Peptidase specificity from the substrate cleavage collection in the MEROPS database and a tool to measure cleavage site conservation

Neil D. Rawlings
Wellcome Trust Sanger Institute and the EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK

Abstract

One peptidase can usually be distinguished from another biochemically by its action on proteins, peptides and synthetic substrates. Since 1996, the MEROPS database (http://merops.sanger.ac.uk) has accumulated a collection of cleavages in substrates that now amounts to 66,615 cleavages. The total number of peptidases for which at least one cleavage is known is 1700 out of a total of 2457 different peptidases. This paper describes how the cleavages are obtained from the scientific literature, how they are annotated and how cleavages in peptides and proteins are cross-referenced to entries in the UniProt protein sequence database. The specificity profiles of 556 peptidases are shown for which ten or more substrate cleavages are known. However, it has been proposed that at least 40 cleavages in disparate proteins are required for specificity analysis to be meaningful, and only 163 peptidases (6.6%) fulfil this criterion. Also described are the various displays shown on the website to aid with the understanding of peptidase specificity, which are derived from the substrate cleavage collection. These displays include a logo, distribution matrix, and tables to summarize which amino acids or groups of amino acids are acceptable (or not acceptable) in each substrate binding pocket. For each protein substrate, there is a display to show how it is processed and degraded. Also described are tools on the website to help with the assessment of the physiological relevance of cleavages in a substrate. These tools rely on the hypothesis that a cleavage site that is conserved in orthologues is likely to be physiologically relevant, and alignments of substrate protein sequences are made utilizing the UniRef50 database, in which in each entry sequences are 50% or more identical. Conservation in this case means substitutions are permitted only if the amino acid is known to occupy the same substrate binding pocket from at least one other substrate cleaved by the same peptidase.

© 2015 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In 2007, Barrett & Rawlings [1] proposed a list of criteria to distinguish one peptidase from another. To be considered different, any one of the following bioinformatics tests can be applied: the two peptidases have similar biochemical characteristics but unrelated sequences; the two peptidases have related sequences but different biochemical properties, different domain architectures or the domains are in a different order; or the two peptidases have greater than 50% sequence identity but are derived from nodes on a phylogenetic tree that are not adjacent. In addition, the following biochemical tests can be applied to distinguish two peptidases: the peptidases act under significantly different conditions; the peptidases have different post-translational modifications; the peptidases are sensitive to different inhibitors; the peptidases act on different substrates, or if they act on the same substrates then the cleavage positions are different. It is the last two criteria with which this paper is concerned.

A peptidase cleaves a substrate at the scissile bond, and substrate residues either side of this bond are known as P1 and P1'. Residues towards the N-terminus of the substrate are on the non-prime side, and are numbered P1, P2, P3, P4 and so on. Residues towards the C-terminus are on the prime side and are numbered P1', P2', P3', P4' and so on. A substrate binding pocket in the peptidase that accommodates a substrate residue is named according to the position the residue occupies in the substrate, except that the “P” is replaced by an “S”. So the S1 binding pocket accommodates the P1 residue, and the S4' binding pocket accommodates the P4' residue [2].

A collection of substrate cleavages has been assembled from the scientific literature, annotated, cross-referenced where applicable.
to the UniProt protein sequence database, and included within the MEROPS database. This collection was originally derived from the CD-ROM version of the first edition of the *Handbook of Proteolytic Enzymes* (1998) [3], which also included a search facility to find the peptidases able to cleave a substrate at a particular position. By knowing where in proteins, peptides or synthetic substrates cleavages occur, it is possible to postulate the specificity of a peptidase. By knowing which amino acids can occupy each substrate binding position, it is also possible to infer whether or not cleavage of a substrate at a particular position is likely to be physiologically relevant from an alignment of protein sequences of closely-related orthologues.

The MEROPS substrate cleavage collection has been widely used to predict cleavages in substrates (for a review see Song et al. (2011) [4]), and to predict what peptidase may be responsible for a known cleavage, for example PROSPER [5]. The MEROPS collection has also been used for the mapping of the human degradome and prediction of “cleavage entropy” as an overall measure of peptidase specificity [6], as well as in the development of the “protease web”, the network of peptidase, substrate and inhibitor interactions [7].

This paper describes the MEROPS substrate cleavage collection and the various displays present on the MEROPS website (http://merops.sanger.ac.uk) which aid in understanding peptidase specificity and the processing and degradation of a protein substrate. In order to help determine whether or not a cleavage is physiologically relevant, a service is also described where a user can upload substrate cleavages and receive by E-mail an analysis to show how well conserved, in terms of peptidase binding, each cleavage is.

2. Materials and methods

2.1. Identification of peptidases, homology searching, sequence alignment and phylogenetic tree generation

A peptidase species was defined according to the principles established in Barrett & Rawlings (2007) [1]. The methods for homology searching, family building, and generation of protein sequence alignments and phylogenetic trees are the same as those described in Rawlings et al. (2014) [8]. In brief, the following methods were used. Only the peptidase domain was used for sequence searching and sequence alignment. For each family a type example was chosen and for each peptidase species a holotype was chosen. The type example and holotype were usually the sequence of the best characterized peptidase in the family or protein species, respectively. A BlastP search [9] of the NCBI non-redundant protein sequence database was performed, using the family type example sequence. Sequences retrieved with an E value of 0.01 or less were considered homologues and included in the family. To find more distant homologues, a HMMER search [10] was performed using a ClustalW alignment [11] of a selection of sequences from the family that included an example from every phylum for which there was a representative. Sequence alignments were built using MAFFT [12]. Phylogenetic trees were built from the family sequence alignment using QuickTree [13].

2.2. Manual substrate cleavage curation

The scientific literature was searched manually for substrate cleavage sites by peptidases. Data were acquired from over 7280 references. The following data were collected, transformed as required and stored in a MySQL database. From the name of the peptidase as given by the authors of the publication, a MEROPS identifier and, if possible, a MERNUM indicating the source organism, were assigned. From the name of the substrate and its source, a UniProt accession was assigned where possible, and the name recommended by UniProt was stored in the MySQL database, unless the substrate was a peptide or was processed, in which case a peptide name or a name to indicate that processing had occurred was stored (for example, “Met-enkephalin” would be stored in preference to “pro-opiomelanocortin” if the substrate was just the peptide). Where more than one UniProt entry existed, the annotated SwissProt accession, name and sequence where used in preference. Where isoforms derived from alternative initiation and alternative splicing were indicated in the UniProt database entry, the sequence chosen as the representative sequence by UniProt was selected unless the original publication indicated that a particular isoform had been used. There was no attempt to map a cleavage to all isoforms on the presumption that a change in sequence could lead to a change in cleavage position. The cleavage position (the position of the P1 residue in the substrate) was converted to the equivalent residue number from the respective UniProt entry. Up to four residues either side of the scissile bond (residues P4 to P4’ were stored for each cleavage. The residue range of the substrate used compared to the sequence in the UniProt entry was also stored. This allowed for annotation of peptide substrates derived from full-length proteins and processing events, such as removal of signal and transit peptides and precursor sequences. The CDC checksum for the UniProt entry was also stored so that any changes to the sequence could be identified subsequently. Kinetic data (Kcat, Kmax, and/or Kcat/Kmax) were stored where available. Annotations to indicate how the peptidase and cleavage position were identified were also stored. The initials of the curator and the date the cleavage was collected were also stored. The reference was stored in a Reference Manager database (Thomson Reuters) and the PubMed accession was obtained and stored where possible. Any additional data that affected where cleavage occurred, such as reactions conditions, where stored as a comment in the MySQL database.

To ensure that curation was consistent, a Perl program was written to aid cleavage data collection and storage. The user (either the author or a summer student) was asked to enter his or her initials; the UniProt accession of the protein substrate in question; the cleavage position; the residue range of the substrate sequence compared to the UniProt entry; the codes for how the cleavage was identified and how the peptidase was identified; whether the cleavage was physiological, non-physiological, pathological or theoretical; whether the substrate was denatured; the reference and its PubMed identifier; and any comment.

Collection of cleavage data from the literature was also outsourced to Molecular Connections, Bangalore, India. Data were returned to the author as an Excel spreadsheet and a pipeline developed to extract data from the spreadsheet and import it into the MySQL database. Existing substrate cleavage collections were also imported into the MEROPS collection. These included data from the CutDB database [14] and the CASBAH database of caspase substrates [15].

A Perl program to check that the P4–P4’ residues around the cleavage position matched the sequence in the UniProt entry was written as a quality control measure and to identify any subsequent changes in the UniProt sequence. Cleavage data were also stored for the cleavage of synthetic substrates. These were manually entered into the database. For a synthetic substrate it was not possible to map the sequence to a UniProt database entry. The P4–P4’ positions around the scissile bond were stored where possible (many synthetic substrates do not have residues beyond P1 or P3), including a unique identifier for each N- or C-terminal blocking or reporter group occurring within that range.

In certain cases, it was not possible to map a cleavage to a single enzyme. This most frequently occurred when cleavage was performed by an enzyme complex, such as the proteasome or
eukaryotic signal peptidase, or where cleavage might have been performed by one or more enzymes, such as removal of the initiating methionine by either of the eukaryotic methionyl aminopeptidases, or where specificity overlapped, for example caspases-3 and -7. Instead of mapping the cleavage to an individual enzyme, the cleavage was mapped to the peptidase family.

2.3. Automated collection of substrate cleavages from SwissProt entries

Entries in the SwissProt section of UniProt are annotated for removal of initiating methionine, signal and transit peptides, and processing events to activate proteins or extract peptides and proteins from polyproteins. These, if not already present in the substrate cleavage collection, were automatically collected using a Perl script. Removal of an initiating methionine was mapped to a MEROPS identifier either for methionyl aminopeptidase (for a prokaryote) or the M24 family (for a eukaryote). Removal of a signal peptidase was mapped to a MEROPS identifier either for signal peptidase or a signal peptidase complex (depending on whether the source species was a prokaryote or a eukaryote). Where the peptidase responsible for the cleavage was not known, the cleavage was not mapped to a specific peptidase or peptidase family. Theoretical cleavages, for example the release of a predicted signal peptide, were not collected.

2.4. Automated collection of substrate cleavages from proteomics experiments

High-throughput identification of proteins in a sample is known as “proteomics”. In many proteomics experiments, proteins are digested with a peptidase so that each protein can be identified from the mass and charge (and sometimes sequence) of its peptides. This requires a prior in silico digestion of all proteins in the proteome, and the choice of the peptidase used for the digestion is important: the peptidase must have a very simple specificity that is easily predictable. For this reason, trypsin, which cleaves all lysyl and arginyl bonds in a denatured protein, with the exception of except Lys-Pro and Arg-Pro bonds, is frequently used. There are many tools to perform theoretical cleavages, for example Peptide-Mass (http://web.expasy.org/peptide_mass/,[16]) which predicts cleaves by peptidases such as trypsin, chymotrypsin, peptidyl-Lys metallopeptidase and glutamylation peptidase I as well as cleavages by chemicals such as CnBr. The neXtProt database includes trypsin digestions of all human proteins and indicates peptides unique to a protein and whether a particular peptide is found in many human proteins [17]. The PRIDE database [[18]; http://www.ebi.ac.uk/pride/archive/] stores the results of proteomics experiments where peptides are identified from a sample by mass spectroscopy. There are a number of different techniques employed of which PICS, TAILS, COFRADIC, Subtiligase, ChaFRADIC and N-terminomics are the most well-known; these have been reviewed by Lai et al. [19] and Schlage & auf dem Keller [20]. TAILS [21] and COFRADIC [22] are techniques that have been employed to identify physiological substrates, including naturally occurring N-termini. Frequently, the peptides derived from a protein can provide evidence of post-translational processing, such as removal of an initiating methionine, signal peptide or transit peptide. A Perl program was written to collect peptides and their correspondingUniProt entries from the PRIDE database and check each to see if the N-terminus matched the site of removal of a hypothetical signal or transit peptide. Because aminopeptidases may further trim an N-terminus once a transit peptidase has been removed, for example aminopeptidase P3 removes an unstable amino terminal tyrosine from mitochondrial proteins following removal of the first transit peptide by mitochondrial processing peptidase but before removal of a second transit peptide by mitochondrial intermediate peptidase [23], the cleavage was only added to the MEROPS collection if the peptide identified and the predicted processing event corresponded. Cleavages from digests performed with more unusual peptidases, such as glutamyl endopeptidase for example, were also collected to boost the number of substrates for these peptidases.

In the last decade or so, there have been a number of attempts to determine peptidase specificity or discover physiological substrates by use of proteomics. In a typical proteomics experiment, a sample, often a human cell lysate, is divided into two portions. The first portion is digested only by trypsin (or a similar peptidase with a known and predictable specificity), and the second with trypsin and a peptidase of choice. The peptides generated are identified by mass spectroscopy. The cleavages that occur only in the second portion are the result of the peptidase of choice. If the second portion is digested first by trypsin and then by the peptidase of choice, then a suite of peptides is prepared to investigate the specificity of the peptidase of choice and the cleavages by this peptidase are not physiologically relevant. Alternatively, if the second portion is digested first with the peptidase of choice and then by trypsin, the digested proteins are likely to include physiological substrates. However, some proteins may be digested because the usual physiological barriers to digestion (compartmentalization, differences in pH, etc) have been removed in the experiment, and some proteins may be “bystanders” which are cleaved under physiological conditions but the cleavage is non-functional [24]. The results from analyses by software such as Mascot [25] or PeptideProphet [26] are exported to Excel spreadsheets. These are often deposited as material supplementary to the paper, but spreadsheets have also been kindly supplied by the authors. In a given sample, there can be hundreds or even thousands of peptide products, so manual curation is not possible. A Perl program was written to extract the accessions or identifiers, translate them into UniProt accessions, and then determine the residue number for the P1 residue in each cleavage. All of the cleavages from the seminal paper by Schilling & Overall [27] were collected, including cleavages by trypsin, but for subsequent papers, trypsin cleavages were not collected.

2.5. Reliability scores

A “reliability score” is calculated which is an average percentage difference for all substrates of a peptidase over the range P4–P4’. All possible pairwise comparisons are made and the number of differences summed for all comparisons. For substrates with non-standard amino acids and blocking and reporter groups such as those found in synthetic substrates, the unusual residues are replaced with “X” before the comparisons are performed. The sum of differences is divided by the number of comparisons times the number of positions considered. For most endopeptidases, the number of positions considered is eight, because at least of one of each of the positions P4 to P4’ will be filled. However, for aminopeptidases, where P4 to P2 are empty, and carboxypeptidases, where P2' to P4' are empty, only five positions are considered. For dipeptidyl-peptidases and peptidyl-dipeptidases only six positions are considered, and for dipeptidylpeptidases only two. The number of positions considered is calculated and not assumed. A peptidase with a reliability score of greater than 75 means that that its substrates are varied in sequence and the calculated preferences are likely to be correct. A peptidase with a score in the range 50–74% had a proportion of substrates with similar sequences and thus the calculated preferences were less reliable, and the calculated preferences for a peptidase with a score of less than 50% should be treated with caution.
3. Results and discussion

3.1. Number of peptidases

By using the criteria of Barrett & Rawlings (2007) [1], it was possible to distinguish 2457 different peptidase species. A holotype was established for each of these. In addition, there are a further 250 different peptidase activities in the literature for which either no or too little sequence information exists to be able to map any of them to a UniProt accession. The breakdown into different catalytic types is shown in Table 1. More serine peptidases (554 or 20%) were found than for any other catalytic type.

The holotypes and their sequence homologues, 502,782 sequences in total, were found to be distributed amongst 254 families. This total includes 3669 sequences that are classified as asparagine lyases rather than peptidases, because proteolysis is dependent upon cyclization of an asparagine residue to a succinimide which does not involve hydrolysis [28].

3.2. Numbers of substrates

The total number of cleavages found for all peptidases was 66,615. This number includes substrates cleaved in the same position by different peptidases. Most of these cleavages (59,276 or 89%) were in peptides or proteins that could be mapped to UniProt identifiers: 5821 (or 9%) were in synthetic substrates. The majority of these cleavages (36,229 or 54%) are non-physiological, but 20,264 (30%) are thought to be physiologically relevant, and a further 1349 (2%) are pathological. The totals include self-cleaving reactions, which in the case of asparagine lyases is all the cleavages identified for this catalytic type.

The breakdown of substrate cleavages per catalytic type is shown in Table 2. Most cleavages (30,711 or 46%) are for serine peptidases, a number boosted by the large number of non-physiological cleavages (21,301 or 32%).

Table 3 shows the number of peptidases in each catalytic type for which at least one substrate cleavage is known. The total number is 1700 or 63% of the total number of different peptidases. Comparing the numbers with those in Table 1 shows that for each catalytic type, a substrate cleavage is known for most peptidases. Substrates may be known for any of the other 1007 peptidases, but cleavages were either not found or were not accessible in the literature, or the site of cleavage is unknown. On average a peptidase has 39 substrate cleavages in the MEROPS collection. Table 4 shows the peptidases with 10 or more known cleavages: the large number of non-physiological cleavages for the serine peptidase trypsin, 12,303 of which are derived from the proteomics paper of Schilling & Overall (2008) [27], where it is used for preparation of samples prior to mass spectroscopy analysis, explains the preponderance of cleavages for serine peptidases. To be able to study peptidase specificity and make predictions about where in a protein cleavage might occur, at least 40 cleavages in substrates are required (Robert Pike, personal communication). The number of peptidases with 40 or more cleavages is 163 (or 6% of the total number of peptidases). It is immediately apparent that except for a small number of peptidases, insufficient numbers of cleavages in substrates are known to be able to draw firm conclusions about the specificity of most peptidases. The distribution of cleavages per peptidase is shown in Fig. 1.

There are 36 peptidases with 100 or more known cleavages in physiological substrates. The peptidases with most physiological cleavages are general processing peptidases, such as the animal signal peptidase complex (a complex containing two peptidases with 1879 cleavages) and methionyl aminopeptidase 1 (964 cleavages). Almost all of these cleavages are derived from annotations in the SwissProt section of the UniProt database, and include theoretical cleavages confirmed by proteomics experiments submitted to the PRIDE database. Other peptidases with many physiological cleavages include peptidases involved in protein turnover such as cathepsins E (1553 cleavages) and D (871) and granzyme M (891). Almost all of which are derived from proteomics studies [29,30]. There are 27 peptidases with 200 or more known cleavages in substrates thought not to be physiological. As mentioned above, the use of trypsin to produce peptides for mass-spectroscopy in proteomics experiments means that more cleavages (14,201) are known for this peptidase rather than any other. Large numbers of non-physiological cleavages are also known for peptidases regularly used in manual protein sequencing such as pepsin A (350 cleavages), peptidyl-lys metallopeptidase (2105), chymotrypsin A (1012), glutamyl peptidase 1 (1273) and lysyl endopeptidase (802). Other peptidases with many non-physiological substrates are again the subjects of proteomic studies: matrix metallopeptidase-2 (2726 cleavages) [27]; human granzyme B (1136) [30,31]; glutamyl peptidase I (1003) [27]; cathepsins L(965), S (694) and B (486) [32]; meprin beta (891) and alpha subunits [33]; and the RC1339 protein from Rickettsia conorii (799) [34]. There are 57 peptidases with 20 or more cleavages in synthetic substrates and neurolysin (139 cleavages) and thimet oligopeptidase (116) have the most. There are 43 peptidases with cleavages in a hundred or more proteins, including cleavages in peptides that can be mapped to protein sequences in the UniProt database. Once again, because of its use in proteomics, more proteins are known that are susceptible to trypsin (3188) than any other peptidase. The methods by which cleavages were identified include mass spectroscopy (40,729 cleavages), N-terminal sequencing (8284), from knowing the consensus cleavage site (1609), amino acid analysis (205) and site-directed mutagenesis (178).

3.3. Substrate specificity

Table 4 shows the specificity of peptidases based on the occurrence of amino acids in the binding pockets P4–P′. The preferences for each peptidase are defined in words so that the table represents a classification by specificity and homology. This also allows classification by amino acid type (acidic, basic, aliphatic, aromatic, etc) which would not be possible if logos were shown. The MEROPS identifier for each peptidase is a link to the peptidase summary page on the MEROPS website where the specificity logo is displayed. Caution should be exercised where these preferences are derived from fewer than 40 substrates. The number of peptidases

| Aspartic | Glutamic | Metallo | Cysteine | Serine | Threonine | Mixed | Asparagine lyases | Unknown | Total |
|----------|----------|---------|----------|--------|----------|-------|-------------------|---------|-------|
| 170 | 7 | 633 | 615 | 942 | 46 | 5 | 23 | 16 | 2457 |
| 118 | 0 | 327 | 297 | 644 | 29 | 0 | 1 | 3 | 1419 |
| 8 | 0 | 86 | 19 | 87 | 0 | 0 | 0 | 49 | 250 |
| 4 | 0 | 103 | 52 | 145 | 26 | 0 | 0 | 0 | 330 |
| 24 | 0 | 5 | 21 | 17 | 3 | 0 | 0 | 0 | 70 |
| 324 | 7 | 1154 | 1004 | 1835 | 105 | 5 | 24 | 68 | 4526 |
For which the cleavage data are considered reliable (highlighted in green) is 319. The number for which the data are less reliable (yellow highlighting) is 123, and the number for which the data should be treated with caution (red highlighting) is 68. The vast majority of peptidases with a low reliability score are those showing apparent specificity in more than three binding pockets. It should be noted, however, that the reliability score will be affected majorly of peptidases with a low reliability score are those showing apparent specificity in more than three binding pockets. It should be noted, however, that the reliability score will be affected by the number of binding pockets in which specificity is shown: this is apparent for the deubiquitinating hydrolases where the substrates positions P4–P1 are occupied in physiological substrates by the highly conserved C-terminus of ubiquitin (Leu–Arg–Gly–Gly). Because proteins substrates are unlikely to be homologous, a peptidase with many protein substrates will have a large reliability score. For a peptidase with cleavages in predominantly synthetic substrates, the reliability score will be low, because synthetic substrates tend to be similar with only the blocking and reporter groups differing. The peptidases with the highest reliability scores (96%) are asclepian A (C01.008); the snake venom enzyme jerdohagin (M12.216) and bpr peptidase from Dichelobacter nodosus (S08.022). In all three examples, protein cleavages are from the insulin B-chain [35–37]. The peptidase for which the reliability score is lowest (12%) is polyglycine endopeptidase (P01.001) also acts as an endopeptidase, processing its own precursor [40]. For many aminopeptidases, only cleavages in synthetic substrates are known and these frequently lack residues in positions P2’–P4’. These apparent absences in binding pockets S2’–S4’ are not shown in Table 4 because they would be misleading.

The majority of peptidases for which specificity can be inferred (107 endopeptidases and 43 exopeptidases) show a preference in a single substrate binding pocket. Most of these peptidases show a preference in the S1 pocket, but there is a preference in each binding pocket S4–S4’ for at least one endopeptidase. There is an overwhelming bias for basic residues in P1, which probably reflects the large number of characterized peptidases in family S1. Exopeptidases, perhaps unsurprisingly, show preference only in pockets P1 and P1’. The TET aminopeptidase is the only exopeptidase to show a preference beyond the residues either side of the cleavage site.

Several peptidases show a preference in more than one binding pocket. There are 79 peptidases with a preference in two binding pockets, 50 peptidases with a preference in three pockets and 29 peptidases with a preference in four pockets. The table also includes examples of peptidases with preferences in more binding pockets, but as with the case for thimet oligopeptidase (see below) these apparent preferences are more likely to reflect the design of the substrates tested rather than true specificities.

Despite the large number of cleavages in substrates, the specificity of some peptidases cannot be explained in terms of preferences in binding pockets. Table 4 shows the 53 peptidases with more than forty substrate cleavages for which no preference is shown. This list includes some well-known and well-characterized peptidases such as cathepsin G, for which even a large proteomics study was unable to show any preference except that lysine is not acceptable in P4 [27].

Table 4 also shows that peptidases with similar specificities can be unrelated to one another (members of different families or even catalytic types), for example gingipain K and bacterial lysyl endopeptidase both have a preference for Lys in P1, yet the former is a cysteine peptidase and the latter a serine peptidase. There are also examples where specificity differs markedly from peptidases with homologous sequences, with examples from family S1 being the most well known in which chymotrypsin A has a preference for aromatic residues in P1, trypsin 1 for basic residues, elastase-2 for aliphatic residues and granzyme B for aspartic acid. It is more unusual for peptidases within a family to utilise different substrate binding pockets, but an example is granzyme B which also shows a preference for an aliphatic residue in P4. A family may include exo- as well as endopeptidases, for example family S9 includes prolyl oligopeptidase and dipeptidyl-peptidase IV, although both have a preference for Pro in P1. These facts emphasize that peptidases cannot be classified to the protein species by specificity or sequence similarity alone, but that both specificity and homology should be taken into consideration.

| Physiological | Aspartic | Glutamic | Metallo | Cysteine | Serine | Threonine | Mixed | Asparagine lyases | Unknown | Total |
|---------------|----------|----------|---------|----------|--------|-----------|-------|-------------------|---------|-------|
| 2780          | 7        | 4113     | 7311    | 5877     | 33     | 2         | 60    | 81                | 20,264  |
| Pathological  | 266      | 345      | 700     | 34       | 0      | 0         | 0     | 0                 | 4       | 1349  |
| Non-physiological | 2893    | 70       | 8762    | 3112     | 21,301 | 84        | 1     | 3                 | 3       | 36,229|
| Synthetic     | 364      | 32       | 1569    | 1179     | 2547   | 42        | 37    | 0                 | 51      | 5821  |
| Theoretical   | 176      | 0        | 545     | 106      | 638    | 0         | 0     | 304               | 0       | 1765  |
| Unclassified  | 77       | 0        | 554     | 194      | 314    | 26        | 1     | 0                 | 21      | 1187  |
| Total         | 6556     | 109      | 15,888  | 12,602   | 30,711 | 185       | 41    | 363               | 160     | 66,615|

**Table 3**
Counts of peptidases with known substrate cleavages by catalytic type.

| Aspartic | Glutamic | Metallo | Cysteine | Serine | Threonine | Mixed | Asparagine lyases | Unknown | Total |
|----------|----------|---------|----------|--------|-----------|-------|-------------------|---------|-------|
| 114      | 4        | 477     | 401      | 520    | 18        | 2     | 17                | 33      | 1586  |
| Sequence not known | 5 | 0 | 41 | 6 | 34 | 0 | 0 | 0 | 28 | 114 |
| Total    | 119      | 4       | 518     | 407     | 554     | 18    | 2     | 17                | 61      | 1700  |
There are many peptidases for which the number of physiological substrates is limited. These include proteins that process themselves to remove a propeptide or to expose a N-terminal nucleophile and then cease to function as peptidases. Examples include the amidophosphoribosyltransferase precursor (C44.001) [41] and the penicillin V acylase precursor (C59.001) [42], for both of which the only known cleavage is to expose a new N-terminal nucleophilic cysteine. Human peptidases such as renin (A01.007) have few natural substrates but the number of cleavages is boosted by synthetic substrates and peptides.

3.4. Displays on the MEROPS website

A variety of displays have been implemented on the MEROPS website to aid understanding of peptidase specificity. The displays include the following.

| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | MEROPS ID | Peptidase name                          | Total cleavages | Reliability |
|----|----|----|----|-----|-----|-----|-----|---------|----------------------------------------|----------------|-------------|
|    |    |    |    |     |     |     |     |         | matrix metallopeptidase-20              | 43             | 67          |
| M10.019 |    |    |    |     |     |     |     |         | cuticle-degrading peptidase of parasitic fungus | 14             | 94          |
| S08.056 |    |    |    |     |     |     |     |         |                                      |                |             |
|    |    |    |    |     |     |     |     |         | matrix metallopeptidase-10 (Homo sapiens-type) | 19             | 83          |
|    |    |    |    |     |     |     |     |         | physarolisin                           | 11             | 94          |
|    |    |    |    |     |     |     |     |         | thermitase                              | 13             | 91          |
|    |    |    |    |     |     |     |     |         |                                      |                |             |
|    |    |    |    |     |     |     |     |         | oligopeptidase A                        | 10             | 68          |
| M03.004 |    |    |    |     |     |     |     |         | chymopapain                             | 12             | 86          |
| C01.002 |    |    |    |     |     |     |     |         | cathepsin V                              | 14             | 83          |
| C01.009 |    |    |    |     |     |     |     |         | cathepsin S                              | 756            | 91          |
| C01.034 |    |    |    |     |     |     |     |         | cathepsin K                              | 182            | 96          |
| C01.036 |    |    |    |     |     |     |     |         | streptopain                             | 27             | 94          |
| C01.001 |    |    |    |     |     |     |     |         | crustapain                              | 10             | 85          |
| C01.030 |    |    |    |     |     |     |     |         |                                      |                |             |
| C01.073 |    |    |    |     |     |     |     |         | peptidase 1 (mite)                       | 29             | 92          |
|    |    |    |    |     |     |     |     |         |                                      |                |             |
|    |    |    |    |     |     |     |     |         | LmaCP1 carboxypeptidase (Leishmania major) | 13             | 42          |
| M32.005 |    |    |    |     |     |     |     |         | carboxypeptidase Y                      | 17             | 50          |
| S10.001 |    |    |    |     |     |     |     |         | serine carboxypeptidase C                | 18             | 48          |
| S10.004 |    |    |    |     |     |     |     |         | serine carboxypeptidase D                | 16             | 51          |
| S10.005 |    |    |    |     |     |     |     |         | carboxypeptidase OcpA                    | 14             | 46          |
| S10.008 |    |    |    |     |     |     |     |         | serine carboxypeptidase III (plant)      | 16             | 48          |

In each peptidase summary, where ten or more substrate cleavages are known, a logo (generated by the Weblogo software, http://weblogo.berkeley.edu/) and a specificity matrix are shown. Fig. 2 shows the logo and specificity matrix for thimet oligopeptidase. In the logo, amino acid residues are shown as single-letter abbreviations, and the greater the height of the letters the greater the preference for the amino acid in that binding pocket [43]. The binding pockets S4−S4′ are numbered along the X axis as 1−8. In the specificity matrix, amino acids are shown in three-letter notation, and the number of times each occurs within the range P4−P4′ for a substrate is shown. The occurrence of each amino acid in each binding pocket is calculated as a percentage of total cleavages, and a different shade of green is used for each tenth percentile. The brighter green the background highlighting, the greater the preference for that amino acid in the corresponding substrate binding pocket. If an amino acid has not been observed in any binding...
| MW | MEROPS ID | Peptidase name | Total cleavages | Reliability |
|----|-----------|----------------|----------------|-------------|
| P4 | S09.006  | dipeptidyl aminopeptidase A | 14 | 67 |
| P3 | M18.002  | aspartyl aminopeptidase | 10 | 55 |
| P2 | M20.010  | DapE peptidase | 15 | 50 |
| P1 | T02.002  | isoaaspartyl dipeptidase (threonine type) | 17 | 38 |
| P1' | M01.003  | aminopeptidase A | 14 | 88 |
| P2' | A01.022  | plasmspepsin-1 | 13 | 83 |
| P3' | M28.004  | aminopeptidase AC | 13 | 52 |
| P4' | S09.070  | acylaminoacyl peptidase (Aeropyrum-type) | 14 | 70 |
| N  | M23.001  | beta-lytic metallopeptidase | 15 | 75 |
| L  | S01.139  | granzyme M | 1363 | 84 |
| K  | A01.085  | candidapepsin SAP10 | 16 | 55 |
| K  | C25.002  | gingipain K | 47 | 82 |
| K  | S01.280  | lysyl endopeptidase (bacteria) | 809 | 82 |
| N  | C13.004  | legumain, animal-type | 81 | 84 |
| P  | M02.006  | angiotensin-converting enzyme-2 | 14 | 82 |
| P  | M9E.004  | membrane Pro-Xaa carboxypeptidase | 20 | 59 |
| P  | S09.001  | prolyl oligopeptidase | 59 | 75 |
| P  | S09.003  | dipeptidyl-peptidase IV (eukaryote) | 28 | 82 |
| P  | S09.013  | dipeptidyl-peptidase 4 (bacteria-type 2) | 17 | 80 |
| P  | S09.018  | dipeptidyl-peptidase 8 | 10 | 83 |
| P  | S09.019  | dipeptidyl-peptidase 9 | 15 | 57 |
| P  | S28.001  | lysosomal Pro-Xaa carboxypeptidase | 11 | 52 |
| P  | S28.002  | dipeptidyl-peptidase II | 30 | 51 |
| R  | C01.064  | RD21 peptidase | 20 | 81 |
| R  | C01.115  | fascipain B | 10 | 62 |
| R  | C14.035  | metacaspase Yca1 (Saccharomyces cerevisiae-type) | 11 | 53 |
| R  | C14.043  | metacaspase Ld (Leishmania-type) | 27 | 76 |
| R  | C25.003  | gingipain R2 | 16 | 83 |
| R  | C69.002  | arginine aminopeptidase (Streptococcus-type) | 11 | 67 |
| R  | M01.014  | aminopeptidase B | 15 | 80 |
| R  | M28.021  | AaaA aminopeptidase | 13 | 50 |
| R  | S01.017  | kallikrein-related peptidase 5 | 75 | 67 |
| R  | S01.021  | DESC1 peptidase | 108 | 81 |
| R  | S01.033  | factor VII-activating peptidase | 11 | 87 |
| R  | S01.047  | human airway trypsin-like peptidase | 95 | 83 |
| R  | S01.071  | kallikrein 1-related peptidase b9 (Mus musculus) | 14 | 70 |
| R  | S01.172  | kallikrein 1-related peptidase c2 (Rattus norvegicus) | 21 | 85 |
| R  | S01.224  | hepsin | 110 | 80 |
| R  | S01.302  | matriptase | 112 | 61 |
| R  | S01.308  | matriptase-2 | 161 | 80 |
| R  | S33.010  | SCO7095-type peptidase | 11 | 62 |
| R  | XP01.001  | tricorn peptidase complex | 14 | 70 |
| W  | M9A.008  | tryptophanyl aminopeptidase (Trichosporon cutaneum) | 15 | 73 |
| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|----|----|----|----|----|----|----|----|----------|---------------------------------------------|----------------|-------------|
|    |    |    |    |    |    |    |    |          | BcepAP aminopeptidase (*Burkholderia cepacia*) | 10             | 63          |
|    |    |    |    |    |    |    |    |          | glutamyl peptidase I                        | 1405            | 85          |
|    |    |    |    |    |    |    |    |          | proteasome catalytic subunit 1              | 15              | 74          |
|    |    |    |    |    |    |    |    |          | elastase-2                                   | 483             | 91          |
|    |    |    |    |    |    |    |    |          | DegP peptidase                              | 15              | 83          |
|    |    |    |    |    |    |    |    |          | HtrA2 peptidase                             | 233             | 85          |
|    |    |    |    |    |    |    |    |          | Htra2 peptidase (*Mycobacterium*-type)       | 39              | 90          |
|    |    |    |    |    |    |    |    |          | chymotrypsin A (cattle-type)                 | 1057            | 91          |
|    |    |    |    |    |    |    |    |          | chymase (*Homo sapiens*-type)                | 106             | 89          |
|    |    |    |    |    |    |    |    |          | mast cell peptidase 4 (mouse numbering)     | 10              | 94          |
|    |    |    |    |    |    |    |    |          | chymotrypsin B                               | 26              | 79          |
|    |    |    |    |    |    |    |    |          | kallikrein-related peptidase 7               | 17              | 88          |
|    |    |    |    |    |    |    |    |          | serine carboxypeptidase A                    | 10              | 47          |
|    |    |    |    |    |    |    |    |          | candidapepsin SAP9                          | 17              | 95          |
|    |    |    |    |    |    |    |    |          | ficain                                       | 15              | 87          |
|    |    |    |    |    |    |    |    |          | cathepsin B-like cysteine peptidase (*Raphanus sativus*) | 24              | 95          |
|    |    |    |    |    |    |    |    |          | interpain A                                 | 22              | 91          |
|    |    |    |    |    |    |    |    |          | metalloprotease B (*Bothrops moojeni*)       | 12              | 85          |
|    |    |    |    |    |    |    |    |          | bontoxylisin                                 | 10              | 87          |
|    |    |    |    |    |    |    |    |          | kallikrein-related peptidase 14              | 74              | 85          |
|    |    |    |    |    |    |    |    |          | matriptase-3                                 | 112             | 85          |
|    |    |    |    |    |    |    |    |          | tryptase alpha                               | 23              | 86          |
|    |    |    |    |    |    |    |    |          | granzyme K                                   | 20              | 84          |
|    |    |    |    |    |    |    |    |          | trypsin 1                                    | 13770           | 88          |
|    |    |    |    |    |    |    |    |          | kallikrein 1                                 | 31              | 87          |
|    |    |    |    |    |    |    |    |          | mesotrypsin                                  | 22              | 86          |
|    |    |    |    |    |    |    |    |          | plasma kallikrein                            | 40              | 84          |
|    |    |    |    |    |    |    |    |          | plasmin                                      | 126             | 89          |
|    |    |    |    |    |    |    |    |          | kallikrein-related peptidase 6               | 75              | 85          |
|    |    |    |    |    |    |    |    |          | lumbrokinase                                 | 25              | 88          |
|    |    |    |    |    |    |    |    |          | kallikrein-related peptidase 8               | 25              | 68          |
|    |    |    |    |    |    |    |    |          | kallikrein-related peptidase 4               | 123             | 85          |
|    |    |    |    |    |    |    |    |          | HAT-like 3 peptidase                          | 11              | 81          |
|    |    |    |    |    |    |    |    |          | polyserase-I A unit 1                         | 11              | 45          |
|    |    |    |    |    |    |    |    |          | kallikrein 1-related peptidase c9 (*Rattus sp.*) | 14              | 85          |
|    |    |    |    |    |    |    |    |          | oligopeptidase B                             | 42              | 54          |
|    |    |    |    |    |    |    |    |          | proteasome catalytic subunit 2               | 14              | 87          |
|    |    |    |    |    |    |    |    |          | matrix metalloprotease-11                    | 21              | 79          |
|    |    |    |    |    |    |    |    |          | atrolysin C                                  | 60              | 91          |
|    |    |    |    |    |    |    |    |          | ruberylsin                                   | 16              | 83          |
|    |    |    |    |    |    |    |    |          | murolysin                                    | 14              | 66          |
|    |    |    |    |    |    |    |    |          | ADAM8 peptidase                              | 23              | 87          |
|    |    |    |    |    |    |    |    |          | carnosine dipeptidase I                      | 17              | 57          |
|    |    |    |    |    |    |    |    |          | carboxypeptidase Q.                           | 33              | 85          |
|    |    |    |    |    |    |    |    |          | glycy1 aminopeptidase (*Actinomucor elegans*) | 10              | 54          |
|    |    |    |    |    |    |    |    |          | peptidyl-D-amino acid hydrolase (cephalopod) | 13              | 65          |
| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | **MEROPS ID** | Peptidase name | Total cleavages | Reliability |
|----|----|----|----|-----|-----|-----|-----|----------------|----------------|---------------|-------------|
| **Specificity at one binding pocket: P1'**
| D  | M72.001 | peptidyl-Asp metallopeptidase | 21 | 84 |
| F  | M04.010 | vimelysin | 12 | 83 |
| K  | M14.023 | CPG70 carboxypeptidase (Porphyromonas gingivalis) | 27 | 72 |
| K  | M35.004 | peptidyl-Lys metallopeptidase | 2105 | 82 |
| P  | M24.003 | Xaa-Pro dipeptidase (bacteria-type) | 13 | 50 |
| P  | M24.009 | aminopeptidase P1 | 18 | 59 |
| P  | M24.033 | aminopeptidase P (Streptomyces-type) | 12 | 57 |
| R  | M14.005 | carboxypeptidase E | 20 | 67 |
| +  | M12.004 | meprin beta subunit | 93 | 91 |
| +  | M96.007 | carboxypeptidase G3 | 12 | 91 |
| A  | M04.001 | thermolysin | 280 | 93 |
| A  | M04.007 | coccolysin | 23 | 93 |
| A  | M04.009 | aureolysin | 24 | 93 |
| A  | M05.001 | mycolysin | 12 | 93 |
| A  | M06.001 | immune inhibitor A peptidase (Bacillus sp.) | 22 | 89 |
| A  | M10.001 | matrix metalloproteinase-1 | 75 | 90 |
| A  | M10.003 | matrix metalloproteinase-2 | 3417 | 90 |
| A  | M10.006 | matrix metalloproteinase-7 | 196 | 81 |
| A  | M12.138 | jararhagin | 40 | 91 |
| A  | M12.140 | bothropasin | 62 | 90 |
| A  | M14.017 | carboxypeptidase A4 | 68 | 90 |
| A  | M14.018 | carboxypeptidase A6 | 42 | 88 |
| A  | M96.034 | metalloproteinase Shpl (Staphylococcus hyicus) | 21 | 94 |
| A  | M13.008 | nepriylisin-2 | 16 | 85 |
| A  | M12.306 | BHRb haemorrhagin (Bitis arietans) | 11 | 95 |
| A  | M13.003 | endothelin-converting enzyme 2 | 17 | 90 |
| A  | M13.009 | Zmp1 peptidase (Mycobacterium-type) | 10 | 93 |
| Q  | M9A.007 | Xaa-Trp aminopeptidase | 13 | 68 |
| -  | M14.004 | carboxypeptidase N | 15 | 82 |
| -  | M14.006 | carboxypeptidase M | 11 | 81 |
| -  | M14.009 | carboxypeptidase B2 | 19 | 84 |
| C  | M12.001 | astacin | 209 | 91 |
| E  | M12.225 | ADAMTS5 peptidase | 38 | 92 |
| **Specificity at one binding pocket: P2'**
| A  | M12.311 | BaP1 peptidase (Bothrops asper) | 45 | 90 |
| A  | M30.001 | hycolysin | 15 | 95 |
| A  | M73.001 | camelysin | 11 | 97 |
| O  | M03.007 | oligopeptidase F | 24 | 88 |
| **Specificity at one binding pocket: P3'**
| O  | A01.042 | syncephapepsin | 18 | 97 |
| O  | M42.005 | TET aminopeptidase | 12 | 94 |
| E  | C01.060 | cathepsin B | 632 | 90 |
| E  | M04.003 | vibriolysin | 17 | 91 |
| E  | M12.210 | ADAM10 peptidase | 40 | 90 |
| **Specificity at one binding pocket: P4'**

| P4 | P3 | P2 | P1 | P1’ | P2’ | P3’ | P4’ | MEROPS ID | Peptidase name                                      | Total cleavages | Reliability |
|----|----|----|----|-----|-----|-----|-----|----------|---------------------------------------------------|----------------|-------------|
|    |    |    |    |     |     |     |     | Omega    | CP790 peptidase *(Lactobacillus helveticus CP790)* | 32             |             |
|    |    |    |    | CR | D  | PR | W   | S9G.114 | moojeni peptidase A                              | 10             |             |
|    | R  | R  |    |    |     |     |     | S08.077 | PCSK7 peptidase                                    | 116            |             |
| A  |    |    |    |     |     |     |     | S01.010  | granzyme B *(Homo sapiens-type)*                   | 1873           |             |
| A  |    |    |    |     |     |     |     | S01.136  | granzyme B, rodent-type                           | 487            |             |
|    |    |    |    |     |     |     |     | A01.084  | yolk cathepsin *(Boophilus sp.)*                   | 17             |             |
|    |    |    |    |     |     |     |     | M12.345  | Specificity at two binding pockets: P4 and P1     | 10             |             |
|    |    |    |    |     |     |     |     | M16.004  | Specificity at two binding pockets: P4 and P1     | 19             |             |
|    |    |    |    |     |     |     |     | M16.004  | Specificity at two binding pockets: P4 and P1     | 19             |             |
|    |    |    |    |     |     |     |     | M16.004  | Specificity at two binding pockets: P3 and P1     | 14             |             |
| A  |    |    |    |     |     |     |     | S01.262  | streptolysin                                    | 201            |             |
| A  |    |    |    |     |     |     |     | C14.005  | caspase-6                                       | 201            |             |
| A  |    |    |    |     |     |     |     | A01.004  | memapsin-2                                      | 24             |             |
| A  |    |    |    |     |     |     |     | S01.034  | transmembrane peptidase, serine 4                 | 10             |             |
| A  |    |    |    |     |     |     |     | S80.001  | prohead peptidase pp175 *(Pseudomonas aeruginosa phage phiK2)* | 25             |             |
| A  |    |    |    |     |     |     |     | S01.271  | glutamyl peptidase BL                            | 72             |             |
| A  |    |    |    |     |     |     |     | S26.008  | thylakoidal processing peptidase                  | 52             |             |
| A  |    |    |    |     |     |     |     | S26.010  | signalase (animal) 21 kDa component               | 363            |             |
| A  |    |    |    |     |     |     |     | XS26.001 | signal peptide complex (animal)                  | 1878           |             |
|    |    |    |    |     |     |     |     | M02.002  | Specificity at two binding pockets: P3 and P1’    | 11             |             |
|    |    |    |    |     |     |     |     | M02.002  | peptidyl-dipeptidase Acer                        | 11             |             |
|    |    |    |    |     |     |     |     | M10.060  | Specificity at two binding pockets: P2 and P1     | 12             |             |
|    |    |    |    |     |     |     |     | S01.405  | bothropain                                       | 10             |             |
|    |    |    |    |     |     |     |     | S01.214  | kallikrein 1 *(Rattus sp.)*                       | 12             |             |
|    |    |    |    |     |     |     |     | S01.438  | fire ant chymotrypsin                            | 17             |             |
|    |    |    |    |     |     |     |     | C01.050  | histolysin                                       | 10             |             |
|    |    |    |    |     |     |     |     | M10.060  | epralysin                                        | 10             |             |
|    |    |    |    |     |     |     |     | S09.073  | Xaa-Pro dipeptidylpeptidase                      | 19             |             |
|    |    |    |    |     |     |     |     | S09.069  | omega peptidase *(Drosophila melanogaster)*       | 18             |             |
|    |    |    |    |     |     |     |     | S01.443  | glutamyl peptidase B I                           | 14             |             |
|    |    |    |    |     |     |     |     | S46.002  | dipeptidyl-peptidase 11 *(Porphyromonas gingivalis-type)* | 10             |             |
|    |    |    |    |     |     |     |     | C01.072  | rhodesain                                        | 11             |             |
|    |    |    |    |     |     |     |     | S08.070  | kexin                                            | 191            |             |
|    |    |    |    |     |     |     |     | M16.011  | falcylisin                                       | 10             |             |
|    |    |    |    |     |     |     |     | S08.072  | PCSK1 peptidase                                   | 92             |             |
|    |    |    |    |     |     |     |     | S09.012  | trypsin                                         | 14             |             |
|    |    |    |    |     |     |     |     | S01.015  | trypsin beta                                     | 22             |             |
|    |    |    |    |     |     |     |     | S01.218  | protein C (activated)                            | 15             |             |
|    |    |    |    |     |     |     |     | S01.231  | Specificity at two binding pockets: P2 and P1’    | 10             |             |
|    |    |    |    |     |     |     |     | S01.231  | Xaa-Pro dipeptidyl-peptidase                      | 10             |             |
|    |    |    |    |     |     |     |     | S01.216  | coagulation factor Xa                            | 59             |             |
|    |    |    |    |     |     |     |     | S01.231  | u-plasminogen activator                          | 17             |             |
|    |    |    |    |     |     |     |     | C01.032  | cathepsin L                                      | 1089           |             |
### Specificity at two binding pockets: P2 and P2′

| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| M16.003   | mitochondrial processing peptidase beta-subunit                              | 81             | 81          |
| A02.007   | feline immunodeficiency virus retropepsin                                    | 28             | 84          |

### Specificity at two binding pockets: P1 and P1′

| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| C14.001   | caspase-1                                                                     | 181            | 83          |
| C14.018   | caspase-14                                                                    | 17             | 79          |
| M12.136   | leucolysin                                                                    | 16             | 71          |
| M14.001   | carboxypeptidase A1                                                           | 16             | 61          |
| M28.003   | aminopeptidase S (Streptomyces-type)                                          | 10             | 50          |
| M24.001   | methionyl aminopeptidase 1 (Escherichia-type)                                 | 978            | 71          |
| M24.002   | methionyl aminopeptidase 2                                                    | 135            | 71          |
| M24.028   | mitochondrial methionyl aminopeptidase                                         | 131            | 67          |
| M98.004   | Met-Xaa dipeptidase                                                           | 14             | 72          |
| S09.017   | prolyl tripeptidyl peptidase                                                  | 13             | 74          |
| A26.001   | omptin                                                                        | 54             | 81          |
| C25.001   | gingipain R                                                                   | 26             | 84          |
| M24.026   | lcpSS peptidase                                                               | 39             | 75          |
| M28.010   | glutamate carboxypeptidase II                                                 | 10             | 50          |
| M13.004   | oligopeptidase O1                                                             | 10             | 50          |
| S01.306   | kallikrein-related peptidase 13                                               | 21             | 82          |
| C01.027   | comosain                                                                       | 13             | 47          |
| S01.135   | granzyme A                                                                    | 441            | 84          |
| S01.217   | thrombin                                                                      | 186            | 82          |
| M9G.018   | neutral endopeptidase (Micrococcus caseolyticus)                              | 12             | 78          |
| M04.005   | pseudolysin                                                                   | 70             | 82          |
| M12.145   | atrolysin E                                                                   | 19             | 84          |
| M12.309   | hemorrhagic toxin I (Gloydius halys blohmoffii)                               | 12             | 90          |
| M12.214   | ADAM19 peptidase                                                              | 13             | 84          |
| P01.001   | DmpA aminopeptidase                                                           | 35             | 82          |

### Specificity at two binding pockets: P1 and P2

| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| S33.008   | prolyl aminopeptidase 2                                                       | 10             | 52          |

### Specificity at two binding pockets: P1 and P3′

| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| S01.281   | arginyl peptidase                                                             | 11             | 82          |
| M10.004   | matrix metallopeptidase-9                                                      | 369            | 88          |
| M10.009   | matrix metallopeptidase-12                                                     | 218            | 86          |

### Specificity at two binding pockets: P1 and P4′

| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| A01.030   | yapsin-1                                                                      | 45             | 84          |
| S01.161   | kallikrein-related peptidase 2                                                | 66             | 82          |

### Specificity at two binding pockets: P1′ and P2′

| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| C01.024   | endopeptidase-B (Hordeum-type)                                                | 70             | 72          |
| M12.178   | brevlysin L6                                                                  | 16             | 81          |
| M24.005   | aminopeptidase P2                                                             | 13             | 44          |
| S53.004   | kumamolisin                                                                   | 11             | 77          |
| M12.032   | LAST peptidase (Limulus-type)                                                 | 78             | 86          |
| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | **MEROPS ID** | **Peptidase name** | **Total cleavages** | **Reliability** |
|----|----|----|----|-----|-----|-----|-----|-------------|------------------|------------------|-----------------|
| λ  | Σ  | M10.002 | matrix metalloproteinase-8 | 114 | 88  |
| λ  | Σ  | M12.217 | ADAM17 peptidase | 60 | 90  |
|   |   | S01.137 | granzyme C | 19 | 88  |
| L  | G  | C30.005 | SARS coronavirus picornain 3C-like peptidase | 26 | 71  |
| R  | -  | S08.074 | PCSK4 peptidase | 103 | 71  |
| R  | -  | S08.075 | PCSK6 peptidase | 105 | 69  |
| R  | -  | S08.076 | PCSK5 peptidase | 129 | 72  |
| Σ  | Σ  | S08.012 | P15UB1 peptidase | 23 | 77  |
| G  | W  | S08.071 | furin | 208 | 72  |
| W  | -  | S08.073 | PCSK2 peptidase | 202 | 77  |
| L  | G  | C03.020 | enterovirus picornain 2A | 12 | 75  |
| Σ  | Σ  | C47.004 | staphopain C | 10 | 91  |
| D  | G  | C14.002 | CED-3 peptidase | 14 | 74  |
| +  | Σ  | C14.003 | caspase-3 | 651 | 77  |
| +  | Σ  | C14.004 | caspase-7 | 179 | 77  |
| Ω  | Σ  | S01.148 | HtrA peptidase (Chlamydia trachomatis-type) | 14 | 73  |
| Σ  | Σ  | S33.005 | tricorn interacting factor F1 | 16 | 73  |
| E  | Ω  | C14.019 | caspase DRONC (Drosophila melanogaster)-type peptidase | 11 | 57  |
| L  | Σ  | S53.007 | aorsin | 10 | 45  |
| Σ  | Σ  | C01.057 | vinckepain-2 | 28 | 64  |
| Σ  | Σ  | S01.150 | mast cell peptidase 5 (mouse numbering) | 44 | 84  |
| Ω  | Σ  | S33.005 | tricorn interacting factor F1 | 16 | 40  |
| E  | Σ  | C14.007 | caspase-4 | 10 | 78  |
| V  | Σ  | S21.002 | cytochromeovirus assemblin | 13 | 58  |
| Ω  | Σ  | M03.006 | mitochondrial intermediate peptidase | 16 | 87  |
| +  | Σ  | C14.009 | caspase-8 | 68 | 76  |
| +  | Σ  | C14.010 | caspase-9 | 14 | 85  |
| +  | Σ  | C14.011 | caspase-10 | 10 | 80  |
| @  | Σ  | S01.526 | SplD peptidase (Staphylococcus aureus) | 22 | 59  |
| @  | Σ  | S01.503 | SplA peptidase (Staphylococcus aureus) | 23 | 57  |
| W  | P  | S26.001 | signal peptidase I | 342 | 75  |
| Σ  | Σ  | Ω  | M10.015 | membrane-type matrix metalloproteinase-2 | 14 | 92  |
| Σ  | Ω  | A  | A01.019 | polyporopepsin | 11 | 90  |
| G  | G  | C23.002 | staphylolysin | 12 | 45  |
| L  | Q  | C30.001 | coronavirus picornain 3C-like peptidase-1 | 44 | 66  |
| L  | Q  | C30.003 | human coronavirus 229E main peptidase | 11 | 69  |
| P4 | P3 | P2 | P1 | P1’ | P2’ | P3’ | P4’ | **MEROPS ID** | Peptidase name | Total cleavages | Reliability |
|----|----|----|----|-----|-----|-----|-----|---------------|----------------|---------------|-------------|
| -  | R  | Σ  |    |     |     |     |     | S07.001      | flavivirin      | 21           | 74          |
| Σ  | Σ  | Σ  | Σ  |     |     |     |     | M9G.029      | MAP1 peptidase (Mycoplasma xanthus, Pimeobacter sp.) | 17           | 83          |

**Specificity at three binding pockets: P2, P1 and P2’**

A  | K  | A  | A  | A01.011 | penicillopepsin | 17          | 66          |
F  | Σ  | A  | A  | A01.023 | plasminspsin-2  | 15          | 83          |

**Specificity at three binding pockets: P1, P1’ and P2’**

A  | Σ  | A  | S53.011 | scytilidolisin  | 20          | 67          |

**Specificity at three binding pockets: P1, P1’ and P2’**

G  | L  | A  | M04.017 | griselysin      | 36          | 66          |
G  | L  | Σ  | M04.014 | bacillioysis     | 41          | 74          |
R  | R  | Σ  | A26.003 | plasminogen activator Pla | 16          | 35          |
Σ  | Σ  | A  | M12.016 | verteoloid-like 1 protein | 17          | 67          |
Σ  | Σ  | Σ  | C13.005 | glycosylphosphatidylinositol:protein transaminase | 19          | 82          |
Σ  | Σ  | Σ  | M12.209 | ADAM9 peptidase | 18          | 84          |

**Specificity at three binding pockets: P1, P1’ and P4’**

E  | Σ  | Σ  | S32.001 | equine arteritis virus serine peptidase | 15          | 68          |

**Specificity at three binding pockets: P1’, P2’ and P3’**

A  | Σ  | Σ  | M08.001 | leishmanolysin | 20          | 65          |

**Specificity at four binding pockets: P4, P3, P2 and P1**

A  | A  | P  | F  | S08.037 | subtilisin DY | 16          | 40          |
L  | R  | G  | G  | C12.003 | ubiquitinyl hydrolase-L3 | 26          | 30          |
L  | R  | G  | G  | C19.001 | ubiquitin-specific peptidase 5 | 10          | 41          |
L  | R  | G  | G  | C65.001 | otubain-1 | 10          | 40          |
Q  | T  | G  | G  | C48.002 | SENP1 peptidase | 11          | 40          |
Ω  | Σ  | Σ  | C54.003 | autophagin-1 | 15          | 37          |

**Specificity at four binding pockets: P4, P3, P1 and P1’**

L  | P  | T  | G  | C60.001 | sortase A (Staphylococcus-type) | 16          | 45          |

**Specificity at four binding pockets: P4, P3, P2’ and P4’**

Ω  | Ω  | Σ  | Ω  | A02.005 | bovine immunodeficiency virus retropepsin | 11          | 91          |

**Specificity at four binding pockets: P4, P2, P1 and P1’**

Ω  | Σ  | Σ  | C14.006 | caspase-2 | 16          | 73          |
Ω  | Γ  | Σ  | C05.001 | ademain | 26          | 76          |

**Specificity at four binding pockets: P4, P2, P1 and P3’**

G  | Σ  | Σ  | Ω  | M03.001 | thimet oligopeptidase | 124         | 72          |

**Specificity at four binding pockets: P4, P2, P1’ and P2’**

Ω  | Ω  | Σ  | C03.021 | rhinovirus picornain 2A | 12          | 81          |

**Specificity at four binding pockets: P4, P1, P1’ and P4’**

Ω  | Ω  | Ω  | C03.001 | polioivirus-type picornain 3C | 14          | 70          |
| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | **MEROPS ID** | **Peptidase name** | **Total cleavages** | **Reliability** |
|----|----|----|----|-----|-----|-----|-----|----------------|-------------------|-------------------|------------------|
|    |    |    |    |     |     |     |     | M10.029       | matrix metallopeptidase-26 | 17               | 87               |
|    |    |    |    |     |     |     |     | S08.109       | KPC2-type peptidase         | 115              | 71               |
|    |    |    |    |     |     |     |     | M10.063       | PrtA peptidase (Photorhabdus-type) | 29               | 77               |
|    |    |    |    |     |     |     |     | M10.066       | karilysin                     | 38               | 60               |
|    |    |    |    |     |     |     |     | M10.013       | matrix metallopeptidase-13    | 147              | 83               |
|    |    |    |    |     |     |     |     | M12.153       | fibrinolytic peptidase (Philodryas aferisi) | 15               | 72               |
|    |    |    |    |     |     |     |     | N10.002       | intein-containing replicative DNA helicase precursor | 158              | 75               |
|    |    |    |    |     |     |     |     | M03.010       | P2-peptidase A                        | 13               | 71               |
|    |    |    |    |     |     |     |     | S01.213       | coagulation factor Xla                     | 12               | 75               |
|    |    |    |    |     |     |     |     | S01.156       | enteropeptidase                             | 21               | 74               |
|    |    |    |    |     |     |     |     | C01.168       | endopeptidase-A (Hordeum-type)             | 20               | 71               |
|    |    |    |    |     |     |     |     | C01.010       | vagnain                                     | 80               | 76               |
|    |    |    |    |     |     |     |     | M09.003       | bacterial collagenase H                     | 18               | 76               |
|    |    |    |    |     |     |     |     | M13.091       | PHEX peptidase                             | 14               | 71               |
|    |    |    |    |     |     |     |     | S01.458       | chymase 1 (Mus musculus-type)              | 44               | 70               |
|    |    |    |    |     |     |     |     | M12.005       | procollagen C-peptidase                    | 41               | 70               |
|    |    |    |    |     |     |     |     | S01.020       | kallikrein-related peptidase 12            | 20               | 80               |
|    |    |    |    |     |     |     |     | C14.044       | metacaspase TbMCA2                         | 11               | 67               |
|    |    |    |    |     |     |     |     | C14.026       | paracaspase                                 | 17               | 64               |
|    |    |    |    |     |     |     |     | S08.063       | site-1 peptidase                           | 30               | 70               |
|    |    |    |    |     |     |     |     | S26.016       | Sps8 signal peptidase                      | 49               | 71               |
|    |    |    |    |     |     |     |     | S01.005       | mast cell peptidase 4 (Rattus)             | 40               | 76               |
|    |    |    |    |     |     |     |     | C39.001       | bacteriocin-processing peptidase            | 13               | 65               |
|    |    |    |    |     |     |     |     | M79.002       | RCE1 peptidase (Homo sapiens-type)          | 22               | 53               |
|    |    |    |    |     |     |     |     | A02.063       | walleye dermal sarcoma virus retropepsin   | 25               | 73               |
|    |    |    |    |     |     |     |     | S14.001       | peptidase Clp (type 1)                     | 15               | 75               |
|    |    |    |    |     |     |     |     | A01.006       | chymosin                                    | 20               | 81               |
| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | MEROPS ID | Peptidase name | Total cleavages | Reliability |
|----|----|----|----|----|----|----|----|---------|--------------|--------------|-------------|
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | C37.001 | calicivirin | 11 | 66 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | N10.005 | intein-containing DNA polymerase II large subunit DP2 precursor Mernname-AA281 | 26 | 68 |
| E | L | S | G | R | S | C | F | C50.001 | separase (yeast-type) | 12 | 69 |
| P | P | F | R | L | C | F | R | C53.002 | sedolisin-B | 47 | 40 |
| & | P | F | R | L | R | & | & | C53.005 | kumamolisin-B | 61 | 50 |
| P | & | & | E | F | R | L | P | C53.001 | sedolisin | 22 | 46 |
| & | & | & | F | R | L | P | A01.002 | pepsin B | 25 | 60 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A22.001 | presenilin 1 | 16 | 88 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | M03.002 | neurolysin | 128 | 60 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | M09.001 | spumeppsin | 11 | 79 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A02.012 | retropepsin (human T-cell leukemia virus) | 12 | 65 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | S50.004 | blotched snakehead birnavirus Vp4 peptidase | 38 | 62 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | N10.004 | intein-containing translation initiation factor IF-2 precursor | 51 | 70 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | S14.003 | peptidase Clp (type 3) | 10 | 75 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | S01.267 | streptogratisin E | 23 | 46 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | G01.001 | scytalidoglutamin peptidase | 37 | 45 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A24.001 | type 4 prepielin peptidase 1 | 16 | 41 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | M12.222 | ADAMTS1 peptidase | 13 | 76 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | S08.112 | ARA12 peptidase | 53 | 58 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | M03.011 | tropolysin | 34 | 53 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | S01.282 | SpiB peptidase (Staphylococcus aureus) | 19 | 62 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | N09.001 | intein-containing V-type proton ATPase catalytic subunit A | 45 | 48 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A01.060 | candidapepsin SAP2 | 22 | 55 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A01.014 | candidapepsin SAP1 | 26 | 66 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | C24.003 | saposin 3C-like peptidase | 10 | 60 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A02.013 | bovine leukemia virus retropepsin | 17 | 58 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A02.010 | mouse mammary tumor virus retropepsin | 13 | 46 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A02.002 | HIV-2 retropepsin | 30 | 75 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A25.001 | gprpepsin | 32 | 51 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A01.060 | candidapepsin SAP2 | 22 | 55 |
| (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | (  ) | A01.014 | candidapepsin SAP1 | 26 | 66 |
| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | MEROPS ID | Peptidase name | Total cleavages | Reliability |
|----|----|----|----|----|----|----|----|----------|---------------|---------------|-------------|
| K  | K  | G  | A  | S  | T  | G5 | IT | A24.017  | PibD peptidase | 18            | 50          |
| K  | L  | R  | S  | S  | K  | O  | C01.038 | cathepsin P   | 10            | 50          |
| K  | L  | R  | F  | S  | K  | O  | C01.074 | CPB peptidase  | 17            | 50          |
| K  | L  | R  | F  | S  | K  | O  | C01.100 | cruzipain 2    | 26            | 54          |
| Y  | -  | -  | E  | E  | A  | A  | A01.031 | yapsin-2       | 16            | 46          |

**Specificity at eight binding pockets**

| A  | L  | F  | Q  | G  | P  | P  | V  | C03.011 | coxsackievirus-type picornain 3C | 10 | 55 |
| I  | A  | Ω  | Ω  | Ω  | F  | G  | S  | S54.004 | aarA-type peptidase | 58 | 26 |
| I  | L  | G  | A  | K  | G  | G  | S  | C39.007 | lacticin 481 processing peptidase (Lactococcus lactis) | 16 | 60 |
| K  | R  | GR | NQ | C  | D  | F  | G  | D N10.006 | intein-containing DNA polymerase III large subunit DP2 precursor Mereana-A282 | 18 | 50 |
| L  | A  | A  | R  | R  | G  | A  | G  | A26.002 | OmpP peptidase | 11 | 57 |
| L  | V  | E  | L  | A  | Y  | L  | V  | A01.061 | candidasepsin SAP3 | 21 | 50 |
| L  | V  | E  | L  | L  | Y  | L  | V  | A01.064 | candidasepsin SAP6 | 22 | 16 |
| P  | A  | L  | N  | S  | P  | Ω  | Ω  | N08.001 | poliovirus capsid VP0-type self-cleaving protein | 38 | 50 |
| P  | F  | H  | L  | L  | V  | H  | Y  | S  | A01.007 | renin | 10 | 57 |
| S  | F  | L  | Y  | Q  | V  | S  | T  | A28.004 | skin SASPase | 15 | 65 |
| S  | Q  | N  | Y  | P  | I  | V  | Q  | A02.004 | equine infectious anaemia virus retroepsin | 13 | 50 |
| S  | Q  | N  | Y  | P  | I  | V  | Q  | A02.009 | Mason-Pfizer leukemia virus retroepsin | 18 | 52 |
| S  | Q  | N  | Y  | P  | I  | V  | Q  | A02.008 | murine leukemia virus-type retroepsin | 20 | 63 |
| S  | Q  | N  | Y  | P  | I  | V  | Q  | A02.015 | Rous sarcoma virus retroepsin | 18 | 52 |
| V  | R  | F  | Q  | S  | G  | T  | R  | C04.003 | tobacco vein motting virus-type Nia peptidase | 13 | 51 |
| V  | R  | F  | Q  | S  | G  | T  | R  | C04.004 | tobacco etch virus Nia peptidase | 14 | 53 |
| O  | L  | O  | λ  | Ω  | O  | L  | O  | M34.001 | anthrax lethal factor | 12 | 78 |
| A  | L  | N  | P  | P  | V  | P  | P  | M34.002 | Zmp1 peptidase (Clostridium difficile-type) | 13 | 20 |
| O  | P  | I  | Ω  | Ω  | Ω  | Ω  | Ω  | S9G.118 | Nysius proteinase | 11 | 67 |

**No positive specificity**

|    | A01.001 | pepsin A | 417 | 33 |
|    | A01.003 | gastricsin | 50 | 33 |
|    | A01.009 | cathepsin D | 897 | 94 |
|    | A01.010 | cathepsin E | 1596 | 90 |
|    | A01.012 | rhizopuspepsin | 249 | 34 |
|    | A01.016 | aspergillopepsin I | 157 | 34 |
|    | A01.017 | endothiapepsin | 11 | 37 |
|    | A01.018 | saccharopepsin | 14 | 34 |
|    | A01.020 | phytpepsin | 62 | 32 |
|    | A01.040 | nephtesinh | 19 | 35 |
|    | A01.041 | memapsin-1 | 33 | 17 |
|    | A01.053 | necpsin-1 | 133 | 33 |
|    | A01.068 | necpsin-2 | 45 | 31 |
|    | A01.071 | pepsin A5 (Homo sapiens) | 10 | 31 |
|    | A02.001 | HIV-1 retroepsin | 1045 | 91 |
|    | A32.002 | RC1339 g.p. (Rickettsia conori) | 785 | 92 |
|    | C01.005 | stem bromelain | 26 | 34 |
|    | C01.007 | actinidain | 19 | 31 |
|    | C01.008 | asclepains A | 10 | 96 |
| MEROPS ID | Peptidase name                                      | Total cleavages | Reliability |
|-----------|----------------------------------------------------|----------------|-------------|
| C01.013   | cathepsin X                                        | 46             | 90          |
| C01.026   | ananain                                            | 17             | 58          |
| C01.033   | cathepsin L1 (Fasciola sp.)                        | 196            | 93          |
| C01.040   | cathepsin H                                        | 48             | 90          |
| C01.046   | falcipain-2                                        | 152            | 92          |
| C01.062   | cathepsin B-like peptidase (platyhelminth)        | 54             | 93          |
| C01.063   | falcipain-3                                        | 126            | 93          |
| C01.070   | dipeptidyl-peptidase I                            | 34             | 89          |
| C01.084   | bleomycin hydrolase (animal)                      | 19             | 59          |
| C01.086   | aminopeptidase C                                  | 20             | 51          |
| C01.092   | cathepsin L1 (arthropod-type)                     | 12             | 93          |
| C01.097   | phytolacain                                        | 13             | 95          |
| C01.101   | cathepsin B-like peptidase, nematode              | 47             | 93          |
| C01.129   | cathepsin L3 (Fasciola sp.)                       | 72             | 90          |
| C01.130   | CSCP3 peptidase (Clonorchis-type)                 | 56             | 93          |
| C01.142   | cathepsin L2 (Fasciola sp.)                       | 54             | 93          |
| C02.001   | calpain-1                                          | 106            | 91          |
| C02.002   | calpain-2                                          | 170            | 93          |
| C02.004   | calpain-3                                          | 19             | 93          |
| C06.001   | cathepsin N                                        | 14             | 92          |
| G01.002   | aspergillo glutamic peptidase                     | 68             | 93          |
| M01.001   | aminopeptidase N                                  | 42             | 58          |
| M01.002   | lysyl aminopeptidase (bacteria)                   | 22             | 75          |
| M01.004   | leukotriene A4 hydrolase                          | 14             | 52          |
| M01.005   | alanyl aminopeptidase (bacterial-type)            | 17             | 56          |
| M01.009   | aminopeptidase N (actinomycete-type)              | 11             | 49          |
| M01.010   | cytosol alanyl aminopeptidase                     | 56             | 68          |
| M01.011   | cystinyl aminopeptidase                           | 11             | 87          |
| M01.013   | aminopeptidase N (insect)                         | 14             | 49          |
| M01.016   | aminopeptidase Eγ                                 | 18             | 73          |
| M01.022   | arginyl aminopeptidase-like 1                     | 14             | 50          |
| M01.024   | ERAP2 aminopeptidase                              | 23             | 89          |
| M01.026   | aminopeptidase Q                                  | 17             | 70          |
| M10.005   | matrix metallopeptidase-3                         | 180            | 90          |
| M10.014   | membrane-type matrix metallopeptidase-1           | 132            | 93          |
| M10.016   | membrane-type matrix metallopeptidase-3           | 20             | 92          |
| M10.017   | membrane-type matrix metallopeptidase-4           | 27             | 92          |
| M10.024   | membrane-type matrix metallopeptidase-6           | 50             | 93          |
| M10.056   | aeruginolysin                                     | 13             | 93          |
| M10.057   | mirabilysin                                       | 32             | 95          |
| M12.002   | meprin alpha subunit                              | 776            | 93          |
| M12.033   | LAST_MAM peptidase (Limulus-type)                 | 432            | 97          |
| M12.137   | BHRa hemorrhagin (Bitis arietans)                 | 14             | 89          |
| M12.164   | lebetase                                          | 14             | 89          |
| M12.216   | jerdohagin (Trimeresurus jerdonii)                 | 15             | 96          |
| M12.221   | ADAMTS4 peptidase                                 | 59             | 97          |
| M12.310   | metallopeptidase MTP-1                            | 18             | 97          |
| MEROPS ID | Peptidase name                                                                 | Total cleavages | Reliability |
|-----------|-------------------------------------------------------------------------------|----------------|-------------|
| M12.327   | leucurolysin-a                                                                 | 25             | 90          |
| M13.001   | nepriylsin                                                                    | 123            | 92          |
| M13.002   | endothelin-converting enzyme 1                                                | 34             | 94          |
| M13.011   | MEP peptidase (nematode)                                                      | 57             | 91          |
| M16.001   | p Kirstein                                                                    | 23             | 94          |
| M16.002   | inssulosin                                                                    | 101            | 98          |
| M16.009   | eukirsilin                                                                    | 60             | 97          |
| M16.016   | YMXG peptidase                                                                | 40             | 90          |
| M16.017   | plinosulosin                                                                  | 17             | 94          |
| M17.015   | aminopeptidase pepZ (Staphylococcus sp.)                                     | 16             | 47          |
| M20.013   | Xaa-methyl-His dipeptidase                                                    | 13             | 74          |
| M28.001   | aminopeptidase Y                                                              | 28             | 41          |
| M28.002   | aminopeptidase Ap1                                                            | 66             | 83          |
| M28.006   | Mernase-AA063 peptidase                                                        | 19             | 57          |
| M28.022   | leucine aminopeptidase-1 (Aspergillus fumigatus)                               | 10             | 50          |
| M35.001   | penicilloysin                                                                 | 20             | 33          |
| M35.002   | deurolysin                                                                    | 22             | 31          |
| M41.001   | FtsH peptidase                                                                | 24             | 27          |
| M41.013   | FtsH peptidase (Thermus-type)                                                 | 32             | 91          |
| M42.003   | PhTET1 aminopeptidase                                                         | 10             | 40          |
| M42.004   | PhTET2 aminopeptidase                                                         | 23             | 83          |
| M49.001   | dipeptidyl-peptidase III                                                      | 24             | 41          |
| M61.001   | glycyl aminopeptidase                                                         | 26             | 84          |
| M9A.011   | neuron-specific aminopeptidase                                                 | 10             | 48          |
| M9G.026   | dactylysin                                                                    | 23             | 25          |
| S01.122   | brachyurin-C                                                                  | 36             | 48          |
| S01.133   | cathepsin G                                                                   | 464            | 89          |
| S01.134   | myeloblastin                                                                  | 52             | 50          |
| S01.142   | duodenase                                                                     | 19             | 27          |
| S01.157   | chymotrypsin C                                                                | 36             | 93          |
| S01.162   | kallikrein-related peptidase 3                                                | 83             | 88          |
| S01.422   | fibrinolytic enzyme A (Annelida-type)                                         | 19             | 90          |
| S08.001   | subtilisin Carlsberg                                                          | 43             | 77          |
| S08.003   | subtilisin lentus                                                             | 16             | 54          |
| S08.008   | Mernase-AA053 peptidase                                                        | 16             | 35          |
| S08.009   | subtilisin Ak1                                                                 | 12             | 63          |
| S08.010   | M-peptidase                                                                  | 14             | 45          |
| S08.014   | ALE1 peptidase                                                                | 16             | 85          |
| S08.019   | lactosepin I                                                                  | 102            | 92          |
| S08.022   | bpr peptidase (Dichelobacter nodosus)                                         | 12             | 36          |
| S08.024   | dentilisin                                                                    | 17             | 95          |
| S08.028   | high alkaline protease (Alkaliphilus transvaalenisi)                          | 28             | 74          |
| S08.034   | subtilisin BPN1                                                               | 17             | 83          |
| S08.051   | aqua lysin 1                                                                  | 11             | 73          |
| S08.052   | cerevisin                                                                     | 20             | 87          |
Obviously, with some amino acids that occur less frequently (Cys and Trp, for example), the number of cleavages should be taken into consideration before any conclusions are drawn about negative preferences by the peptidase.

The example in Fig. 2 is based on 125 cleavages. However, 80 of these cleavages are derived from one reference [44] in which the substrates based on Abz-GFSPFRQ-EDDnp were synthesized to help distinguish thimet oligopeptidase from its close relative neurolysin.

It is immediately obvious that the preferences shown in binding pockets S4–S3 (Gly, Phe, Ser, Pro, Phe, Arg, Gln) reflect the residues that were kept constant during the residue-scanning experiments. This example serves as a cautionary note to show that apparent preferences may reflect the methodology employed rather than represent the true specificity of the peptidase.

For each peptidase where substrate cleavages are known there is a page dedicated to listing the substrates and the known cleavage positions. Fig. 3 shows sections of the table of substrates for thimet oligopeptidase. For each cleavage, the substrate name is given, the UniProt accession if the substrate is from a naturally occurring protein, the residue range of the substrate with respect to the Uniprot entry, the cleavage site, the nature of the substrate ("cleavage type"), the evidence for the cleavage position (if known), a reference and a cross-reference to the CutDB database [14].

| P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | MEROPS ID | Peptidase name | Total cleavages | Reliability |
|----|----|----|----|----|----|----|----|----------|----------------|---------------|-------------|
|    |    |    |    |    |    |    |    | S08 054  | peptidase K    | 47            | 54          |
|    |    |    |    |    |    |    |    | S08 057  | thermomycolin  | 28            | 55          |
|    |    |    |    |    |    |    |    | S08 092  | cucumisin      | 31            | 56          |
|    |    |    |    |    |    |    |    | S08 116  | lactopepin-3   | 158           | 53          |
|    |    |    |    |    |    |    |    | S08 135  | tenglonsin     | 19            | 54          |
|    |    |    |    |    |    |    |    | S08 142  | collagenolytic endopeptidase (Geobacillus sp. MO-1) | 11 | 55 |
|    |    |    |    |    |    |    |    | S08 148  | keratinase (Doratomyces microsporus) | 19 | 55 |
|    |    |    |    |    |    |    |    | S08 151  | SISBT3 g.p. (Solanum lycopersicum) | 19 | 51 |
|    |    |    |    |    |    |    |    | S09 004  | acylaminocarbonyl-peptidase | 14 | 59 |
|    |    |    |    |    |    |    |    | S09 072  | puromycin hydrolase | 17 | 56 |
|    |    |    |    |    |    |    |    | S10 006  | Membrane-A083 peptidase | 72 | 54 |
|    |    |    |    |    |    |    |    | S10 010  | serine carboxypeptidase Z (Abadia zachae) and similar | 16 | 74 |
|    |    |    |    |    |    |    |    | S10 014  | carboxypeptidase O | 89 | 58 |
|    |    |    |    |    |    |    |    | S12 001  | D-Ala-D-Ala carboxypeptidase B | 11 | 62 |
|    |    |    |    |    |    |    |    | S12 002  | aminopeptidase DmpB | 17 | 51 |
|    |    |    |    |    |    |    |    | S12 003  | alkaline D-peptidase | 12 | 56 |
|    |    |    |    |    |    |    |    | S16 001  | Lon-A peptidase | 21 | 57 |
|    |    |    |    |    |    |    |    | S41 007  | ctpB peptidase | 18 | 53 |
|    |    |    |    |    |    |    |    | S46 001  | dipeptidyl-peptidase 7 (Porphyromonas gingivalis-type) | 21 | 69 |
|    |    |    |    |    |    |    |    | S53 003  | tripeptidyl-peptidase I | 45 | 51 |
|    |    |    |    |    |    |    |    | S53 010  | grifolisin | 13 | 55 |
|    |    |    |    |    |    |    |    | T01 002  | archaean proteasome, beta component | 44 | 53 |
|    |    |    |    |    |    |    |    | T01 005  | bacterial proteasome, beta component | 31 | 53 |
|    |    |    |    |    |    |    |    | T01 012  | proteasome catalytic subunit 3 | 24 | 60 |
|    |    |    |    |    |    |    |    | XM02-001 | angiotensin-converting enzyme compound peptidase | 24 | 62 |
|    |    |    |    |    |    |    |    | XM12-001 | memprin A complex peptidase | 71 | 54 |
|    |    |    |    |    |    |    |    | XT01-001 | 20 S constitutive proteasome peptidase complex (eukaryote) | 468 | 54 |
|    |    |    |    |    |    |    |    | XT01-002 | 26 S proteasome peptidase complex (eukaryote) | 67 | 51 |
| C  | C  | C  | C  | C  | CW | C  | C  | XT01-003 | 20 S immunoproteasome peptidase complex (eukaryote) | 228 | 54 |
physiologically relevant is whether the cleavage site is conserved in close homologues. A cleavage site that is not conserved is unlikely to be physiologically relevant, although it may be pathologically relevant in the species in which it occurs. Conservation should be in terms of which residues are accepted in the binding pockets of the peptidase and not just sequence conservation in the substrate orthologues. This is only possible if a large number of substrate cleavages are known for the peptidase in question. Fig. 4 shows part of the alignment for orthologues of the Ebola virus envelope glycoprotein. The known cleavage is from the Zaire strain (UniProt P87671) and cleavage is by ADAM17 (MEROPS ID M12.217) at residue 637; the cleavage results in shedding of the ectodomain which circulates in the blood of the patient and may interfere with antibodies thus helping to prolong infection [45]. The alignment is dynamically generated from the sequences clustered in the UniRef50 database entry that contains this sequence. The UniRef50 database entry contains sequences that share 50% or more sequence identity [46]. Alignments are generated using MUSCLE [47]. The sequence containing the known cleavage site is highlighted in green. Residues in the range P4−P4′ are highlighted in pink if they are identical to that from the Zaire strain; substituted residues are highlighted in orange if the amino acid from another ADAM17 substrate is known to occupy the same binding pocket; and substituted residues are shown as white on black if the amino acid is not known to occupy the same binding pocket from any ADAM17 substrate (“unacceptable replacements”).

In this example, in one sequence the residues around the cleavage site have not been determined and are replaced by Xs, and in sequences from a number of strains of the virus the P1′ residue is replaced by His and the P2′ residue by Asp or Asn, which have not been observed in any of the 59 other substrates for ADAM17. The possible conclusions are: 1) ADAM17 is not the physiological peptidase that performs this cleavage, 2) some Ebola virus envelope glycoproteins are processed by a different peptidase, 3) the specificity of ADAM17 has not been fully explored and His is permissible in P1′ and Asp or Asn in P2′, or 4) it doesn’t matter if the cleavage is inefficient in some cases provided some processing occurs.

3.5. A service to test conservation of cleavage sites

The production of this cleavage collection has allowed the installation of a service on the MEROPS website to help researchers assess whether or not a particular cleavage is physiologically relevant. Following on from the display described above, the philosophy behind this service is that a cleavage site in a protein substrate is most likely to be physiologically relevant if it is conserved. The user can submit a list of cleavage sites in a file the structure of...
### Substrates for peptidase M03.001: thimet oligopeptidase

| Substrate | Uniprot ID | Residue range | Cleavage Site | Cleavage Type | Evidence | P4 | P3 | P2 | P1 | P1' | P2' | P4' | Reference | CutDB |
|-----------|------------|---------------|---------------|---------------|-----------|-----|----|----|----|-----|-----|-----|-----------|------|
| Abz-Ala-Lys-Pro-Arg-| P01149 | 24-33 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |
| Abz-Ala-Lys-Pro-Arg- | P02036 | 90-107 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |
| Abz-Ala-Lys-Pro-Arg- | P02036 | 90-107 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |
| Abz-Ala-Lys-Pro-Arg- | P03990 | 151-163 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |
| Abz-Ala-Lys-Pro-Arg- | P04307 | 141-157 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |
| Abz-Ala-Lys-Pro-Arg- | P04307 | 141-157 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |
| Abz-Ala-Lys-Pro-Arg- | P04307 | 141-157 | Abz-Ala-Lys-Pro-Arg- | | | | | | | | | | |
| Pro-Gln-EDDnp | | | Pro-Gln-EDDnp | | | | | | | | | | |

#### Fig. 3. Example of a substrate page

Part of the substrates page for thimet oligopeptidase is shown. For each substrate the following are shown: name; a cross-reference and link to the entry in the UniProt database where appropriate; the residue range of the substrate as used in the experiment with reference to the numbering in the UniProt entry; a description of the cleavage where the scissile bond is represented by the symbol ‘+’; whether the cleavage is physiological, non-physiological, pathological or in a synthetic substrate; the evidence by which the cleavage site was determined; the residues occupying residues P4 to P4' in the substrate; the source reference; and a cross-reference and link to the CutDB database [14]. By default substrates are listed alphabetically, but the order can be changed by clicking the column heading. It is possible to filter the results for physiological, non-physiological, pathological or cleavages in synthetic substrates by clicking on the appropriate letter in the table legend.
which is shown in Table 5. This should be a text file created with software such as Notepad and not a document file created with a word processing package such as Word. The data required per line are: MEROPS identifier of the peptidase performing the cleavage, the UniProt accession for the substrate protein, and the position of the residue occupying the S1 binding pocket (the residue after which cleavage occurs) taken from the full coding sequence in the UniProt entry. For each line, an alignment is generated from the sequences in the UniRef50 entry containing the UniProt accession of the substrate protein, and the number of unacceptable replacements is counted. A file is generated and is E-mailed to the user. An example of a results file is shown in Table 6. For each line in the file the following are included: the MEROPS identifier of the peptidase; the total number of cleavages known for that peptidase; the UniProt accession of the substrate protein; the number of homologues aligned from the UniRef50 entry; the cleavage position; the number of unacceptable replacements at P4, P3, P2, P1, P1', P2', P3' and P4'; and a URL to display the substrate alignment at the MEROPS website (to conserve space, the URL column is not shown in Table 6). The number of known substrate cleavages for the peptidase is returned because this will help the user assess the reliability of the analysis: the more cleavages the more reliable the analysis. Similarly, the number of sequences in the alignment will help the user assess the results: if the alignment contains many sequences and the cleavage site is well conserved, then the likelihood that the cleavage site is physiologically relevant is greater. If on the other hand the cleavage site is not conserved, and there are many sequences in the alignment, then it is much more likely that the cleavage is not physiologically relevant.

If the number of unacceptable replacements in positions P4—P4' is zero, then the cleavage site is extremely well conserved and is likely to represent a physiological cleavage by the peptidase concerned. If the number of unacceptable replacements is small, then the cleavage may still be physiological. The most common explanation is that the UniRef50 entry on which the alignment is based contains two or more very closely related paralogues, one of which is a physiological substrate, but the other is not. A second possibility is that the sequences in the UniRef50 entry represent species variants of only one protein, but one or more is derived from genomic sequencing and is either a fragment (perhaps because the initiating methionine has been misidentified or an anomalous frame-shift has been introduced which truncates the C-terminus), or is missing the exon which codes for all or part of the cleavage site. The user is advised to consult the alignment via the URL provided to check. If the number of unacceptable replacements is high, then the cleavage site is not conserved and the cleavage is unlikely to be physiologically significant. However, the user should check the number of cleavages for the peptidase: if this is less than 40, then it is possible that there is not enough variation amongst the known cleavage site sequences to account for the substitutions that have occurred. It is possible that many homologues in the alignment have the same replacement which has not been observed in any substrate for the peptidase, simply because not enough substrates have been found.

The user should also be aware that if there is a high number of unacceptable replacements in one position, this requires further investigation, because substitution to a rare amino acid may have taken place. For a peptidase with no preference in a binding pocket, for example S3', almost any amino acid can occupy the P3' position.

In the alignment of substrate homologues, the count of unacceptable replacements is therefore likely to be small. However, if no substrate is known with cysteine in P3', for example, yet many homologues of a known substrate have a replacement cysteine in P3', then the unacceptable replacement count will be high. In such an example, it is simply not known if cysteine is acceptable in P3', so it is impossible to say if the cleavage is physiological or not. The user should check the alignment for such a replacement.

I would like to reiterate that a cleavage site that is not conserved could still be important pathologically. The user is advised to check the preferences for the peptidase on the peptidase summary page in MEROPS to see if, for example, a large number of mismatches in P4' is relevant.

This service is known as “Analyze Substrates” and is accessed from the main menu on the website. It is particularly useful for proteomics studies in which hundreds of potential physiological substrates are found. The limitations on the service are that the maximum number of cleavages per file uploaded is 5000 and the maximum file size is 10 Mbytes.
4. Conclusions

Cleavages in substrates (proteins, peptides and synthetic substrates) by proteolytic enzymes have been collected from the literature. In total, 66,615 cleavages have been annotated for 1700 different peptidases (69% of the 2457 different peptidases so far identified). The number of cleavages per peptidase varies greatly, from zero to 13,770 for trypsin 1. Peptidases with the most known cleavages are derived from proteomics experiments.

Users should be aware of the biases that may exist in the substrate cleavage data for a peptidase. If the data are predominantly derived from amino acid scanning experiments in which all but one amino acid is changed in a variety of synthetic peptides, then the amino acid variety may be limited with changes only to blocking and reporter groups. To address this problem, a bioinformatics approach that has been adopted here is to associate whether the cleavage is conserved in orthologues of the substrate protein. For each cleavage site, the UniProt accession of the substrate protein is specified.

Consider whether the cleavage is conserved in orthologues of the substrate protein. For each cleavage site, the UniProt accession of the substrate protein is specified.

By analysing the residues that occupy the P4–P4' substrate binding sites, it has been possible to categorize the specificity of some peptidases. A logo and specificity matrix is shown on the MEROPS website for each of the 556 peptidases (22.7%) with ten or more substrate cleavages. Of these, 107 endopeptidases and 43 exopeptidases show a preference in one substrate binding pocket, 79 peptidases show a preference in two binding pockets, 50 show a preference at three binding pockets, 29 show a preference at four binding pockets and 53 show no positive preference in any substrate binding site. However, it has been proposed that at least 40 cleavages in disparate proteins are required for specificity analysis to be meaningful, and only 163 peptidases (6.6%) fulfil this criterion.

Proteomics experiments in which cleavage sites are identified by mass spectroscopy provides the bulk of the substrate cleavage data. However, a common problem is that in order to extract the proteins from the sample, the natural, physical boundaries that separate peptide from substrate may have been removed, and a large number of substrate cleavages are generated. While this is not a problem if no claims are made about the physiological relevance of the data and the suite of peptides are generated just to examine peptidase specificity, there are issues when it is claimed that the experiment will reveal physiological substrates. The biggest concern is that not all substrates will be physiological, but distinguishing these from artefactual and bystander substrates is difficult. A bioinformatics approach that has been adopted here is to consider whether the cleavage is conserved in orthologues of the substrate protein. For each cleavage site, the UniProt accession of...
the substrate protein is determined, as well as the UniRef50 cluster of sequences to which it belongs. All protein sequences from the UniRef50 cluster are then aligned, and the residues P4–P4 for each cleavage site are examined for conservation. The number of replacements that are not found to occupy the same binding pocket in any of the substrates for the peptidase is counted. If this number is high, then the likelihood that the substrate is physiologically relevant is low. A server has been set up at the MEROPS website by which results from the experiments to study peptidase specificity can be uploaded, and the conservation around each cleavage site examined. Results are returned by E-mail.

Acknowledgements

The author would like to thank Dr Alan Barrett for maintaining the literature collection for the MEROPS database; Dr Usha Mahadevan (Molecular Connections) and summer students Matthew Jenner, Jack Feltham and Danielle Weaver for all their hard work in collecting substrate cleavages; Jung Kong for implementing the WebLogo software; Wei Yang for help with collecting data from the PRIDE database; and Matthew Waller and Paul Bevan from the Wellcome Trust Sanger Institute web team and former colleagues Emmett O’Brien, Dominik Tolle, Fraser Morton and Joyce Kok for help with the design, development and maintenance of the MEROPS website. I would also like to thank John Marioni, statistician at the OPS website. I would also like to thank John Marioni, statistician at the MEROPS website. I would also like to thank John Marioni, statistician at the MEROPS website. I would also like to thank John Marioni, statistician at the MEROPS website. I would also like to thank John Marioni, statistician at the MEROPS website.
[36] R.Q. Chen, Y. Jin, J.B. Wu, X.D. Zhou, D.S. Li, Q.M. Lu, W.Y. Wang, Y.L. Xiong, A novel high molecular weight metalloproteinase cleaves fragment F1 of activated human prothrombin, Toxicon 44 (2004) 281–287.

[37] A.A. Kortt, J.B. Caldwell, G.G. Lilley, R. Edwards, J. Vaughan, D.J. Stewart, Characterization of a basic serine proteinase (pl approximately 9.5) secreted by virulent strains of Dichelobacter nodosus and identification of a distinct, but closely related, proteinase secreted by benign strains, Biochem. J. 299 (1994) 521–525.

[38] T.A. Naumann, D.T. Wicklow, N.P. Price, Polyglycine hydrolases secreted by Pleosporineae fungi that target the linker region of plant class IV chitinases, Biochem. J. 460 (2014) 187–198, http://dx.doi.org/10.1042/BJ20140268.

[39] D. Chandu, A. Kumar, D. Nandi, PepN, the major Suc-LLVY-AMC-hydrolyzing enzyme in Escherichia coli, displays functional similarity with downstream processing enzymes in Archaea and Eukarya. Implications in cytosolic protein degradation, J. Biol. Chem. 278 (2003) 5548–5556, http://dx.doi.org/10.1074/jbc.M207926200.

[40] L. Fanuel, C. Goffin, A. Cheggour, B. Devreese, G. Van Driessche, B. Joris, J. Van Beeumen, J.M. Frère, The DmpA aminopeptidase from Ochrobactrum anthropi LMG7991 is the prototype of a new terminal nucleophile hydrolase family, Biochem. J. 341 (1999) 147–155, http://dx.doi.org/10.1042/0264-6021:3410147.

[41] S.J. Vollmer, R.L. Switzer, M.A. Hermodson, S.G. Bower, H. Zalkin, The glutamine-utilizing site of Bacillus subtilis glutamine phosphoribosylpyrophosphate amidotransferase, J. Biol. Chem. 258 (1983) 10582–10585.