombin-exosite I binding peptide from PAR1 (LRNPNDKYEPFWEDEEK) or a glycine-serine peptide (GSGGSGGSGGSGGSGG) was inserted in the exosite ligand position. For the thrombin and caspase-7 libraries, a 12-mer randomized region was constructed using NNS codons. Detailed cloning methods on the display constructs and library construction are included in the Supplementary Information.

Optimizing screening conditions for cell displayed thrombin substrates

To identify the screening conditions for the cell displayed substrates, the overnight bacterial cultures of the clones (thrombin substrate and glycine-serine or PAR1 control exosite ligands) were subcultured 1:50 (v:v) for 2 h at 37°C in LB medium (pH 5.5) and 0.5% glycerol, then induced with 0.2% (w:v) L-(+)-arabinose and 1 mM EDTA for 3.5 h at 37°C. Lower pH media was necessary to improve the display of substrates containing anionic residues. Cells were pelleted, washed with reaction buffer (50 mM Tris-Cl, pH 7.5 supplemented with 20 mM NaCl, 2 mM CaCl₂ and 10 µM ZnCl₂), and pelleted again. The pelleted cells were then resuspended in 20 µl of reaction buffer and the volume was then divided into four 5 µl aliquots. The cells were incubated with thrombin concentrations of 20, 2, 1, and 0 nM in a total of volume of 10 µl. The reaction samples were incubated at 37°C with shaking for 1 h. The samples were then diluted 1:100 in phosphate-buffered saline (PBS pH 7.4) to impede the reaction, pelleted, and then resuspended in PBS supplemented with 50 nM SA-PE and 250 nM *Alaj*GFP-Mona SH3 to label the cells and were incubated at 4°C for 1 h. The labeled cells were then washed with PBS and analyzed on a FACSAria (BD Biosciences, San Jose, CA, USA) with 488 nm excitation and fluorescence detection at 530 nm and 576 nm. To also optimize the thrombin incubation time, the cells were incubated with 250 pM thrombin at 37°C and washed after 30, 45, and 60 min. Then the cells were labeled and analyzed as described previously.

Screening the thrombin and caspase-7 exosite libraries

The libraries were grown and expression induced using the same conditions as the control constructs. For the library sorting against thrombin, approximately 10⁸ cells were washed with reaction buffer and then pelleted. The pelleted cells were resuspended in 20 µl of reaction buffer and 10 µl were incubated either with 500 pM thrombin or no protease at 37°C with shaking for 1 h. The samples were washed with PBS and labeled with SA-PE and *Alaj*GFP-Mona SH3 with the same protocol as the individual clones. Using a BD FACSAria, the cleaved cells [green⁽⁺⁾ red⁽⁻⁾] were sorted. Following the initial round of sorting, eight additional sorts were performed. The rounds

alternated between thrombin-free isolation of cells that properly displayed substrates [green⁽⁺⁾ red⁽⁺⁾] and sorting for cells with cleaved substrates after incubating with 125 pM thrombin [green⁽⁺⁾ red⁽⁻⁾].

Sorting of the caspase-7 library for enhanced substrates was performed in a similar manner to the thrombin library. The caspase-7 substrate with the control glycine-serine ligand did not display efficiently and was therefore not used for optimizing the sorting conditions. The reaction buffer used with caspase-7 was 25 mM HEPES pH 7.4, 20 mM NaCl, and 2 mM CaCl₂. The library was sorted for a total of ten rounds, alternating between sorts for cleavage and sorts for display using 1 μM caspase-7 in the initial round and decreasing the concentration to 100 nM caspase-7 in the later rounds. During the sorting of both the thrombin and caspase-7 exosite libraries, approximately 5% of the library population was sorted for cleavage and 10% of the population exhibiting high red and green fluorescence were recovered during the display sorts. Cells from the final round of sorting were plated to isolate single clones for DNA sequencing. The plasmids from approximately 100 clones from the final round of sorting of both the thrombin and caspase-7 exosite libraries were sequenced using the standard pBAD reverse primer.

Cleavage kinetics of the cell displayed substrates

To measure the extent of conversion of cell displayed substrates, the overnight bacterial cultures of the clones from the final round of sorting were subcultured and induced as before in triplicate samples. Cells were washed and resuspended in reaction buffer (15 μl) and the volume was then divided in three (each 5 μl); the samples were incubated with 250, 125, and 0 pM thrombin for 1 h at 37°C. Cells were washed and labeled with 50 nM SA-PE for 1 h at 4°C. The labeled cells were again washed in PBS and their fluorescence was measured by flow cytometry. Conversion values were calculated using the following equation:

$$\text{Conversion} = \frac{FL_{-}FL_{+}}{FL_{-}FL_{0}}$$

where FL₋ is the red fluorescence of the bacterial cell population without the protease, FL₊ is the red fluorescence of the cells with protease, and FL₀ is the fluorescence of unlabeled cells (Boulware and Daugherty, 2006).

Measurement of thrombin and caspase-7 activity using FRET probes

To determine the second-order rate constants of the substrates in soluble form, CyPet-YPet FRET pairs were constructed with the substrate followed by a linker (GGSGSGGS) and the exosite ligand. The FRET pairs were expressed and purified as described previously (Jabaiah and Daugherty, 2011). The hydrolysis reaction for the thrombin substrates was performed in triplicate in 100 μl total volume with 80 nM FRET substrate and 125 pM of thrombin at 37°C for 250 min. With the caspase-7 FRET substrates, the reactions included 120 nM FRET substrate and 8 nM caspase-7 and were run at 37°C for 200 min. To monitor the cleavage of the peptides, the fluorescence emissions of the FRET probes (485 nm and 527 nm for thrombin substrates, 475 nm and 532 nm for caspase substrates) were measured every 5 min upon excitation at 433 nm using a Tecan Safire™ fluorescence spectrophotometer (Tecan, Durham, NC, USA). The second-order rate constants (k_{cat}/K_m) were then calculated by fitting the substrate conversion to the simplified Michaelis-Menten equation using non-linear regression:

$$\text{Conversion} = 1 - \exp\left(-\frac{k_{cat}}{K_m}[E]t\right)$$

where [E] is the protease concentration and t is time (Boulware and Daugherty, 2006).

Acknowledgments

This work was funded by grants 1R01GM097114-01A1 and 1R21CA159232-01. The authors also thank Amol Shivange for constructing the thrombin-PAR1 and caspase-7 images.

References

- Agniswamy, J., Fang, B., and Weber, I.T. (2007). Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis. *FEBS J.* 274, 4752–4765.
- Ala-aho, R. and Kahari, V. (2005). Collagenases in cancer. *Biochimie* 87, 273–286.
- Backes, B.J., Harris, J.L., Leonetti, F., Craik, C.S., and Ellman, J.A. (2000). Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin. *Nat. Biotechnol.* 18, 187–193.
- Bergstrom, R.C., Coombs, G.S., Ye, S., Madison, E.L., Goldsmith, E.J., and Corey, D.R. (2003). Binding of nonphysiological protein and peptide substrates to proteases: differences between urokinase-type plasminogen activator and trypsin and contributions to the evolution of regulated proteolysis. *Biochemistry* 42, 5395–5402.
- Bock, P.E., Panizzi, P., and Verhamme, I.M.A. (2007). Exosites in the substrate specificity of blood coagulation reactions. *J. Thromb. Haemost.* 5, 81–94.
- Bode, W. and Stubbs, M.T. (1993). Spatial structure of thrombin as a guide to its multiple sites of interaction. *Semin. Thromb. Hemost.* 19, 321–333.
- Borgoño, C.A., Michael, I.P., and Diamandis, E.P. (2004). Human tissue kallikreins: physiologic roles and applications in cancer. *Mol. Cancer Res.* 2, 257–280.
- Boulware, K.T., and Daugherty, P.S. (2006). Protease specificity determination by using cellular libraries of peptide substrates (CLiPS). *Proc. Natl. Acad. Sci. USA* 103, 7583–7588.
- Boulware, K.T., Jabaiah, A., and Daugherty, P.S. (2010). Evolutionary optimization of peptide substrates for proteases that exhibit rapid hydrolysis kinetics. *Biotechnol. Bioeng.* 106, 339–46.
- Breidenbach, M.A. and Brunger, A.T. (2004). Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* 432, 925–929.
- Burrage, P.S., Mix, K.S., and Brinckerhoff, C.E. (2006). Matrix metalloproteinases: role in arthritis. *Front. Biosci.* 11, 529–543.
- Cornille, F., Martin, L., Lenoir, C., Cussac, D., Roques, B.P., and Fournie-Zaluski, M.-C. (1997). Cooperative exosite-dependent cleavage of synaptobrevin by tetanus toxin light chain. *J. Biol. Chem.* 272, 3459–3464.
- Deperthes, D. (2002). Phage display substrate: a blind method for determining protease specificity. *Biol. Chem.* 383, 1107–1112.
- Gandhi, P.S., Chen, Z., and Di Cera, E. (2010). Crystal structure of thrombin bound to the uncleaved extracellular fragment of PAR1. *J. Biol. Chem.* 285, 15393–15398.
- Hardy, J.A., Lam, J., Nguyen, J.T., O'Brien, T., and Wells, J.A. (2004). Discovery of an allosteric site in the caspases. *Proc. Natl. Acad. Sci. USA* 101, 12461–12466.

- Harris, J.L., Backes, B.J., Leonetti, F., Mahrus, S., Ellman, J.A., and Craik, C.S. (2000). Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. USA* *97*, 7754–7759.
- Huntington, J.A. (2005). Molecular recognition mechanisms of thrombin. *J. Thromb. Haemostasis* *3*, 1861–1872.
- Jabaiah, A. and Daugherty, P.S. (2011). Directed evolution of protease beacons that enable sensitive detection of endogenous MT1-MMP activity in tumor cell lines. *Chem. Biol.* *18*, 392–401.
- Lutgens, S.P.M., Cleutjens, K.B.J.M., Daemen, M.J.A.P., and Heeneman, S. (2007). Cathepsin cysteine proteases in cardiovascular disease. *FASEB J.* *21*, 3029–3041.
- Maun, H.R., Eigenbrot, C., and Lazarus, R.A. (2003). Engineering exosite peptides for complete inhibition of factor VIIa using a protease switch with substrate phage. *J. Biol. Chem.* *278*, 21823–21830.
- Nguyen, A.W. and Daugherty, P.S. (2005). Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* *23*, 355–360.
- Ohkubo, S., Miyadera, K., Sugimoto, Y., Matsuo, K., Wierzb, K., and Yamada, Y. (2001). Substrate phage as a tool to identify novel substrate sequences of proteases. *Comb. Chem. High Throughput Screening* *4*, 573–583.
- Rice, J.J. and Daugherty, P.S. (2008). Directed evolution of a biterminal bacterial display scaffold enhances the display of diverse peptides. *Protein Eng. Des. Sel.* *21*, 435–442.
- Schneider, E.L. and Craik, C.S. (2009). Positional scanning synthetic combinatorial libraries for substrate profiling. *Methods Mol. Biol.* *539*, 59–78.
- Stamenkovic, I. (2000). Matrix metalloproteinases in tumor invasion and metastasis. *Semin. Cancer Biol.* *10*, 415–433.
- Stubbs, M.T. and Bode, W. (1993). A model for the specificity of fibrinogen cleavage by thrombin. *Semin. Thromb. Hemost.* *19*, 344–351.
- Talanian, R.V., Quinlan, C., Trautz, S., Hackett, M.C., Mankovich, J.A., Banach, D., Ghayur, T., Brady, K.D., and Wong, W.W. (1997). Substrate specificities of caspase family proteases. *J. Biol. Chem.* *272*, 9677–9682.
- Tenore, S.B. and Ferreira, P.R.A. (2009). The place of protease inhibitors in antiretroviral treatment. *Braz. J. Infect. Dis.* *13*, 371–374.
- Turk, B.E., Huang, L.L., Piro, E.T., and Cantley, L.C. (2001). Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol.* *19*, 661–667.
- Verhamme, I.M., Olson, S.T., Tollefsen, D.M., and Bock, P.E. (2002). Binding of exosite ligands to human thrombin. *J. Biol. Chem.* *277*, 6788–6798.
- Vu, T.K., Hung, D.T., Wheaton, V.I., and Coughlin, S.R. (1991a). Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* *64*, 1057–1068.
- Vu, T.K., Wheaton, V.I., Hung, D.T., Charo, I., and Coughlin, S.R. (1991b). Domains specifying thrombin-receptor interaction. *Nature* *353*, 674–677.
- Walker, C.P. and Royston, D. (2002). Thrombin generation and its inhibition: a review of the scientific basis and mechanism of action of anticoagulant therapies. *Br. J. Anaesth.* *88*, 848–863.
- Walsh, J.G., Cullen, S.P., Sheridan, C., Lüthi, A.U., Gerner, C., and Martin, S.J. (2008). Executioner caspase-3 and caspase-7 are functionally distinct proteases. *Proc. Natl. Acad. Sci. USA* *105*, 12815–12819.
- Warkentin, T.E. (2004). Bivalent direct thrombin inhibitors: hirudin and bivalirudin. *Best Pract. Res. Clin. Haematol.* *17*, 105–125.
- Yasuda, Y., Kaleta, J., and Brömme, D. (2005). The role of cathepsins in osteoporosis and arthritis: Rationale for the design of new therapeutics. *Adv. Drug Delivery Rev.* *57*, 973–993.

Received March 20, 2012; accepted May 6, 2012