AD_____

Award Number: DAMD17-99-1-9091

TITLE: Diagnosing Breast Cancer Using Protease Fingerprint

PRINCIPAL INVESTIGATOR: Emily Chen

CONTRACTING ORGANIZATION: The Burnham Institute
La Jolla, California 92037

REPORT DATE: June 2001

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2001	Annual Summary (1 Jun 00 - 31 May 01)

4. TITLE AND SUBTITLE
Diagnosing Breast Cancer Using Protease Fingerprint

5. FUNDING NUMBERS
DAMD17-99-1-9091

6. AUTHOR(S)
Emily Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Burnham Institute
La Jolla, California 92037

E-Mail: echen@burnham-inst.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

In the original fellowship proposal, I planed to profile protease activity using substrate phage display library in the biological samples from mice at different stages of breast cancer. I first used the substrate phage display library I constructed to characterize substrate recognition profiles of two tumor-related proteases, Metalloproteinase-2 (MMP-2) and 9 (MMP-9). Three groups of MMP-2 substrate sequences were novel and found highly selective for MMP-2 over MMP-9. This result supports the hypothesis that substrate phage display library can be used to differentiate diminutive structural difference of proteases. However, I found that phage display library had limited application for in vivo protease profiling because only small quantity of proteases present in the biological samples. Hence, I modified my approach of carrying out the functional study of disease-related proteases and developed a one-pot phage selection system that yield the substrate recognition profile of multiple purified proteases from a single round of selection. This method allows analysis of multiple proteases simultaneously, and prior knowledge of substrate preference is not required. As an illustration, a phage selection with a mixture of thrombin and factor Xa (serine proteases) along with MMP-9 and atrolysin C (metalloproteinases). Peptide substrates were identified that *i*) have high kcat/Km ration, *ii*) are selective for individual proteases, and *iii*) match the sequences of known physiological substrates. Ultimately, highly selective peptide substrates for disease-related proteases can be obtained, and an array of selective peptide substrates can be assembled and profile protease activity in the biological samples from mice at different stages of breast cancer.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	67
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction:

The purpose of my study is to obtain a collective knowledge of proteases which activity is differentially overexpressed during breast cancer progression using phage display technology. The profile of proteases or "protease fingerprint" will then be validated as a diagnostic tool for predicting breast cancer progression.

Body:

Proteases are key regulators of a wide range of physiological processes (1,2), and are recognized as important and tractable drug targets. The activity of proteases is tightly regulated by inhibitors, and some of them by protein activation. Many proteases such as matrix-metalloproteinase are synthesized as zymogen (inactive form) and are activated later by other proteases (3-5). Upon activation, protease activity is then regulated by either the non-specific inhibitors such as α1-macroglobulin or specific inhibitors such as tissue specific metalloproteinase inhibitors (TIMPs) for MMPs (6). Mistakes in the regulation of protease activity can lead to various pathological conditions such as cancer growth, tumor metastasis, inflammation, cardiovascular, and autoimmune diseases (7-11). Consequently, information of protease activity in diseases and progression of diseases will be important in discovering biomarkers and potential therapeutic targets. In this fellowship proposal, I hypothesized that a profile of protease activity can be used as a diagnostic tool to predict breast cancer progression. Three aims were designed in this fellowship proposal to test this hypothesis. So far, studies toward completing these three aims have lead to three publications that I am either listed as the first or the second author. The manuscripts are included in this annual summary repot.

Aim#1. Construct A Substrate Phage Display Library To Use As A Tool For Studying Protease Fingerprint.

I aimed in the fellowship proposal to profile protease activity in vivo using substrate phage display library. Construction of the substrate phage library was reported in the annual summary report last year. To demonstrate that this substrate phage display library can be used to distinguish catalytic activity of proteases, I used the substrate phage display library to distinguish the catalytic two cancer-related proteases, Metalloproteinase-2 (MMP-2) and 9 (MMP-9). MMP-2 and 9 belong to a subfamily of matrix metalloproteinases that is historically referred as gelatinases. They are highly homologous in their overall structure and catalytic domains. These two proteases have received great attention continuously because their involvement in tumor progression and angiogenesis (8,12-14) and are of interest as pharmaceutical targets. Up to now, no functional assay is available to differentiate their catalytic activity. Hence, the study of MMP-2 and MMP-9 substrate profile will demonstrate the power of substrate phage display library to discriminate highly similar proteases. Additionally, new understanding of MMP-2 and MMP-9 substrate recognition profile will provide information of distinct substrate recognition to aid the development of a functional assay that leads to uncover their functions in the normal and pathological processes.

Result 1: Four Groups Of MMP-2 Substrates Were Found Via Substrate Phage Selection.

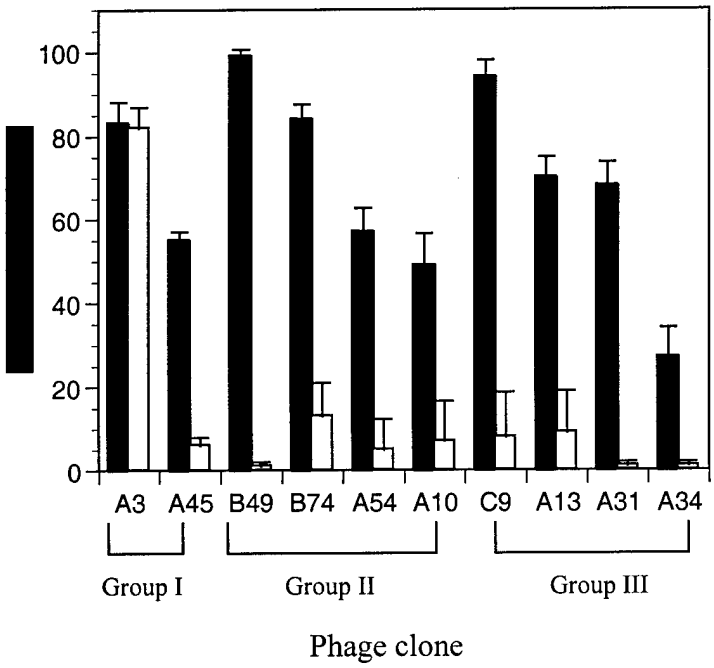
After two rounds of substrate selections of MMP-2, thirty individual clones were selected for sequencing based on the fact that they were cleaved by more than 25% when incubated for two hours

with activated MMP-2. Four distinct groups of substrates are found among these clones. The first group of substrate contains the P-X-X-X_{Hy} motif, where X_{Hy} represents a large hydrophobic residue. This motif appears to be a substrate for a number of different MMPs (15-17). Substrates in group II-IV represent novel recognition motifs for MMP-2. Group II substrates contain the I/L-X-X- X_{Hy} motif, in which the last hydrophobic residue is usually Ile or Leu. Group III substrates contain the X_{Hy}-S-X-L motif, where Ser and Leu are invariant. Group IV substrates are comprised of the H-X-X-X_{Hy} motif, which is similar to the cleavage site for MMP-2 within laminin-5. A list of group I-IV substrate sequences can be found in the manuscript " A Unique Substrate Recognition Profile for Matrix Metalloproteinase-2" included with the annual summary report.

Result 2: Three Groups of MMP-2 Substrates Were Selective For MMP-2 over MMP-9.

Comparison of the rate of hydrolysis of a set of representative phage substrates by MMP-2 and MMP-9 using the substrate ELISA was made to determine the selectivity of the MMP-2 substrates. The ability of MMP-2 (dark bars) and MMP-9 (open bars) to cleave substrate selected from the phage library were compared using the phage ELISA and illustrated in figure 1. Results are presented as the percentage of hydrolysis compared to non-treated control phage. This experiment was repeated three times, yielding nearly identical results in each repetition.

Figure 1.



With the exception of one clone, A45, the phage from group I lacked selectivity for MMP-2 over MMP-9. In contrast however, all of the phage substrates from group II and III were highly selective for MMP-2 over MMP-9. The extent of hydrolysis of phage substrates from group IV was not compared using this assay because their rate of hydrolysis by MMP-2 was generally low. These substrates were characterized in more detailed with the aid of synthetic peptides.

(This figure can also be found in the manuscript " A Unique Substrate Recognition Profile for Matrix Metalloproteinase-2" included with the annual summary report as figure 1.)

Result 3: Selectivity Of MMP-2 Substrates Was Confirmed By Charactering Synthetic Peptides.

Representative peptides from groups I-IV were synthesized and used to characterized substrate hydrolysis in greater detail. The peptides were initially used to determine the position of the scissile bond with each motif by analyzing the cleavage products by MALDI-TOF mass spectrometry. The rate of hydrolysis was measured by kinetic parameters, K_m, k_{cat} and k_{cat}/K_m ratio.

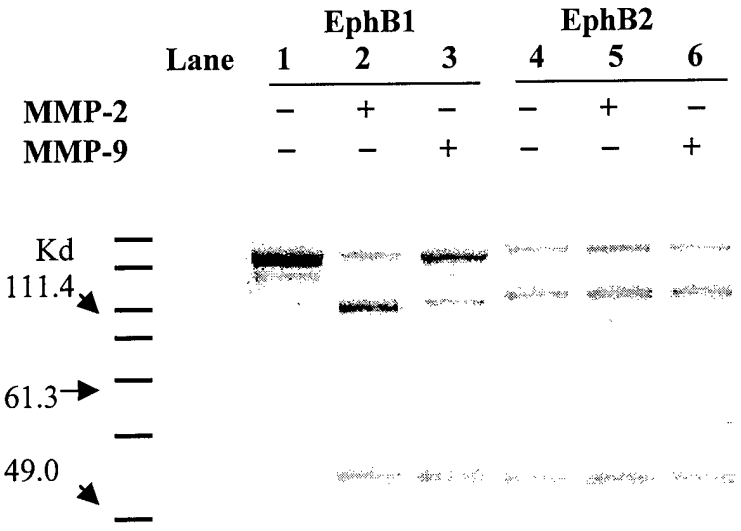
In virtually all cases, the scissile bond directly precedes a large hydrophobic residue, which is frequently Ile or Leu. This feature is consistent with the presence of a deep binding pocket at the corresponding $S_{1'}$ subsite within MMP-2. Most significantly, substrates from group II, III, and IV exhibit k_{cat}/K_m ratios that are eight to almost two hundred-fold higher for MMP-2 than MMP-9. The selectivity of each of the peptide substrates for MMP-2 compared to MMP-9 is represented by dividing the k_{cat}/K_m ratio for MMP-2 by the same value for the MMP-9 as shown in the following table.

Peptide	MMP-2/MMP-9 (Selectivity Ratio)
Group I **(P-X-X-X_{Hy})**	
A3	2
Group II **(I/L-X-X-X_{Hy})**	
B49	**14**
B74	**108**
Group III **(X_{Hy}-S-X-L)**	
C9	**198**
A13	**15**
B37	**71**
A34	**40**
Group IV **(H-X-X-X_{Hy})**	
A21	**8**

Result 4: Hydrolysis of A Protein Substrate Containing the S-X-L Motif.

Ultimately, one would hope to be able to use the substrate recognition profiles obtained from substrate phage, and other substrate profiling strategy (18,19) to generate hypothesis regarding physiologic substrates. As an initial step in this direction, I compared the ability of MMP-2 and 9 to cleave Eph B1 and Eph B2. Eph B1 and B2 are tyrosine kinase receptors that are responsible for cell-cell signaling during neuronal development. Only EphB1 has a putative cleavage site of MMP-2 corresponding to the motif of X_{Hy}-S-X↓L and EphB2 lacks the predicted cleavage. The Eph B1 and Eph B2 fusion proteins were incubated with equimolar amounts (280 nM) of MMP-2 and MMP-9 for four hours. The extent of hydrolysis was gauged by SDS-PAGE (Figure 2). The Eph B1-Fc fusion protein was almost quantitatively cleaved by MMP-2 (Figure 2, lane 2). The extent of cleavage by MMP-9 was far lower (Figure 2, Lane 3), which is not surprising considering the selectivity exhibited by the group II substrates. Neither protease cleaved the Eph B2 fusion protein. These findings illustrate that motifs found to be selective for MMP-2 by the substrate phage display library, can act as selective substrates within the context of whole proteins.

Figure 2.



Summary of results in Aim#1:
MMP-2 and 9 were found to have distinct substrate recognition profiles. Novel and highly selective MMP-2 substrate sequences were found in this study. Data from this study supports the hypothesis that substrate phage display library can be used to differentiate diminutive structural difference of proteases. In addition, these novel and highly selective MMP-2 substrates will be valuable in developing in vivo activity assay for MMPs and mechanistic design of specific MMP inhibitors.

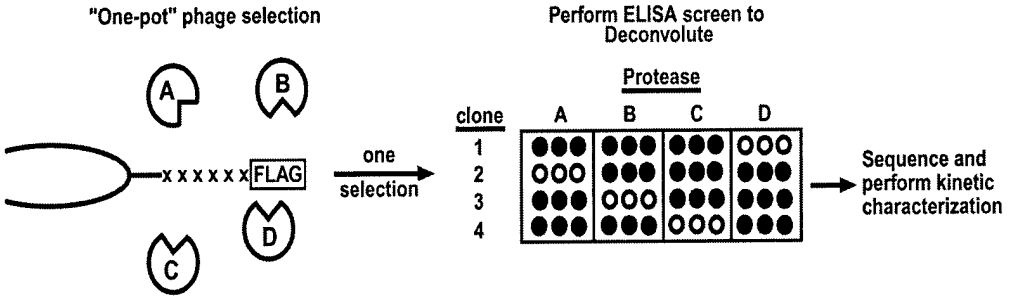
It is also encouraging to illustrate a new technology that has a great utility in revealing potential functions of proteases. To understand the reeling involvement of proteases in pathological conditions, one must approach from the fundamental biology of the proteases. Therefore, structure-functional information obtained from this study (detailed in the manuscript of "a unique substrate profile recognition of MMP-2") will help advancing our knowledge of the biology of matrix metalloproteinases.

Aim#2. Charcterize profile of protease activity present in the blood of mice bearing non-metastatic breast cancer tumors and mice bearing metastatic breast cancer tumors.

To test whether substrate phage display library can be use to profile protease activity in vivo, I exposed phage to serum from normal SCID mice, mice bearing non-metastatic MDA-MD-435 tumor, and mice bearing metastatic MDA-MD-435 tumor. No significant protease activity was detected in any samples. Subsequently, the serum was concentrated by either acetone or TCA precipitation to increase the concentration of the protein used in the reaction volume. Also, several experiments were conducted to find out the lowest concentration of active protease that can be detected by phage display library in the presence of nature inhibitors in the serum. I found that no protease activity could be detected below 100ng of active protease in the serum from a normal mouse. Hence, a range of protein concentration from 100ng to 10ug (from 5ml of mouse serum) was tested for detection of protease activity in serum samples. Fair amount of protease activity was only detected when using the highest amount of concentrated protein.

It was expected that the amount of proteases present in the samples might be small. Low protease concentration poses a difficulty to drive the kinetic reaction between proteases and substrates and might potentially explain why no significant protease activity was detected when the serum was not concentrated. Although protease activity can be detected in 10ug of total protein from the serum, it points to the problem of low sensitivity when exposing phage display library directly to biological samples. Furthermore, the amount of serum that would need to screen and validate protease activity poses another technical difficulty to pursue this method. Hence, I modified my approach of carrying out the functional study of disease-related proteases and developed a one-pot phage selection system that yield the substrate recognition profile of multiple purified proteases from a single round of selection.

I reasoned that a phage display based approach could be modified to characterize the substrate recognition profile of all proteases. Phage display, when used in the traditional manner, is a heritable system, allowing one to evolve optimal substrates (or binders) through multiple rounds of selective pressure. It is this property, however, that typically prevents phage from being used to obtain substrate recognition profiles for multiple proteases at once. Here, I developed a strategy for obtaining the substrate recognition profile of multiple proteases, simultaneously. It couples one round of substrate phage display with a substrate ELISA to deconvolute the specificity of protease substrates. The system allows one to avoid problems caused by the heritable nature of phage display by activity profiling even the rarest of substrates after a single round of protease selection. Furthermore, it allows all possible substrates to be tested by any proteases individually. It allows one to compare and deconvolute the substrate preferences for large panels of proteases in single experiment. A diagram of the "one-pot" protease substrate phage selection scheme is listed below to illustrate the expected outcome from this method.



Four proteases representing two catalytic classes were used to test the system's ability to discriminate the catalytic activity of each protease. Within this set, I included two classical serine proteases, thrombin and factor Xa, for which the substrate specificity has been well characterized. A test of the system with "knowns" would indicate the ability of the system to properly report substrate recognition. MMP-9 (matrix metalloproteinase) and atrolysin C (snake venom metalloproteinase) were used as representatives of metalloproteinases, whose recognition profiles are not well established (20). Within each class, the proteases share structural similarity and, reportedly, have overlapping substrate preferences. A comparison of the substrates selected for each protease to the reported substrate preference of each enzyme demonstrates the power of the approach. As the result of a single round of substrate selection using the mixture of these four proteases, peptide substrates were identified that *i*) have high k_{cat}/K_m ration, *ii*) are selective for individual proteases, and *iii*) match the sequences of

known physiological substrates. Please refer to the paper that I published this year in Analytical Biochemistry included in this annual summary update for more detail information.

Result1: Selective substrates were obtained from a single round of "one-pot" phage selection.

The substrate phage library was incubated with a mixture of thrombin (Th), factor Xa (Xa), MMP-9 (M9) and atrolysin C (At). Phage substrates were selected in one round of selection. Individual phage clones were tested as substrates in an ELISA assay. The extent of hydrolysis is measured by the loss of the Flag epitope at the N-terminus before the substrate insert and expressed as a percentage of the total relative to untreated controls, and is the average of two independent experiments. The hexamer sequence displayed by individual phages was derived from the DNA sequence of the phage inserts. Highly selective peptides for individual proteases tested were identified by substrate phage ELISA. A list of the substrate sequences is in the table I of the paper title "Protease Profiling by Phage Display" in the journal Analytical Biochemistry included in the fellowship update.

Result 2:Phage substrates identified for thrombin and factor Xa matched their physiologic substrate sequences.

By aligning the substrates cleaved more than 60% by thrombin, two groups of substrates become apparent (Table 2). The consensus motif derived from a combination of the groups, (Gly/Pro)-(Are/Lys)-↓-Ser-(Trp/Phe), is in agreement with the known substrate preference of thrombin and matches the sequence surrounding the scissile bond in several physiologic substrates for thrombin, including the thrombin receptor (protease-activated receptor), factor V and factor VII [ ]. Similarly, phage substrates identified for factor Xa contain a distinct recognition motif (Gly-Arg-↓-X) that matches well with the prothrombin activation site, the main physiologic target of factor Xa (21). A comparison table is listed in table II of the paper title "Protease Profiling by Phage Display" in the journal Analytical Biochemistry included in this annual summary update.

Result 3: The substrate phage ELISA as a semi-quantitative indicator of substrate hydrolysis.

Experiments were conducted to determined how well the substrate phage ELISA predicts the rank-order preference of analogous synthetic peptides. This feature was explored because of it potential to be adapted to high throughput screening. Screening protease substrates using synthetic peptides is costly and time-consuming. Therefore, to speed up the protease substrate screening process, I explored the possibility of using substrate phage ELISA as an indicator of the relative hydrolysis rate of the substrate sequence by the targeted protease.

The hydrolysis of a series of phage substrates for MMP-9 was measured as a function of time (figure 2A) and compared to the k_{cat}/K_m ratio for the corresponding synthetic peptides (figure 2B). Results from phage ELISA and actual k_{cat}/K_m of the peptide were compared. Both method yielded the same rank order of preference of substrate, A6 > A7 > A16. Similar observations were made when phage substrates for thrombin were compared to the analogous synthetic peptides. Figure 2A and 2B are in the paper title "Protease Profiling by Phage Display" in the journal Analytical Biochemistry included in the fellowship update.

Conclusions and future plans:

Large body of work on studying protease expression in diseases, especially in cancer progression, has been reported and continues to grow. Combining the information of protease expression and one-pot phage selection system, one can obtain highly selective peptide substrates for disease-related proteases rapidly. Then, an array of selective peptide substrates can be assembled and synthesized with fluorescent tags for example to increase the detection sensitivity. Large screening of protease activity profile can then be carried out to study pathological conditions in vivo such as the breast cancer mouse model as designed in the fellowship proposal to identify diagnostic markers in breast cancer progression. Therefore, my future plans are the following.

1. Profiling activity of more breast cancer relevant proteases such as MT1-MMP to obtain highly selective peptide substrates.

2. Explore methods to assemble the array of selective peptides such as fluorescent tagging for in vivo screening.

3. Establish the orthotopic breast cancer model as described in the original fellowship proposal to obtain the serum as well as the tumor tissue for in vivo protease activity profiling.

KEY RESEARCH ACCOMPLISHMENTS:

- Highly selective and novel MMP-2 substrate sequences were found via substrate phage display library.
- The selectivity of the novel MMP-2 peptide substrates was confirmed by the rate of hydrolysis of the synthetic peptides.
- A method based on the strategy of substrate phage display library was developed to obtain the substrate recognition/selectivity profile of multiple proteases simultaneously.

REPORTABLE OUTCOMES:

Two published papers and one accepted manuscript are listed as the following:

Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., and Smith, J. W. (2001) *Journal of Biological Chemistry* **276**(23), 20572-20578

Kridel, S. J., Chen E., Smith, J.W. (2001) *Analytical Biochemistry* (Currently only online version is available).

Chen, E., Kridel, S.J., Smith, J.W. (2001) the manuscript titled: "A unique substrate recognition profile for Matrix metalloproteinase-2" was accepted in May by Journal of Biological Chemistry (M1:03823)

CONCLUSIONS:

MMP-2 and 9 were found to have distinct substrate recognition profiles. Novel and highly selective MMP-2 substrate sequences were found in this study. Data from this study supports the hypothesis that substrate phage display library can be used to differentiate diminutive structural difference of proteases. In addition, these novel and highly selective MMP-2 substrates will be valuable in developing in vivo activity assay for MMPs and mechanistic design of specific MMP inhibitors.

It is also encouraging to illustrate a new technology that has a great utility in revealing potential functions of proteases. To understand the reeling involvement of proteases in pathological conditions, one must approach from the fundamental biology of the proteases. Therefore, structure-functional information obtained from this study (detailed in the manuscript of "a unique substrate profile recognition of MMP-2") will help advancing our knowledge of the biology of matrix metalloproteinases.

In additional to the work of MMP-2 and MMP-9 substrate profile comparison, I developed a method of profiling the activity of multiple proteases simultaneously using the substrate phage display library that I constructed last year. Large body of work on studying protease expression in diseases, especially in cancer progression, has been reported and continues to grow. Combining the information of protease expression and one-pot phage selection system, one can obtain highly selective peptide

substrates for disease-related proteases rapidly. Then, an array of selective peptide substrates can be assembled and synthesized with fluorescent tags for example to increase the detection sensitivity. Large screening of protease activity profile can then be carried out to study pathological conditions in vivo such as the breast cancer mouse model as designed in the fellowship proposal to identify diagnostic markers in breast cancer progression.

REFERENCES:

1. Basbaum, C. B., and Werb, Z. (1996) *Curr. Opin. Cell Biol.* **8**(5), 731-8
2. Werb, Z., and Yan, Y. (1998) *Science* **282,** 1279-1280
3. Nagase, H., and Woessner, J. F., Jr. (1999) *Journal of Biological Chemistry* **274**(31), 21491-4
4. Van Wart, H. E., and Birkedal-Hansen, H. (1990) *Proc Natl Acad Sci U S A* **87**(14), 5578-82
5. Nagase, H. (1997) *Biol. Chem.* **378,** 151-160
6. Kahari, V. M., and Saarialho-Kere, U. (1999) *Annals of Medicine* **31**(1), 34-45
7. Leppert, D., Ford, J., Stabler, G., Grygar, C., Lienert, C., Huber, S., Miller, K. M., Hauser, S. L., and Kappos, L. (1998) *Brain* **121**(Pt 12), 2327-34
8. Poulsom, R., Hanby, A. M., Pignatelli, M., Jeffery, R. E., Longcroft, J. M., Rogers, L., and Stamp, G. W. (1993) *Journal of Clinical Pathology* **46**(5), 429-36
9. Davies, B., Miles, D. W., Happerfield, L. C., Naylor, M. S., Bobrow, L. G., Rubens, R. D., and Balkwill, F. R. (1993) *British Journal of Cancer* **67**(5), 1126-31
10. Boag, A. H., and Young, I. D. (1994) *American Journal of Pathology* **144**(3), 585-591
11. Sier, C. F., Kubben, F. J., Ganesh, S., Heerding, M. M., Griffioen, G., Hanemaaijer, R., van Krieken, J. H., Lamers, C. B., and Verspaget, H. W. (1996) *British Journal of Cancer* **74**(3), 413-7
12. Brooks, P. C., Stromblad, S., Sanders, L. C., von Schalscha, T. L., Aimes, R. T., Stetler-Stevenson, W. G., Quigley, J. P., and Cheresh, D. A. (1996) *Cell* **85**(5), 683-93
13. Itoh, T., Tanioka, M., Yoshida, H., Yoshioka, T., Nishimoto, H., and Itohara, S. (1998) *Cancer Research* **58**(5), 1048-51
14. Bergers, G., Brekken, R., Mcmahon, G., Vu, T. H., Itoh, T., Tamaki, K., Tanzawa, K., Thorpe, P., Itohara, S., Werb, Z., and Hanahan, D. (2000) *Nature Cell Biology* **2,** 737-744
15. Smith, M. M., Shi, L., and Navre, M. (1995) *J Biol Chem* **270**(12), 6440-9
16. Deng, S. J., Bickett, D. M., Mitchell, J. L., Lambert, M. H., Blackburn, R. K., Carter, H. L., Neugebauer, J., Pahel, G., Weiner, M. P., and Moss, M. L. (2000) *Journal of Biological Chemistry* **40**(6), 31422-31427
17. Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., and Smith, J. W. (2001) *Journal of Biological Chemistry* **276**(23), 20572-20578
18. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) *Nature Biotechnology* **18**(2), 187-93
19. McGeehan, G. M., Bickett, D. M., Wiseman, J. S., Green, M., and Berman, J. (1995) *Methods in Enzymology* **248,** 35-46
20. Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (eds) (1998) *Handbook of Proteolytic Enzymes*, Academic Press, San Diego
21. Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K., and Engh, R. A. (1996) *Journal of Biological Chemistry* **271**(47), 29988-92

APPENDICES:

Two original copies of the paper I published this year and one manuscript are included to supplement and clarify the annual summary update.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 23, Issue of June 8, pp. 20572–20578, 2001
Printed in U.S.A.

Substrate Hydrolysis by Matrix Metalloproteinase-9*

Received for publication, January 30, 2001, and in revised form, March 12, 2001
Published, JBC Papers in Press, March 14, 2001, DOI 10.1074/jbc.M100900200

Steven J. Kridel‡, Emily Chen‡, Lakshmi P. Kotra§¶, Eric W. Howard‖, Shahriar Mobashery§, and Jeffrey W. Smith‡**

From the ‡Program On Cell Adhesion and the Cancer Research Center, Burnham Institute, La Jolla, California 92037, the §Institute for Drug Design and the Department of Chemistry, Wayne State University, Detroit, Michigan 48202, and the ‖Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104

The catalytic clefts of all matrix metalloproteinases (MMPs) have a similar architecture, raising questions about the redundancy in substrate recognition across the protein family. In the present study, an unbiased phage display strategy was applied to define the substrate recognition profile of MMP-9. Three groups of substrates were identified, each occupying a distinct set of subsites within the catalytic pocket. The most prevalent motif contains the sequence Pro-*X*-*X*-Hy-(Ser/Thr) at P_3 through $P_{2'}$. This sequence is similar to the MMP cleavage sites within the collagens and is homologous to substrates the have been selected for other MMPs. Despite this similarity, most of the substrates identified here are selective for MMP-9 over MMP-7 and MMP-13. This observation indicates that substrate selectivity is conferred by key subsite interactions at positions other than P_3 and $P_{1'}$. This study shows that MMP-9 has a unique preference for Arg at both P_2 and P_1, and a preference for Ser/Thr at $P_{2'}$. Substrates containing the consensus MMP-9 recognition motif were used to query the protein data bases. A surprisingly limited list of putative physiologic substrates was identified. The functional implications of these proteins lead to testable hypotheses regarding physiologic substrates for MMP-9.

Matrix metalloproteinase-9 (MMP-9)[1] is a member of the matrixin family of metallo-endopeptidases (1–3). MMP-9 is historically referred to as gelatinase B because of its ability to cleave gelatin, a denatured form of collagen, *in vitro*. Along with MMP-2, MMP-9 differs from other MMPs because it contains three fibronectin type II repeats that have high binding affinity for collagen. These repeats are thought to mediate the binding of MMP-2 and -9 to collagen (1, 2). This binding interaction brings the catalytic pocket of the MMP in proximity to collagen, thereby enhancing its rate of hydrolysis. Despite these well characterized biochemical interactions, it is now clear that MMP-9 is also able to cleave a number of other proteins and may have a rather wide range of physiologic substrates (4–8).

Much of our understanding of the biological function of MMP-9 comes from the study of mice lacking this gene. For example, MMP-9-deficient mice have impaired ossification of the skeletal growth plate, a defect that has been partially attributed to poor vascularization of developing bone (9). Studies on these mice also show that MMP-9 is essential for the recruitment of osteoclasts into developing bones (10). Other work indicates that MMP-9-deficient mice are resistant to dermal blistering in a bullous pemphigoid model, an effect that has been attributed to the inability of these mice to cleave the SERPIN α1-proteinase inhibitor (5). Finally, recent studies in the RIP1-Tag2 transgenic mouse model of multistage carcinogenesis indicate that MMP-9 is part of the angiogenic "switch" that is essential for tumor growth (11, 12). Other reports suggests that MMP-9 may play a role in inflammation in the nervous system. MMP-9 is elevated in encephelomyelitis (7, 8), in the cerebrospinal fluid of patients with multiple sclerosis (13), and in patients with AIDS-related dementia (14).

A first approximation of the substrate recognition specificity of MMP-9 has been gleaned from alignments of its cleavage site within a number of different proteins (4, 8, 9, 15–17). Nevertheless, a detailed understanding of subsite preferences for MMP-9 is lacking. Such an analysis is particularly important because the catalytic cleft of MMP-9 is closely related to other MMPs, raising questions about the distinction among substrates for these proteases. These issues are particularly important because many of the current pharmacologic antagonists of the MMPs have overlapping inhibition profiles. Furthermore, in the post-genomic era, where the sequences of all proteins will soon be available, information on substrate recognition could help identify important physiologic substrates.

With these ideas in mind, we applied an unbiased approach to define the substrate recognition preference of MMP-9. MMP-9 was used to cleave substrates within a vastly complex phage display library of random hexamers. Substrates within this library can be cleaved at any position within the hexamer, allowing information on the substrate specificity on both sides of the scissile bond to be obtained. Three families of substrates were identified. The largest group contains a Pro-*X*-*X*-Hy-(Ser/Thr) motif (*X* is any residue, and Hy is a hydrophobic residue) that occupies positions P_3 through $P_{2'}$. This general motif is cleaved by a number of MMPs and is presumed to represent a collagen-like substrate (18–23). Nevertheless, substrates within this family that were selected show considerable selec-

This workstudy was supported by California Breast Cancer Research Program Grant 5JB0033, National Institutes of Health Grants AR42750 and CA69036, Cancer Center Support Grant CA30199 (to J. W. S.), United States Army Grant DAMD17-97-17174 (to S. M.), Fellowships AR08505 from the National Institutes of Health and 2PD0182 from the California Cancer Research Program (both to S. J. K.), and a graduate fellowship from the United States Breast Cancer Research Program (to E. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Current address: Faculty of Pharmacy, University of Toronto, Toronto, Ontario M5S 2S2, Canada.

** To whom correspondence should be addressed: Program on Cell Adhesion and Cancer Research Center, Burnham Inst., 10901 N. Torrey Pines Rd., La Jolla, CA 92037. E-mail: jsmith@burnham-inst.org.

[1] The abbreviations used are: MMP, matrix metalloproteinase; PCR, polymerase chain reaction; BSA, bovine serum albumin; ELISA, enzyme-linked immunosorbent assay; TBS-T, Tris-buffered saline with Tween 20; MALDI-TOF, matrix-assisted laser desorption ionization/time of flight; APMA, p-Aminophenylmercuric acetate.

tivity for MMP-9. The second group of substrates are defined by a Gly-Leu-(Lys/Arg) motif at positions P_1 through $P_{2'}$. Members of the third family of substrates are unique in that they contain Arg residues at both P_1 and P_2. Altogether, these findings reveal multiple modes of substrate recognition by MMP-9 and provide important insights into the hydrolysis of physiologic substrates that may be important in biology and pathology.

EXPERIMENTAL PROCEDURES

Purified forms of full-length MMP-7, MMP-13, and TIMP-2 were purchased from Chemicon (Temecula, CA). Ilomastat (GM6001) was purchased from AMS Scientific (Concord, CA). Restriction enzymes were from Roche Biosciences or New England Biolabs. Oligonucleotides were synthesized by Integrated DNA Technologies, Inc. Tissue culture media and reagents were from Irvine Scientific (Irvine, CA). All other reagents, chemicals, and plasticware were from Sigma or Fisher.

Construction of Substrate Phage Display Library—Substrate phage libraries were generated using a modified version of the fUSE5 phagemid (24–26). A FLAG epitope was engineered at the NH_2 terminus of the geneIII protein by annealing oligonucleotides 5′-CCGGGTTTGTCGTCGTCGTCTTTGTAGTCGGTAC-3′ and 5′-CGACTACAAAGACGACGACGACAAAC-3′ and ligating them into fUSE5 at the *Kpn*I and *Xba*I restriction sites. The random hexamers were generated by PCR extension of the template oligonucleotide 5′-GGGGAGGCCGACGTGGCCGTCATCAGGCGGCTCAGGC(NNK)$_6$ACGG CCTCTGGGGCCGAAAC-3′, where N is any nucleotide and K is either G or T. The template oligonucleotide also encodes an SGGSG linker positioned in between the FLAG epitope and the random hexamer. A primer oligonucleotide 5′-AATTTCTAGTTTCGGCCCCAGAGGC-3′ and the template oligonucleotide were mixed and heated at 65 °C for 2 min. The heating block was switched off and allowed to cool passively to 40 °C to allow annealing of the extension oligonucleotide to the template oligonucleotide. Elongation of the template oligonucleotide was performed using Sequenase (United States Biochemical Corp.) (27). The final cDNA product was precipitated with ethanol, re-suspended in water, and digested with *Sfi*I. The DNA insert and fUSE5 were mixed and ligated at a 5:1 molar ratio and electroporated into *Escherichia coli* MC1061(F−). Several phage were selected for sequencing to confirm randomness in the insert sequences and the correct reading frame.

Expression and Purification of MMP-9—The cDNA encoding the catalytic domain of MMP-9 (28, 29) was generated by PCR and cloned into pCDNA3 (Invitrogen). The expression vector was used to transfect HEK 293 cells by electroporation, and individual antibiotic-resistant colonies were isolated with cloning rings, expanded, and then screened by reverse transcription-PCR for MMP-9 mRNA. Expression of the proteinaceous catalytic domain was determined by zymography using conditioned medium from the transfected cells. Cells expressing the MMP-9 catalytic domain were grown in 150-mm dishes and conditioned in Dulbecco's modified Eagle's medium with 10% fetal bovine serum supplemented with G418 (200 μg/ml) for 5 days. The catalytic domain of MMP-9 was purified from the conditioned medium by gelatin-Sepharose chromatography as described previously (28, 29). Fractions containing MMP-9 were pooled and dialyzed against 50 mM Tris, pH 7.8, 10 mM NaCl. The dialyzed material was then purified further using a 1-ml Hi-Trap Source Q ion exchange column (Amersham Pharmacia Biotech). Fractions containing MMP-9 were pooled and dialyzed against 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$ and then concentrated in a dialysis bag using Aquacide II (Calbiochem). The final concentration of purified material was typically 1.5–3 mg/ml.

Activation and Active Site Titration of Proteases—The purified catalytic domain of MMP-9 was activated by incubation with 2 mM APMA for 16 h at room temperature (29). Activation was monitored by altered migration of the protein on SDS-polyacrylamide gel electrophoresis and zymography. Pro-MMP-13 was activated with APMA for 3 h at room temperature, and MMP-7 required no activation.

The level of active protease was always quantified by active site titration prior to kinetic studies. The active site of MMP-9 was titrated using the hydroxamate inhibitor Ilomastat (30). Briefly, 25 nM MMP-9 was incubated with a range of Ilomastat. Inhibition of MMP-9 by Ilomastat was then measured by determining the rate of cleavage of three concentrations of the peptide, m1A11, using the fluorescamine incorporation assay (21, 31). The active sites of MMP-13 and MMP-7 were titrated with human TIMP-2. Briefly, 5–15 nM of each protease were pre-incubated with a range of TIMP-2 for 5 h at room temperature. Residual MMP-13 activity was measured by cleavage of MCA-Pro-Cha-Gly-Nva-His-Ala-Dpa-NH_2 (Calbiochem) as substrate, and residual MMP-7 activity was measured by cleavage of MCA-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2 (Calbiochem) with monitoring at λ_{ex} 320 nm and λ_{em} 405 nm.

Phage Selection of MMP-9 Substrates—The substrate phage library (2×10^{10} phage) was incubated with 2.5 μg/ml MMP-9 in 50 mM Tris, pH 7.4, 100 mM NaCl, 10 mM $CaCl_2$, 0.05% Brij-35, and 0.05% BSA for 1 h at 37 °C. A control selection was performed without protease. The cleaved phage were separated from the non-cleaved phage by immunodepletion. 100 μg of anti-FLAG monoclonal antibody (Sigma) was added to the phage samples and then incubated for 18 h with rocking at 4 °C. The phage-antibody complexes were precipitated by the addition of 100 μl of Pansorbin (Calbiochem). The cleaved phage remaining in the supernatant were amplified using K91 *E. coli* as described previously (32–34) and were then used for subsequent rounds of substrate selection.

Substrate Phage ELISA—Hydrolysis of individual phage substrates was measured using a modified ELISA. 96-well microtiter plates were coated with anti-M13 antibody (Amersham Pharmacia Biotech, 2.5 μg/ml), in phosphate-buffered saline, overnight at 4 °C. After coating, the wells were blocked for 1 h at room temperature in TBS-T (50 mM Tris, pH 7.8, 150 mM NaCl, 0.2% Tween 20) containing 10 mg/ml BSA. After blocking, 150 μl of supernatant from an overnight phage culture was added to each well and incubated for 2 h at 4 °C to allow for phage capture. Unbound phage were removed with three washes of ice-cold TBS-T. To asses hydrolysis, MMP-9 (2.5 μg/ml) was added to the appropriate wells in Incubation Buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, 0.05% BSA, 0.05% Brij-35, 50 μM $ZnCl_2$) for 2 h at 37 °C. Control wells lacked protease. The protease solution was removed, and the wells were washed four times with ice-cold TBS-T. To measure hydrolysis of the peptides on phage by MMP-9, anti-FLAG polyclonal antibody (1.8 μg/ml in TBS-T with 1 mg/ml BSA) was added to each well and the plates were incubated at 4 °C for 1 h. Binding of anti-FLAG antibody to FLAG epitope was measured with a horseradish peroxidase-conjugated goat anti-rabbit IgG antibody (Bio-Rad) followed by detection at 490 nm. The extent of hydrolysis, taken as a measure of substrate hydrolysis, was calculated by the ratio of the optical density at 490 nm of the protease-treated samples *versus* samples lacking protease.

Determination of Scissile Bonds—The cleavage site for MMP-9 within peptide substrates was determined using MALDI-TOF mass spectrometry. MMP-9 (25 nM) was incubated with 100 μM amounts of each peptide, independently, in 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$ for 2 h at 37 °C. The mass of the cleavage products was determined using a Voyager DE-RP MALDI-TOF mass spectrometer (PerSeptive Biosystems, Framingham, MA). Following hydrolysis, the peptide samples were prepared according to methods described previously (35–38). In all cases, no other peaks that corresponded to other potential cleavage products were observed.

Kinetic Measurements of Peptide Hydrolysis—The kinetic parameters of substrate hydrolysis were measured using a fluorescamine incorporation assay that has been described previously (21, 31, 34, 39). Briefly, MMP-9, MMP-7, or MMP-13 were incubated with individual peptide substrates at concentrations ranging from 100 to 800 μM in 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, 50 μM $ZnCl_2$. At selected time points, the reactions were stopped by the addition of 1,10-phenanthroline. Peptide hydrolysis was determined by the addition of fluorescamine followed by detection at λ_{ex} 355 nm and λ_{em} 460 nm. The data were transformed to double-reciprocal plots ($1/[S]$ *versus* $1/v_i$) to determine K_m and k_{cat} (21, 31, 34, 39). Similar results were obtained using different batches of protease. For some substrates, K_m and k_{cat} could not be determined individually, but the specificity constant, k_{cat}/K_m, was derived by the equation: $k_{cat}/K_m = v_i/(E_0)(S_0)$ (22).

Computational Molecular Modeling—The terminology established by Schechter and Berger is used to describe the subsites in the protease active site and the correlating positions in substrates (40). This method describes the former as a set of subsites designated as S (S_1, S_2, etc.) and the latter as P (P_1, P_2, etc.). The primed sites are to the carboxyl-terminal and unprimed sites are to the amino-terminal side of the scissile bond. The x-ray structure of MMP-9 has not been reported to date. We have previously reported a computational three-dimensional model for the catalytic domain of MMP-9, which was generated by homology modeling taking advantage of sequence alignment and conservation of secondary structural elements (41–43). This model shows high sequence and three-dimensional homology to the structure of MMP-2 (42), and it indicates an extended cleft for binding of substrate that can accommodate at least six amino acids from the substrate. We considered molecular models for representative members of each class of substrates (peptide A10 and peptide C11), but we concluded that there is not sufficient sequence information for the group II substrates

to draw meaningful conclusions. Therefore, the modeling work was concentrated on groups I and III. The molecular models of the sequences flanking the sites of MMP-9 cleavage in clones A10 and C11 (Ac-SGPLFYSVTA-NH_2 and Ac-SGRRLIHHTA-NH_2, respectively) were constructed using the SYBYL molecular modeling program and were docked into the active site of MMP-9 (42). Each enzyme-substrate complex contained the hydrolytic water molecule positioned in the active site near the Glu-402 side chain, the general base that promotes it for the hydrolytic reaction. The enzyme-substrate complexes were energy-minimized using the AMBER 5.0 software package (44, 45). The protocol for the energy minimization was described previously (42).

RESULTS

Construction of the Phage Library—We created a system for displaying random hexamer substrates on the surface of phage. The fUSE5 polyvalent phage display vector (24–26) was modified to express random hexamers at the amino terminus of the geneIII protein, and an octapeptide FLAG epitope at the amino-terminal end of the random hexamers. The library comprises 2.4×10^8 independent transformants, giving a 75% confidence that each of the 6.4×10^7 possible random hexamer sequences are represented in the library. Sequencing of phage confirmed the randomness of the hexamer insert. Under the selection conditions, greater than 95% of phage could be immunodepleted using an anti-FLAG antibody (data not shown).

Selection of Peptide Substrates for MMP-9—Optimal substrates were selected by exposing the phage library to a recombinant form of the catalytic domain of MMP-9 expressed in HEK 293 cells (29). The recombinant catalytic domain of MMP-9 was purified on gelatin-Sepharose, followed by ion exchange chromatography (29). The protease was activated with 2 mM APMA for 16 h at room temperature, and the active site was titrated with the hydroxamate inhibitor Ilomastat (29, 30). Phage selections were performed with 2.5 μg/ml (56 nM) active MMP-9. Following three rounds of exposure to MMP-9, individual phage clones were selected for sequencing. An alignment of the motifs revealed three groups of structurally distinct substrates (Table I). Substrates in group I contain the motif with sequence Pro-*X*-*X*-Hy-(Ser/Thr). Further analysis shows that, within this larger motif, Arg is favored at P_2 and Ser/Thr is favored at P_1 (consensus motif noted in Table I). Substrates in group II contain a Gly-Leu-(Lys/Arg) motif. Interestingly, both groups I and II are related to substrates that have been described previously for other MMPs and are somewhat related to sites of cleavage in collagen (21–23). Group III substrates, however, appear to represent a novel recognition sequence as it contains the Arg-Arg-Hy-Leu (group IIIA) and Arg-X-Leu (group IIIB) motifs. The two subclasses that comprise group III have not been described previously as a substrate motif for MMP-9.

The ability of MMP-9 to hydrolyze each of the phage clones was assessed in a semiquantitative manner using a modified ELISA. Individual phage clones were captured into microtiter plates using anti-M13 antibody. Captured clones were exposed to MMP-9 (2.5 μg/ml). A polyclonal anti-FLAG antibody was used to measure the liberation of the FLAG epitope by hydrolysis of the substrate with MMP-9. Results are expressed as the extent of hydrolysis relative to untreated phage clones (Table I). In general, the group I substrates were hydrolyzed most efficiently. However, a number of the substrates within groups II and III were cleaved to the same degree as group I substrates.

Identification of Scissile Bonds within MMP-9 Substrates—To identify the position of the scissile bonds, 10 peptides, representing each of the three groups of substrates, were designed and synthesized. Following exposure to MMP-9, the scissile bond was determined by analyzing the cleavage products by MALDI-TOF analysis (Table II). This analysis revealed that each substrate contains a hydrophobic residue at the $P_{1'}$ position. The substrates in group I contain Pro at the P_3 position. The single peptide from group II contained a Gly at P_1, Leu at $P_{1'}$, and a Lys at $P_{2'}$. The motifs of groups I and II, and locations of their scissile bonds, are similar to MMP cleavage sites in collagen and gelatin (21–23). Interestingly, however, the peptides from group III were all cleaved after an Arg residue, and in several cases an Arg-Arg occupied the P_2 and P_1 positions.

Kinetic Characterization of Substrate Hydrolysis by MMP-9—The Michaelis constant (K_m) and first-order rate constant of substrate peptide turnover (k_{cat}) were measured by incubating a range of each peptide with MMP-9. Peptide hydrolysis was measured by incorporation of fluorescamine onto newly formed amino termini as previously described (31, 33, 34). From these measurements, k_{cat} and K_m were derived for each peptide using double-reciprocal plots of 1/[S] *versus* $1/v_i$. Results from this

TABLE I

Sequence alignment and grouping of phage selected by MMP-9

Phage from the third round MMP-9 selection were selected for their ability to be cleaved by MMP-9 and sequenced. The substrates are separated into three groups based on distinguishable motifs. Residues that define each group are in bold. The ability of MMP-9 to cleave the substrates on phage was measured in a modified ELISA format. A rank order of preference within each group is demonstrated as a function of phage cleaved by MMP-9 relative to a non-treated control. This is expressed as percentage of hydrolysis. The results of two independent experiments are shown. Hy, hydrophobic amino acids.

Group	Peptide sequence								Clone	% Hydrolysis	
										1	2
I			**P**	L	F	**Y**	**S**	V	A10	100	100
	K	I	**P**	**R**	**T**	**L**	**T**		A11	100	100
			P	L	R	**L**	S	W	A12	100	100
			P	**R**	A	**V**	S	T	A6	100	99
	K	G	**P**	**R**	Q	**I**	T		C15	100	98
			P	**R**	P	**L**	S	G	D12	82	94
		W	**P**	L	G	**L**	A		D35	89	91
	F	R	**P**	**R**	**S**	**I**	T		D36	87	91
	R	L	**P**	V	G	**L**	T		D20	89	90
	R	S	**P**	K	**S**	**L**	T		D21	88	90
			P	V	W	**L**	A	A	D22	85	90
	I	H	**P**	S	**S**	**L**	T	A	D4	87	89
			P	A	**S**	**F**	T	S	D29	87	86
	G	Q	**P**	H	Y	**L**	T		B1	79.6	85
	M	K	**P**	A	**S**	**W**	T		B2	77.5	77
	T	H	**P**	Y	**T**	**M**	T		B10	74.1	82
		T	**P**	A	Y	**M**	L	T	A13	78.9	72
			P	L	Y	**L**	T		D2	79	82
			P	G	L	**I**	G	T	D1	75	85
			P	**R**	**S**	**I**	**S**	N	D24	82	82
	R	L	**P**	A	**S**	**Y**	**T**		D30	82	60
	N	P	**P**	**R**	Y	**L**	**T**		D5	70	67
			P	K	**T**	Q	I	S	D8	73	65
	S	V	**P**	**R**	H	**F**	**T**		D17	70	52
	L	L	**P**	A	W	**L**	**T**		D3	56	48
	T	H	**P**	Y	**T**	**M**	**T**		D14	50	47
	L	R	**P**	A	K	S	**T**		A18	28.9	30
	S	G	**P**	S	**T**	S	**T**		A9	13.1	5.8
Consensus			**P**	**R**	**S/T**	**Hy**	**S/T**				
II			G	S	G	L	K	A	A7	79.8	76
			A	M	G	L	K	S	B9	73.3	86
			K	V	G	L	R	T	B8	13.3	12
IIIA			G	R	R	L	I	H	C11	94.9	74
		H	P	R	R	S	I	T	C1	13.9	29
			G	R	R	L	L	S	A16	20.2	11
		A	L	R	R	L	E	T	C13	8.3	24
		F	Y	K	R	V	L	T	B6	11.7	11.1
			F	R	R	I	C	V	C7	17.9	0
	V	F	F	R	R	Q	T	A	C8	8.9	0
IIIB	G	L	A	R	N	I	T	A	B3	66	70
	F	G	S	R	Y	L	T	A	A19	45	50
		Q	D	R	Y	L	N	T	C6	40	27

TABLE II

Kinetic analysis of MMP-9 peptide substrates based on phage clones

Each peptide was incubated with protease and hydrolysis was detected as described under "Experimental Procedures." The peptides are named according to their phage clone designation. All peptides were synthesized with acetylated NH_2 termini and amidated COOH-termini. A11m1 and A11m2 are analogues of peptide A11. Changes are indicated in bold. Scissile bonds are designated with ↓. Amino acids to the right of the ↓ are primed residues (. . .$P_{1'}$, $P_{2'}$, etc.), and amino acids to the left are non-primed residues (. . .P_2, P_1) (40). **, measurements were not made; *, not detected. Standard deviations of triplicate experiments ranged from 11% to 22%.

Peptide	Sequence	MMP-9			MMP-13	MMP-7
		k_{cat}/K_m	k_{cat}	K_m	k_{cat}/K_m	k_{cat}/K_m
		$M^{-1}\ s^{-1}$	s^{-1}	mM	$M^{-1}\ s^{-1}$	$M^{-1}\ s^{-1}$
C15	SGKGPRQ↓ITA	188,000	703	3.7	15,400	13,800
A11	SGKIPRT↓LTA	67,500	86.4	1.3	23,500	7,000
A6	SGPRA↓VSTTA	61,000	178	2.9	12,400	1,300
A11m1	SGKIPR**R**↓LTA	160,000	224	1.4	15,200	24,000
A11m2	SGKIPRT↓**A**TA	2000	9.4	4.7	**	**
A10	SGPLF↓YSVTA	25,600	20	0.5	**	**
B1	SGQPHY↓LTTA	8800	9	1	**	**
A18	SGLRPAK↓STA	5800	78	13.5	**	**
A7	SG↓LKALMITA	12,400	30	2.4	12,300	41,000
A19	SGFGSRY↓LTA	13,200	11	0.8	*	1,000
A16	SGRR↓LLSRTA	1200	**	**	**	**
C11	SGRR↓LIHHTA	*	**	**	**	**

analysis are shown in Table II. Among the peptides, the k_{cat} values ranged from 9 s^{-1} to 703 s^{-1}, and K_m values were generally in the high micromolar to low millimolar range. Overall, peptides from group I appeared to represent the best substrates, with peptide C15 having a k_{cat}/K_m ratio of 180,000 M^{-1} s^{-1}, the highest of any of the substrates. Peptide A11, which closely matches the consensus recognition motif of group I, also exhibits a high k_{cat}/K_m ratio (67,500 $M^{-1}\ s^{-1}$) and considerable selectivity for MMP-9.

Interestingly, the three peptides cleaved most efficiently by MMP-9 were from group I and contained Arg at the P_2 position. This led us to hypothesize that group I and III substrates might represent a larger set of substrates whose relationship to one another is not entirely evident from the sequences of only 100 clones. To test this idea, we synthesized a mutant peptide (A11m1) that contains a Thr → Arg substitution at P_1 within the context of the sequence of the A11 peptide. Hence, the mutant peptide contained elements of both group I (Pro at P_3 and Leu at $P_{1'}$) and group III (Arg at P_2 and P_1). Peptide A11m1 had a k_{cat}/K_m twice that of the parent peptide, an effect resulting primarily from an increase in k_{cat}. This finding suggests that the substitution of Arg at P_1 lowers the transition state energy of the protease-substrate interaction. This observation also indicates that, within the Pro-*X*-*X*-Hy-(Ser/Thr) motif, Arg residues at P_2 and P_1 are favored for MMP-9, and that the substrates from groups I and III may have a related binding mechanism.

In almost every case, the MMP-9 substrates selected from the phage library contain a hydrophobic residue at $P_{1'}$, a finding that is entirely consistent with the fact that MMPs are known to have a deep hydrophobic $S_{1'}$ pocket (1, 2, 46). To assess the contribution of this hydrophobic residues to substrate binding and to substrate turnover, we synthesized another mutant peptide (A11m2) based on the sequence of peptide A11, but containing Ala, rather than Leu, at $P_{1'}$ (Ac-SGKIPRT↓ATA-NH_2). This substitution had deleterious effects on both k_{cat} and K_m, and reduced the k_{cat}/K_m ratio nearly 30-fold. These results are consistent with the idea that the $S_{1'}$ subsite of MMP-9 coordinates substrate binding and also influences the rate of hydrolysis. Importantly, however, even this mutant peptide had a measurable k_{cat}/K_m ratio (2000 $M^{-1}\ s^{-1}$), indicating that efficient hydrolysis can be enacted by MMP-9 if the rest of the substrate sequence is optimal.

Kinetic Characterization of Substrate Hydrolysis by MMP-7 and MMP-13—Since many of the selected substrates contained a motif similar to that described for other MMPs (P-*X*-*X*-Hy), we measured the degree to which MMP-7 and MMP-13 could cleave these MMP-9 substrates. These two MMPs were used for comparison because substrates for both proteases have been selected using a similar phage display approach. Interestingly, most of the substrates we tested were cleaved more efficiently by MMP-9. The k_{cat}/K_m ratios ranged from 2.6- to 47-fold higher for MMP-9 than for either MMP-7 or MMP-13 (Table II). Only peptide A7 deviated from this trend. It is also worth noting that, although the substrates in group IIIA (Arg-Arg) were not cleaved rapidly by MMP-9, we were unable to detect any hydrolysis of these peptides by either MMP-7 or -13.

Modeling Substrate Interactions with MMP-9—Molecular modeling studies were conducted to help visualize how substrates might dock into the enzymatic cleft of MMP-9. Energy-minimized models of MMP-9 were constructed according to procedures described under "Experimental Procedures," and then docked with peptides A10 (Ac-SGPLFYSVTA-NH_2) representing group I and C11 (Ac-SGRRLIHHTA-NH_2) representing group III.

A primary feature of the group I substrates is a Pro residue at the P_3 position. The unique conformational features of Pro introduce the appropriate "bend" needed for the optimal substrate binding at this site. This Pro residue, occupying the S_3 subsite of MMP-9, is illustrated in Fig. 1*A* (*white arrow*). The group I substrates also contain an invariant hydrophobic residue at the $P_{1'}$ position. This residue protrudes into the deep $S_{1'}$ pocket of the protease, as depicted by the *orange arrows*.

Amino acids with long basic side chains at P_2 and P_1, such as Arg, are the defining features of the substrates in group III. Although the presence of these residues at P_2 and P_1 is somewhat surprising, the energy-minimized models support the observation that these residues bind favorably. An Arg at P_2 is likely to interact with the backbone carbonyl moieties of His-405, Gly-408, and the side chain of Asp-410, all of which contribute to the S_2 subsite within MMP-9 (*white arrow* in Fig. 1*B*). These electrostatic interactions are predicted to contribute to favorable binding of this class of substrates. Many of the group III substrates also contain an Arg residue at P_1. The favorable interaction of Arg into the S_1 subsite can be explained by the somewhat unusual nature of this subsite. It is essentially a hydrophobic binding surface that would be predicted to accommodate the hydrophobic side chains of amino acids such as Ala, Phe, and Tyr. However, the docking studies of peptides with Arg at P_2 show that the hydrophobic surface of the S_2 subsite could also bind to the extended methylene group in the

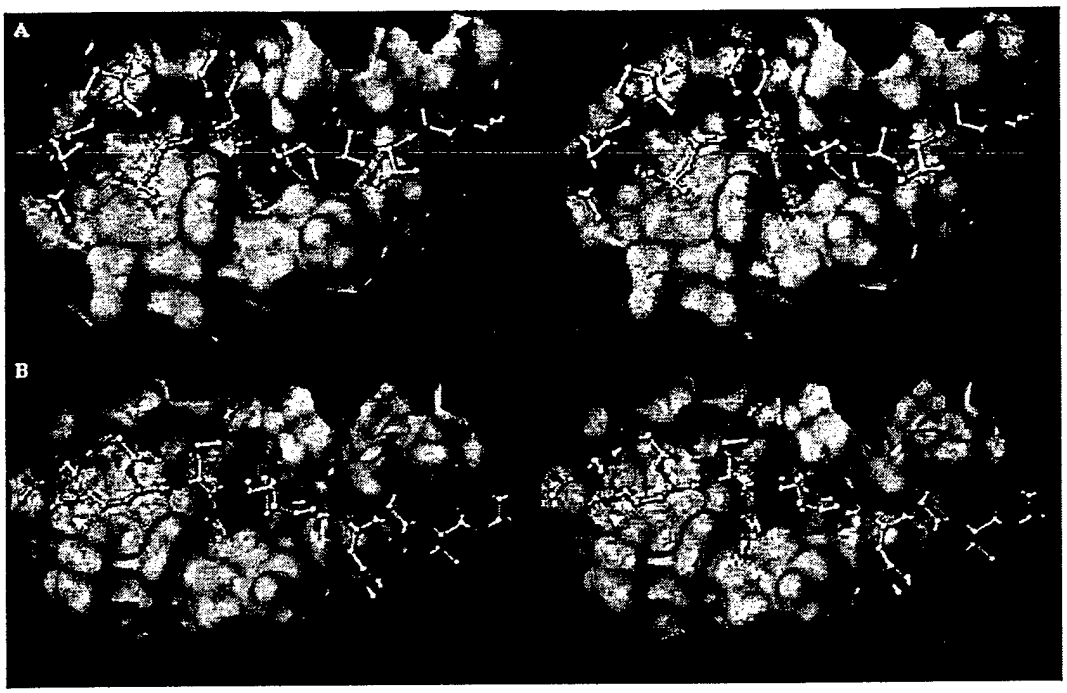


FIG. 1. **Three-dimensional models of distinct binding modes between substrate and MMP-9.** To ascertain whether peptides from different substrate families exhibit different modes of binding in the active site of MMP-9, models with two representative peptides binding to MMP-9 were constructed. The stereo view represents the energy-minimized complexes of peptide A10 (SGPLFYSVTA, *panel A*) and peptide C11 (SGRRLLSRTA, *panel B*) in the active site of the MMP-9 catalytic domain. The active site of MMP-9 is shown as a *green* Conolly surface, and the peptide atoms are colored according to type (carbon, *white*; oxygen, *red*; nitrogen, *blue*). The catalytic zinc ion is represented as a *yellow sphere*, and the backbone of MMP-9 outside the active site is shown in *magenta*. The *orange arrows* in both panels show the deep hydrophobic pocket at the $S_{1'}$ subsite. In *panel A*, the *white arrow* shows the Pro of substrate A10 penetrating well into the S_3 pocket. In contrast, a Gly from substrate C11 occupies this subsite (*panel B*).

side chain of Arg. In addition, the hydrophobic channel of the S_2 subsite contains the backbone carbonyl of Pro-180, which is likely to engage in electrostatic interaction with the basic side chain of Arg, stabilizing the interaction.

DISCUSSION

Because of their association with a number of diseases, the MMPs have received considerable attention as drug targets (15–17). Much of the effort in this area has focused on the design of small molecule antagonists with two primary features; 1) a hydroxamate moiety that binds to the proteases catalytic zinc, and 2) a rather large hydrophobic moiety that fits into the deep $S_{1'}$ pocket present in all MMPs (30). This synthetic strategy has focused structure-activity studies to essentially two positions within the catalytic pocket. An understanding of the interactions between the substrate and other key subsites within the catalytic pocket is lacking.

We have identified three families of peptide substrates for MMP-9 that each appear to interact differently with the catalytic cleft of the protease. Substrates in group I are cleaved most efficiently by MMP-9. These peptides all contain a Pro at P_3 and a large hydrophobic residue at $P_{1'}$. In this respect, the group I substrates are similar to collagen-like sequences that are known to be cleaved by the MMPs (19, 20). A prior analysis of a small series of synthetic peptides based on collagen, and containing a similar Pro-*X*-*X*-Hy core, showed k_{cat}/K_m ratios ranging from 340 to 1000 $mM^{-1}\ h^{-1}$. These values are ~100 fold lower than those exhibited by the best peptides reported here. Consequently, the added diversity afforded by phage libraries allowed us to identify better substrates and a wider range of substrates. Interestingly, many different MMPs recognize the Pro-*X*-*X*-Hy core sequence. Can subsite interactions at positions other than P_3 and $P_{1'}$ generate selectivity within this family of sequences?

Apparently, individual MMPs do exhibit a great deal of selectivity for peptides containing the P-*X*-*X*-Hy core sequence. For example, most of the MMP-9 substrates within group I are selective for MMP-9 over MMP-7 and MMP-13. Some of the k_{cat}/K_m ratios for these peptides are up to 47-fold higher for MMP-9 than for the other MMPs tested. These findings suggest that substrate selectivity can be conferred by subsite interactions outside of the dominant P_3 and $P_{1'}$ subsites that are common among MMP substrates.

Since phage substrate selections have been performed for MMP-3, -7, -9, and -13, an analysis of the frequency by which individual residues occupy distinct subsites can be used as a first test of this idea. This comparison reveals considerable distinction in the residues that occupy the P_2, $P_{1'}$, and $P_{2'}$ subsites. Nearly one-third of all group I substrates for MMP-9 contain Arg at P_2. Although Arg can also be found at P_2 in peptide substrates for MMP-13, its frequency is much lower than in the MMP-9 substrates we selected (47). Furthermore, Arg is rarely, if ever, found at P_2 in MMP-3 or -7 peptide substrates (32). Ser or Thr most frequently occupies the $P_{1'}$ position of the MMP-9 substrates. However, a Gly residue is preferred at this position by MMP-13, and Asp or Glu are preferred by MMP-3 and -7 (32, 47). Significant differences are also observed at $P_{2'}$. In the group I MMP-9 substrates, 23 of 28 substrates have Ser or Thr at $P_{2'}$, but neither residue is prevalent at this subsite in the substrates selected for the other MMPs (32, 47). These observations support the contention that subsite interactions other than P_3 and $P_{1'}$ have a significant impact on substrate selectivity among the MMP family.

Here we observed two additional families of substrates for MMP-9 that are distinct from the P-*X*-*X*-Hy family. Group II substrates were selected least frequently and contain only three members, making it difficult to identify subsite prefer-

TABLE III

A comparison of the MMP-9 substrate consensus with cleavage sites in known protein substrates

The group I MMP-9 consensus motif is compared to the cleavage sites in proteins known to be cleaved by MMP-9. The protein substrates are divided into two groups: collagen and non-collagen. The residues in the protein substrates that match the MMP-9 consensus are shown in bold. Scissile bonds are identified by ↓.

	P	R	S/T	↓	Hy	S/T
MMP-9 consensus						
collagen substrates						
α1(I) collagen	**P**	Q	G	↓	**V**	R
	P	S	G	↓	**L**	HyP
	P	A	G	↓	**V**	Q
CB3 α1(IV) collagen	**P**	S	G	↓	R	D
	P	P	G	↓	**I**	V
α1(V) collagen	**P**	P	G	↓	**V**	V
α2(V) collagen	**P**	G	G	↓	**V**	V
α1(XI) collagen	**P**	P	G	↓	**L**	R
Non-collagen substrates						
Aggrecan	**P**	E	N	↓	**F**	F
Tissue factor pathway inhibitor	**P**	L	K	↓	**L**	M
Galectin 3	**P**	G	A	↓	**Y**	H
Cartilage link protein	A	I	H	↓	**I**	Q
Myelin basic protein	H	F	F	↓	K	N
	G	L	**S**	↓	**L**	**S**
	A	S	D	↓	**Y**	K

ences outside of the Gly-Leu-(Lys/Arg) consensus motif that occupies the P_2, $P_{1'}$, and $P_{2'}$ positions. Group III represents a novel substrate preference for MMP-9. As a whole, group III substrates were cleaved far less efficiently than either group I or group II. Nevertheless, in the context of the whole phage, at least one of the phages in this family (C11) was cleaved nearly as well as most of the group I substrates. An overriding feature of the group III substrates is the presence of Arg at P_2 and often at P_1. Although this recognition specificity is unexpected, it is corroborated by three other pieces of data. First, eight substrates from group I, including the four best peptides in this study, also have Arg at the P_2 position. Second, when peptide A11 from group I was mutated to contain Arg at both P_2 and P_1, the k_{cat}/K_m ratio for the peptide nearly tripled. Third, docking studies with a representative group III peptide illustrate the favorable interaction of Arg with the S_2 subsite (Fig. 1).

A walk across the model of the catalytic pocket of MMP-9, subsite by subsite, illustrates many of the features that are likely to guide substrate interactions, and confer recognition specificity. On the non-primed sides of the scissile bond, the structural features of the S3 and S2 subsites are consistent with the corresponding substrates that we selected. As with most MMPs, the S_3 subsite is a hydrophobic pocket that introduces a deviation from linearity into the active-site cleft. This deviation explains why Pro is favored at position P_3. At S_2, one finds the side chain of Asp-410 contributing its hydrophilic characteristics, and providing a potential salt bridge for the guanidinium group of Arg that is often found at P_2. MMP-13 also has an Asp residue at this position, and Arg occasionally occupies the P_2 positions of its substrates. In contrast, however, MMP-7 displays an Ala at this site rather than an Arg (3, 42), and none of the MMP-7 peptide substrates contain Arg at P_2 (32).

Moving across the cleft, one encounters the S_1 and $S_{1'}$ pockets. The S_1 space of MMP-9 is an extended hydrophobic surface that can accommodate a wide variety of amino acids (41). In fact, this is what we find. Like other MMPs, the $S_{1'}$ subsite is a deep hydrophobic pocket, and it is likely to be the most important substrate recognition point in the active site. This notion is supported by the fact that Leu and Ile residues dominate the $P_{1'}$ position. As discussed previously, the depth of this subsite also guides the binding of most of the small molecule antagonists of the MMPs (1, 3, 42).

The $S_{2'}$ subsite, a small hydrophobic surface, is the final

TABLE IV

Identification of potential MMP-9 substrates

The group I MMP-9 consensus motif was used to query the SWISSPROT data base to identify potential physiological substrates of MMP-9. The search was limited to proteins that contained these potential MMP-9 cleavage sites. The putative protein substrates and potential cleavage sites are shown.

MMP-9 consensus:	P	R	S/T	Hy	S/T	Potential protein substrate (residues)
	P	R	T	I	T	Kallikrein 14 (180–184)
	P	R	T	I	S	Ladinin 1 (371–375)
	P	R	T	V	T	Endoglin (231–235)
	P	R	T	I	S	Endothilin receptor (82–86)
	P	R	S	L	T	Laminin α3 chain (1568–1572)
	P	R	S	L	S	Phosphate regulating neutral endopeptidase (566–570)
	P	R	T	I	S	ADAM 2 (299–302)
	P	R	S	L	T	Desmoglein 3 (579–583)
	P	R	S	I	T	Integrin β_5 (57–61)

significant recognition point for substrate in the cleft of MMP-9. The $S_{2'}$ subsite provides opportunity for flexible amino acids to adapt conformations that would allow interactions with the solvent on substrate binding at this position, including amino acids with small hydrophobic or large hydrophilic side chains. We find that Ser/Thr are preferred at this position. Even though two-thirds of the MMP-9 substrates derive their $P_{2'}$ position from an invariant Thr encoded by the vector, the remaining one-third derive Thr or Ser at their $P_{2'}$ positions by selection. Interestingly, the character of the $S_{2'}$ subsite differs among MMP-9, MMP-7, and MMP-13. In MMP-9 this subsite essentially comprises Asp-188, Gly-189, Leu-190, and Met-422 (D-G-L-M) (3, 42). The corresponding residues in MMP-7 are G-N-T-T, and in MMP-13 are S-G-L-I. These distinctions are consistent with the fact that the substrates for each protease also differ at $P_{2'}$ (see above).

Given this new understanding of substrate recognition by MMP-9, one can arrive at an optimal substrate consensus (Pro-Arg-(Ser/Thr)- ↓ -(Leu/Ile)-(Ser/Thr)). A comparison of this motif to previously reported substrates for MMP-9 reveals some interesting similarities and distinctions (Table III). For example, all of the collagen substrates and even aggrecan, tissue factor pathway inhibitor and galectin 3 contain a Pro at the P_3 position and a hydrophobic amino acid at the $P_{1'}$ position (6, 48, 50–52). However, none of the substrates contain residues found to be optimal at other positions. The remaining substrates have significantly lower homology to the MMP-9 consensus recognition motif of group I (8, 53).

These comparisons raise questions as to whether all of the physiologically relevant substrates for MMP-9 have been identified. We reasoned that a query of the protein data bases with the optimal recognition motif might reveal a short list of putative substrates. Indeed, this search revealed only nine human proteins that contain this sequence within their extracellular domain (Table IV). Although at this juncture these proteins can only be considered hypothetical substrates for MMP-9, many of them are functionally linked to MMP-9-related pathology. For example, MMP-9 influences skin blistering in the autoimmune disorder bullous pemphigoid (5, 55). Interestingly, two of the hypothetical substrates for MMP-9, ladinin 1 and desmoglein 3, are also associated with autoimmune skin blistering disorders (56, 57). Similarly, MMP-9 has been suggested as a regulator of the angiogenic switch in tumor development (11). Two of the hypothetical substrates for MMP-9, integrin β_5 and endoglin 1, are also involved in angiogenesis (49, 54, 58). Along with the presence of a potential MMP-9 cleavage site, these functional associations lead to important and testable hypotheses about physiologic substrates for MMP-9.

REFERENCES

1. Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F. X., McKay, D. B., and Bode, W. (1995) *Protein Sci.* **4**, 823–840
2. Bode, W., Fernandez-Catalan, C., Tschesche, H., Grams, F., Nagase, H., and Maskos, K. (1999) *Cell. Mol. Life Sci.* **55**, 639–652
3. Massova, I., Kotra, L. P., Fridman, R., and Mobashery, S. (1998) *FASEB J.* **12**, 1075–1095
4. Senior, R. M., Griffin, G. L., Fliszar, C. J., Shapiro, S. D., Goldberg, G. I., and Welgus, H. J. (1991) *J. Biol. Chem.* **266**, 7870–7875
5. Liu, Z., Zhou, X., Shapiro, S. D., Shipley, J. M., Twining, S. S., Diaz, L. A., Senior, R. M., and Werb, Z. (2000) *Cell* **102**, 647–655
6. Belaaouaj, A. A., Li, A., Wun, T. C., Welgus, H. J., and Shapiro, S. D. (2000) *J. Biol. Chem.* **275**, 27123–27128
7. Gijbels, K., Proost, P., Carton, M. H., Billiau, C. A., and Opdenakker, G. (1993) *J. Neurosci. Res.* **36**, 432–440
8. Proost, P., Damme, J. V., and Opdenakker, G. (1993) *Biochem. Biophys. Res. Commun.* **192**, 1175–1181
9. Vu, T. H., Shipley, J. M., Bergers, G., Berger, J. E., Helms, J. A., Hanahan, D., Shapiro, S. D., Senior, R. M., and Werb, Z. (1998) *Cell* **93**, 411–422
10. Engsig, M. T., Chen, Q. J., Vu, T. H., Pedersen, A. C., Therkidsen, B., Lund, L. R., Henriksen, K., Lenhard, T., Foged, N. T., Foged, N. T., Werb, Z., and Delaisse, J. M. (2000) *J. Cell Biol.* **151**, 879–889
11. Bergers, G., Brekken, R., Mcmahon, G., Vu, T. H., Itoh, T., Tamaki, K., Tanzawa, K., Thorpe, P., Itohara, S., Werb, Z., and Hanahan, D. (2000) *Nat. Cell Biol.* **2**, 737–744
12. Coussens, L. M., Tinkle, C. L., Hanahan, D., and Werb, Z. (2000) *Cell* **103**, 481–490
13. Leppert, D., Ford, J., Stabler, G., Grygar, C., Lienert, C., Huber, S., Miller, K. M., Hauser, S. L., and Kappos, L. (1998) *Brain* **121**, 2327–2334
14. Conant, K., McArthur, J. C., Griffin, D. E., Sjulson, L., Wahl, L. M., and Irani, D. N. (1999) *Ann. Neurol.* **46**, 391–398
15. Birkedal-Hansen, H. (1995) *Curr. Opin. Cell Biol.* **7**, 728–735
16. Birkedal-Hansen, H., Moore, W. G., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A., and Engler, J. A. (1993) *Crit. Rev. Oral Biol. Med.* **4**, 197–250
17. Basbaum, C. B., and Werb, Z. (1996) *Curr. Opin. Cell Biol.* **8**, 731–738
18. Lauer-Fields, J. L., Tuzinski, K. A., Shimokawa, K., Nagase, H., and Fields, G. B. (2000) *J. Biol. Chem.* **275**, 13282–13290
19. Seltzer, J. L., Akers, K. T., Weingarten, H., Grant, G. A., McCourt, D. W., and Eisen, A. Z. (1990) *J. Biol. Chem.* **265**, 20409–20413
20. Xia, T., Akers, K. T., Eisen, A. Z., and Seltzer, J. L. (1996) *Biochim. Biophys. Acta* **1293**, 259–266
21. Netzel-Arnett, S., Fields, G. B., Birkedal-Hansen, H., and Van Wart, H. E. (1991) *J. Biol. Chem.* **266**, 6747–6755
22. Netzel-Arnett, S., Sang, Q. X., Moore, W. G., Navre, M., Birkedal-Hansen, H., and Van Wart, H. E. (1993) *Biochemistry* **32**, 6427–6432
23. McGeehan, G. M., Bickett, D. M., Green, M., Kassel, D., Wiseman, J. S., and Berman, J. (1994) *J. Biol. Chem.* **269**, 32814–32820
24. Smith, G. P. (1993) *Gene (Amst.)* **128**, 1–2
25. Scott, J. K., and Smith, G. P. (1990) *Science* **249**, 386–390
26. Smith, G. P. (1985) *Science* **228**, 1315–1317
27. Koivunen, E., Arap, W., Valtanen, H., Rainisalo, A., Medina, O. P., Heikkila, P., Kantor, C., Gahmberg, C. G., Salo, T., Konttinen, Y. T., Sorsa, T., Rouslahti, E., and Pasqualini, R. (1999) *Nat. Biotechnol.* **17**, 768–774
28. Knauper, V., Cowell, S., Smith, B., Lopez-Otin, C., O'Shea, M., Morris, H., Zardi, L., and Murphy, G. (1997) *J. Biol. Chem.* **272**, 7608–7616
29. O'Connell, J. P., Willenbrock, F., Docherty, A. J., Eaton, D., and Murphy, G. (1994) *J. Biol. Chem.* **269**, 14967–14973
30. Galardy, R. E., Cassabonne, M. E., Giese, C., Gilbert, J. H., Lapierre, F., Lopez, H., Schaefer, M. E., Stack, R., Sullivan, M., and Summers, B. (1994) *Ann. N. Y. Acad. Sci.* **732**, 315–323
31. Fields, G. B., Van Wart, H. E., and Birkedal-Hansen, H. (1987) *J. Biol. Chem.* **262**, 6221–622 6
32. Smith, M. M., Shi, L., and Navre, M. (1995) *J. Biol. Chem.* **270**, 6440–6449
33. Madison, E. L., Coombs, G. S., and Corey, D. R. (1995) *J. Biol. Chem.* **270**, 7558–7562
34. Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R., and Madison, E. L. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7627–7631
35. Karas, M. (1996) *Biochem. Soc. Trans.* **24**, 897–900
36. Landry, F., Lombardo, C. R., and Smith, J. W. (2000) *Anal. Biochem.* **279**, 1–8
37. Vorm, O., Roepstorff, P., and Mann, M. (1994) *Anal. Chem.* **66**, 3281–3287
38. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal. Chem.* **68**, 850–858
39. Coombs, G. S., Bergstrom, R. C., Pellequer, J. L., Baker, S. I., Navre, M., Smith, M. M., Tainer, J. A., Madison, E. L., and Corey, D. R. (1998) *Chem. Biol.* **5**, 475–488
40. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162
41. Olson, M. W., Bernardo, M. M., Pietila, M., Gervasi, D. C., Toth, M., Kotra, L. P., Massova, I., Mobashery, S., and Fridman, R. (2000) *J. Biol. Chem.* **275**, 2661–2668
42. Massova, I., Fridman, R., and Mobashery, S. (1997) *J. Mol. Model.* **3**, 17–30
43. Morgunova, E., Tuuttila, A., Bergmann, U., Isupov, M., Lindqvist, Y., Schneider, G., and Tryggvason, K. (1999) *Science* **284**, 1667–1670
44. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997) Amber 5, University of California, San Francisco
45. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, DeBolt, S., Ferguson, D. M., Seibel, G. L., and Kollman, P. A. (1995) *Comput. Phys. Res.* **91**, 1–41
46. Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S., and Tschesche, H. (1994) *EMBO J.* **13**, 1263–1269
47. Deng, S. J., Bickett, D. M., Mitchell, J. L., Lambert, M. H., Blackburn, R. K., Carter, H. L., Neugebauer, J., Pahel, G., Weiner, M. P., and Moss, M. L. (2000) *J. Biol. Chem.* **40**, 31422–31427
48. Fosang, A. J., Neame, P. J., Last, K., Hardingham, T. E., Murphy, G., and Hamilton, J. A. (1992) *J. Biol. Chem.* **267**, 19470–19474
49. Wong, N. C., Mueller, B. M., Barbas, C. F., Ruminski, P., Quaranta, V., Lin, E. C., and Smith, J. W. (1998) *Clin. Exp. Metastasis* **16**, 50–61
50. Ochieng, J., Fridman, R., Nangia-Makker, P., Kleiner, D. E., Liotta, L. A., Stetler-Stevenson, W. G., and Raz, A. (1994) *Biochemistry* **33**, 14109–14114
51. Niyibizi, C., Chan, R., Wu, J. J., and Eyre, D. (1994) *Biochem. Biophys. Res. Commun.* **202**, 328–333
52. Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (eds) (1998) *Handbook of Proteolytic Enzymes*, pp. 1199–1211, Academic Press, San Diego
53. Nguyen, Q., Murphy, G., Hughes, C. E., Mort, J. S., and Roughley, P. J. (1993) *Biochem. J.* **295**, 595–598
54. Lin, E. C. K., Ratnikov, B. I., Tsai, P. M., Carron, C. P., Myers, D. M., Barbas, C. F. I., and Smith, J. W. (1997) *J. Biol. Chem.* **272**, 23912–23920
55. Liu, Z., Shipley, J. M., Vu, T. H., Zhou, X., Diaz, L. A., Werb, Z., and Senior, R. M. (1998) *J. Exp. Med.* **188**, 475–482
56. Amagai, M., Klaus-Kovtun, V., and Stanley, J. R. (1991) *Cell* **67**, 869–877
57. Marinkovich, M. P., Taylor, T. B., Keene, D. R., Burgeson, R. E., and Zone, J. J. (1996) *J. Invest. Dermatol.* **106**, 734–738
58. Gougos, A., and Letarte, M. (1990) *J. Biol. Chem.* **265**, 8361–8364

Analytical Biochemistry
doi:10.1006/abio.2001.5167, available online at http://www.idealibrary.com on IDEAL®

A Substrate Phage Enzyme-Linked Immunosorbent Assay to Profile Panels of Proteases

Steven J. Kridel, Emily Chen, and Jeffrey W. Smith[1]
Program on Cell Adhesion, Cancer Research Center, The Burnham Institute, La Jolla, California 92037

Received March 29, 2001

It is estimated that proteases comprise nearly 2% of the human genome. Given that the primary structure of all known proteases will soon be available, an important challenge is to define the structure-activity relationships that govern substrate hydrolysis. Ideally this would be accomplished on a genome-wide scale. To this end, we have developed a one-pot phage selection system that yields the substrate recognition profile of multiple proteases from a single round of selection. The system meets five key criteria: (i) multiple proteases can be analyzed simultaneously, (ii) prior knowledge of substrate preference is not required, (iii) information regarding substrate preferences on both side of the scissile bond is obtained, (iv) the system yields selective substrates that distinguish closely related proteases, and (v) semiquantitative information on substrate hydrolysis is obtained, allowing for the assignment of initial rank-order preferences. As an illustration, a phage selection with a mixture of thrombin and factor Xa (serine proteases) along with matrix-metalloproteinase-9 and atrolysin C (metalloproteinases) was performed. Peptide substrates were identified that (i) have high k_{cat}/K_m ratios, (ii) are selective for individual proteases, and (iii) match the sequences of known physiological substrates. © 2001 Academic Press

Proteases are key regulators of a wide range of physiologic processes (1, 2), and are recognized as important and tractable drug targets. A primary aspect of protease function is substrate recognition and hydrolysis, and information on substrate recognition is an excellent starting point for elucidating a protease's biological function. Such information is key for three reasons, (i) it provides structure-activity relationships that guide our fundamental understanding of catalysis, (ii) it can be used to interrogate gene and protein databases to identify putative physiologic substrates, and (iii) it can also guide the synthesis of pharmaceutical inhibitors. Traditional methods for establishing substrate recognition profiles involve the identification of the scissile bond within a given protein substrate. Then, synthetic peptide variants of this sequence can be screened in an iterative fashion to arrive at more optimized substrates. Even though it can be fruitful, this iterative strategy is biased toward the original substrate sequence and is also tremendously cumbersome. Furthermore, it is not amenable to high-throughput analysis.

Given that the first draft of the human genome is now complete (3, 4), the sequences of all proteases are essentially known. Current estimates indicate that proteases account for 2% of the human genome (5). Since this equates to less than 1000 genes, the creation of a library of all human proteases in recombinant expression systems and/or in purified form is well within current technological capabilities. Do we biochemically analyze each protease one at a time to derive function? Or, is there a functional genomics strategy that can be applied to rapidly characterize the substrate recognition profile of all proteases?

We reasoned that a phage display-based approach could be modified to characterize the substrate recognition profile of all proteases. Substrate phage display has already been employed to characterize the substrate preferences of a number of individual proteases (6–11). Phage display is advantageous for several reasons. First, libraries containing all possible combinations of a random hexamers/heptamers can be constructed. Second, substrates displayed on phage allow one to gain information on substrate preference on both sides of a scissile bond. This is not possible with synthetic libraries of fluorogenic substrates. Lastly,

[1] To whom correspondence should be addressed at The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037. E-mail: jsmith@burnham-inst.org.

0003-2697/01 $35.00

phage display precludes the identification of protein substrates that are then analyzed in an iterative fashion.

Phage display, when used in the traditional manner, is a heritable system, allowing one to evolve optimal substrates (or binders) through multiple rounds of selective pressure. It is this property, however, that typically prevents phage from being used to obtain substrate recognition profiles for multiple proteases at once. Using conventional phage selections, a few phage substrates that best meet the selection criteria ultimately dominate the phage population. Consequently, a conventional selection against a mixture of proteases would yield a small subset of substrates that are cleaved by only one or a few of the proteases within the mixture. Information on substrate recognition by other proteases in the mixture, or on rare substrates, will be lost as a consequence of enrichment.

Here we put forth a strategy for obtaining the substrate recognition profile of multiple proteases, simultaneously. It couples one round of substrate phage display with a substrate ELISA[2] to deconvolute the specificity of protease substrates. The system allows one to avoid problems caused by the heritable nature of phage display by activity profiling even the rarest of substrates after a single round of protease selection. Furthermore, it allows all possible substrates to be tested for any protease, individually, allowing one to compare and deconvolute the substrate preferences for large panels of proteases in single experiments. This report describes the ELISA-based analysis of phage from one round of selection, in a semiquantitative fashion, by four proteases.

MATERIALS AND METHODS

Restriction enzymes were purchased from New England Bio-Labs and Roche Molecular Biosciences. Thrombin and factor Xa were from Calbiochem. Oligonucleotides were synthesized by IDT Inc. (Iowa). Peptides were synthesized by Anaspec Inc. (San Jose, Ca.). All chemical and buffers were from Sigma and Fisher. The catalytic domain of MMP-9 was expressed and purified as described previously (12). Atrolysin C was provided by Dr. Jay. W. Fox (University of Virginia).

Construction of phage library. The substrate phage library was generated using a modified version of the fUSE5 phagemid (13). A FLAG epitope was engineered at the N-terminus of the geneIII protein by annealing oligonucleotides 5′-CCGGGTTTGTCGTCGTCGTCTTTGTAGTCGGTAC-3′ and 5′-CGACTACAAAGACGACGACGACAAAC-3′ and ligating them into fUSE5 at the *Kpn*I and *Xba*I restriction sites. The random hexamers were generated by PCR extension of the template oligonucleotide 5′-GGGGAGGCCGACGTGGCCGTCATCAGGCGGCTCAGGC-(NNK)$_6$ACGGCCTCTGGGGCCGAAAC-3′, where N is any nucleotide and K is either G or T. The template oligonucleotide also encodes a peptide linker (SGGSG) positioned in between the FLAG epitope and the random hexamer. An extension oligonucleotide 5′-AATTTCTAGTTTCGGCCCCAGAGGC-3′ and the template oligonucleotide were mixed and heated at 65°C for 2 min. The heating block was allowed to cool passively to 40°C to facilitate annealing of the extension oligonucleotide to the template oligonucleotide. Elongation of the template oligonucleotide was performed using Sequenase (USB) (14). The final cDNA product was precipitated with ethanol, resuspended in water, and digested with *Sfi*I. The DNA insert and fUSE5 were mixed and ligated at a 5:1 molar ratio and electroporated into *Escherichia coli* MC1061(F−). Several phage were selected for sequencing to confirm the random nature of the phage insert.

Selection of phage substrates by protease mixture. An aliquot of the substrate phage library (2×10^{10} phage) was incubated with the protease mixture (thrombin, factor Xa, MMP-9, and atrolysin C, at 2 μg/ml each) in 50 mM Tris, pH 7.4, 100 mM NaCl, 10 mM $CaCl_2$, 0.05% Brij-35, and 0.05% BSA for 1 h at 37°C. A control selection contained no protease. The phage that were cleaved by protease were separated from the noncleaved phage by immuno-depletion by the addition of 100 μg of M1 anti-FLAG monoclonal antibody (Sigma) followed by incubation with Pansorbin (Calbiochem). The cleaved phage remaining in the supernatant were amplified using K91 *E. coli* as described previously (6–8, 15).

Analysis of substrate hydrolysis by phage ELISA. Hydrolysis of individual phage substrates by each protease was measured using a modified ELISA. The wells of a 96-well microtiter plate (Titertek) were coated with anti-M13 antibody (Pharmacia, 2.5 μg/ml in PBS), overnight at 4°C. After coating, the wells were blocked for 1 h at room temperature in TBS-T (50 mM Tris, pH 7.8, 150 mM NaCl, 0.2% Tween 20) containing 10 mg/ml BSA. Phage were captured from 150 μl of overnight phage culture supernatant in the antibody-coated wells for 2 h at 4°C. The unbound phage were removed with three washes of ice-cold TBS-T. To asses hydrolysis, protease (2 μg/ml) was added to the appropriate wells in Incubation Buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, 0.05% BSA, 0.05% Brij-35) for 2 h at 37°C. Control wells lacked protease. The protease solution was removed and the wells were washed four times with ice-cold TBS-T. To measure protease hydrolysis of the peptides on phage, anti-FLAG polyclonal antibody was added to each well and

[2] Abbreviations used: ELISA, enzyme-linked immunosorbent assay; MALDI-TOF, matrix-assisted laser desorption ionization-time of flight; MM9, matrix metalloproteinase-9.

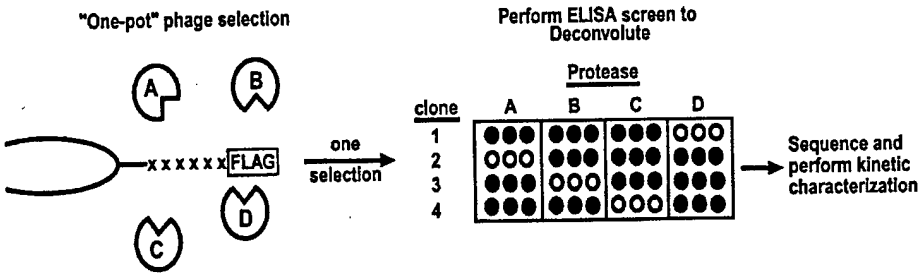


FIG. 1. Diagram of the "one-pot" protease substrate phage selection scheme. Protease substrates are selected from the phage library by incubation with four proteases, simultaneously. The phage cleaved by protease within the random hexamer substrate region of the phage are separated from the uncleaved phage with anti-FLAG antibody and Pansorbin precipitation. Individual phage clones representing protease substrates are then tested in a phage ELISA by each protease, independently, to identify which protease(s) cleaves which phage substrate. Clear represents positive signal and solid represents negative signal.

the plates were incubated at 4°C for 1 h. Binding of anti-FLAG antibody to FLAG epitope was measured with HRP-conjugated goat anti-mouse IgG (Bio-Rad) followed by detection at 490 nm. The extent of hydrolysis, taken as a measure of substrate efficiency, was calculated by the ratio of the OD at 490 nm of the protease-treated samples versus matched samples lacking protease. Each clone that was hydrolyzed more than 50% by any of the proteases was sequenced.

Determination of scissile bonds using synthetic peptides. The scissile bond within synthetic peptide substrates was determined using MALDI-TOF mass spectrometry. Each peptide (100 μM) was incubated with protease in 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$ for 2 h at 37°C. The mass of the cleavage products was determined using a Voyager DE-RP MALDI-TOF mass spectrometer (PerSeptive Biosystems, Framingham, MA). Following hydrolysis, the peptide samples were prepared according to methods described previously (16–18). In all cases, no peaks corresponding to alternative cleavage products were observed. Assignment of the scissile bond and numbering of residues on either side of the scissile bond were performed according to nomenclature established by Schechter and Berger (19).

Kinetic measurements of peptide hydrolysis. The kinetic parameters of substrate hydrolysis were measured using a fluorescamine incorporation assay that has been described previously (20–22). Briefly, protease was incubated with individual peptide substrates at concentrations ranging from 100 to 800 μM in 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$. At selected time points, reactions were stopped by the addition of 1,10-phenanthroline. Peptide hydrolysis was determined by the addition of fluorescamine followed by detection at λ_{ex} 355 nm and λ_{em} 460 nm. The data were transformed to double reciprocal plots ($1/[S]$ vs $1/v_i$) to determine K_m and k_{cat}. For some substrates, K_m and k_{cat} were not determined individually, but the specificity constant, k_{cat}/K_m, was derived by the equation: $k_{cat}/K_m = v_i/(E_0)\ (S_0)$.

RESULTS

A Phage System Designed for a One-Pot Selection

Our objective was to create a phage system capable of profiling the substrate specificity of multiple proteases in a single mixture. The crux of this phage system is a phagemid that was constructed from fUSE5, a vector that encodes the polyvalent expression of geneIII (13). The phagemid is modified such that a random hexamer is displayed as an N-terminal fusion with the geneIII protein. In addition, the phagemid was engineered to encode a FLAG epitope linked to the random hexamer via a peptide linker (SGGSG). The library comprises 2.4×10^8 independent transformants, assuring a 75% confidence that every combination of hexamer is represented (data not shown). Twenty clones were sequenced to verify the random nature of the hexamer insert. Greater than 95% of all uncleaved phage could be precipitated from solution with an anti-FLAG antibody (data not shown).

We reasoned that the advantages of phage display could be preserved, and the disadvantages overcome, by coupling a substrate phage selection to a semiquantitative ELISA (Fig. 1). The purpose of the ELISA is twofold. First, it allows the assignment of individual substrates to their cognate protease. Second, the ELISA gives a first approximation of the rate of hydrolysis by each protease. This allows the identification of optimal substrates, obviating the need for further rounds of selection.

One-Pot Selection of Substrates with Serine and Metalloproteinases

We reasoned that the substrate preference of multiple proteases could be determined from a one-pot selection provided that: (i) only one round of selection was used, and (ii) accurate measures of the hydrolysis of each phage clone could be derived from a separate assay. To test the one-pot phage system, we combined two serine proteases, thrombin and factor Xa, with two metalloproteinases, MMP-9 and atrolysin C.

The hexamer library described above (2×10^{10} phage) was incubated with an equimolar amount of the protease mixture for 1.5 h at 37°C. An anti-FLAG antibody was used to separate the cleaved and noncleaved phage by precipitation with Pansorbin. Cleaved phage were recovered from the supernatant and used to infect K91 bacteria. Phage clones were selected randomly from this pool for further characterization. The enrichment from one round of treatment was fivefold.

To assess the hydrolysis of individual clones, phage were captured from bacterial supernatants into microtiter plates using an anti-M13 antibody. The immobilized phage were treated with each protease and the extent of hydrolysis was measured by the release of the FLAG epitope. Results from the average of two such ELISAs are shown in Table 1. The extent of hydrolysis of each phage clone is expressed as the percentage of FLAG epitope removed by protease treatment. Approximately 17% of phage tested were scored as nonsubstrates (<20% hydrolysis), as expected from the one round of enrichment. From this table, the preferred substrates of each protease are clearly evident. Clones 81 and 89 are highly specific substrates for thrombin, whereas clones 74 and 86 exhibit similar selectivity for MMP-9. Substrates selective for factor Xa and atrolysin C are also evident. Some phage are cleaved by multiple proteases. For example, phage 1 and 17, which both contain an Arg–Ser bond, are cleaved by thrombin and factor Xa. This overlapping recognition is not surprising given the structural homology between the two proteases.

From the results shown in Table 1, groups of preferred substrates can be visualized for each protease. By aligning the substrates cleaved more than 60% by thrombin, two groups of substrates become apparent (Table 2). The consensus motif derived from a combination of the groups, (Gly/Pro)-(Arg/Lys)-↓-Ser-(Trp/Phe), is in agreement with the known substrate preference of thrombin and matches the sequence surrounding the scissile bond in several physiologic substrates for thrombin, including the thrombin receptor (protease-activated receptor), factor V and factor VII (23–25). Similarly, phage substrates identified for factor Xa contain a distinct recognition motif (Gly-Arg-↓-X) that matches well with the prothrombin activation site, the main physiologic target of factor Xa (Table 2) (26).

Comparison of One-Pot Selection to Conventional Substrate Phage Selection

Since phage selections are traditionally performed by three rounds of enrichment, we compared the one-pot approach to a conventional substrate phage selection. A group of phage from the one-pot selection, that were hydrolyzed at least 55% by MMP-9, were compared to phage we selected through three rounds of hydrolysis by MMP-9. As shown in Table 3, both methods yielded substrates with a consensus of Pro-X-X-X_{Hy}, where X is any amino acid and X_{Hy} is a hydrophobic amino acid. This motif is nearly identical to a motif that is reported to represent the optimal cleavage sequence of a number of other metalloproteinases (7, 20, 21). These observations show that a single round of selection is capable of yielding excellent substrates that match well with those obtained from conventional selections.

The Substrate ELISA as a Semiquantitative Indicator of Substrate Hydrolysis

Experiments were conducted to determine how well the substrate phage ELISA predicts the rank-order preference of analogous synthetic peptides. The hydrolysis of a series of phage substrates for MMP-9 was measured as a function of time (Fig. 2A) and compared to the k_{cat}/K_m ratio for the corresponding synthetic peptides (Fig. 2B). Phage A6 was hydrolyzed by 85% at 2 h and the matching peptide had a k_{cat}/K_m ratio of 61,000 M^{-1} s^{-1}. Phage A7 was hydrolyzed to a lesser degree, 50% at 2 h. Correspondingly, the A7 peptide had a k_{cat}/K_m ratio of 12,000 M^{-1} s^{-1}. The worst substrate, phage A16, was hydrolyzed by only 20% at 2 h, and its peptide exhibited the lowest k_{cat}/K_m ratio, only 1100 M^{-1} s^{-1}. Both methods then yield the same rank order of preference of substrates, A6 > A7 > A16. Similar observations were made when phage substrates for thrombin (Fig. 2C) were compared to the analogous synthetic peptides (Fig. 2D). Clearly the use of synthetic peptide substrates provides a more precise measure of K_m and k_{cat}, but the ELISA is sufficient to indicate a rank order of preference and a greater number of substrates can be explored quickly. Our experiments indicate that the substrate phage ELISA can distinguish substrates across a range of k_{cat}/K_m extending from approximately 1000 to 60,000 M^{-1} s^{-1}. Beyond these lower and upper limits, distinction is more difficult, but possible. It is likely, however, that manipulation of the ELISA conditions could alter the dynamic range of detection.

DISCUSSION

Two decades ago, the Holy Grail of biology appeared to be the knowledge of the sequence of the entire human genome. As the prospect of success in this endeavor increased, emphasis shifted toward knowing, or predicting, the structure of all human proteins (27–30). Now, even these genome-wide structural predictions appear within the realm of possibility (31). Consequently, emphasis should now shift toward large-scale

TABLE 1

Substrates Obtained from a "One-Pot" Selection

Clone	Peptide	Protease Th	Xa	M9	At
1	Q G R S W Q	100	49	9	0
2	T Q R K R S	25	55	13	96
3	A R G F R Q	100	26	49	35
4	L I N A V S	3	18	78	33
5	T D L R N A	35	16	29	31
6	H T L R K A	12	29	11	100
7	E F Q I N S	23	39	33	83
8	L S G F G L	11	19	88	45
9	Y R R V S A	22	55	13	49
10	S Q K G E T	80	43	27	77
11	R T Q D G P	45	49	27	51
12	P I P K M T	100	33	49	47
13	R T D R L N	26	29	20	37
14	T N L N T K	29	39	31	3
15	F S F R L T	13	20	32	99
16	N M S G R S	26	27	16	85
17	R S F L N L	100	52	42	3
18	R H F N G G	45	15	12	9
19	S H L A S S	10	16	19	55
20		21	20	15	32
21	L V Q R F H	14	7	8	35
22	D T W I L I	26	20	17	52
23	R H F N G G	7	16	30	63
24	N M L E Q T	1	8	7	61
25	V Y D S Q K	8	11	2	92
26		41	19	35	41
27	I P L R T V	100	14	24	100
28	S L L A I P	0	4	4	91
29	V P R M M C	100	4	100	29
30	S Q A H A S	0	0	4	86
31	S F M R L S	6	20	19	100
32	W V L S I R	9	36	95	20
33	V Q R G L T	10	7	13	83
34	P L R M N R	100	19	15	100
35	I A Q L N R	7	12	9	100
36	L L L L Y I	5	4	8	12
37	F H T E T H	13	25	16	76
38	D F T S W S	7	20	18	14
39	T R S Y S K	23	13	1	4
40	P P M T L Q	2	10	55	96
41	R S A H S Q	96	30	26	0
42	S A S L I R	100	11	10	77
43	V G Y P T L	1	8	4	0
44	P N T I Q L	1	3	9	8
45	N S T S S Y	6	6	7	33
46	M P L S S M	1	9	71	45
47	V D T L P F	16	18	12	54
48		17	19	12	21
49	K P K S N T	84	30	24	37
50		4	12	7	13
51		9	18	7	21
52	Q R Q G A L	5	3	9	62
53		18	21	21	28
54		21	23	21	31
55	C M G M T S	20	24	32	51
56		24	28	29	34
57	I V G V L Y	0	28	21	40
58	S Q A I T G	33	0	22	21
59		3	23	23	15
60	A L V N D N	20	24	14	57
61		25	0	12	48
62	T Q K V I S	4	6	9	96
63		3	6	4	12
64		0	7	1	8
65	A A M K K Q	0	8	2	97
66	F K A L S I	0	13	75	16
67		0	10	5	12
68	V A F K L Q	0	5	2	52
69		2	1	6	8
70	Y M K W M K	0	0	5	80
71		0	0	4	4
72		5	15	17	22
73		20	12	15	34
74	P Y S L R K	4	1	95	11
75		5	8	0	10
76		2	6	0	25
77		9	12	17	22
78	T M S R V L	14	7	2	73
79		0	2	2	5
80		0	4	0	4
81	Q P R H W S	100	3	8	3
82		0	0	0	19
83	R G R L F N	62	100	10	0
84		4	2	0	27
85	L S F M L L	0	4	0	21
86	P W N I Y L	0	0	97	2
87	I M P K G V	100	2	92	5
88		2	5	0	6
89	R S K M R M	100	2	0	4
90		0	11	6	19
91	R S S G V R	78	34	6	27
92	H S V L A L	0	9	0	20

Note. The substrate phage library was incubated with a mixture of thrombin (Th), factor Xa (Xa), MMP-9 (M9), and atrolysin C (At). Phage substrates were selected as described under Materials and Methods. Individual phage clones were tested as substrates in an ELISA assay. The extent of hydrolysis is expressed as a percentage of total relative to untreated controls, and is the average of two independent experiments. The hexamer sequence displayed by individual phages was derived from the DNA sequence of the phage inserts. Standard error between duplicate experiments was 15% or less of the measured values. Shaded clones are referred to in text, and nonsubstrates are outlined.

TABLE 2

Comparison of Phage Substrates to Known Physiologic Substrates

Thrombin substrates									
Group I	Q	G	R	G	R	S	W		
				G	R	S	F	L	N L
				A	R	G	F	T	Q T
			R	G	R	L	F	N	
				G	R	S	A	H	S Q
				G	R	S	K	M	R M
				G	R	S	S	G	V R
Group II		I	M	P	K	G	V		
		P	I	P	K	M	T		
			Q	P	R	H	W	S	
			V	P	R	M	M	C	
			K	P	K	S	N	T	
Consensus				(G/P)	(R/K)	S	(W/F)		
Physiologic substrates									
Thrombin receptor			D	P	R	S	F	L	
Factor VIII (site 1)			S	P	R	S	F	Q	
Factor VIII (site 2)			S	P	R	S	F	S	
Factor V			S	P	R	T	F	H	
Factor Xa substrates									
			R	G	R	L	F	N	
			Y	R	R	V	S	A	
				G	R	T	Q	D	G P
	T	Q	R	K	R	S			
				G	R	S	F	L	N L
Consensus				G	R	X			
Physiologic subtrate									
Prothrombin			E	G	R	T	A		

Note. Several of the optimum substrates for thrombin and factor Xa were aligned by sequence identity. The position of the scissile bond was then surmised based upon the hydrolysis of representative synthetic peptides. From this alignment, general consensus motif was derived by visual inspection, which in turn was compared to the amino acid sequence within known physiologic substrates for each protease.

measures and predictions of function. Here, we set forth a strategy to profile the catalytic space of large sets of proteases. In principle, this strategy makes it possible to define, and directly compare, the substrate recognition profile of nearly all *endo*-peptidases.

Four proteases representing two catalytic classes were used to test the system's ability to profile catalytic space. Within this set, we included two classical serine proteases, thrombin and factor Xa, for which the substrate specificity has been well characterized. We reasoned that a test of the system with "knowns" would indicate its ability to properly report substrate recognition. MMP-9 (matrix metalloproteinase) and atrolysin C (snake venom metalloproteinase) were used as representatives of metalloproteinases, whose recognition profiles are not well established (5). Within each class, the proteases share structural similarity and, reportedly, have overlapping substrate preferences. A comparison of the substrates selected for each protease to the reported substrate preference of each enzyme demonstrates the power of the approach.

A semiquantitative readout of phage hydrolysis is required to establish rank-order preferences among substrates for a single protease. Such information is also important for comparing hydrolysis of a single phage by more than one protease. To date, the analysis of phage substrates has been limited to cumbersome Western blots of protease-treated phage, or analysis of synthetic peptide corresponding to individual phage clones. A modified ELISA was devised to measure the hydrolysis substrates on phage. The substrate phage ELISA allowed for the identification of substrates for each protease (Table 1). In addition, when equal protease concentrations are used to cleave phage substrates, a rank-order preference for each substrate by its cognate protease can be assigned. The trends of phage hydrolysis observed in the ELISA correlate well with the k_{cat}/K_m ratios of the corresponding synthetic peptides (Fig. 2). The semiquantitative ELISA will discern substrates whose k_{cat}/K_m ratios differ by less than an order of magnitude. In addition, the ELISA can be modified into a time course or a protease dose-response curve (data not shown) to further analyze substrates of interest. Therefore, the scheme described in this study can establish accurate first approximations of substrate preferences between related proteases.

Given its importance to hemostasis, considerable effort has been devoted to defining the substrate recognition properties of thrombin (25, 32–35). Therefore, we felt it served as an excellent example as a "known" test case. Thrombin preferentially cleaves substrates with a sequence of Gly-Pro-Arg-↓-Ser-Phe-$X_{charged}$. Using the one-pot approach, we have identified two groups of substrates that generally match this structural preference (Table 2). In fact, a consensus motif derived from the two groups of substrates matches many of thrombin's physiologic substrates. Thrombin is somewhat unique in that it has a strict substrate

TABLE 3

Comparison of MMP-9 Substrates Obtained from "One-Pot" and Conventional Selections

	"One-pot" selection							Conventional selection						
	V	P	R	M	M	C		F	R	P	R	S	I	T
	I	M	P	K	G	V	T	L	L	P	L	Y	L	T
		P	Y	S	L	R	K			P	A	S	F	T S
		P	W	N	I	Y	I			P	R	Q	I	T A
	P	P	M	T	L	Q		R	S	P	K	S	L	T
	M	P	L	S	S	S	M			P	R	S	I	S N
								F	K	P	K	H	F	T
Consensus	P	X	X	Hy	Po			P	(R/K)	(S/X)	Hy	(S/T)		

Note. Phage substrates for MMP-9 obtained from the "one-pot" selection were compared to substrates selected from the same library using three rounds of enrichment. Sequences were aligned by inspection, based primarily on the P-X-X-Hy motif that has been reported as a common recognition motif for metalloproteases. Identical and conserved residues are in bold text. Hy; hydrophobic residues, Po, polar residues.

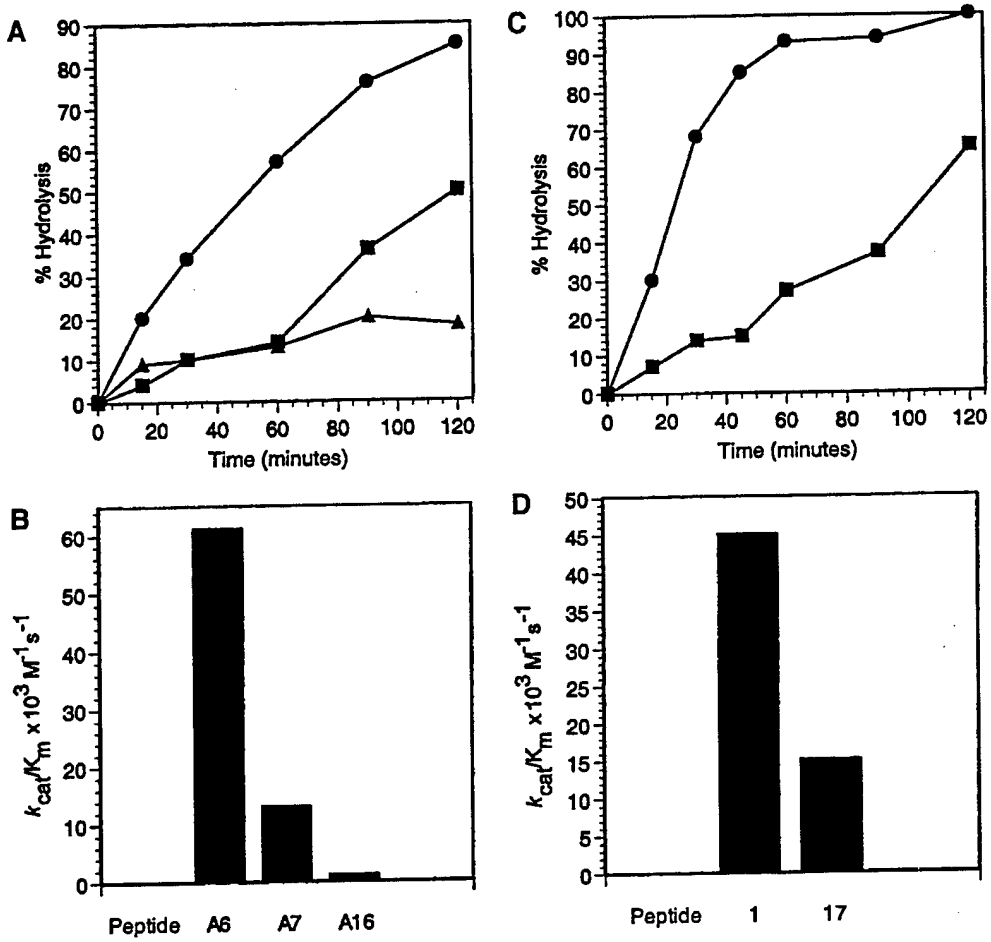


FIG. 2. Comparative analysis of substrate phage ELISA vs synthetic peptide hydrolysis. A comparison of relative rates of hydrolysis of substrate peptides on phage with the hydrolysis of the corresponding synthetic peptides was performed. Phage clones from a conventional selection with MMP-9, A6 (●), A7 (■), and A16 (▲) were cleaved by MMP-9 (A) and phage clones 17 (●) and 1 (■) from the "one-pot" system were cleaved by thrombin (C) as described under Materials and Methods. The rates of hydrolysis of the time course ELISA (A and C) and the k_{cat}/K_m of the corresponding synthetic peptides (B, MMP-9; and D, thrombin) were analyzed to arrive at a quantitative comparison. Each substrate ELISA was performed twice, and the standard error ranged from 5 to 20% between experiments. Each k_{cat}/K_m value was measured in three separate experiments with standard deviations ranging from 7 to 17%.

preference across multiple substrate subsites. However, our observations illustrate that the one-pot selection strategy provides the same type of information that has been culled from the analysis of thrombin's individual substrates and from profiling with peptide libraries, even with the background of three other proteases.

We also included factor Xa in the selection because of its close relationship to thrombin. Factor Xa is similar to thrombin in both fold and substrate preference (5, 26, 33). Perhaps not surprisingly, we identified a number of substrates that are cleaved by both factor Xa and thrombin. Similar to the Group I thrombin substrates, most of the substrates selected by factor Xa contain Gly-Arg at P_2 and P_1. Although the number of "good" substrates identified for factor Xa was small, the residues at $P_{1'}$ appear to be more variable than for thrombin. This finding is consistent with the fact that, in four physiologic substrates of factor Xa, there are three different $P_{1'}$ residues (26).

Even though thrombin and factor Xa have a similar substrate recognition profile, the one-pot selection revealed substrates that distinguish the two proteases. For example, substrates 81 and 89 are completely hydrolyzed by thrombin, but factor Xa has little effect on either. A further illustration of this ability is provided by a comparison of the substrates selected by MMP-9 vs atrolysin C. Although atrolysin C and MMP-9 are structurally related, phage substrates selective for each of the two proteases were obtained from the one-pot selection. Phage 74 and 86 are cleaved by MMP-9, but not by atrolysin. Conversely, phage 6 and 35 are cleaved by atrolysin but not MMP-9. The ability to identify substrates that distinguish closely related proteases is of great significance, especially if the strategy is to be applied to large groups of related proteases.

Prior work with substrate phage has employed subtraction protocols to obtain peptides that distinguish two closely related proteases (15). Yet, to obtain substrates selective for one protease over four or five related proteases, one would have to employ a matrix of subtractions. Therefore, the manipulations involved in subtractive substrate phage display are too complicated and involved to apply to a large set of proteases. The fact that the one-pot approach reveals substrates that are selective for related proteases, and can be applied to multiple proteases at once, indicates that it would be of great value in obtaining a panel of substrates that distinguish proteases in one family. In this regard, the one-pot strategy is unique.

We also compared the one-pot selection approach to a conventional substrate phage selection involving multiple rounds of selection. Using a conventional selection with MMP-9, we identified substrates with a consensus of Pro-Arg-Ser/Thr-↓-X_{Hy}-Ser/Thr, where X is any amino acid and X_{Hy} is a hydrophobic amino acid (preferably Leu or Ile) (11). This is a motif that appears to be a general recognition motif for MMPs (7, 20, 21, 36). Using the one-pot scheme, a similar motif was also identified for MMP-9 (Table 3). Equally as significant, the consensus we observed for MMP-9 is closely related to several known substrates of this protease (7, 37–39). In addition, none of these substrates were cleaved well by atrolysin, indicating specificity between related enzymes.

In principle one could profile a group of proteases individually to find substrates that describe the substrate preference of each protease and compare those substrates with each protease in the group. The one-pot system described here streamlines this process in several ways. In using only one round of selection, the number of phage manipulations is greatly reduced. Others have noted that, after multiple rounds of substrate phage selection, phage clones that are resistant to antibody depletion begin to dominate a population (7). This process avoids such inherent problems. In addition, this method reduces the number of individual experiments required to obtain the same data, thereby reducing the amount of reagents and time required for data acquisition. Perhaps most significant is that the ELISA-based approach gives a first approximation of a substrates efficiency by protease for individual substrates.

The one-pot system does have limitations. Even though it provides information on recognition specificity on both sides of the scissile bond, the position of that bond must be inferred from the known recognition specificity of the enzyme, or determined empirically with synthetic peptides, as done in this study. We envision that determination of the scissile bond could be streamlined by coupling to sensitive mass spectrometry to identify the mass of the peptide released by protease. In combination with the cDNA sequence of the phage insert, this measurement could rapidly yield the position of the scissile bond.

Another limitation centers on the quality of the protease used in the selection and ELISAs. As is true with any enzyme, precise kinetic measurements can only be made with enzymes in which the active site has been titrated. In this study we were unable to active-site titrate factor Xa. This inability likely explains why fewer substrates were obtained for factor Xa. However, this limitation does not preclude the identification of substrates that are of great value in further analysis of the protease functions. Empirical assessments of how many proteases could be profiled simultaneously will need to be determined as well. Despite these minor limitations, we envision that the one-pot approach can be applied on a larger scale, with greater throughput, to define substrates for large panels of proteases. It is also conceivable that, with modifications, a similar strategy could be applied to screen for substrates for other enzymes, like kinases. Likewise, we envision that other phage display systems, including M13 and T7, expressing cDNA libraries could be modified for use in this type of system, thus allowing for a more rapid determination of potential physiologic substrates for many classes of enzymes.

ACKNOWLEDGMENTS

The authors thank Dr. Jay W. Fox for the atrolysin C used in this study. This work was supported by Postdoctoral Fellowships AR08508 from the National Institutes of Health and 2PD0182 from the California Cancer Research Program (to S.J.K.), a graduate fellowship from the United States Army Breast Cancer Research Program (to E.C.), and by California Breast Cancer Research Program Grant 5JB0033, National Institutes of Health Grants AR42750 and CA69036, and Cancer Center Support Grant CA30199 (to J.W.S.).

REFERENCES

1. Basbaum, C. B., and Werb, Z. (1996) *Curr. Opin. Cell Biol.* **8,** 731–738.
2. Werb, Z., and Yan, Y. (1998) *Science* **282,** 1279–1280.
3. Consortium, I. H. G. S. (2001) *Nature* **409,** 860–921.
4. Venter, J. C., *et al.* (2001) *Science* **291,** 1304–1351.
5. Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (Eds.) (1998) Handbook of Proteolytic Enzymes, Academic Press, San Diego.
6. Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R., and Madison, E. L. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7627–7631.
7. Smith, M. M., Shi, L., and Navre, M. (1995) *J. Biol. Chem.* **270,** 6440–6449.
8. Coombs, G. S., Bergstrom, R. C., Pellequer, J. L., Baker, S. I., Navre, M., Smith, M. M., Tainer, J. A., Madison, E. L., and Corey, D. R. (1998) *Chem. Biol.* **5,** 475–488.
9. Madison, E. L., Coombs, G. S., and Corey, D. R. (1995) *J. Biol. Chem.* **270,** 7558–7562.
10. Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A., and Craik, C. S. (1998) *J. Biol. Chem.* **273,** 27364–27373.

11. Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., and Smith, J. W. (2001) *J. Biol. Chem.*, in press.
12. O'Connell, J. P., Willenbrock, F., Docherty, A. J., Eaton, D., and Murphy, G. (1994) *J. Biol. Chem.* **269,** 14967–14973.
13. Smith, G. P. (1985) *Science* **228,** 1315–1317.
14. Koivunen, E., Arap, W., Valtanen, H., Rainisalo, A., Medina, O. P., Heikkila, P., Kantor, C., Gahmberg, C. G., Salo, T., Konttinen, Y. T., Sorsa, t., Rouslahti, E., and Pasqualini, R. (1999) *Nature Biotechnol.* **17,** 1999.
15. Ke, S. H., Coombs, G. S., Tachias, K., Navre, M., Corey, D. R., and Madison, E. L. (1997) *J. Biol. Chem.* **272,** 16603–16609.
16. Vorm, O., Roepstorff, P., and Mann, M. (1994) *Anal. Chem.* **66,** 3281–3287.
17. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal. Chem.* **68,** 850–858.
18. Landry, F., Lombardo, C. R., and Smith, J. W. (2000) *Anal. Biochem.* **279,** 1–8.
19. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27,** 157–162.
20. Netzel-Arnett, S., Sang, Q. X., Moore, W. G., Navre, M., Birkedal-Hansen, H., and Van Wart, H. E. (1993) *Biochemistry* **32,** 6427–6432.
21. Netzel-Arnett, S., Fields, G. B., Birkedal-Hansen, H., and Van Wart, H. E. (1991) *J. Biol. Chem.* **266,** 6747–6755.
22. Fields, G. B., Van Wart, H. E., and Birkedal-Hansen, H. (1987) *J. Biol. Chem.* **262,** 6221–6226.
23. Kahn, M. L., Nakanishi-Matsui, M., Shapiro, M. J., Ishihara, H., and Coughlin, S. R. (1999) *J. Clin. Invest.* **103,** 879–887.
24. Shapiro, M. J., Weiss, E. J., Faruqi, T. R., and Coughlin, S. R. (2000) *J. Biol. Chem.* **275,** 25216–25221.
25. Le Bonniec, B. F., Myles, T., Johnson, T., Knight, C. G., Tapparelli, C., and Stone, S. R. (1996) *Biochemistry* **35,** 7114–7122.
26. Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K., and Engh, R. A. (1996) *J. Biol. Chem.* **271,** 29988–29992.
27. Davidson, A. R., Lumb, K. J., and Sauer, R. T. (1995) *Nature Struct. Biol.* **2,** 856–864.
28. Ku, J., and Schultz, P. G. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6552–6556.
29. Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K., and Baker, D. (1995) *Protein. Sci.* **4,** 1108–1117.
30. O'Neil, K. T., and Hoess, R. H. (1995) *Curr. Opin. Struct. Biol.* **5,** 443–449.
31. Zhang, B. H., Rychlewski, L., Pawlowski, K., Fetrow, J. S., Skolnick, J., and Godzik, A. (1999) *Protein Sci.* **8,** 1104–1115.
32. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) *Nature Biotechnol.* **18,** 187–193.
33. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7754–7759.
34. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) *Nature Biotechnol.* **18,** 187–193.
35. Vindigni, A., Dang, Q. D., and Di Cera, E. (1997) *Nature Biotechnol.* **15,** 891–895.
36. Deng, S. J., Bickett, D. M., Mitchell, J. L., Lambert, M. H., Blackburn, R. K., Carter, H. L., Neugebauer, J., Pahel, G., Weiner, M. P., and Moss, M. L. (2000) *J. Biol. Chem.* **40,** 31422–31427.
37. Patterson, B. C., and Sang, Q. A. (1997) *J. Biol. Chem.* **272,** 28823–28825.
38. Seltzer, J. L., Akers, K. T., Weingarten, H., Grant, G. A., McCourt, D. W., and Eisen, A. Z. (1990) *J. Biol. Chem.* **265,** 20409–20413.
39. Belaaouaj, A. A., Li, A., Wun, T. C., Welgus, H. J., and Shapiro, S. D. (2000) *J. Biol. Chem.* **275,** 27123–27128.

A Unique Substrate Recognition Profile for Matrix Metalloproteinase-2

[@]Emily I. Chen, [@]Steven J. Kridel, [&]Eric W. Howard [@]Weizhong Li, [@]Adam Godzik and

*[@]Jeffrey W. Smith

[@]Cancer Research Center

The Burnham Institute

10901 North Torrey Pines Road

La Jolla, Ca 92037

[&]Department of Pathology, BMSB 434,

University of Oklahoma Health Sciences Center

940 Stanton L. Young Blvd., Oklahoma City, Oklahoma 73104

*Corresponding author: jsmith@burnham.org

Running Title: Substrate Recognition by Matrix Metalloproteinases

Summary

The catalytic domains of the matrix metalloproteinases (MMPs) are structurally homologous, raising questions as to the degree of distinction, or overlap, in substrate recognition. The primary objective of the present study was to define the substrate recognition profile of MMP-2, a protease that was historically referred to as gelatinase A. By cleaving a phage peptide library with recombinant MMP-2, four distinct sets of substrates were identified. The first set is structurally related to substrates previously reported for other MMPs. These substrates contain the P-X-X↓X_{Hy} (X_{Hy} = Hydrophobic residues) consensus motif and are not generally selective for MMP-2 over the other MMPs tested. Two other groups of substrates were selected from the phage library with similar frequency. Substrates in group II contain the L/I-X-X↓X_{Hy} consensus motif. Substrates in group III contain a consensus motif with a sequence of X_{Hy}-S-X↓L, and the fourth set of substrates contain the H-X-X↓X_{Hy} sequence. Substrates in Group II, III and IV were found to be eight to almost two hundred-fold more selective for MMP-2 over MMP-9. To gain an understanding of the structural basis for substrate selectivity, individual residues within substrates were mutated, revealing that the P_2 residue is a key element in conferring selectivity. These findings indicate that MMP-2 and MMP-9 exhibit different substrate recognition profiles, and point to the P_2 subsite as a primary determinant in substrate distinction.

INTRODUCTION

The matrix metalloproteinases (MMPs) have important roles in a number of normal tissue remodeling events (1). The MMPs are also of interest as pharmaceutical targets because of their association with a number of pathological conditions (2). One key area of interest is tumor progression and tumor angiogenesis. Two closely related MMPs, MMP-2 and MMP-9, have been strongly linked to these events. For example, MMP-2 is up-regulated in invasive tumors, where it has been suggested to enact matrix degradation during tumor growth and invasion (3-5). Other reports link MMP-2 with the $\alpha_v\beta_3$ integrin in the process of tumor angiogenesis (6). The fact that mice lacking the gene for MMP-2 have reduced tumor angiogenesis also supports the idea that this protease has a causal role in tumor progression (7). MMP-9 is also implicated in angiogenesis, and is suggested to be part of the "angiogenic switch" that enacts the vascularization of tumors (8). Despite the fact that both MMP-2 and MMP-9 are linked to angiogenesis, there are strong indications that their mechanisms of action are distinct. There are circumstances where both proteases are present within tumors, but only one participates in the angiogenic switch (8). At present there is no mechanistic explanation for this distinction, but it raises the possibility that the two MMPs operate by cleaving distinct substrates.

The factors that govern substrate recognition and substrate distinction by the MMPs have not been fully elucidated. The structural features of the catalytic clefts of all MMPs are generally similar. All of the catalytic clefts contain a zinc ion and glutamic acid residue involved in catalysis (9). Furthermore, the MMPs contain a deep $S_{1'}$ pocket (10-15). This subsite has been exploited as a docking point for the vast majority of pharmaceutical inhibitors of the MMPs. Perhaps not surprisingly then, many of these antagonists are broad spectrum inhibitors (16), and

show evidence of unwanted side effects (17,18). There is also evidence to indicate that the common structural features of the catalytic cleft lead to an overlap in substrate recognition (19). The vast majority of known MMP substrates contain a large hydrophobic residue, which is frequently leucine, at the $P_{1'}$ position. This is consistent with the deep $S_{1'}$ pocket. Further similarity is observed at the P_3 position, where proline is often preferred. In fact, the P-X-X↓X_{Hy} motif appears to be an excellent substrate for a wide range of MMPs (20-22). However, recent work counters the notion that MMPs have an overlapping substrate recognition profile. Recent studies on both MMP-13 and MMP-9 show that a high degree of selectivity can be obtained for individual MMPs, even among substrates comprised of the P-X-X↓X_{Hy} motif (21,22).

Given these recent observations, we wondered whether very closely related MMPs would exhibit distinctions in substrate recognition. Thus, we focus on substrate recognition by MMP-2 and its closely related homolog, MMP-9. These two MMPs are unique in that they contain three type II fibronectin domains intercalated within their catalytic domains. Although these fibronectin domains are oriented away from the catalytic cleft, they are believed to mediate docking interactions with substrates like gelatin and with the natural inhibitors, TIMPs. Until now it was assumed that MMP-2 and MMP-9 have overlapping substrate recognition profiles. Here we use an unbiased substrate phage approach to obtain the substrate recognition profile of MMP-2. We find that like other MMPs, MMP-2 can cleave peptides containing the P-X-X-X_{Hy} sequence. However, three novel substrate motifs were also identified. These novel substrates are highly selective for MMP-2 over MMP-9. Consequently, the results of the present study challenge the idea that the two enzymes are functionally similar and provide a potential basis for their distinct biological roles.

Experimental Procedures

Source of Commercial Proteins and Reagents

MMP-7 (active enzyme), MMP-13 (pro-enzyme) and TIMP-2 were purchased from Chemicon (Temecula, Calif.). Ilomastat was purchased from AMS Scientific (Concord, Calif.). Restriction enzymes were from Roche Bioseciences or New England Biolabs. Oligonucleotides were synthesized by Integrated DNA Technologies, Inc (Iowa). Tissue culture media and reagents were from Irvine Scientific (Irvine, Calif.). All other reagents, chemicals, and plastic ware were from Sigma or Fisher.

Construction of the substrate phage library

The substrate phage library used in this study was generated using a modified version of the fUSE5 phagemid (23), as we have previously described (22). The library's primary features are a random peptide hexamer on the N-terminus of the gene III protein and a FLAG epitope positioned to the N-terminus of the hexamer. The substrate phage library represents 2.4×10^8 individual sequences, ensuring with 75% confidence that all possible sequences are represented (23).

Expression and purification of the catalytic domain of human MMP-2 and 9

The cDNA encoding the catalytic domain of MMP-2 was generated by PCR and cloned into the pCDNA3 expression vector (Invitrogen). The expression vector was used to transfect HEK 293 cells. Individual antibiotic resistant colonies were isolated with cloning rings, expanded, and then screened by RT-PCR for MMP-2 mRNA. Expression of MMP-2 protein was assessed by zymography of conditioned medium. For purification, cells were grown for five days in DMEM

with 10% FBS. The catalytic domains of MMP-2 and 9 were purified from the conditioned medium by gelatin-Sepharose chromatography as described previously (24,25) followed by ion exchange on Q-sepharose. Fractions containing MMP-2 or 9 were concentrated in a dialysis bag using Aquacide II (Calbiochem). The purity of MMP-2 and 9 was greater than 90% judging by silver stained acrylamide gels. The purified MMP-2 and 9 were stored at -70°C at concentrations ranging from 0.4 -1.3mg/ml.

Activation and active site titration of proteases

MMPs were activated by 2mM p-Aminophenylmercuric Acetate(APMA) at room temperature as previously described (24). Although APMA may be a less efficient way to activate MMP-2 and 9 than utilization of other MMPs as it leaves a bit of the propeptide (including the free cysteine) still attached, we chose APMA activation to avoid contamination of other proteases. Besides the presence of leftover propetide is less likely to interfere with the catalytic activity against peptide substrates. After activation, the activities of MMP-2 and MMP-9 were titrated using the hydroxamate inhibitor Ilomastat as previously described (22,26). The active sites of full length MMP-13 and MMP-7 were titrated with human TIMP-2. Briefly, 5-15 nM of each protease was pre-incubated with a range of TIMP-2 for five hours at room temperature. Residual MMP-13 activity was monitored by cleavage of MCA-Pro-Cha-Gly-Nva-His-Ala-Dpa-NH_2 (Calbiochem). Residual MMP-7 activity was monitored by cleavage of MCA-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2 (Calbiochem).

Selection of MMP-2 substrates from the phage library

The substrate phage library (2×10^{10} phage) was incubated with 2.5 µg/ml of MMP-2 in 50 mM Tris, pH 7.4, 100 mM NaCl, 10 mM $CaCl_2$, 0.05% Brij-35, and 0.05% BSA for 1 hour at 37°C.

A control reaction was performed without protease. The cleaved phage were separated from the non-cleaved phage by immuno-depletion. 100 μg of an anti-FLAG monoclonal antibody (Sigma) was added to the phage samples and then incubated for 18 hours with rocking at 4°C. The phage-antibody complexes were twice precipitated by the addition of 100 μl Pansorbin (Calbiochem). The cleaved phage remaining in the supernatant were amplified using K91 *E. coli* , and were then used for one additional round of substrate selection.

Monitoring phage hydrolysis by ELISA

Hydrolysis of individual phage substrates was measured using a modified ELISA that we have previously described (22). Briefly, phage from overnight cultures were captured into microtiter plates coated with anti-M13 antibody (Pharmacia, 2.5 μg/ml). The captured phage were incubated with MMP-2 (2.5 μg/ml) in Incubation Buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, 0.05% BSA, 0.05% Brij-35, 50 μM $ZnCl_2$) for 2 hours at 37°C. Control wells lacked protease. Following hydrolysis and extensive washing, anti-FLAG polyclonal antibody (1.8 μg/ml in TBS-T with 1 mg/ml BSA) was added to the wells and incubated for 1 hr. Following additional washing, the level of bound anti-FLAG antibody was quantified with an HRP-conjugated goat anti-rabbit IgG antibody (BioRad) followed by detection at 490 nm. The extent of hydrolysis of each phage was calculated by the ratio of the O.D. at 490 nm of the protease-treated samples versus samples lacking protease.

Mapping the position of scissile binds within peptide substrates

Peptides representing the phage inserts were synthesized to our specifications by Annaspec Inc.. The N-termini were acetylated and C-termini were synthesized as amides. The cleavage site

within peptide substrates was determined using MALDI-TOF mass spectrometry. MMP-2 (25 nM) was incubated with 200 μM of each peptide, independently, in 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$ for 2 hours at 37°C. The mass of the cleavage products was determined using a Voyager DE-RP MALDI-TOF mass spectrometer (PerSeptive Biosysytems, Framingham, MA). Following hydrolysis, the peptide samples were prepared for MALDI analysis according to methods described previously (27), and subsites within the peptide were designated according to the nomenclature of Schechter and Berger (28).

Quantifying the kinetics parameters of peptide hydrolysis

The kinetic parameters of substrate hydrolysis were measured using a fluorescamine incorporation assay that has been previously described (29-32). Briefly, MMP-2, MMP-9, MMP-7, or MMP-13 were incubated with individual peptide substrates at concentrations ranging from 100-800 μM in 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $CaCl_2$, 50 μM $ZnCl_2$. At selected time points the reactions were stopped by the addition of 1,10-phenanthroline. Peptide hydrolysis was determined by the addition of fluorescamine followed by detection at λ_{ex} 355 nm and λ_{em} 460 nm. The data were transformed to double reciprocal plots (1/[S] vs $1/v_i$) to determine K_m and k_{cat} (29-32). For some substrates, K_m and k_{cat} could not be determined individually, but the specificity constant, k_{cat}/K_m, was derived by the equation: $k_{cat}/K_m = v_i/(E_0)(S_0)$ assuming [S] is significantly lower than the K_m (19).

Assessing cleavage of Recombinant Eph receptors by MMP-2 and MMP-9

Recombinant fusion proteins between EphB1, EphB2 and the Fc domain of IgG were purchased from R&D Systems Inc.. In these constructs, the extracellular domain of rat EphB1 (amino acid

residues 1-538) and mouse EphB2 (amino acid residues 1-548) are fused to the Fc region of human IgG via a short polypeptide linker (33). The fusion proteins (1.8uM) were incubated for 4 hours at 37°C with 280 nM of either MMP-2 or MMP-9. Following incubation, samples were resolved by 10% SDS-PAGE, and samples were visualized by Coomassie staining. The N-terminus of the cleaved Eph B1 was determined by automated Edman degradation of protein blotted to PVDF membranes.

RESULTS

Activated MMP-2 was used to select optimal substrates from a phage display library that we have previously described (22). Following two rounds of phage hydrolysis and subsequent amplification, several individual phage clones were selected randomly and assessed for hydrolysis by MMP-2. The phage vector encodes a FLAG epitope that is positioned to the N-terminal side of the randomized peptide hexamer. Therefore, cleavage within the hexamer by MMP-2 was gauged by measuring the liberation of the FLAG epitope using an ELISA. Thirty individual clones were selected for sequencing based on the fact that they were cleaved by more than 25% when incubated for two hours with activated MMP-2 (60 nM). Four distinct groups of substrates are found among these clones (Table I).

The first group of substrates contains the P-X-X-X_{Hy} motif, where X_{Hy} represents a large hydrophobic residue. This motif appears to be a substrate for a number of different MMPs (20-22). Substrates in group II-IV represent novel recognition motifs for MMP-2. Group II substrates contain the I/L-X-X-X_{Hy} motif, in which the last hydrophobic residue is usually Ile or Leu. Group III substrates contain the X_{Hy}-S-X-L motif, where Ser and Leu are invariant. Group IV substrates are comprised of the H-X-X-X_{Hy} motif, which is similar to the cleavage site for MMP-2 within laminin-5 (34).

Assessing the selectivity of the MMP-2 Substrates.

We compared the rate of hydrolysis of a set of representative phage substrates by MMP-2 and MMP-9 (Figure 1). With the exception of one clone, A45, the phage from group I lacked selectivity for MMP-2 over MMP-9. In contrast however, all of the phage substrates from

groups II and III were highly selective for MMP-2 over MMP-9. The extent of hydrolysis of phage substrates from group IV was not compared using this assay because their rate of hydrolysis by MMP-2 was generally low. These substrates were characterized in more detail with the aid of synthetic peptides (see below).

Characterization of Synthetic Peptide Substrates

Representative peptides from groups I-IV were synthesized and used to characterize substrate hydrolysis in greater detail. The peptides were initially used to determine the position of the scissile bond within each motif by analyzing the cleavage products by MALDI-TOF mass spectrometry (Table II). In virtually all cases, the scissile bond directly precedes a large hydrophobic residue, which is frequently Ile or Leu (Table I). This feature is consistent with the presence of a deep binding pocket at the corresponding $S_{1'}$ subsite within MMP-2 (14).

The synthetic peptide substrates were also used to gain more detailed information on the selectivity of the substrate motifs for individual MMPs. We compared the rate of hydrolysis of each peptide by MMP-2, MMP-9, MMP-7 and MMP-13. Each enzyme was quantified by active site titration to ensure that accurate comparisons were obtained. The initial velocity of hydrolysis was measured across a range of peptide concentrations and reciprocal plots were used to derive K_m and k_{cat} for MMP-2, and the k_{cat}/K_m ratio for each enzyme (Table III). A double reciprocal plot of B74 peptide hydrolysis serves as an illustration of how K_m, and k_{cat} for MMP-2 were derived (figure 3). When cleaved by MMP-2, most of the substrates exhibited Michaelis constants in the mM range, with turnover rates ranging from 1 to 750 s^{-1}. Correspondingly, the k_{cat}/K_m ratios were all between 1.6 x 10^4 $M^{-1}s^{-1}$ and 2 x 10^5 $M^{-1}s^{-1}$. Most significantly, substrates

from groups II, III and IV exhibit k_{cat}/K_m ratios that are eight to three hundred and fifty-fold higher for MMP-2 than for the other MMPs. As a group, the substrates within group II show the most selectivity. The selectivity of these substrates for MMP-2 is conveniently illustrated by the ratio of k_{cat}/K_m for MMP-2 divided by the same value for the other MMPs (Table IV).

The Role of Residues at P_3 and P_2 in Conferring Substrate Selectivity

Several experiments were conducted to gain a better understanding of the structural basis for the selectivity of the substrates in group II-IV. Two hypotheses were tested. The first centered on the fact that neither group II nor group III substrates contain a Pro at the P_3 position. Since this residue is frequently found in substrates for other MMPs, and is also present at this position in group I substrates, we reasoned that its absence may be a defining feature of the selective substrates. To test this hypotheses, variants of three peptides from groups II and III were synthesized to contain a Pro at P_3, and their hydrolysis by each MMP was measured (Table III). Interestingly, the substitution of Pro into the P_3 position of peptide B74 and A13 increased the k_{cat}/K_m ratio for all MMPs. However, Pro substitution did not reduce the selectivity exhibited by the peptides completely. Hence, the lack of Pro at P_3 may partially involve in determining the selectivity but is not the primary determinant in the selectivity of these substrates for MMP-2 because peptides with Pro substitution still confer good selectivity to MMP-2 over MMP-9 (15 fold).

The second hypothesis was based on our recent characterization of the substrate recognition profile of MMP-9 (22). That study revealed a critical role or preference for Arg at the P_2 position. Since Arg is rarely present at P_2 in the substrates selected for MMP-2, we hypothesized

that substitution of Arg into P_2 might shift selectivity away from MMP-2 and toward MMP-9. Indeed, in the three peptides tested, B74, A13 and C9, the substitution of Arg into the P_2 position dramatically increased hydrolysis by MMP-9. This substitution also significantly decreased hydrolysis by MMP-2. In combination, these effects completely switch the selectivity ratio of the mutated peptides (Table IV). These findings underscore the significance of Arg at P_2 in facilitating substrate recognition by MMP-9, and also point to the important role of the S_2 subsite in distinguishing the activity of MMP-2 and MMP-9. Interestingly, the substitution of Arg at P_2 had minimal effects on the k_{cat}/K_m ratio of MMP-7 or MMP-13 (Table III), indicating that other features, that are still not understood, confer selectivity of these peptides for MMP-2 over MMP-7 and MMP-13.

We noticed that the sequence of substrate A21 is similar to an MMP-2 selective cleavage site within laminin-5 (34). Peptide A21 has a k_{cat}/K_m ratio that is eight-fold higher for MMP-2 over MMP-9. Interestingly though, it contains a rather large residue, Lys, at P_2. Because smaller residues are favored at P_2 by MMP-2, we mutated this Lys to Ala as an additional test of the P_2 residue in conferring selectivity. In addition, this mutation makes the sequence of the peptide match more closely with the cleavage site within laminin-5, which contains the HAAL sequence (34). The mutated peptide is hydrolyzed by MMP-2 more than ten-fold better than the A21 parent peptide. This mutation was without significant effect on hydrolysis by MMP-9. These observations support the idea that the P_2/S_2 interaction is key to distinguishing substrate recognition by MMP-2 and MMP-9.

Selective Hydrolysis of A Protein Substrate Containing the S-X-L Motif.

Ultimately, one would hope to be able to use the substrate recognition profiles obtained from substrate phage, and other substrate profiling strategies (35,36), to generate hypotheses regarding physiologic substrates. However, there have not been enough test cases to establish the rules for such extrapolations. As an initial step in this direction, we compared the ability of MMP-2 and MMP-9 to cleave Eph B1 and Eph B2. Eph B1 and B2 are tyrosine kinase receptors that are responsible for cell-cell signaling during neuronal development. These proteins contain putative cleavage sites that correspond rather closely to the substrates selected from the phage library by MMP-2. There are two potential cleavage sites within Eph B1. The first motif contains the sequence is S-I-S-S-L-W, which matches well with the X_{Hy}-S-X↓L motif of the group III substrates. This motif is positioned within a predicted β-strand within a fibronectin repeat in Eph B1 (37). A second potential cleavage site with a sequence of K-S-E-L, is located in a ten residue linker between the membrane spanning segment and the second type III repeat of Eph B1. This sequence does not precisely match the motif of the group III substrates, but it does contain the core S-X-L motif. In the recombinant form of Eph B1 used here, this putative cleavage site is positioned in a short segment between the second fibronectin type III repeat, and the Fc domain of IgG. Eph B2 contains only a single putative cleavage site, with a sequence of is Y-I-S-D↓L-L. This motif is positioned in a predicted β-strand in the second type III fibronectin repeat and it corresponds to the X_{Hy}-S-X↓L motif of group III substrates. EphB2 lacks the second predicted cleavage site, even though the recombinant protein contains an analogous linker.

The Eph B1 and Eph B2 fusion proteins were incubated with equimolar amounts (280 nM) of MMP-2 and MMP-9 for four hours. The extent of hydrolysis was gauged by SDS-PAGE (Figure 2). The Eph B1-Fc fusion protein was almost quantitatively cleaved by MMP-2 (Figure 2, lane

2). The extent of cleavage by MMP-9 was far lower (Figure 2, Lane 3), which is not surprising considering the selectivity exhibited by the group II substrates. Neither protease cleaved the Eph B2 fusion protein. The site of hydrolysis within Eph B1 was determined by sequencing the N-terminus of one of the released fragments. The amino acid sequence of this fragment indicates that the protein was cleaved at the sequence D-D-Y-K-S-E ↓ L-R-E, that is found within the ten residue linker of Eph B1. This is one of the predicted cleavage sites in EphB1. These findings illustrate that motifs found to be selective for MMP-2 by substrate phage display, can act as selective substrates within the context of whole proteins. Equally as important however, this experiment illustrates that the three dimensional conformation of the putative cleavage site will, to a large degree, control the extent of hydrolysis. Therefore, we conclude that meaningful genome-wide predictions on putative physiologic substrates will require the incorporation of additional information and computational filters (see Discussion).

Discussion

Despite their high structural homology, MMP-2 and MMP-9 are reported to have distinct biological roles. For example, in tumors in the RIP-Tag mouse, where both proteases are present, only MMP-9 seems to be causally involved in the angiogenic switch of these tumors (8). Similarly, in the process of platelet aggregation, MMP-2 promotes aggregation, but MMP-9 inhibits aggregation (38). The reasons for these differences in function are not clear. One possible explanation could be that the two enzymes have distinct substrate recognition profiles. Until now however, there has been no systematic and unbiased comparison of these profiles. Here, we defined the substrate recognition profile of MMP-2, and compared it to that of the closely related MMP-9. The findings of the study show that although two MMPs can cleave a common set of substrates with the P-X-X-X_{Hy} sequence, MMP-2 hydrolyzes an additional array of peptide substrates. These observations support the idea that differences in biological function of MMP-2 and MMP-9 could stem from their action on distinct physiologic substrates.

We found that like other MMPs, MMP-2 efficiently cleaves peptide substrates that contain the P-X-X-X_{Hy} motif. This observation is consistent with the idea that this motif is recognized universally by the MMP family. The sequence of this motif is generally complementary with the structural features within the catalytic clefts of the MMPs, including a deep $S_{1'}$ binding pocket, and a general deviation from linearity extending from the S_3 position, which is occupied by the invariant Pro in this group of substrates. Even within the context of this commonly recognized motif, some substrate selectivity is evident. Substrates for MMP-13 that contain this motif exhibit good selectivity for MMP-13 over three other MMPs (21). Similarly, the MMP-9 substrates with this motif that we recently characterized are selective for MMP-9 over MMP-7

and MMP-13. Interestingly, those substrates were found to be cleaved equally well by MMP-9 and MMP-2 (data not shown). Here we observed one substrate with the P-X-X-X_{Hy} sequence, phage A45, that is selective for MMP-2 over MMP-9, but such selectivity among the two gelatinases does not appear to be a general property of substrates containing P-X-X-X_{Hy}.

Three of the novel groups of substrates for MMP-2 were characterized in detail and were found to be selective for MMP-2 over the other MMPs tested. These substrates contain consensus motifs of L/I-X-X-X_{Hy}, X_{Hy}-S-X-L, and H-X-X-X_{Hy}. All three sets of substrates have Michaelis constants in the low mM range, and turnover rates of several hundred per second. They also contain large hydrophobic residues at the $P_{1'}$ position, a feature that is consistent with the depth of the $S_{1'}$ pocket. The dominant role of this interaction to substrate recognition by the MMPs is underscored by presence of a hydrophobic residues at this position in virtually all substrates selected for MMPs out of randomized libraries (20-22). The P_3 residue of group II and III substrates also has a hydrophobic character. In group II this position is an invariant Ile or Leu. Interestingly, an invariant Ser at P_2 is the feature that distinguishes substrates in group II and III. The group IV substrates are unique in that the P_3 residue is occupied by histidine. We did not identify a sufficient number of substrates within this group to arrive at a consensus at other positions.

The findings presented here also begin to provide a structural basis for substrate selectivity between MMP-2 and MMP-9. In our selection of substrates for MMP-9 (22), we found Arg to be preferred at P_2, whereas Arg was rarely observed at this position in the substrates selected for MMP-2. Rather, the MMP-2 substrates frequently display a relatively small residue at P_2. For

example, this position occupied by an invariant Ser in the group III substrates. When Arg is substituted into the P_2 position in the MMP-2 substrates, a marked shift in substrate recognition was observed. In all cases the rate of hydrolysis by MMP-2 was decreased, and hydrolysis by MMP-9 was increased. In fact, this substitution increased the k_{cat}/K_m ratios for MMP-9 between 25 and 60-fold. Further support for the role of the P_2 residue in substrate recognition comes from analysis of the group IV substrate A21, which contains Lys at P_2. We converted this Lys to Ala in order to more closely mimic the natural sequence in laminin-5, which is known to be selectively cleaved by MMP-2. Substitution of this Lys to Ala dramatically increased the k_{cat}/K_m ratio for MMP-2, providing further support for the idea that the S_2 subsite within MMP-2 is sterically hindered. It is interesting to note that the substitution of Arg into P_2 of the substrate had little effect on the already poor recognition of these motifs by MMP-7 and MMP-13. Hence, the primary effect of substitutions at P_2 is a switch in substrate recognition by the closely related gelatinases.

Analysis of the structure of the S_2 subsite within MMP-2 (14) and MMP-9 reveals a potential structural basis for this distinction in substrate recognition at the P_2 position (Figure 4). The S_2 pocket is remarkably similar in both enzymes, save the presence of Glu-412 in MMP-2 which is replaced by an Asp in MMP-9. Our prior docking of substrates containing Arg at P_2 into the catalytic cleft of MMP-9 indicated a favorable interaction between the positively charged guanidino group of Arg with the acidic side chain of Asp-410 (22). Since Glu contains an additional methylene group in its side chain, Glu-412 would be expected to extend further into the S_2 pocket of MMP-2 (orange arrow). Thus Glu-412 is expected to occlude the S_2 subsite, and hinder the docking of substrates which contain larger residues, like Arg, at the P_2 position.

Another distinction between the two proteases is observed at Ala-196 of MMP-2, which is replaced by Pro-193 in MMP-9. The pyrrolidine ring of Pro-193 of MMP-9 extends further out into the S_3 space. This Pro could potentially interfere with docking of extended residues like Leu and Ile that are often found at P_3 in the MMP-2 substrates.

Aside from the differences in structure within the catalytic clefts of MMP-2 and MMP-9, we must also consider the possibility that these proteases can assume distinct conformations that are not revealed by the existing crystal structure (10-13,39) and models (40). This possibility is supported by the fact that unbiased searches reveal four separate sets of substrates for MMP-2 (this report) and three families of substrates for MMP-9 (22). Thus, it is conceivable that the two pockets, although containing virtually the same resides, could adopt different conformations because of constraints imposed at ancillary regions of the protease. Although there is little direct evidence for this possibility, it does provide an alternative hypothesis to explain their differences in substrate recognition. Such conformational modulation of the catalytic pocket could also relate to the biology behind the two MMPs because their binding to other proteins, like the $\alpha v\beta 3$ integrin (6), could conceivably regulate substrate recognition. One might also hypothesize that the binding of either protease to collagen through its type II fibronectin repeats could provide similar conformational regulation.

Ideally, one would like to use the information obtained from substrate profiling approaches, like the one reported here, and elsewhere (35,36), to predict the physiologic and pathophysiologic substrates for proteases. Our findings indicate that such predictive strategies have merit but that they will require additional refinement. For example, it is encouraging that the differences in the

substrate phage profiles for MMP-2 and MMP-9 are reflected by the selective cleavage of Eph B1 by MMP-2. It is equally encouraging that the closely related homologue Eph B2, which lacks the predicted MMP-2 cleavage site, is not cleaved by MMP-2. The group IV substrates that contain the H-X-X-X_{Hy} motif match closely with the selective MMP-2 cleavage site in laminin-5 (sequence of H-A-A ↓ L-T-S) (34), an observation that provides additional support for the idea that substrate profiling could ultimately have predictive utility.

However, it is evident that if they are to be applied on a genome-wide scale these predictive methods require further refinement. Even though the sequences within Eph B1 and laminin-5 are similar to the substrate motifs from phage, they are not identical. Consequently, using the phage profiles to arrive at a consensus recognition motif that is based on the physical properties of preferred residues (e.g. large hydrophobic vs. small hydrophilic) rather than actual residues, is worthy of exploration. Building from here though, additional filters or constraints will need to be applied to narrow the number of putative substrates. For example, one could limit searches for putative substrates to proteins expressed in the appropriate cellular compartment or extracellular space. Further constraints could be imposed based on information obtained from gene expression profiling, which will reveal all genes that are co-expressed with any given protease. Our findings also indicate that the degree to which a predicted cleavage site is exposed to solvent should be taken into consideration. Since automated medium resolution structural predictions can now be made across entire genome (41), it is not unreasonable to suggest that this information could be used as an additional filter to identify proteins with accessible cleavage sites.

Acknowledgments

This study was supported by NIH grants AR42750, CA82713, CA69306, and a grant 5JB003 from the California Breast Cancer Research Program to JWS. EC was supported by a pre-doctoral fellowship from the U.S. D.O.D. Breast Cancer Research Program. SJK was supported be postdoctoral fellowship ZPD0812 from the NIH. Additional support was derived from NIH grant GM60049 to AG and Cancer Center Support Grant CA30199.

References

1. Parks, W. C. a. M., R. P. (1998) *Matrix Metalloproteinases.* Biology of Extracellular Matrix Series (Mecham, R. P., Ed.), Academic Press, San Diego

2. Greenwald, R. A., Zucker, S., and Golub, L. M. (1999) *Inhibition of Matrix Metalloproteinases Therapeutic Applications*, 878, Annals of the New York Academy of Sciences, New York

3. Poulsom, R., Hanby, A. M., Pignatelli, M., Jeffery, R. E., Longcroft, J. M., Rogers, L., and Stamp, G. W. (1993) *Journal of Clinical Pathology* 46(5), 429-36

4. Davies, B., Miles, D. W., Happerfield, L. C., Naylor, M. S., Bobrow, L. G., Rubens, R. D., and Balkwill, F. R. (1993) *British Journal of Cancer* 67(5), 1126-31

5. Boag, A. H., and Young, I. D. (1994) *American Journal of Pathology* 144(3), 585-591

6. Brooks, P. C., Stromblad, S., Sanders, L. C., von Schalscha, T. L., Aimes, R. T., Stetler-Stevenson, W. G., Quigley, J. P., and Cheresh, D. A. (1996) *Cell* 85(5), 683-93

7. Itoh, T., Tanioka, M., Yoshida, H., Yoshioka, T., Nishimoto, H., and Itohara, S. (1998) *Cancer Research* 58(5), 1048-51

8. Bergers, G., Brekken, R., Mcmahon, G., Vu, T. H., Itoh, T., Tamaki, K., Tanzawa, K., Thorpe, P., Itohara, S., Werb, Z., and Hanahan, D. (2000) *Nature Cell Biology* 2, 737-744

9. Birkedal-Hansen, H., Moore, W. G., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A., and Engler, J. A. (1993) *Critical Reviews in Oral Biology & Medicine* 4(2), 197-250

10. Grams, F., Reinemer, P., Powers, J. C., Kleine, T., Pieper, M., Tschesche, H., Huber, R., and Bode, W. (1995) *European Journal of Biochemistry* 228(3), 830-41

11. Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S., and Tschesche, H. (1994) *EMBO Journal* 13(6), 1263-9

12. Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F. X., McKay, D. B., and Bode, W. (1995) *Protein Sci* 4(5), 823-40

13. Gomis-Ruth, F. X., Maskos, K., Betz, M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G. P., Bartunik, H., and Bode, W. (1997) *Nature* 389(6646), 77-81

14. Morgunova, E., Tuuttila, A., Bergmann, U., Isupov, M., Lindqvist, Y., Schneider, G., and Tryggvason, K. (1999) *Science* 284(5420), 1667-70

15. Lovejoy, B., Cleasby, A., Hassell, A. M., Longley, K., Luther, M. A., Weigl, D., McGeehan, G., McElroy, A. B., Drewry, D., and Lambert, M. H. (1994) *Science* 263(5145), 375-7

16. Brown, P. D., and Giavazzi, R. (1995) *Annals of Oncology* 6(10), 967-74

17. Parsons, S. L., Watson, S. A., and Steele, R. J. (1997) *European Journal of Surgical Oncology* 23(6), 526-31

18. Nemunaitis, J., Poole, C., Primrose, J., Rosemurgy, A., Malfetano, J., Brown, P., Berrington, A., Cornish, A., Lynch, K., Rasmussen, H., Kerr, D., Cox, D., and Millar, A. (1998) *Clinical Cancer Research* 4(5), 1101-9

19. Netzel-Arnett, S., Sang, Q. X., Moore, W. G., Navre, M., Birkedal-Hansen, H., and Van Wart, H. E. (1993) *Biochemistry* 32(25), 6427-32

20. Smith, M. M., Shi, L., and Navre, M. (1995) *J Biol Chem* 270(12), 6440-9

21. Deng, S. J., Bickett, D. M., Mitchell, J. L., Lambert, M. H., Blackburn, R. K., Carter, H. L., Neugebauer, J., Pahel, G., Weiner, M. P., and Moss, M. L. (2000) *Journal of Biological Chemistry* 40(6), 31422-31427

22. Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., and Smith, J. W. (2001) *Journal of Biological Chemistry* 276(23), 20572-20578

23. Scott, J. K., and Smith, G. P. (1990) *Science* 249(4967), 386-90

24. O'Connell, J. P., Willenbrock, F., Docherty, A. J., Eaton, D., and Murphy, G. (1994) *Journal of Biological Chemistry* 269(21), 14967-73

25. Knauper, V., Cowell, S., Smith, B., Lopez-Otin, C., O'Shea, M., Morris, H., Zardi, L., and Murphy, G. (1997) *J Biol Chem* 272(12), 7608-16

26. Galardy, R. E., Cassabonne, M. E., Giese, C., Gilbert, J. H., Lapierre, F., Lopez, H., Schaefer, M. E., Stack, R., Sullivan, M., and Summers, B. (1994) *Annals of the New York Academy of Sciences* 732, 315-23

27. Landry, F., Lombardo, C. R., and Smith, J. W. (2000) *Anal. Biochem.* 279, 1-8

28. Schechter, I., and Berger, A. (1967) *Biochemical & Biophysical Research Communications* 27(2), 157-62

29. Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R., and Madison, E. L. (1995) *Proc.Natl.Acad.Sci.U.S.A.* 92, 7627-7631

30. Coombs, G. S., Bergstrom, R. C., Pellequer, J. L., Baker, S. I., Navre, M., Smith, M. M., Tainer, J. A., Madison, E. L., and Corey, D. R. (1998) *Chemistry and Biology* 5(9), 475-488

31. Fields, G. B., Van Wart, H. E., and Birkedal-Hansen, H. (1987) *Journal of Biological Chemistry* 262(13), 6221-6

32. Netzel-Arnett, S., Fields, G. B., Birkedal-Hansen, H., and Van Wart, H. E. (1991) *Journal of Biological Chemistry* 266(11), 6747-6755

33. Lhotak, V., Greer, P., Letwin, K., and Pawson, T. (1991) *Molecular & Cellular Biology* 11(5), 2496-502

34. Giannelli, G., Falk-Marzillier, J., Schiraldi, O., Stetler-Stevenson, W. G., and Quaranta, V. (1997) *Science* 277(5323), 225-8

35. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) *Proceedings of the National Academy of Sciences of the United States of America* 97(14), 7754-9

36. McGeehan, G. M., Bickett, D. M., Wiseman, J. S., Green, M., and Berman, J. (1995) *Methods in Enzymology* 248, 35-46

37. Tang, X. X., Biegel, J. A., Nycum, L. M., Yoshioka, A., Brodeur, G. M., Pleasure, D. E., and Ikegaki, N. (1995) *Genomics* 29(2), 426-37

38. Fernandez-Patron, C., Martinez-Cuesta, M. A., Salas, E., Sawicki, G., Wozniak, M., Radomski, M. W., and Davidge, S. T. (1999) *Thrombosis & Haemostasis* 82, 1730-5

39. Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K., and Engh, R. A. (1996) *Journal of Biological Chemistry* 271(47), 29988-92

40. Massova, I., Fridman, R., and Mobashery, S. (1997) *J. Mol. Model.* 3(1), 17-30

41. Zhang, B. H., Rychlewski, L., Pawlowski, K., Fetrow, J. S., Skolnick, J., and Godzik, A. (1999) *Protein Science.* 8(5), 1104-1115

42. Sali, A., and Blundell, T. L. (1993) *Journal of Molecular Biology* 234(3), 779-815

Footnotes:

APMA = p-Aminophenylmercuric Acetate

BSA = Albumin, Bovine

MALDI-TOF = Matrix-Assisted Laser Desorption Ionization Time of Flight

Figure Legends

Figure 1. Phage substrates are selective for MMP-2 over MMP-9.

The ability of MMP-2 (solid bars) and MMP-9 (open bars) to cleave substrate selected from the phage library were compared using the phage ELISA procedure described under Methods. Immoblized anti M13 antibody was used to capture individual phages onto 96-well microtiter wells. The captured phage were cleaved with 2.5ug/ml of MMP-2 or MMP-9. The extent of cleavage within the phage insert was assessed by measuring the release of the FLAG epitope. Results are presented as the percentage of hydrolysis compared to non-treated control phage. This experiment was repeated three times, yielding nearly identical results in each repetition.

Figure 2. EphB1 is cleaved by MMP-2 but not by MMP-9.

The ability of MMP-2 to cleave predicted sites within EphB1 was tested using a recombinant fusion protein between Eph B1 and the Fc domain of IgG. The Eph B1-Fc fusion protein and the corresponding fusion protein encoding the Eph B2 homologue (1.8uM of each) were incubated for 4 hours at 37°C with 280nM of MMP-2 or MMP-9. Following this incubation, samples were resolved by 10% SDS-PAGE, and the proteins were visualized by Coomassie blue staining. The position of the fragment of Eph B1 generated by MMP-2 is shown by an arrow. The recombinant EphB2 migrates at 80 kDa. A contaminant is present at 100kd.

Figure 3. The Double Reciprocal Plot of B74 Peptide Cleaved by MMP-2.

The initial velocity of B74 peptide hydrolysis was measured by incubating 12.5nM of active MMP-2 with 100, 200, 400, and 800uM of peptide. The double reciprocal plot of 1/[S] vs. 1/V

was then generated and used to derive an equation from the best-fit line. Value of K_m is equal to −1/X-intercept and value of k_{cat} is equal to Vmax/[E][S].

Figure 4. Model of the Catalytic Cleft of MMP-2 and MMP-9.

Models of the catalytic domain of MMP-2 (top panel) and MMP-9 (bottom panel) were constructed using the Modeller4 software from Rockefeller University (42). Models were constructed using the coordinates of the catalytic domain of MMP-2 from the reported crystal structure (14)(PDB accession number 1qibA). The model of MMP-9 is based on the crystal structure of MMP-2. A view looking into the catalytic cleft is illustrated, with the zinc ion colored orange. The side chains of the three His residues that ligand with zinc are evident. The S_2 and S_3 subsites fall within the cleft to the left of the zinc ion. Glu-412 protrudes into the S_2 subsite of MMP-2 (orange arrow). In MMP-9 the corresponding residue is Asp-410 (orange arrow). One other notable difference among the two proteases is Pro-193 in MMP-9, whose pyrrolidine ring protrudes away from the upper rim of the catalytic cleft and into the interface between the S_2 and S_3 subsites (white arrow). In MMP-2 this Pro is substituted by Ala, creating a relatively unobstructed surface across the S_3 space.

Table I. Sequences of phage substrates for MMP-2. Phage from the second round of selection were screened for hydrolysis by MMP-2 using an ELISA. Two separate experiments were performed. Based on this result, inserts of individual phages were sequenced to obtain the amino acid sequences of the displayed peptides. The substrates are separated into four structural motifs. Phage are also ranked according to the extent of hydrolysis by MMP-2 at 2 hours. (n.d. = not determined)

	Peptide Sequence									% Hydrolysis	
Group			P_3	P_2	P_1	$P_{1'}$	$P_{2'}$	$P_{3'}$	Clone	exp1	exp2
I	S	G	**P**	Y	V	**I**	W	L	A45	100	100
	R	P	**P**	M	A	**K**	T	A	A50	100	100
	S	G	**P**	V	W	**Y**	M	S	A26	100	98
	S	G	**P**	V	R	**Y**	M	V	A35	100	90
	G	L	**P**	R	W	**L**	L	T	A47	91	90
	G	R	**P**	I	R	**M**	K	T	A46	51	49
II	L	R	**L**	A	A	**I**	T	A	B49	100	100
	I	Y	**L**	G	W	**A**	T	A	A29	100	100
	E	S	**L**	A	Y	**Y**	T	A	B74	81	90
	G	S	**L**	H	S	**I**	I	T	B54	72	75
	S	D	**I**	R	M	**L**	T	A	B79	64	70
	F	N	**L**	Y	N	**L**	T	A	B66	51	50
	G	Y	**L**	Q	V	**L**	L	T	B46	48	50
	S	G	**I**	V	N	**L**	Y	P	A43	38	40
	E	Y	**L**	H	M	**R**	T	A	A10	36	40
	V	G	**L**	I	A	**I**	T	A	A6	30	44
III	R	S	**L**	**S**	R	**L**	T	A	C9	100	100
	N	R	**Y**	**S**	S	**L**	T	A	A34	100	100
	W	T	**S**	**S**	W	**L**	T	A	A31	100	100
	G	A	**V**	**S**	W	**L**	L	T	A13	100	93
	P	M	**I**	**S**	V	**L**	T	A	A54	98	96
	A	N	**I**	**S**	D	**L**	T	A	B37	97	95
	T	I	**L**	**S**	L	**L**	T	A	B53	73	75
	G	F	**N**	**S**	M	**L**	K	T	C3	28	30
IV	G	L	**H**	R	R	**I**	D	T	C5	60	n.d.
	G	M	**H**	S	R	**P**	P	T	A9	50	63
	H	M	**H**	K	A	**L**	T	A	A21	30	49
	G	R	**H**	L	G	**L**	Q	T	A4	30	45
	G	L	**H**	K	K	**V**	H	T	C6	19	n.d.
	G	A	**H**	A	K	**H**	W	T	D3	13	n.d.

Table II. Identification of the position of scissile bonds within MMP-2 substrates. Representative peptides were synthesized from each group of substrates. Using MALDI-TOF mass spectrometry the position of scissile bonds was determined by analyzing the mass of cleaved peptide fragments generated by MMP-2. The position of the scissile bond is noted by placing the P_1 and $P_{1'}$ residues in bold.

	Peptide Sequences								Peptide Mass	Observed fragments (Da)
Group I			**P_3**	**P_2**	**P_1**	**$P_{1'}$**	**$P_{2'}$**	**$P_{3'}$**		
A3	A	K	P	R	**A**	**L**	T	A	1013.7	727.9
Group II										
B49	L	R	L	A	**A**	**I**	T	A	1013.8	728.73
B74	E	S	L	A	**Y**	**Y**	T	A	1089	804.27
Group III										
C9	R	S	L	S	**R**	**L**	T	A	1103.4	790.29[a]
A13	G	A	V	S	**W**	**L**	L	T	1046.9	670.06[a]
B37	A	N	I	S	**D**	**L**	T	A	1012.9	729.03
A34	N	R	Y	S	**S**	**L**	T	A	1097.2	812.41
Group IV										
A21	H	M	H	K	**A**	**L**	T	A	1093.7	809.61

[a] denotes the mass of Na^+ adducts of the peptide fragments.

Table III. Measuring peptide hydrolysis by a panel of MMPs. The hydrolysis of synthetic peptides by different MMPs was quantified using procedures outlined in Methods. For MMP-2, hydrolysis was measured over a concentration range of peptide. Values for k_{cat} and K_m were derived from Lineweaver-Burk plots. For other proteases only the k_{cat}/K_m ratio was measured (see Methods). Additional synthetic peptides were synthesized with substitution of Pro at P_3 and Arg at P_2 to test the influence of these two subsites on substrate specificity. Each measurement was repeated three times and the relative error of k_{cat}/K_m in these measurements was less or equal to 10%.

		MMP-2			MMP-9	MMP-7	MMP-13
Peptide	**Sequence**	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Group I							
A3	SGAKPRA ↓ LTA	400	3.6	1.1E+05	4.9E+04	5.2E+03	7.9E+03
Group II							
B49	SGLRLAA ↓ ITA	510	4.5	1.1E+05	7.9E+03	1.9E+04	8.5E+03
B74	SGESLAY ↓ YTA	622	2.2	2.8E+05	2.6E+03	7.9E+02	2.0E+03
B74P	SGES**P**AY ↓ YTA	264	0.6	4.4E+05	3.0E+04	1.5E+03	5.3E+03
B74R	SGESL**R**Y ↓ YTA	61	0.8	7.6E+04	1.0E+05	7.1E+02	1.1E+03
C9	SGRSLSR ↓ LTA	740	4.4	1.7E+05	8.6E+02	4.8E+03	3.3E+03
C9R	SGRSL**R**R ↓ LTA	302	6.4	4.7E+04	5.3E+04	2.2E+03	3.9E+03
Group III							
A13	SGAVSW ↓ LLTA	105	1.3	8.1E+04	5.4E+03	6.4E+03	2.9E+03
A13P	SGA**P**SW ↓ LLTA	204	0.4	5.1E+05	1.7E+04	5.4E+03	4.9E+03
A13R	SGAV**R**W ↓ LLTA	28	0.4	7.0E+04	2.0E+05	3.5E+03	2.9E+03
B37	SGANISD ↓ LTA	353	2.4	1.5E+05	2.1E+03	8.6E+03	2.6E+03
A34	SGNRYSS ↓ LTA	202	2.2	9.2E+04	2.3E+03	1.1E+03	3.9E+03
Group IV							
A21	SGHMHKA ↓ LTA	124	7.7	1.6E+04	2.1E+03	n.d.	n.d.
A21A	SGHMHAA ↓ LTA	601	3.2	1.9E+05	1.9E+03	n.d.	n.d.

n.d.= not determined

Table IV. Selectivity of individual substrates forMMP-2. The selectivity of each of the peptide substrates for MMP-2 compared to other MMPs is represented by dividing the k_{cat}/K_m ratio for MMP-2 by the same value for the other MMPs that were tested. In several cases mutation of the P_2 residue gave rise to dramatic shifts in the selectivity ratio between MMP-2 and MMP-9. These values are in bold text.

	Selectivity Ratio		
Peptide	MMP-2/MMP-9	MMP-2/MMP-7	MMP-2/MMP-13
Group I **[P-X-X-X_{Hy}]**			
A3	2	21	14
Group II **[I/L-X_{po}-X-X_{Hy}]**			
B49	14	6	13
B74	108	354	140
B74P	15	293	83
B74R	**0.8**	107	69
Group III **[X_{Hy}-S-X-L]**			
C9	198	35	52
C9R	**0.9**	21	12
A13	15	13	28
A13P	30	94	104
A13R	**0.4**	20	24
B37	71	17	58
A34	40	84	24
Group IV **[H-X-X-X_{Hy}]**			
A21	8	n.d	n.d.
A21A	**100**	n.d.	n.d.

n.d. = not determined

Figure 1.

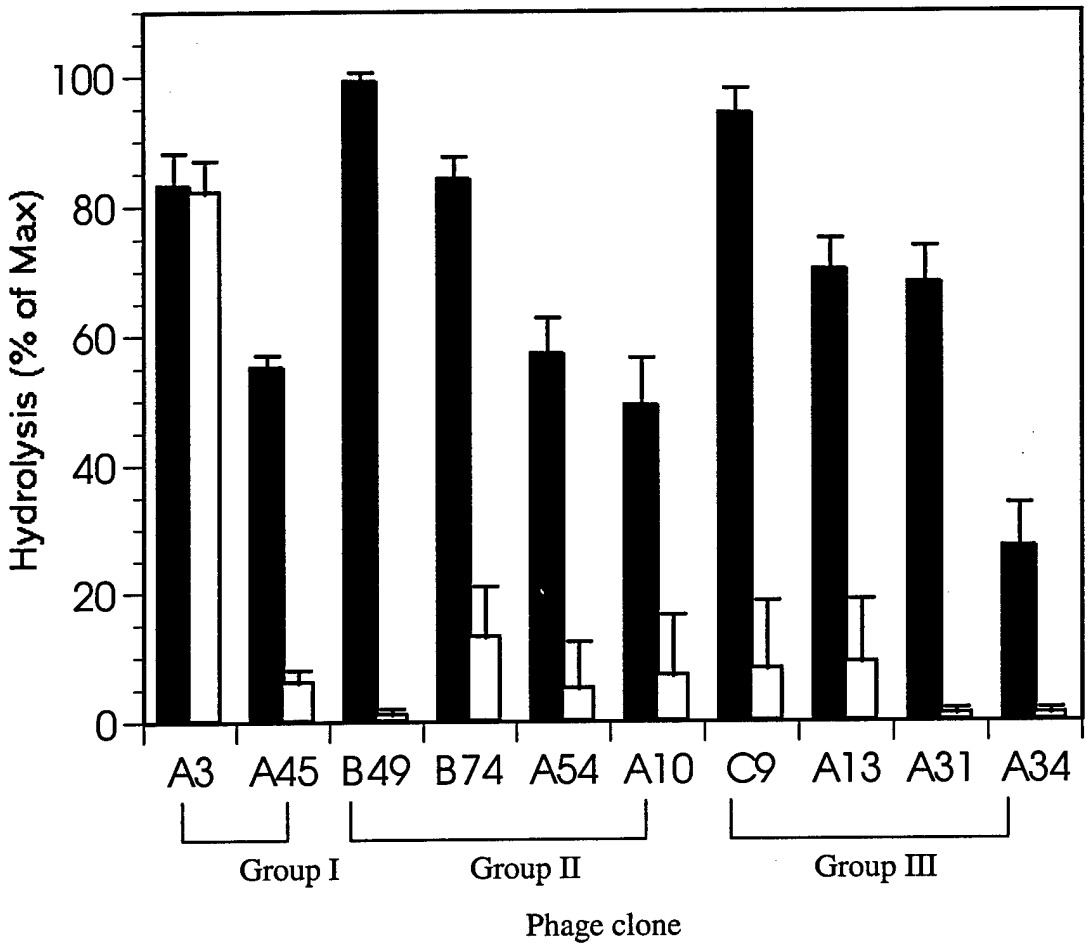

Hydrolysis (% of Max)
100
80
60
40
20
0
A3
A45
B49
B74
A54
A10
C9
A13
A31
A34
Group I
Group II
Group III
Phage clone

Figure 2.

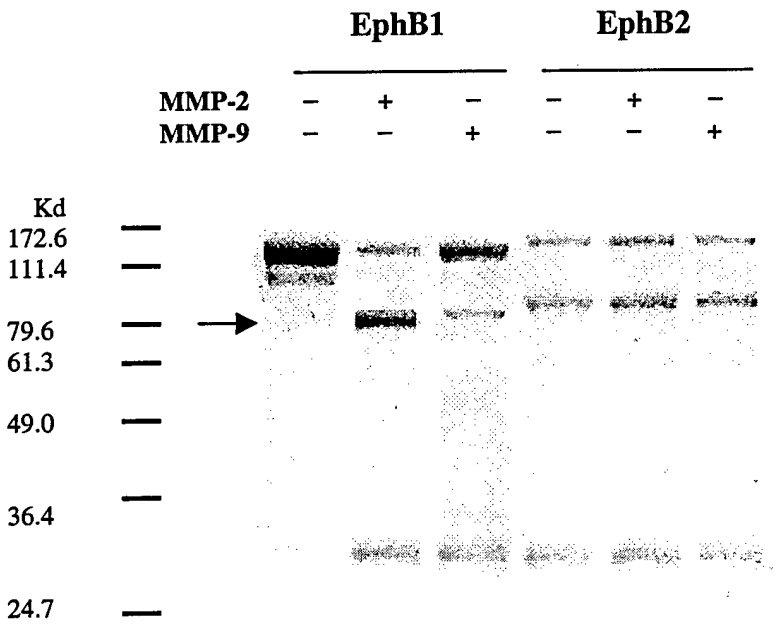
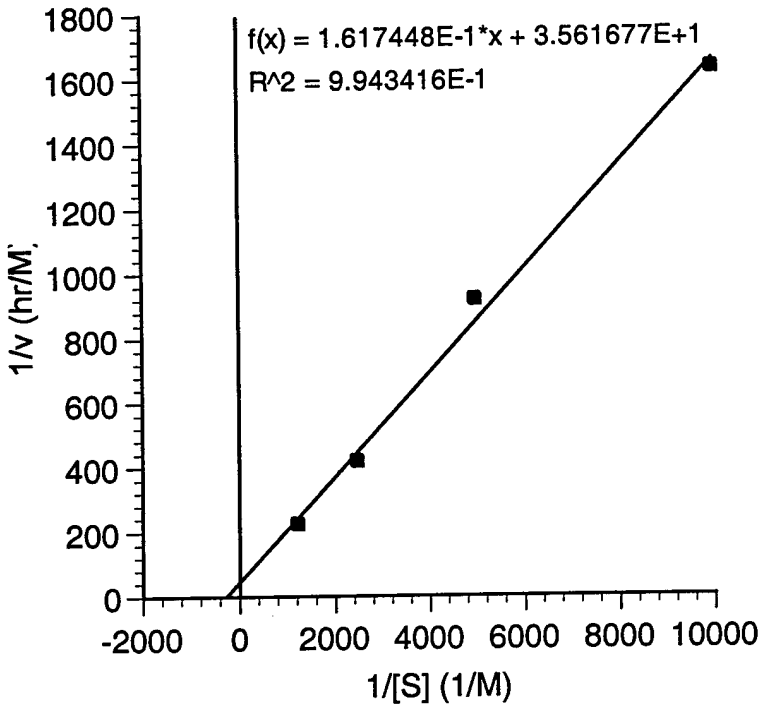


Figure 3.


f(x) = 1.617448E-1*x + 3.561677E+1
R^2 = 9.943416E-1
1800
1600
1400
1200
1000
800
600
400
200
0
1/v (hr/M)
-2000
0
2000
4000
6000
8000
10000
1/[S] (1/M)

MMP-2

MMP-9