

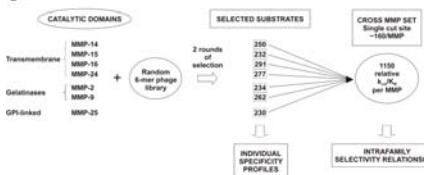
Toward accurate prediction of natural substrates: family based protease specificity profiling.

Boris Ratnikov, Piotr Cieplak, James Pierce, Alexei Eroshkin, Sergey Shiryayev, Alex Strongin, and Jeffrey W. Smith

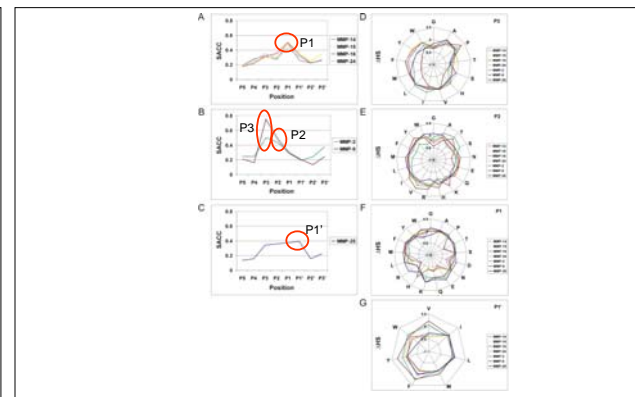
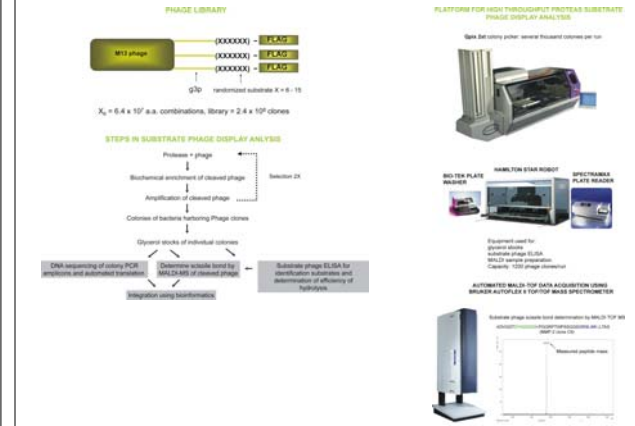
Burnham Institute for Medical Research, La Jolla, CA,

Abstract

Currently, there are more than 90,000 known peptidase sequences reported in the MEROPS database. This is a staggering number for those trying to understand how these enzymes work and what they do. On the other hand, since proteases have evolved as families of closely related members, this diversity could be exploited to elucidate their functionality. We describe an approach for delineation of catalytic cleft specificity and selectivity in protease families. We studied seven matrix metalloproteinases (MMPs) representing the three distinct groups in the MMP family using high throughput substrate phage display. We demonstrate that catalytic domain primary structure relationships hold at the level of substrate selectivity, allowing elucidation of the structural features underlying differences in specificity. For the first time we have demonstrated cooperative interactions between pairs of subsites in the catalytic groove. This information can be very useful for the prediction of differential physiologic and pathologic substrates linking individual family members to particular aspects of biology and pathology, and for the design of selective inhibitors.



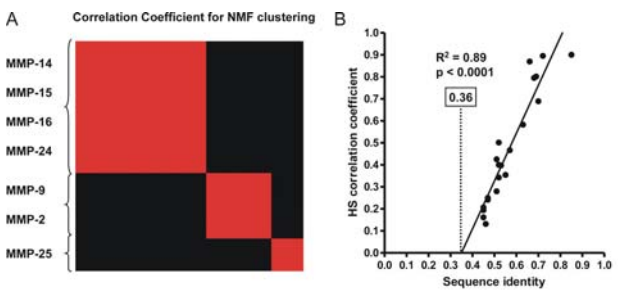
A. NJ tree for almost 600 MMPs catalytic domains from different species created using the HyperTree algorithm. Human members are labeled according to SWISS-PROT. B. Substrate selectivity of matrix metalloproteinases belonging to three distinct subfamilies was addressed. Catalytic domains of transmembrane domain-containing enzymes (MMP-14, 15, 16, and 24), Gelatinases (MMP-2 and 9), and a GPI-linked MMP-25 were used for selection of phage substrates for each protease. Specificity profiles for each enzyme were derived based on frequency of occurrence of amino acid residues at individual positions relative to the scissile bond. Selectivity relationships among the MMPs were determined using a set of substrates combined from individually selected ones containing a single cut site by determining catalytic efficiencies for each enzyme for every substrate in the set.



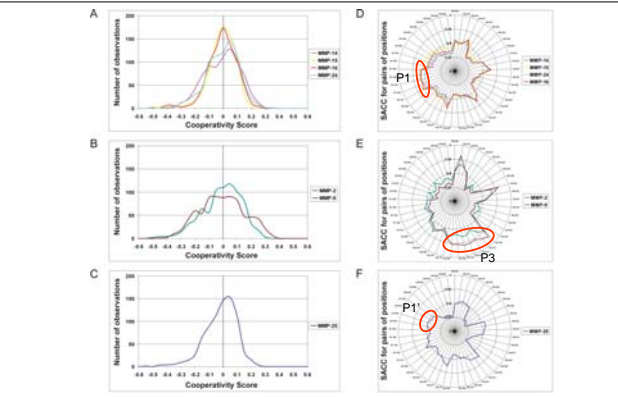
A - C. Sequence Activity Correlation Coefficient (SACC) for the catalytic domains of the MMPs was calculated based on the hydrolysis scores obtained using the cross MMP set of substrates. A - for MT-MMPs, B - for Gelatinases, and C - for MMP-25. D - G. The most impactful subsites are circled in red. Differences between the average values of hydrolysis scores for the substrates containing a particular amino acid at positions of high sensitivity (SACC ≥ 0.4) and the average hydrolysis score for a given protease were calculated: D - for P3, E - for P2, F - for P1, and G for P1' positions respectively. The sign of the values indicates the direction of the impact on hydrolysis.

Conclusions:

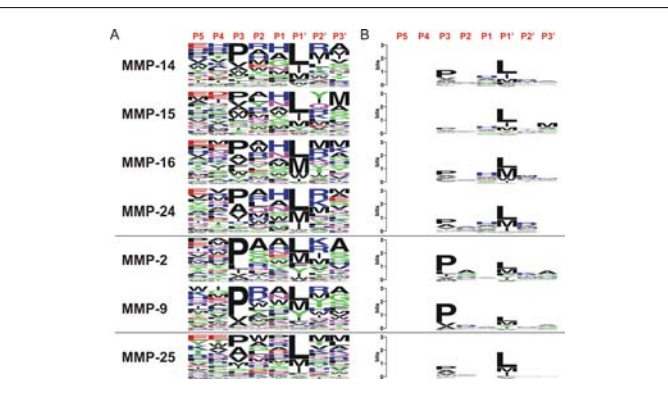
We have developed a family based approach for high resolution analysis of substrate specificity of proteolytic enzymes. We determined substrate specificity profiles of 7 members of Matrixin family of metalloproteinases using high throughput phage display analysis. The enzymes were selected to represent the 3 major branches of the Matrixin family tree and belong to three structurally related groups. We then determined selectivity relationships among the 7 by analyzing catalytic efficiencies of hydrolysis of a set of 1150 phage substrates by each of the MMPs. The observed selectivity relationships closely match those obtained on the basis of amino acid sequence similarity of the respective catalytic domains. This observation validates the results of the study and provides grounds for finding structural basis for the observed selectivity profiles. In all cases we observed group related specificity and selectivity relationships. In addition, the wealth of functional data also allows us to analyze the impact of cooperative interactions between pairs of catalytic cleft subsites on hydrolysis. One of the major practical implications of this work is the use of the specificity models obtained in this work for accurate prediction of cut sites in protein substrates, thus allowing to rationalize the study of the protease biology.



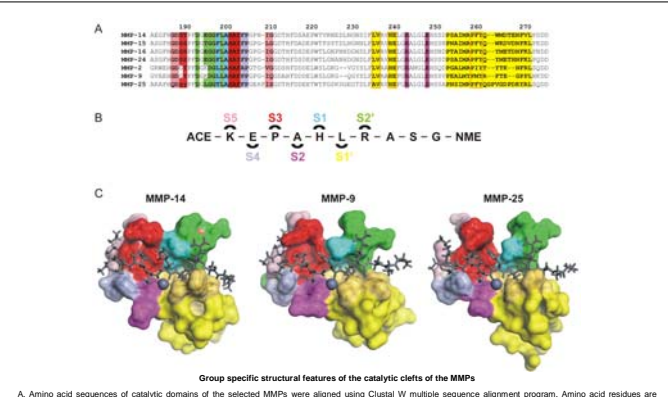
A. Nonnegative matrix factorization clustering of the MMPs used in the study based on hydrolysis score correlation for phage substrates in the cross MMP set. 3 groups of MMPs with cophenetic correlation coefficient of 1.0 were derived. B. Analysis of correlation between sequence homology and hydrolysis score correlation coefficient was done as described in the materials and methods. A linear correlation is observed ($R^2 = 0.89$, $p < 0.0001$) with a threshold of 0.36.



Impact of catalytic cleft subsite interactions on efficiency of hydrolysis. Distributions of net cooperative contributions to the hydrolysis scores. A - for the MT-MMPs, B - for the Gelatinases, and C - for MMP-25. Structure activity correlation coefficients were determined for pairs of subsites in the catalytic clefts of the studied MMPs: D - MT-MMPs, E - Gelatinases, and F - MMP-25. Major cooperativity centers are shown with red circles.



Web logo representation of position weight matrix analysis of sets of substrates selected by phage display for individual MMP catalytic domains. A. Frequency plot: heights of the symbols and their vertical positions indicate the frequency of occurrence of the residues at individual positions relative to the scissile bond (P5 - P3). B. Information content plot: the overall height of the stack indicates sequence conservation at a given position, while the height of symbols within the stack indicates the relative frequency of each amino acid residue at that position.



A. Amino acid sequences of catalytic domains of the selected MMPs were aligned using Clustal W multiple sequence alignment program. Amino acid residues are numbered based on MT1-MMP sequence NP_004586. Residue coloring represents its contribution to subsites in the catalytic cleft shown in C. Dark yellow boxes around the S1' residues highlighted in yellow denote the amino acids forming the entrance of the S1' pocket colored in peach in C. B. A schematic diagram of a substrate peptide ACE-K-E-P-A-H-L-R-A-S-G-NME (ACE-K-E-P-A-H-L-R-A-S-G-NME) docked into the catalytic cleft in extended conformation. C. Three dimensional structures of the catalytic clefts of MMP-14, 9, and 25 with the peptide shown in 1 modeled into it. Residues contributing to individual subsites are shown in color.