Title

Basis for substrate recognition and distinction by matrix metalloproteinases

Journal

Proceedings of the National Academy of Sciences of the United States of America, 111(40)

ISSN

0027-8424

Authors

Ratnikov, Boris I
Cieplak, Piotr
Gramatikoff, Kosi
et al.

Publication Date

2014-10-07

DOI

10.1073/pnas.1406134111

Peer reviewed

Basis for substrate recognition and distinction by matrix metalloproteinases

Boris I. Ratnikov[a,1], Piotr Cieplak[a,1], Kosi Gramatikoff[a], James Pierce[a], Alexey Eroshkin[a], Yoshinobu Igarashi[a], Marat Kazanov[b], Qing Sun[a], Adam Godzik[a,c], Andrei Osterman[a,c], Boguslaw Stec[a], Alex Strongin[a,c], and Jeffrey W. Smith[a,2]

[a]The Cancer Center and [c]The Inflammatory and Infectious Disease Center, Sanford–Burnham Medical Research Institute, La Jolla, CA 92037; and [b]Research and Training Center on Bioinformatics, Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia

Edited by Perry Allen Frey, University of Wisconsin–Madison, Madison, WI, and approved August 22, 2014 (received for review April 3, 2014)

Genomic sequencing and structural genomics produced a vast amount of sequence and structural data, creating an opportunity for structure–function analysis in silico [Radivojac P, et al. (2013) *Nat Methods* 10(3):221–227]. Unfortunately, only a few large experimental datasets exist to serve as benchmarks for function-related predictions. Furthermore, currently there are no reliable means to predict the extent of functional similarity among proteins. Here, we quantify structure–function relationships among three phylogenetic branches of the matrix metalloproteinase (MMP) family by comparing their cleavage efficiencies toward an extended set of phage peptide substrates that were selected from ∼64 million peptide sequences (i.e., a large unbiased representation of substrate space). The observed second-order rate constants [$k_{(obs)}$] across the substrate space provide a distance measure of functional similarity among the MMPs. These functional distances directly correlate with MMP phylogenetic distance. There is also a remarkable and near-perfect correlation between the MMP substrate preference and sequence identity of 50–57 discontinuous residues surrounding the catalytic groove. We conclude that these residues represent the specificity-determining positions (SDPs) that allowed for the expansion of MMP proteolytic function during evolution. A transmutation of only a few selected SDPs proximal to the bound substrate peptide, and contributing the most to selectivity among the MMPs, is sufficient to enact a global change in the substrate preference of one MMP to that of another, indicating the potential for the rational and focused redesign of cleavage specificity in MMPs.

protease | specificity-determining positions | MMPs

A paramount objective of biological research is to understand how sequence encodes function. Previously, functional regions in proteins were identified using large-scale mutagenesis (e.g., alanine scanning) (1). More recently, our insights are largely gained by computational approaches aimed at comparing sequences and structures of large protein sets across multiple genomes. The vast increase in the number of available sequences makes it possible to compare homology between sequences from genome projects to proteins of known structure and function and, as a result, identify functional similarities in silico (2–4). Because the global fold of most ordered proteins can be reliably predicted (5, 6) and because the catalytic residues of most classes of enzymes are either known or can be inferred (7, 8), protein sequences can now be directly used to elucidate and classify major protein functions (9–12), and are even being extended to predict enzyme substrates (13).

However, such classifications fail to explain the specialization and expansion of function that is required for organismal plasticity, complexity, and adaptability, all of which are normally driven by gene duplications and subsequent divergence. Computational approaches aimed at identifying functional distinctions across protein families have primarily focused on phylogeny. Fryxell (14) put forth the idea that protein function could be assigned based on correlations among evolutionary trees of interacting proteins and that different branches of a phylogenetic tree are expected to have different functions. To support this idea, Goh et al. (15) constructed evolutionary trees of chemokines and their cognate receptors, and, based on correlations in the phylogenetic trees, predicted potential ligands for orphan G proteins. The approach was then extended to large-scale studies, including the use of evolutionary trees to predict interactions among 2,742 *Escherichia coli* proteins (16).

Questions on functional specialization are not limited to phylogeny of the full-length proteins. Similar questions can also be asked at the molecular level: are there the amino acid positions within a protein family that selectively govern distinctions in substrate interaction and ligand binding? These residues are often called specificity-determining regions or specificity-determining positions (SDPs). The SDPs are distinct from a relatively large binding interface. Although mutations across the binding interface can ablate ligand binding, only a subset of residues in the interface confers distinctions in ligand binding. SDPs are normally identified using mutagenesis and protein engineering (17–20). Some efforts have been made to perform this type of analysis across a protein family, but because of the experimental challenge, only a limited number of family members were analyzed (21).

Several computational approaches have been developed to predict SDPs (22–26). In most cases, these methods are based on the ideas of Mirny and Gelfand (27–30), who suggested that functional specificity is determined by protein regions that are more similar within a group having a common function than between groups with distinct functions. Previous analysis of this

Significance

Specificity-determining positions (SDPs) account for distinctions in function across a protein family. Many theories on the evolution of functional specificity have led to approaches for predicting SDPs in silico, but large experimental datasets allowing a statistical assignment are lacking. Here, the SDPs of matrix metalloproteinases are elucidated by querying the proteolytic efficiency of eight matrix metalloproteinases, representing three phylogenetic branches, in an extended and diverse substrate space. More than 10,000 measures of cleavage efficiency reveal a near-perfect correlation between similarity in proteolytic function and sequence identity at 50–57 positions on the front face of the catalytic domain. These positions are assigned as SDPs. Transmutation of proteolytic function is possible by swapping SDPs nearest to bound substrate.

Author contributions: B.I.R., P.C., and J.W.S. designed research; B.I.R., J.P., and Q.S. performed research; J.W.S. contributed new reagents/analytic tools; B.I.R., P.C., K.G., A.E., Y.I., M.K., A.G., A.O., B.S., A.S., and J.W.S. analyzed data; and B.I.R., P.C., A.G., A.O., B.S., A.S., and J.W.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]B.I.R. and P.C. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: jsmith@sanfordburnham.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406134111/-/DCSupplemental.

type used branch points in phylogenetic trees to indicate functional distinction, but, in principle, the idea could be expanded to any indicator of functional distinction. Despite the expansion in methods for predicting SDPs, there are few quantitative datasets on function for benchmarking these predictions. Furthermore, the current computational methods are unable to provide any information on the relative impact of the individual SDPs, so it is not clear which SDPs are most important for functional distinction.

To address these issues, we use matrix metalloproteinases (MMPs) as model system. Eight representative MMPs, representing three different phylogenic subfamilies, were used in our study. We measured the ability of these MMPs to cleave substrates in a large set of phage peptide substrates (~6.4×10^7 hexapeptide sequences). Interrogating this large sequence space with several MMPs is akin to simultaneously assessing the impact of many mutations in both protease and substrate. This type of quantitative analysis could not be accomplished with so called "physiologic substrates." Ultimately, we arrive at ~10,500 measures of $k_{(obs)}$ that indicate functional similarity and distinction in MMPs. Hierarchical clustering of $k_{(obs)}$ values for the large substrate space mirrors a phylogenetic tracing of the MMP catalytic domains, proving the connection between function and sequence phylogeny.

In addition, the $k_{(obs)}$ values for each substrate for one MMP were compared with those for all other MMPs, yielding a series of correlation coefficients that are quantitative indicators of MMP functional similarity. We observed a remarkable and near-perfect correlation between these coefficients and sequence identity among 57 discontinuous residues on the front face of the MMP catalytic domain. We conclude that these residues are the SDPs that confer specialization in substrate recognition. This assignment is supported by the fact that swapping the selected SDPs switches substrate recognition between two individual MMPs.

Results

Substrate Preference Is a Mirror of Sequence Phylogeny. The phylogenetic landscape of the MMP catalytic domains was determined by aligning 660 vertebrate MMP sequences using the MUSCLE program (31). HyperTree (32) was used to draw a near joint tree (Fig. 1*A*). Eight human MMPs that represent three different phylogenetic branches were chosen as test proteases. The latter include (*i*) MMP-2 and -9, which belong to the gelatinase subfamily; (*ii*) MMP-14, -15, -16, and -24, which belong to the transmembrane MMPs (MT-MMPs); and (*iii*) MMP-17 and -25, which contain a GPI cell surface anchor (GPI-MMPs) (33).

A large set of test substrates was created using substrate phage display. Substrates specific for each of the eight MMPs were selected from a library of hexameric peptides expressed on the surface of M13 phage as illustrated in *SI Materials and Methods* and described in ref. 34. Approximately 300 phage substrates were selected for each MMP (Fig. S1*A*). The position of the scissile bond was determined from the monoisotopic mass of the cleavage products determined by MALDI-TOF MS (Fig. S1*B*). The sequence of the peptide was derived from cDNA sequencing. From this original set of ~2,400 phage substrates, a test set of 1,369 (~170 substrates from each MMP) was selected. Substrates in the test set were (*i*) nonredundant in sequence, (*ii*) contained a single cleavage site, and (*iii*) had a value of $k_{(obs)} > 3{,}000\ M^{-1}{\cdot}s^{-1}$ for the target MMP (Fig. S1*A*). The $k_{(obs)}$ values for each substrate were determined for each MMP as described in *Materials and Methods* and ranged from exceedingly low (undetectable) to the maximum value of over $1 \times 10^4\ M^{-1}{\cdot}s^{-1}$ (Dataset S1). From over 1,300 substrates, only ~100 were cleaved efficiently by all of the MMPs, and almost all of these contained the canonical P-X-X-↓L motif.

To visualize the way the MMPs function in substrate space that is selective across the eight MMPs, the $k_{(obs)}$ values for each

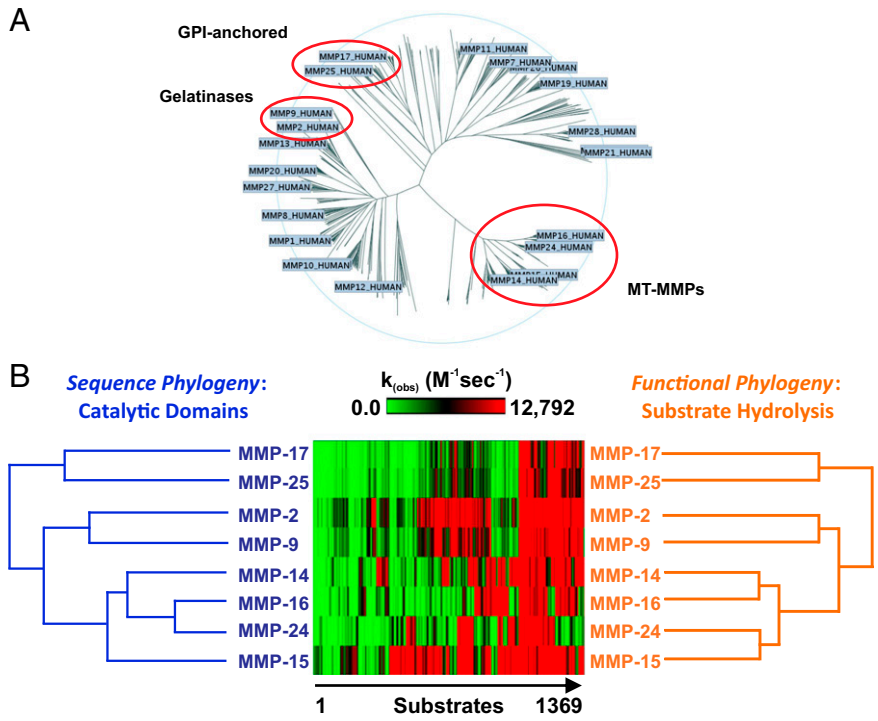

Fig. 1. MMP substrate recognition profiles correlate with phylogenetic distances between catalytic domains. (*A*) Phylogeny of the catalytic domains of mammalian MMPs. Amino acid sequences of 600 MMPs from different species were aligned using JalView, and a near joint tree was generated using HyperTree algorithm for the catalytic domains (see *Materials and Methods* for details). Representative members from each of the three major branches were selected for the study: membrane type I MMPs (transmembrane) MMP-14, -15, -16, and -24; gelatinases MMP-2 and -9; and membrane GPI-anchored MMP-17 and -25. (*B*) Average linkage trees of sequence identity among the catalytic domains of the MMPs (blue) and $k_{(obs)}$ values obtained for each MMP using the 1,369 phage substrates (orange) were generated using JalView and hierarchical clustering editor (HCE) programs as described in *Materials and Methods*. The heat map displays the $k_{(obs)}$ of each MMP for each phage peptide substrate.

substrate were subjected to 2D clustering using average linkage, which provided a relative Euclidean distance measure of similarity and distinction in substrate recognition (Fig. 1*B*, *Right*). This tree is in complete concordance with the linkage tree for the peptide sequences of the catalytic domain of each MMP (Fig. 1*B*, *Left*). We conclude that, at a global level, distinctions in substrate recognition across the MMP family mirror sequence phylogeny.

Distinction in Substrate Recognition Among MMPs. Differences in substrate preference were further characterized in three ways: (*i*) based on the frequency of a residue at a particular position in substrate, (*ii*) the impact of variability at each position, and (*iii*) the impact of particular residues at each position on $k_{(obs)}$. Differences in amino acid frequency at different positions are visualized by a Logo plot (Fig. 2*A*). MT-MMPs and GPI-anchored MMPs frequently exhibit the P1′ Leu, whereas gelatinases do not. In gelatinases, the P3 position displays the highest frequency residue, which is predominantly Pro. This observation suggested that particular positions contribute to substrate recognition in varying degrees. This idea was tested by calculating the sequence–activity correlation coefficient (SACC), which determines the impact of sequence variability at each position on substrate cleavage (35). An SACC score of 1 indicates the total dominance of substrate recognition by that position, whereas a 0 score indicates a lack of any effect. The SACC profiles for the individual P5–P2′ positions, shown as a heat map (Fig. 2*B*), are remarkably similar within each phylogenetic branch, but distinct between branches.

Residues preferred at each substrate position were highlighted by calculating residue preference (RP). RP is the ratio of the average $k_{(obs)}$ of all substrates with given residue at a specified position, divided by the average $k_{(obs)}$ for the entire test set of substrates (see equations in *SI Materials and Methods*). RP values >1 indicate that substrate hydrolysis is enhanced by the presence of a particular residue at a specific position. Conversely, a value <1 indicates that the residue is unfavorable at the specified positions. Amino acid residues absent from the list for each position were observed with very low frequency or were absent in the phage substrates, indicating that such residues are unfavorable, especially because no positional bias was observed in the sequences from the 766 randomly selected phage inserts. The full spectrum of RPs is displayed in a heat map format generated by one-dimensional clustering. Fig. 2*C* shows that RP generally correlates with sequence phylogeny of the MMPs. Certain notable deviations were also observed. For example, there is a strong bias against aromatic residues at S3 in both MMP-9 and MMP-17, which belong to the different phylogenetic branches. This feature is more specific for MMP-9 and MMP-17 than for any other MMPs, including those in the respective phylogenetic branches. This observation suggests that this substrate subsite has evolved in a modular manner, perhaps independently from the other subsites.

Assignment of the MMP SDPs. Our measures of $k_{(obs)}$ for >1,300 substrates by eight individual MMPs is akin to a large and reciprocal mutagenesis experiments, in which sequence changes are made in both the substrate and the protease. This allowed us to propose a set of SDPs: a set of residues in protease that account for differences in substrate specificity across the MMP family. This was done by establishing the functional relationship between each pair of MMPs expressed as pairwise hydrolysis correlation coefficient (H*R*) for the $k_{(obs)}$ values of each test substrate (Table S1 and *SI Materials and Methods*). This value indicates the degree of similarity in substrate recognition between each MMP pair in the given substrate space. For example,

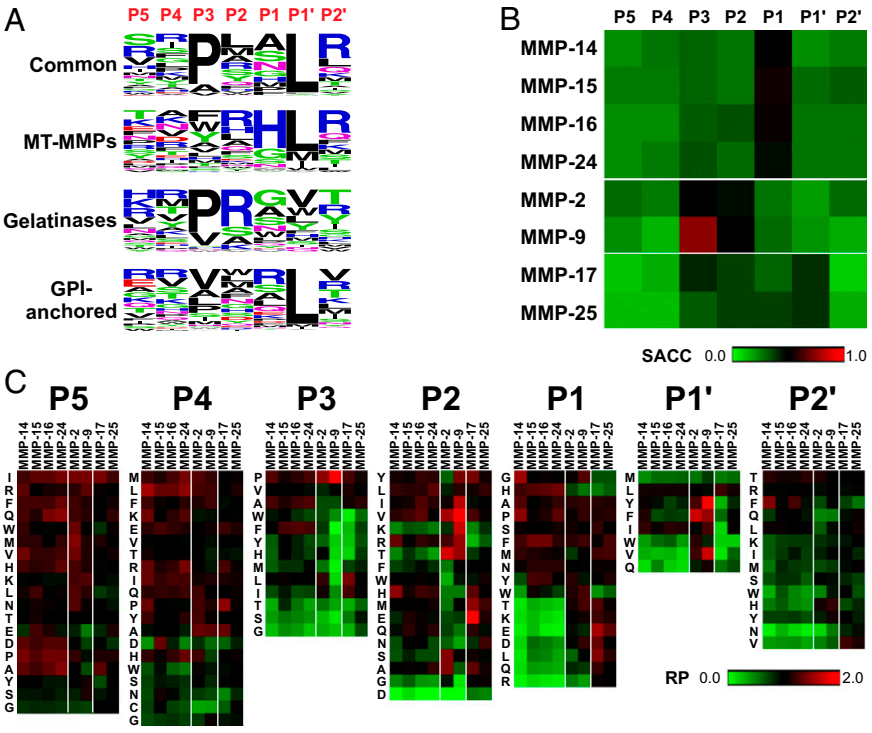

Fig. 2. Distinctions in substrate recognition profiles follow MMP phylogeny. (*A*) Sequence logos (37) were generated for phage substrates cleaved equally well by the entire set of MMPs (Common) and those selective for individual subfamilies (MT-MMPs, gelatinases, and GPI-anchored MMPs). The height and vertical position of each letter is proportional to the frequency of occurrence of amino acid residues at each position of substrates from P5 to P2′. (*B*) The contribution of individual subsites to substrate recognition by the test MMPs was determined based on sequence–activity correlation coefficient (SACC) (see *SI Materials and Methods* for equations) values calculated using $k_{(obs)}$ data obtained for each MMP against 1,369 phage substrates. SACC ranges from 0 (low) to 1 (high). (*C*) Heat maps of residue preference (RP) (described in *SI Materials and Methods*) show how individual residues at specific positions from P5 to P2′ influence $k_{(obs)}$. The heat map was generated by hierarchical clustering. RP ranges from 0 (low) to 2 (high).

the correlation coefficient between MMP-14 and MMP-16 is ~0.77, indicating their considerable similarity in substrate recognition. However, the correlation coefficient between MMP-16 and MMP-17 is 0.31, indicating substantial distinction in their substrate recognition. When combined, correlation coefficients mathematically describe the relationship between each MMP relative to other MMP in cleaving the given substrate set.

The series of H*R*s for all pairs of MMP were plotted against the sequence identify of the catalytic domains to quantify any potential relationships. We noticed that there is a linear correlation ($R^2 = 0.9$) between sequence identity of the entire catalytic domains and the set of H*R* (Fig. 3*A*). However, the line describing the relationship intersects the *y* axis at a point corresponding to a ~30% sequence identity. This offset suggested that residues identical among the MMPs induce a bias to the correlation. Consequently, we tested the relationship between H*R* and the sequence identity of 50–57 residues (depending on the MMP) that belong to the variable loops surrounding the catalytic cleft, and constitute ~55% of the residues on the binding face of the catalytic domain. This group of residues was chosen based on the ideas of Mirny and Gelfand (27, 30). In this case, we simply eliminated the positions that are conserved on the front face of the catalytic domain in all MMPs. When the variable positions alone are considered, the slope improves to 1.01 with an intercept at the origin while the correlation coefficient continues to remain high ($R^2 = 0.87$; Fig. 3*B*). Statistical significance of the results was tested by identifying the minimum number of substrates sufficient to support robust correlation (Fig. S2). Only 400 to 600 substrates were necessary for optimal correlation. When sequence identity of the remainder of the catalytic domain is plotted against the set of H*R*, a correlation levels remains acceptably high ($R^2 = 0.79$), but the slope deteriorates to 0.44 with a *y* intercept at 0.40 (Fig. 3*C*). The connection between sequence differences and the regions on the MMP structure that comprise the SDPs is illustrated in Fig. 4. The remarkable correlation between the 50–57 SDPs and H*R* is also supported by a statistical analysis showing that any reduction in the number of assigned SDPs led to a significant loss in the correlation (Fig. S3). The correlation between sequence of the SDPs and H*R* suggests that the catalytic behavior of a single MMP across the substrate space can be predicted based on information from the other seven MMPs. Indeed, such predictions can be made for each of the eight representative MMPs with remarkable precision (Fig. S4). This observation suggests that, in the future, it may be possible to predict the catalytic efficiency of MMPs for a subset of substrates, or even a single substrate.

Transmutation of Substrate Preference. Although the total number of SDPs is represented by 50–57 individual residue positions, we reasoned that some aspects of substrate preference could be altered by changing only a few selected SDPs. We tested this idea by transforming the substrate preferences of MMP-16 into those of MMP-17. Selection of the transmutated residues were based on the idea that the MMP residues near the key positions in the substrate (as indicated by SACC score) have the highest impact. The dominant substrate position for MMP-17 is S1′, but both MMP-16 and MMP-17 prefer S1′ Leu. We then excluded this position from our analysis, and focused on the S1 to S3 subsites, which have the highest SACC scores in MMP-16, and which are also important in MMP-17.

Models of substrates docked with MMPs (Fig. 5*A*) were used to select those MMP residues, which are positioned at a short, a few angstroms, distance from a high SACC score substrate position, and which are not conserved at the equivalent position in MMP-16. Only four MMP-17 SDP residues meet these criteria: Tyr71, Thr79, Ala92, and Ser130. The corresponding positions in MMP-16 are Ser197, Phe205, and Glu255 (Fig. 5*A*), which are also close to key positions in the bound optimal substrate. Ser197 is within a 4-Å distance from the S3 substrate residue, which has a high SACC score in MMP-16 substrates. Both Ser197 and Thr79 are within a 5-Å distance from the S1 position, which has the highest SACC score in the MMP-16 substrates. The carboxyl of Glu255 is likely within a few angstroms distance from the S2 residue, and this parameter is likely to explain the preference for the S2 Arg in MMP-9 (36).

Based on this analysis, we mutated Ser197, Phe205, Gly217, and Glu255 in MMP-16 into the corresponding MMP-17 residues. The ability of the transmutated hybrid to cleave ~70 MMP-17–selective substrates, and ~70 MMP-16–selective substrates was quantified. The $k_{(obs)}$ of each substrate for MMP-16, MMP-17, and the hybrid is shown in Dataset S2. The average $k_{(obs)}$ for the selective sets of substrates for each MMP shows the functional transmutation to be successful (Fig. 5*B*). Although MMP-16 was barely capable of cleaving the selective MMP-17 substrates, the hybrid cleaved these

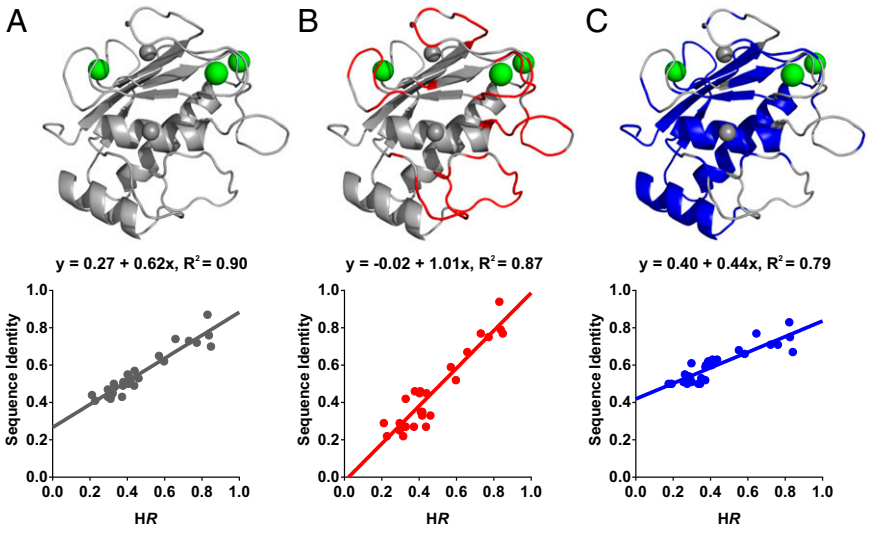

Fig. 3. Sequence phylogeny of the variable regions of the front face of the catalytic cleft of MMPs directly correlates with functional phylogeny. The degree of sequence identity vs. HR—the pairwise hydrolysis correlation coefficient [defined as Pearson correlation coefficients of $k_{(obs)}$ values]—was plotted for each pair of MMPs using the sequence of the whole catalytic domain (*A*), 50–57 variable residues on the binding face of the catalytic domain (red) (*B*), and the entire catalytic domain exclusive of the areas identified in *B* (blue) (*C*). Equations describing individual correlations are shown above each plot. The green spheres in structures represent calcium ions, and the gray ones, zinc ions.

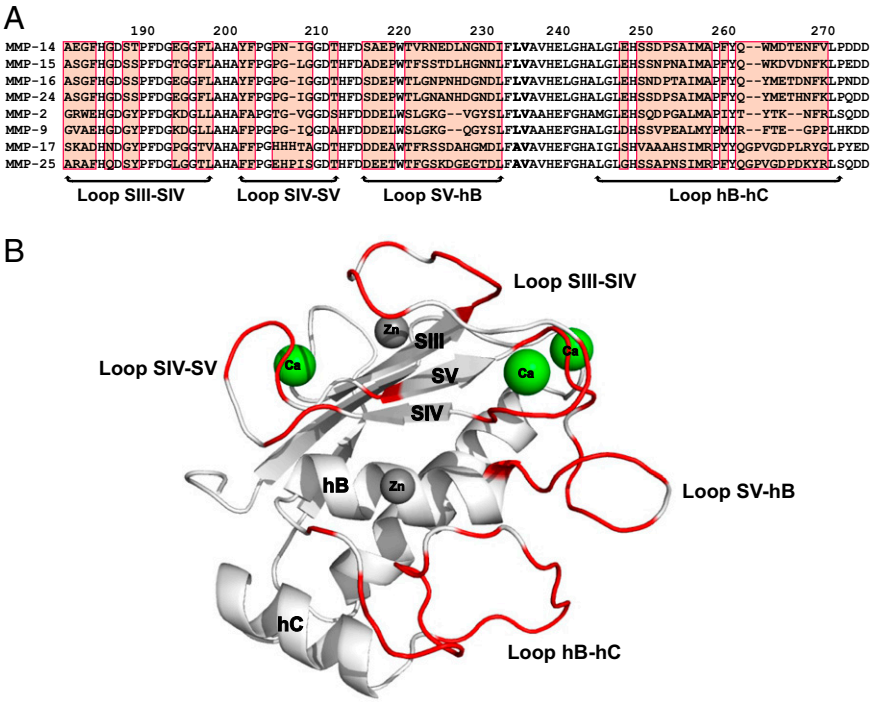

Fig. 4. SDPs of the MMP catalytic domains are located in the variable surface loops. (*A*) Sequences of catalytic domains of human MMPs were aligned using JalView and variable residues in the surface loop regions surrounding the catalytic cleft were identified using sequence variability at each position, along with existing structural information or structural models as described in *Materials and Methods*. The colored regions correspond to the residues comprising the SDPs. Location of the residues relative to the catalytic cleft surface loops is indicated by brackets below the alignment. (*B*) Localization of the SDPs in 3D space is marked in red on a ribbon diagram of MMP-2.

substrates nearly as efficiently as intact MMP-17. The MMP-16/17 hybrid also lost most of its ability to cleave substrates selective for MMP-16. This observation is consistent with the idea that SDPs proximal to key positions in the bound substrate have a major impact on substrate distinctions. To further illustrate the nature of the change in substrate preference, Logo plots (37) indicating the frequency of each amino acid residue at particular positions in substrates selective for MMP-16 (*Left*) and selective for MMP-17 and the hybrid (*Right*) are shown in Fig. 5*C*.

Discussion

A number of computational approaches based on the hypothesis that functional specificity is determined by protein regions that are more similar within a common functional group than between distinct functional groups have been put forth for predicting SDPs (22–26). Because we determined that distinctions in MMP substrate recognition directly correlate with the phylogenetic distance in the protein sequence of the MMP catalytic domains, our findings are generally consistent with this hypothesis. To the best of our knowledge, our work is the first to achieve the quantitative experimental confirmation of the correlations between functional and phylogenetic distinctions. More significantly, we observe a remarkable correlation between substrate recognition and the sequence identity of 50–57 (depending on the MMP) noncontiguous residues that line the mouth of the MMP catalytic pocket. Our results support the idea that these residues are the SDPs responsible for substrate distinction among the MMP catalytic domains. Together, these results can serve as a benchmark for both testing algorithms aimed at identifying SDPs, and developing new approaches to assign a quantitative value for the significance of the individual SDPs. Of course, the use of our large dataset as a benchmark comes with the proviso that it measures substrate cleavage, rather than substrate binding.

There are prior studies that lend support to our assignment of SDPs in MMPs. For example, we previously identified the role of Glu412 in MMP-2 and the corresponding Asp in MMP-9 as a key to substrate distinction by these MMPs. Position 412 in MMP-2 is among the SDPs we identified in our current study. Second, a SDP subset we identified here in the SV-hB loop (Fig. 4) colocalizes with the region that is required for the collagenolytic activity of several MMPs (35, 38, 39). This colocalization raises questions as to whether these particular SDPs distinguish the collagenolytic vs. noncollagenolytic MMPs (see below).

Our data suggest that there is a distinct level of contribution of individual SDPs to substrate recognition. Although the SDPs comprise 50–57 residues, crystal structures indicate that only a handful of these positions are within a few angstroms of bound substrate and could directly bind to it. Our results also show that transmutation of a few residues that are close to the bound substrate is sufficient to transform, at least qualitatively, the substrate recognition function of one MMP into that of another. It is important, however, to emphasize that our transmutation test is based on the use of MMP-16 and MMP-17 and the peptide substrates that distinguish these two MMPs by between 3- and 10,000-fold. The resulting average $k_{(obs)}$ for these substrates proves the functional transmutation of MMP-16 into MMP-17 via the hybrid. However, the $k_{(obs)}$ for certain individual substrates did not precisely revert in the hybrid to the levels observed in MMP-17. So, although, in general, the functional transmutation of MMP-16 into MMP-17 was successful, it did not result in a precise transmutation of the kinetic behavior for each and every substrate. It is reasonable to conclude that other SDPs, which are distinct from those we mutated, provide the fine-tuning of the MMP substrate specificity, and could even be involved in inducing the torsion thought to be involved in cleavage of the peptide bond.

The catalytic cleft of MMPs binds two additional types of ligands: the MMP propeptide, which maintains the proenzyme status of MMPs, and the natural inhibitors of MMPs called tissue inhibitors of metalloproteinases (TIMPs). It is likely that many of the same residues that make contact with the substrate are also important in binding to these key ligands. Consequently, evolutionary pressure to acquire a wider, or different, substrate

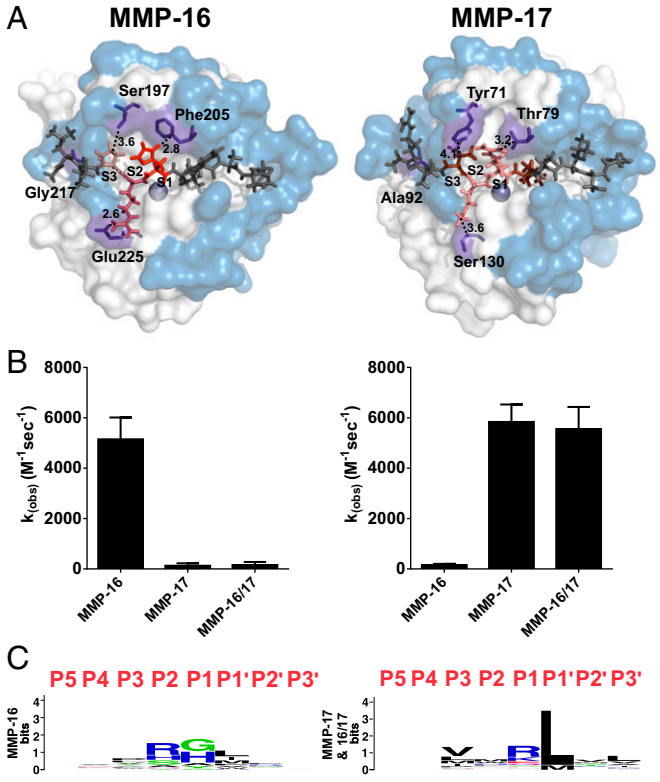

Fig. 5. Swapping dominant SDPs transmutes selectivity of MMP-16 into that of MMP-17. (*A*) Structures of the catalytic domain of MMP-16 (*Left*) and MMP-17 (*Right*) with optimal peptide substrates (acetyl-Leu-Val-Pro-Arg-His↓Leu-Phe-Ala-Ser-Gly-*N*-methyl and acetyl-Arg-Val-Val-Met-Arg↓Leu-Val-Leu-Ser-Gly-*N*-methyl, respectively) docked into the active site groove as described in *Materials and Methods*. Regions of the catalytic domains that are SDPs are colored light blue or purple if they were chosen for mutation. The substrate side chains are colored to indicate the SACC score for the corresponding positions in the catalytic cleft (S3–S1). So residues in darker red are at substrate positions with high SACC scores (Fig. 2*B*), and lighter red indicates residues at positions with lower SACC scores. The remainder of the docked substrate (outside of S3–S1) is colored dark gray. (*B*) The effect of the transmutation on substrate selectivity was determined using ~140 phage substrates about one-half of which were selective for each protease. The $k_{(obs)}$ for each substrate was measured as described in *Materials and Methods*. Error bars represent SEMs. (*C*) Sequence logos indicating the frequency of residues at each position calculated for substrates selective for MMP-16 (*Left*) and those preferentially cleaved by the 16/17 transmutant and MMP-17 (*Right*).

preference was probably counterbalanced by the pressure to preserve binding to the prodomain and TIMPs, which, in general, are less sequence divergent than substrates. This may explain why all of the MMPs examined here retain the ability to efficiently cleave the substrates with the P-X-X-↓L motif. It may be that cleavage of this sequence is a "passenger function" of the evolutionary pressure to maintain propeptide and TIMP binding.

The mechanistic relationship between cleavage of a linear sequence, as studied here, and cleavage of 3D triple helical collagen by some MMPs is not entirely clear. In almost all cases, the hemopexin domain, which is adjacent to the catalytic domain, is required for cleaving collagen (40–42), probably by providing an ancillary binding surface. Furthermore, recent work indicates that the MMP-1 collagenase is engaged in an intermittent diffusion alongside the collagen fibril, with the bias component being attributed to proteolysis of substrate (43). These pauses may allow for unwinding of the fibril to make the cut sites more accessible (44). In addition to the impact of ancillary domains in binding collagen, cleavage of this complex substrate is fundamentally different from hydrolysis of linear peptide/protein substrates by MMPs. This distinction is illustrated by work with triple helical peptides that model collagen structure (45–47). It will be interesting to know how the SDPs identified here contribute to this multistep process of collagen cleavage, which is unique to only a few MMPs.

There are many noncollagenous substrates for the MMPs, including membrane proteins that are shed from the cell surface (48–51). MMPs can also cleave proteins like the monocyte chemoattractant proteins, converting them from agonists into antagonists (52, 53). The substrate preferences identified here may be more reflective of this type of activity by MMPs, especially because our prior computational analysis on all MMP cleavage sites reported in the literature showed the proteases to cleave preferentially in exposed loops and in the disordered regions of proteins (54). It may ultimately be possible to use the results on substrate preference to assign a "fitness" score for each peptide bond in the proteome for its probability of being susceptible to cleavage by any MMP. However, this will require extensive statistical analysis, and experimental validation in the future. Ultimately such fitness scores could help formulate new hypotheses and to guide proteomic searches for proteins that have been cleaved in a complex mixture of proteins (4, 33, 34).

Our findings also indicate that there could be substantial distinction in the substrates preferences of MMPs present even in the same phylogenetic branch. For example, 22 from 50 SDPs differ in MMP-2 vs. MMP-9. This divergence is likely to be the foundation for the distinct substrate preference of MMP-2 and MMP-9. In agreement with this statement, we are able to detect substrates that readily discriminate MMP-2 and MMP-9 (Fig. 1*B* and Dataset S1). Thus, 15 substrates are ≥10-fold more selective for MMP-9 vs. MMP-2, whereas 174 substrates are ≥10-fold selective for MMP-2 vs. MMP-9. Moreover, several MMP-2 substrates were >1,000-fold more selective for MMP-2 compared with MMP-9. Proteomic studies also suggest that MMP-2 and MMP-9 cleave distinct subsets of cellular substrates (55).

Other reports show that exosites on the surface of the catalytic domains of the MMPs can modify proteolytic function. One study used a library of nonnatural polymeric ligands to identify 14 residues in MMP-14 that could be physiologically relevant allosteric inhibitory sites (48). Interestingly, of these 14 positions only 3 overlapped with the 50 SDPs identified by us, suggesting little functional, or evolutionary, overlap among the two functional sites. Consequently, it seems that substrate selectivity and allosteric inhibition probably evolved separately. This line of thought has some broader implications as one might envision using diverse libraries targeted to different functional sites to elucidate the structural basis, and evolutionary relationship, of many different types of ligands for any given family of proteins. It may even be possible to extend this analysis to identify areas in multidomain proteins, or even protein complexes that coevolve with a particular function.

Although all of the SDPs are likely required for the precise kinetic behavior of the individual MMPs toward the extended substrate space, our results show that, on an aggregate level, substitution of a limited number of SDPs can transform proteolytic function of one MMP to that of another. Thus, transmutation of only four SDP residues changed the proteolytic function of MMP-16 into that of MMP-17. We believe that this and similar knowledge may contribute to the redesign of MMPs to exhibit unconventional substrate specificity, which could have therapeutic, diagnostic, or other biotechnology applications (56, 57). The substrate specificity of several proteases distinct from MMPs has been manipulated (58–60), but this was thought to be difficult to achieve with MMPs because of their presumed overlapping substrate specificity. MMPs not only offer a different set of potential target sequences but also have a longer substrate binding cleft than most serine proteases. Consequently, if the subsite recognition preference can be altered simultaneously at several positions, it may be possible to create mutant MMPs with extremely selective target sequences. In addition, the MMP substrate recognition motifs identified here are likely to be valuable for the interpretation of proteomics data and to the assignment of the individual protein substrates to a cognate protease (61–63).

In sum, our compendium of cleavage data constitutes a truly valuable resource. The results provide a large dataset for benchmarking computational predictions of SDPs, give information for subsequent use in predicting physiologic substrates, indicate directed strategies for designing proteases with novel substrates, and may help design new diagnostic assays based on proteolytic function.

Materials and Methods

Expression and Purification of Recombinant Catalytic Domains. The recombinant catalytic domains of MMP-2 and -9 were expressed, purified from culture medium from transfected HEK 293 cells, and their active site was titrated as previously described (36, 64). The catalytic domains of MMP-14 (spanning Tyr112 to Gly284), MMP-15 (Tyr132 to Gly304), MMP-16 (Tyr120 to Gly291), MMP-17 (Gln129 to Gly298), MMP-24 (Tyr156 to Gly327), and MMP-25 (Tyr108 to Gly280) were amplified from a mixture of three first-strand cDNA libraries (Human Panel I, Fetal Panel, and Tumor Panel; all from Clontech). The amplified fragments were ligated into the NdeI–XhoI sites of the expression vector pET21a(+) to generate constructs tagged with a C-terminal His×6 tag. The authenticity of the constructs was confirmed by DNA sequencing. Recombinant catalytic domains of all MMPs (except MMP-2 and -9) were expressed as inclusion bodies in BL-21 DE3 Codon Plus RIL *E. coli* cells, purified using Ni-NTA agarose (GE Healthcare) after solubilization in 8 M urea, and refolded. To ensure proper quantification of active proteases, their active sites were titrated with active site inhibitors. MMP-2 and -9 were titrated with GM6001 (Ilomastat) (65), and the AG3340 (66) was used to titrate the other MMPs.

Phylogeny of MMP Catalytic Domains. Phylogenetic trees of the MMPs catalytic domains were built from protein families developed by ENSEMBL (www.ENSEMBL.org) using the Markov clustering (MCL) package (http://micans.org/mcl/) (67). Two gene families, ENSF00000000368 (358 protein sequences) and ENSF00000000425 (279 sequences), were downloaded from the website and supplemented with sequences of 23 human MMPs from SWISS-Prot (660 sequences altogether). The sequences were aligned using the MUSCLE program with its default parameters (31). Sequences lacking complete catalytic domains were then removed from the analysis. HyperTree (32) was used to draw NJ tree for the nearly 600 MMPs catalytic domains used for the comparison.

Substrate Phage Selections. The substrate phage selection was performed essentially as in ref. 68 using the semiautomated substrate phage platform we recently described (39), and which is illustrated in *SI Materials and Methods*. Briefly, 5×10^{11} phage particles are mixed with a protease at 200 nM of active enzyme in the Reaction Buffer in a total volume of 0.5 mL and incubated for 2 h at 37 °C. A reaction without the protease was performed as a control. Following hydrolysis of phage, the reaction was halted by addition of Ilomastat or AG3340. The uncleaved phage were separated by immunodepletion using M2 anti-FLAG monoclonal antibody (Sigma-Aldrich) coupled to epoxy-activated M-450 magnetic beads (Invitrogen). Following two rounds of selection, individual phage colonies were picked using QPix2 robot, transferred into deep-well 96-well culture plates (Simport), and grown overnight. All subsequent characterizations were performed with these individual phage cultures. Through the phage selections, we arrive at 1,369 test substrates that were used in this study. However, statistical analyses indicate that the number of substrates could probably be reduced to ~600 and still provide the same information.

Quantifying Cleavage of Phage Substrates. Hydrolysis of individual phage substrates was performed using a modified substrate phage ELISA as previously described (34). In brief, phage from overnight 1-mL cultures were captured in 96-well microtiter plates coated with anti-M13 antibody and incubated in the Reaction Buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 5 mM $CaCl_2$, 0.005% Brij 35) with or without addition of 50 nM individual protease catalytic domain at 37 °C for 2 h. After removal of the protease by washing, the FLAG content of the captured phage was determined by incubation with HRP-conjugated M-2 anti-FLAG monoclonal antibody at 4 μg/mL for 1 h (Sigma-Aldrich) followed by detection at 490 nm with *o*-phenylenediamine substrate. The extent of hydrolysis or the hydrolysis score (HS) was calculated as 1 minus the ratio between OD at 490 nm of the protease-treated and the matching protease-untreated samples. Because the substrate concentrations ($\sim 10^{-10}$ M) were much below the typical K_M ($\sim 10^{-4}$ M for peptide substrates), we determined the $k_{(obs)}$ values for each reaction. We chose to use the term $k_{(obs)}$ because the substrates were immobilized in wells of a microtiter plate when measures were made, so we cannot meet the precise requirements of Michaelis–Menten kinetics, which dictate that both substrate and enzyme are in solution. Therefore, instead of using the term k_{cat}/K_M, to quantify cleavage efficiency, we instead use $k_{(obs)}$ as a relative measure of cleavage efficiency. $k_{(obs)}$ was calculated using the integrated Michaelis–Menten equation:

$$[P] = [S]_0\left(1 - e^{-kt}\right),$$

where $[P]$ is the product concentration, $[S]_0$ represents the initial substrate concentration, and $k = (k_{(obs)})[E]$. $[E]$ represents the enzyme concentration, and t is the reaction time at which substrate concentration is $[S]_t$. Because HS is defined as

$$\mathrm{HS} = \frac{[S]_0 - [S]_t}{[S]_0} = \frac{[P]}{[S]_0} = 1 - e^{-kt},$$

therefore

$$k(\mathrm{obs}) = -\ln(1 - \mathrm{HS})/(t \cdot [E]).$$

Determination of Scissile Bonds Within Phage Substrates. The position of scissile bonds within individual phage substrates was determined using methods we previously described (34), and as illustrated in *SI Materials and Methods*. Briefly, phage from overnight cultures were PEG precipitated and then dissolved and incubated with 50 nM protease for 2 h at 37 °C. For poor substrates, the reaction times were extended to 8 h. The reactions were stopped by addition of MMP inhibitors and the portion of the substrate released from the phage (the C-terminal portion of the cleavage product) was precipitated with anti-FLAG monoclonal antibody-coupled M-280 magnetic beads (Invitrogen). The cleavage product was eluted from the beads with 0.1% TFA and spotted a MALDI-target plate (AnchorChip 400 from Bruker Daltonics). The masses of cleaved peptides were determined using UltraFLEX II MALDI-TOF/TOF mass spectrometer (Bruker Daltonics). The position of the scissile bond in each phage substrate was derived by comparing the mass of the released product to the predicted mass of the phage insert (sequence determined by cDNA sequencing).

Docking Substrates to MMP-16 and MMP-17. Models of substrate docked into the catalytic pocket of MMP-16 and MMP-17 were built by AMBER molecular modeling program using TIMP-2 amino acids from MMP-14–TIMP-2 crystal structures (1BQQ) as guiding template. The structures of either MMP-17 or MMP-16 were superimposed onto the crystal structure of MMP-14 by the FATCAT program. In these models, we maintained the backbone peptide chain of substrate (TIMP-2) according to the crystal structure, but replaced the side chains with those of the corresponding optimal substrate. These were acetyl-LVPRHL-FASG-*N*-methyl amide for MMP-16 and acetyl-RVVMRL-VLSG-*N*-methyl amide for MMP-17. Side chains were positioned as closely to residues in MMP by combination of limited molecular-dynamics simulations and force field minimization.

ACKNOWLEDGMENTS. We thank Guy S. Salvesen for thoughtful discussions. The research in this paper was supported by National Institutes of Health Grants CA83017, CA157328, DE022757 (to A.S.), GM098835 (to P.C.), and RR080243 (to J.W.S.).

1. Cunningham BC, Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244(4908):1081–1085.
2. Teichmann SA, Grishin NV (2008) Sequences and topology: From genome structure to protein structure. *Curr Opin Struct Biol* 18(3):340–341.
3. Friedberg I, Jaroszewski L, Ye Y, Godzik A (2004) The interplay of fold recognition and experimental structure determination in structural genomics. *Curr Opin Struct Biol* 14(3):307–312.
4. Jones DT (2000) Protein structure prediction in the postgenomic era. *Curr Opin Struct Biol* 10(3):371–379.
5. Pirovano W, Heringa J (2010) Protein secondary structure prediction. *Methods Mol Biol* 609:327–348.
6. Xu D, Jaroszewski L, Li Z, Godzik A (2014) FFAS-3D: Improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* 30(5):660–667.
7. Kirshner DA, Nilmeier JP, Lightstone FC (2013) Catalytic site identification—a web server to identify catalytic site structural matches throughout PDB. *Nucleic Acids Res* 41(Web Server issue):W256–W265.
8. Fajardo JE, Fiser A (2013) Protein structure based prediction of catalytic residues. *BMC Bioinformatics* 14:63.
9. Standley DM, Yamashita R, Kinjo AR, Toh H, Nakamura H (2010) SeSAW: Balancing sequence and structural information in protein functional mapping. *Bioinformatics* 26(9):1258–1259.

10. Wang K, Horst JA, Cheng G, Nickle DC, Samudrala R (2008) Protein meta-functional signatures from combining sequence, structure, evolution, and amino acid property information. *PLoS Comput Biol* 4(9):e1000181.
11. McDowall J, Hunter S (2011) InterPro protein classification. *Methods Mol Biol* 694: 37–47.
12. Cai XH, Jaroszewski L, Wooley J, Godzik A (2011) Internal organization of large protein families: Relationship between the sequence, structure, and function-based clustering. *Proteins* 79(8):2389–2402.
13. Röttig M, Rausch C, Kohlbacher O (2010) Combining structure and sequence information allows automated prediction of substrate specificities within enzyme families. *PLoS Comput Biol* 6(1):e1000636.
14. Fryxell KJ (1996) The coevolution of gene family trees. *Trends Genet* 12(9):364–369.
15. Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE (2000) Co-evolution of proteins with their interaction partners. *J Mol Biol* 299(2):283–293.
16. Pazos F, Valencia A (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng* 14(9):609–614.
17. Lin EC, et al. (1997) Identification of a region in the integrin beta3 subunit that confers ligand binding specificity. *J Biol Chem* 272(38):23912–23920.
18. Yuan L, Voelker TA, Hawkins DJ (1995) Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. *Proc Natl Acad Sci USA* 92(23):10639–10643.
19. de Prat Gay G, Duckworth HW, Fersht AR (1993) Modification of the amino acid specificity of tyrosyl-tRNA synthetase by protein engineering. *FEBS Lett* 318(2): 167–171.
20. Bradshaw JM, Mitaxov V, Waksman G (2000) Mutational investigation of the specificity determining region of the Src SH2 domain. *J Mol Biol* 299(2):521–535.
21. Kaustov L, et al. (2011) Recognition and specificity determinants of the human cbx chromodomains. *J Biol Chem* 286(1):521–529.
22. Teppa E, Wilkins AD, Nielsen M, Buslje CM (2012) Disentangling evolutionary signals: Conservation, specificity determining positions and coevolution. Implication for catalytic residue prediction. *BMC Bioinformatics* 13:235.
23. Pirovano W, Feenstra KA, Heringa J (2006) Sequence comparison by sequence harmony identifies subtype-specific functional sites. *Nucleic Acids Res* 34(22):6540–6548.
24. Chakrabarti S, Panchenko AR (2009) Coevolution in defining the functional specificity. *Proteins* 75(1):231–240.
25. Mazin PV, et al. (2010) An automated stochastic approach to the identification of the protein specificity determinants and functional subfamilies. *Algorithms Mol Biol* 5:29.
26. Ye K, Feenstra KA, Heringa J, Ijzerman AP, Marchiori E (2008) Multi-RELIEF: A method to recognize specificity determining residues from multiple sequence alignments using a Machine-Learning approach for feature weighting. *Bioinformatics* 24(1): 18–25.
27. Mirny LA, Gelfand MS (2002) Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors. *J Mol Biol* 321(1): 7–20.
28. Li L, Shakhnovich EI, Mirny LA (2003) Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proc Natl Acad Sci USA* 100(8):4463–4468.
29. Kolesov G, Mirny LA (2009) Using evolutionary information to find specificity-determining and co-evolving residues. *Methods Mol Biol* 541:421–448.
30. Ivanisenko VA, Eroshkin AM, Kolchanov NA (2005) WebProAnalyst: An interactive tool for analysis of quantitative structure-activity relationships in protein families. *Nucleic Acids Res* 33(Web Server issue):W99–W104.
31. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
32. Bingham J, Sudarsanam S (2000) Visualizing large hierarchical clusters in hyperbolic space. *Bioinformatics* 16(7):660–661.
33. Massova I, Kotra LP, Fridman R, Mobashery S (1998) Matrix metalloproteinases: Structures, evolution, and diversification. *FASEB J* 12(12):1075–1095.
34. Ratnikov B, Cieplak P, Smith JW (2009) High throughput substrate phage display for protease profiling. *Methods Mol Biol* 539:93–114.
35. Bhaskaran R, Palmier MO, Lauer-Fields JL, Fields GB, Van Doren SR (2008) MMP-12 catalytic domain recognizes triple helical peptide models of collagen V with exosites and high activity. *J Biol Chem* 283(31):21779–21788.
36. Chen EI, Li W, Godzik A, Howard EW, Smith JW (2003) A residue in the S2 subsite controls substrate selectivity of matrix metalloproteinase-2 and matrix metalloproteinase-9. *J Biol Chem* 278(19):17158–17163.
37. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14(6):1188–1190.
38. Chung L, et al. (2000) Identification of the (183)RWTNNFREY(191) region as a critical segment of matrix metalloproteinase 1 for the expression of collagenolytic activity. *J Biol Chem* 275(38):29610–29617.
39. Pelman GR, Morrison CJ, Overall CM (2005) Pivotal molecular determinants of peptidic and collagen triple helicase activities reside in the S3′ subsite of matrix metalloproteinase 8 (MMP-8): The role of hydrogen bonding potential of ASN188 and TYR189 and the connecting *cis* bond. *J Biol Chem* 280(3):2370–2377.
40. Sela-Passwell N, Rosenblum G, Shoham T, Sagi I (2010) Structural and functional bases for allosteric control of MMP activities: Can it pave the path for selective inhibition? *Biochim Biophys Acta* 1803(1):29–38.
41. Arnold LH, et al. (2011) The interface between catalytic and hemopexin domains in matrix metalloproteinase-1 conceals a collagen binding exosite. *J Biol Chem* 286(52):45073–45082.
42. Lauer-Fields JL, et al. (2009) Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity. *J Biol Chem* 284(36): 24017–24024.
43. Sarkar SK, Marmer B, Goldberg G, Neuman KC (2012) Single-molecule tracking of collagenase on native type I collagen fibrils reveals degradation mechanism. *Curr Biol* 22(12):1047–1056.
44. Chung L, et al. (2004) Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO J* 23(15):3020–3030.
45. Minond D, et al. (2006) The roles of substrate thermal stability and P2 and P1′ subsite identity on matrix metalloproteinase triple-helical peptidase activity and collagen specificity. *J Biol Chem* 281(50):38302–38313.
46. Akers WJ, et al. (2012) Detection of MMP-2 and MMP-9 activity in vivo with a triple-helical peptide optical probe. *Bioconjug Chem* 23(3):656–663.
47. Minond D, et al. (2007) Differentiation of secreted and membrane-type matrix metalloproteinase activities based on substitutions and interruptions of triple-helical sequences. *Biochemistry* 46(12):3724–3733.
48. Sanderson MP, Dempsey PJ, Dunbar AJ (2006) Control of ErbB signaling through metalloprotease mediated ectodomain shedding of EGF-like factors. *Growth Factors* 24(2):121–136.
49. Kumar S, et al. (2014) Antibody-directed coupling of endoglin and MMP-14 is a key mechanism for endoglin shedding and deregulation of TGF-β signaling. *Oncogene* 33(30):3970–3979.
50. Rizzo R, et al. (2013) Matrix metalloproteinase-2 (MMP-2) generates soluble HLA-G1 by cell surface proteolytic shedding. *Mol Cell Biochem* 381(1-2):243–255.
51. Rahman M, et al. (2013) Platelet shedding of CD40L is regulated by matrix metalloproteinase-9 in abdominal sepsis. *J Thromb Haemost* 11(7):1385–1398.
52. McQuibban GA, et al. (2002) Matrix metalloproteinase processing of monocyte chemoattractant proteins generates CC chemokine receptor antagonists with anti-inflammatory properties in vivo. *Blood* 100(4):1160–1167.
53. McQuibban GA, et al. (2000) Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3. *Science* 289(5482):1202–1206.
54. Kazanov MD, et al. (2011) Structural determinants of limited proteolysis. *J Proteome Res* 10(8):3642–3651.
55. Prudova A, auf dem Keller U, Butler GS, Overall CM (2010) Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol Cell Proteomics* 9(5):894–911.
56. Craik CS, Page MJ, Madison EL (2011) Proteases as therapeutics. *Biochem J* 435(1): 1–16.
57. Li Q, Yi L, Marek P, Iverson BL (2013) Commercial proteases: Present and future. *FEBS Lett* 587(8):1155–1163.
58. Pogson M, Georgiou G, Iverson BL (2009) Engineering next generation proteases. *Curr Opin Biotechnol* 20(4):390–397.
59. Yi L, et al. (2013) Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc Natl Acad Sci USA* 110(18):7229–7234.
60. Varadarajan N, Rodriguez S, Hwang BY, Georgiou G, Iverson BL (2008) Highly active and selective endopeptidases with programmed substrate specificities. *Nat Chem Biol* 4(5):290–294.
61. Agard NJ, et al. (2012) Global kinetic analysis of proteolysis via quantitative targeted proteomics. *Proc Natl Acad Sci USA* 109(6):1913–1918.
62. Lange PF, Overall CM (2013) Protein TAILS: When termini tell tales of proteolysis and function. *Curr Opin Chem Biol* 17(1):73–82.
63. Kleifeld O, et al. (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* 6(10): 1578–1611.
64. Kridel SJ, et al. (2001) Substrate hydrolysis by matrix metalloproteinase-9. *J Biol Chem* 276(23):20572–20578.
65. Galardy RE, et al. (1994) Low molecular weight inhibitors in corneal ulceration. *Ann N Y Acad Sci* 732:315–323.
66. Shalinsky DR, et al. (1999) Broad antitumor and antiangiogenic activities of AG3340, a potent and selective MMP inhibitor undergoing advanced oncology clinical trials. *Ann N Y Acad Sci* 878:236–270.
67. van Dongen S, Abreu-Goodger C (2012) Using MCL to extract clusters from networks. *Methods Mol Biol* 804:281–295.
68. Matthews DJ, Wells JA (1993) Substrate phage: Selection of protease substrates by monovalent phage display. *Science* 260(5111):1113–1117.