Determination of the Substrate Specificity of Tripeptidyl-peptidase I Using Combinatorial Peptide Libraries and Development of Improved Fluorogenic Substrates*

Yu Tian, Istvan Sohar, John W. Taylor, and Peter Lobel

From the Center for Advanced Biotechnology and Medicine and the Department of Chemistry and Chemical Biology, Rutgers, the State University of New Jersey and the Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Classical late-infantile neuronal ceroid lipofuscinosis is a fatal neurodegenerative disease caused by mutations in CLN2, the gene encoding the lysosomal protease tripeptidyl-peptidase I (TPP I). The natural substrates for TPP I and the pathophysiological processes associated with lysosomal storage and disease progression are not well understood. Detailed characterization of TPP I substrate specificity should provide insights into these issues and also aid in the development of improved clinical and biochemical assays. To this end, we constructed fluorogenic and standard combinatorial peptide libraries and analyzed them using fluorescence and mass spectrometry-based activity assays. The fluorogenic group 7-amino-4-carbamoylmethylcoumarin was incorporated into a series of 7-amino-4-carbamoylmethylcoumarin tripeptide libraries using a design strategy that allowed systematic evaluation of the P1, P2, and P3 positions. TPP I digestion of these substrates liberates the fluorescence group and results in a large increase in fluorescence that can be used to calculate kinetic parameters and to derive the substrate specificity constant $k_{cat}/K_m$. In addition, we implemented a mass spectrometry-based assay to measure the hydrolysis of individual peptides in peptide pools and thus expand the scope of the analysis. Nonfluorogenic tetrapeptide and pentapeptide libraries were synthesized and analyzed to evaluate P1 and P2 residues. Together, this analysis allowed us to predict the relative specificity of TPP I toward a wide range of potential biological substrates. In addition, we evaluated a variety of new fluorogenic peptides with a P3 Arg residue, and we demonstrated their superiority compared with the widely used substrate Ala-Ala-Phe-AMC for selectively measuring TPP I activity in biological specimens.

The neuronal ceroid lipofuscinoses (NCLs) are a group of hereditary neurodegenerative diseases primarily affecting children and adolescents who share similar clinical features, including visual loss, seizures, mental regression, behavioral changes, movement disorders, and shortened life expectancy (1). At a cellular level, NCLs are characterized by accumulation of autofluorescent storage material in neurons and other cell types (1, 2).

Classical late-infantile NCL (LINCL), one of the most common forms of NCL, typically has an age of onset of about 2–4 years and results in death at around age 10. The gene defective in LINCL was found by using a proteomic approach that compared the spectrum of lysosomal proteins in control and LINCL brain autopsy specimens (3). This gene, designated CLN2, encodes a lysosomal protease that shares sequence similarities with bacterial pepstatin-insensitive peptidases, and a pepstatin-insensitive protease activity was found to be missing in LINCL specimens (3, 4).

The CLN2 gene product was later demonstrated to be identical to the lysosomal enzyme tripeptidyl-peptidase I (TPP I) (5, 6). This enzyme is a serine protease (7) that, based on the structure of the bacterial homolog, has an unusual Ser, Asp, and Glu catalytic triad, allowing activity at acidic pH (8, 9). TPP I has two proteolytic activities, a weak endopeptidase activity with a pH optimum of 3.0 (10) that may be important for the low pH-triggered intramolecular autoactivation and autoprocessing of the 66-kDa inactive proenzyme to the 46-kDa mature enzyme (7, 11), and a stronger exopeptidase activity with a pH optimum of 4.5. The exopeptidase activity of TPP I enables the cleavage of tripeptides sequentially from unsubstated N termini of polypeptides or proteins. A variety of protease assays has been developed for enzyme-based post- and prenatal diagnosis (4, 12–16).

In terms of a potential physiological substrate, subunit c of mitochondrial ATP synthase was found to be a major constituent of the storage material in LINCL (17). However, this extremely hydrophobic protein also accumulates in a number of different other NCLs (17–22). Without detailed understanding of TPP I substrate specificity and the combined actions of other lysosomal proteinases, it remains difficult to distinguish whether subunit c accumulation in LINCL is a primary or secondary effect of TPP I deficiency.

Although there is some information regarding TPP I substrate specificity (reviewed in Ref. 23), no comprehensive study has been performed to date. Originally, TPP I was found to release Gly-Pro-Xaa tripeptides from collagen-related synthetic polypeptides (24). Pro was not tolerated at P1 and P1’ (25–27), and the substrate specificity was related to peptide length (27). Detailed knowledge of the substrate specificity of TPP I is important for two reasons. First, identification of the true physiological substrates of TPP I may provide important insights into LINCL pathophysiology and may reveal targets for therapeutic intervention. Second, there is a need for a more specific substrate to assay TPP I activity for improved clinical enzyme-based diagnostic and carrier testing assays. In addition, a highly specific substrate would allow...
measurement of low levels of enzyme activity, which is important for determining residual activity that may be associated with some CLN2 mutants as well as screening for drugs that could potentially rescue the function of mutant CLN2 gene products (28).

Several recent methodological developments facilitate the construction and analysis of combinatorial peptide libraries for evaluation of peptidase substrate specificity. First, 7-amino-4-carbamoylmethylcoumarin (ACC) has been developed as a novel reagent that allows a fluorogenic molecule to be incorporated as the C-terminal residue of peptides using standard solid phase synthetic methods (29, 30). Hydrolysis of the C-terminal amide bond results in release of free ACC and an ~900-fold increase in fluorescence that is largely independent of pH, thus providing an efficient method to assay enzyme activity. Second, use of an “isokinetic mixture” of Fmoc amino acids, where concentrations of reagents are adjusted so that they have similar pseudo-first order reaction kinetics, allows rapid incorporation of multiple residues at a given position (30, 31). Third, recent advances in LC/MS/MS technologies allow quantification of relatively complex mixtures of peptides for enzyme kinetic studies (32).

In this study, we used combinatorial fluorogenic and standard peptide libraries combined with fluorescence and LC/MS/MS-based assays to analyze the substrate specificity of TPP I. We determined the relative contribution of all non-sulfur-containing amino acids and noreleucine to each of the P3-, P2-, P1-, and P2’-substituted residues (see Ref. 48 for peptide subite nomenclature). This approach allowed us to obtain a comprehensive substrate specificity map (landscape) of TPP I that should be useful in identifying true biological substrates. In addition, this information has enabled synthesis of new fluorogenic peptides that exhibit greater selectivity for TPP I compared with substrates used previously. These new TPP I substrates represent superior reagents for use in diagnostic and other activity assays.

**EXPERIMENTAL PROCEDURES**

Reagents—Rink amide 4-methylbenzhydrylamine resin and PyBOP were from Novabiochem. Fmoc-His(t-butyloxycarbonyl)-OH was from Bachem. All other Fmoc amino acids and HOBt were from Advanced ChemTech. 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate was from Applied Biosystems. Recombinant human TPP I proenzyme containing a C-terminal hexahistidine tag was expressed and purified as described (7). All peptide synthesis and HPLC solvents were purchased from Aldrich.

Fmoc-ACC Resin and ACC Synthesis—7-Fmoc-aminocoumarin-4-acetic acid was prepared as described (29, 30). In a reaction vessel, 150 ml of 20% piperidine/DMF solution was added to rink amide methylbenzhydrylamine resin (10 g, 5 mmol). The mixture was then shaken for 20 min. After washing (DMF and dichloromethane, 100 ml, three times each, alternately), 7-Fmoc-aminocoumarin-4-acetic acid (4.4 g, 10 mmol), PyBOP (5.2 g, 10 mmol), HOBt (1.53 g, 10 mmol), DIEA (2.61 ml, 15 mmol), and DMF (30 ml) were added. The mixture was shaken overnight, filtered, and washed (four times with 100 ml of DMF). The solution was removed, and fresh reagents were added to the resin, and the procedure was repeated (“double coupling”) to achieve sufficient substitution level. The resin was dried under vacuum. The substitution level was 0.48 mmol/g by Fmoc analysis. The unsubstituted free amines were capped by reaction with excess acetic anhydride. The final product was dried over P₂O₅ before weight measurement.

Free ACC Synthesis—Free ACC was obtained by deprotection of the Fmoc-ACC resin followed by trifiuoroacetic acid cleavage and cold ether precipitation. ACC was subject to reversed-phase HPLC purification and lyophilization. The final product was dried over P₂O₅ before weight measurement.

**Synthesis of Combinatorial Peptide Library**—The splitting step in the combinatorial peptide synthesis was achieved by dividing the resin into multiple 1.5-ml polypropylene microcentrifuge tubes, and the synthesis of each pool of peptides was carried out manually. Reactions were conducted at room temperature with continuous mixing on an orbital shaker. Randomization was achieved during the coupling step by using the isosokinetic Fmoc amino acid mixture described previously (31). All peptides synthesized in this study contain a free N terminus and a C-terminal amide.

**P1 Peptide Library I**—This synthesis followed the procedure described previously (29, 30). In brief, the Fmoc-ACC resin was divided to 19 portions (60 mg, 0.029 mmol for each portion). Upon deprotection of the resins, one of the 19 Fmoc amino acids (0.15 mmol, omitting Cys and substituting Met by Nle) along with DMF (0.5 ml), 2-(7-aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (56 mg, 0.15 mmol), and collidine (39 µl, 0.29 mmol) were added to the designated tubes and coupled for 20 h. Double couplings were performed to improve yield. The unresolved aniline group on the ACC resin was partially blocked by adding AcOH (17 µl, 0.29 mmol), diisopropylcarbodiimide (46 µl, 0.29 mmol), and nitrotriazole (32 mg, 0.29 mmol). P2 and P3 positions were coupled using the isokinetic mixture (0.29 mmol of total Fmoc amino acids) along with PyBOP (150.9 mg, 0.29 mmol), HOBt (44.4 mg, 0.29 mmol), and DIEA (75.8 µl, 0.44 mmol). After removing the Fmoc group, the resin was dried and treated with 95% trifluoroacetic acid, 2.5% trisopropylsilane, 2.5% H₂O to cleave the peptide from the resin. The peptides were lyophilized and dried under vacuum. The composition of the libraries was evaluated through peptide sequencing and LC/MS/MS. Both methods showed that the 19 amino acids were all present at approximately the same levels at both the P2 and P3 positions (data not shown). The molar yield of each mixture was calculated from the dry weight and the theoretical average molecular weight of the 361 peptides in the mixture. Peptide stock solutions (25 mM) were prepared by dissolving the lyophilized peptide pools in Me₂SO.

**P2 Library**—Four peptidyl ACC resins with the P1 position fixed as Leu, Nle, Phe, or Tyr (0.6 g, 0.29 mmol for each) were constructed following the procedure described above. Each library was divided into 19 portions (30 mg, 0.014 mmol each). The P2 residues were coupled individually by standard solid phase Fmoc chemistry. Fmoc amino acid (0.042 mmol), PyBOP (22 mg, 0.042 mmol), HOBt (7 mg, 0.042 mmol), and DIEA (11 µl, 0.078 mmol) were added to the designated tube and coupled for 3 h. Complete coupling was confirmed by the Kaiser test (33). The P3 residues were coupled using the isokinetic mixture method with the same synthesis, cleavage, and stock solution preparation steps as described above. The composition of the P2 library was evaluated by LC/MS, which demonstrated similar levels of the expected 19 peptides in each pool.

**P3 Library**—The P2 and P1 positions (P2-P1) were fixed as Nle-Leu, Nle-Nle, Pro-Nle, Ala-Phe, Nle-Phe, Pro-Phe, and Thr-Phe. Fmoc-ACC resin synthesis and coupling of both P1 and P2 residues followed the procedures described above. Batches of resin representing the seven P1-P2 sequences were divided into 19 portions each, and the 19 × 7 peptides were then synthesized individually. The identities of the peptides were confirmed by LC/MS, and the purity of all the crude peptides was analyzed to be greater than 85% by analytical reversed-phase HPLC. Peptide stock solutions were prepared by using the crude peptides without further purification.

**P1 Library II**—Fmoc-ACC resin was divided into 19 portions, and 19 P1-specific peptidyl resins were synthesized. Each of these resins was then divided into two portions and reacted to introduce either Ala-Ala
or Arg-Nle as the P3-P2 sequence. These 38 ACC tripeptides were purified by semi-preparative reversed-phase HPLC to a purity greater than 90%.

P1′ Tetrapeptide and P1′ P2′ Pentapeptide Libraries—These libraries were synthesized by standard Fmoc chemistry using the procedure described elsewhere (34). The P3, P2, and P1 residues were fixed as Ala, Ala, and Phe. For the P1′ tetrapeptides, the 19 Ala-Ala-Phe-Xaa peptides were synthesized individually. A similar strategy was used for the P1′-P2′ pentapeptide library, but after the initial coupling of the 19 P2′ residues to the resin, a mixture of 19 P1′ residues was introduced by isokinetic coupling. The 19 peptide pools were then individually coupled to introduce Ala-Ala-Phe at the P3-P1 positions. However, LC/MS/MS and Edman degradation analyses showed that in the 19 pools of the P1′ P2′ library, the peptides containing Lys at P1′ were not present, reflecting failed incorporation of Lys from the isokinetic mixture. Thus, each P1′ P2′ peptide pool contains 18 different peptides.

Fluorescence-based Enzyme Assay—The conditions for the enzyme kinetic assay were described previously (7, 13). Briefly, purified proenzyme of TPP I (1 mg/ml, 16.6 μM) was 10-fold diluted into 0.1 M sodium formate buffer, 0.15 M NaCl, 0.1% Triton X-100, pH 3.5, and incubated overnight at room temperature to allow for complete activation. The activated TPP I was stored at −80 °C and thawed shortly before use. The Me2SO stock solutions of the peptide substrates were diluted into Assay Buffer (0.1 M sodium acetate buffer, 0.15 M NaCl, 0.1% Triton X-100, pH 4.5) to the appropriate concentration before incubation with TPP I.

Calibration curves were constructed to obtain a conversion factor to transform output data obtained using a fluorescence plate reader (CYTOFLUOR 4000, PerSeptive Biosystems, excitation filter at 360 to transform output data obtained using a fluorescence plate reader to determine the digestion of substrate was 200, 100, 50, and 25 μM. (Note: when multiple substrates were operated at low resolution (2 μM) was 10-fold diluted into 0.1 M sodium formate buffer, 0.15 M NaCl, 0.1% Triton X-100, pH 3.5, and incubated overnight at room temperature to allow for complete activation. The activated TPP I was stored at −80 °C and thawed shortly before use. The Me2SO stock solutions of the peptide substrates were diluted into Assay Buffer (0.1 M sodium acetate buffer, 0.15 M NaCl, 0.1% Triton X-100, pH 4.5) to the appropriate concentration before incubation with TPP I.

Calibration curves were constructed to obtain a conversion factor to transform output data obtained using a fluorescence plate reader (CYTOFLUOR 4000, PerSeptive Biosystems, excitation filter at 360 nm and emission filter at 460 nm) to units of ACC concentration. Free ACC was obtained by complete digestion of 80, 40, 20, 10, 5, 2.5, and 1.25 μM of Ala-Ala-Phe-ACC by 4.1 nM activated TPP I at 30 °C for 24 h. Dilutions of synthetic free ACC (described above) were also analyzed directly to construct calibration curves. Both methods gave equivalent results. Preparation of tissue extract for enzymatic measurements was performed according to a reported procedure (4). The procedure for measurement of TPP I activity at different pH values was described previously (13). CLN2(−/−) mouse specimens were from the gene-targeted model described previously (35).

Derivation of Vmax and Ks—Activated TPP I and ACC peptides were serially diluted using Assay Buffer. Reactions were conducted in flat-bottom polystyrene 96-well plates and initiated by adding 80 μl of peptide solution to 20 μl of TPP I solution. The final concentration of TPP I during incubation was 4.10, 2.05, and 1.00 nM, and the concentration of substrate was 200, 100, 50, and 25 μM. (Note: when multiple peptides are present in the mixture, the concentration given refers to the sum of all peptides.) Fluorescence was monitored at 1-min intervals for at least 30 min at 30 °C. In some cases, the reactions were repeated using a 4-fold lower range of substrate concentrations to achieve optimum reaction conditions. Digestion rate was determined by measuring the initial rate of digestion using conditions where the rate of digestion was linear with respect to time and enzyme concentration. Vmax and Ks values were obtained by fitting the digestion rate with the corresponding substrate concentration to the Michaelis-Menten function. kcat was calculated by dividing Vmax by the TPP I concentration, [E]0. For substrates where the initial rate did not approach Vmax, at increasing substrate concentrations (Ks > 400 μM), the direct method was used to determine kcat/Ks (see below).

Derivation of kcat/Ks from Low Substrate Concentration—Substrate concentration for the TPP I digestion was 10, 5, 2.5, and 1.25 μM during the digestion. Incubation and fluorescent monitoring followed the procedure described above. The slope (k) of digestion rate (Vobs/endo) plotted against the substrate concentration was obtained by linear fitting. The kcat/Ks value was calculated by dividing k with the TPP I concentration [E].

TPP I Digestion for LC/MS/MS Analysis—All incubations were carried out in Assay Buffer. Peptide substrates were diluted in Assay Buffer to the desired concentration from stock solutions (25 μM) in Me2SO. All incubations were carried out at 30 °C. To check that substrate disappearance corresponded to cleavage of the expected P1′−P1′ bond, the proteolysis products from the peptide pool Xaa-Nle-Phe-ACC were identified by LC/MS/MS.

For the P2 library, peptide solutions were added to the activated TPP I to give a final concentration of 100 μM total peptides and 6.9 nM TPP I in a final volume of 250 μl. Samples (100 μl) were removed after 40 and 480 min, and the reaction was terminated using 10 μl of 15% trifluoroacetic acid. The initial substrate concentration was determined as described above for the 40-min time point but omitting enzyme from the reaction mixture. More detailed studies were initially conducted with the Xaa-Nle-Phe-ACC library to monitor the time course of the reaction for method validation.

The P1′ P2′ pentapeptide library was digested as described above, except for the following. 1) The TPP I concentration during incubation was 2.1 nM. 2) Three ACC peptides (Ala-Ala-Phe-ACC, Asn-Nle-Nle-Nle, and Ala-Ala-Pro-ACC) were included as internal standards with a final concentration of 1 μM. 3) Samples (50 μl) were removed at 4, 40, and 480 min, and the reactions were terminated by adding 5 μl of 15% trifluoroacetic acid. 4) Terminated reactions were diluted 5-fold with distilled water before LC/MS/MS analysis. Before the digestion of the P1′ tetrapeptide library, equimolar amounts of 19 P1′ tetrapeptides were mixed. The peptide mixture was digested under the same conditions as the P1′ P2′ pentapeptide library, but 7 time points were monitored (20 s and 4, 10, 40, 120, 280, and 480 min) to measure the time course of digestion.

LC/MS/MS Instrumentation—A Qtrap hybrid mass spectrometer (MDS/Sciex) was used for analysis of the P2 libraries. An LTQ linear ion trap mass spectrometer (Thermo) was used for analysis of the P1′ tetrapeptide and the P1′-P2′ pentapeptide libraries. The mass spectrometers were interfaced with an LC Packings HPLC system consisting of a solvent pumping system (Ultimate), a column switching device (Switchos II), and an autosampler (Famos).

LC/MS/MS Analysis of P2 Library—For the analysis of selected peptide pools from the P2 library (each pool containing 19 peptides), 5 μl of incubation mixture were injected for each LC/MS/MS run. Samples were loaded onto a trapping column (5 × 0.3 mm inner diameter, polymeric C-18; LC Packings) and washed for 5 min at a flow rate of 50 μl/min using 0.1% formic acid/H2O. The valve was then switched so that the flow to the trapping column was reversed and brought in line with an analytical column (15 cm × 1 mm inner diameter, 3-μm C-18 resin; LC Packings) and the mass spectrometer. The peptides were separated by using a linear gradient of 5–25% of buffer B (2% H2O, 98% ACN, 0.1% formic acid) in buffer A (2% ACN, 98% H2O, 0.1% formic acid) for 40 min at a flow rate of 40 μl/min. Multiple reaction monitoring (MRM) was used as the scan mode to identify and quantify each of the components in the peptide mixture. The first and third quadrupoles were operated at low resolution (2 m/z) to achieve maximum ion passage. The parent masses were selected based on the calculated molecular weight of each component in the mixture. The m/z values for singly charged MH+ ions were monitored for the majority of the peptides, whereas doubly charged (M + 2H)+/2 ions were monitored for the peptides containing positively charged residues (Arg, His, and Lys). The
Substrate Specificity of Tripeptidyl-peptidase I

Table 1: Peptide libraries synthesized and analyzed in this study

| Peptide Library | Number of Peptides | Number of Pools | Number of Peptide/pool | Residue* |
|-----------------|--------------------|-----------------|------------------------|---------|
| P1 Library I    | 6859               | 19              | 361                    | P3      |
| P1 Library II (38 total) | 19                |                 |                        | P4      |
| P2 Library      | 2527               | 76              | 19                     | P1      |
| P3 Library (133 total) | 38                |                 |                        | P1, P2  |
| P1’ Library     | 19                 | 19              | 18                     | P1’     |
| P2’ Library     | 342                | 19              | 18                     | P2’     |

* Indicated residue at a given position. When designated as “defined,” separate syntheses were used to introduce the indicated amino acid at a given position. When designated as “mixed,” an approximately equimolar mixture of amino acids at a given position was introduced by isokinetic coupling during synthesis. The 19 amino acids used consist of all natural amino acids excluding Cys and Met and including Nle. All peptides contain a free N terminus and a C-terminus amide.

Collision energy for each component was optimized based on direct-infusion experiments of the peptide mixture using the “Enhanced MS/MS” scan mode. Briefly, Arg at the P3 position required a higher collision energy of 45–55 (arbitrary value); Pro, His, Glu, Gln, Phe, Trp, and Tyr required 35–40; Ile, Leu, Lys, and Nle required 30; Ala, Asn, Asp, Ser, Thr, and Val, required 25; and Gly required 20. The most frequently occurring fragments based on the enhanced MS/MS experiments were a2 and b2 ions and the immonium ions of the residues at the Asp, Ser, Thr, and Val, required 25; and Gly required 20. The most frequently occurring fragments based on the enhanced MS/MS experiments were a2 and b2 ions and the immonium ions of the residues at the Asp, Ser, Thr, and Val, required 25; and Gly required 20.

RESULTS

Library Design—We used combinatorial fluorogenic ACC tripeptides as well as tetrapeptide and pentapeptide libraries to evaluate systematically the substrate specificity of TPP I. Construction of fluorogenic peptides was based on a procedure where we sequentially evaluated each P1, P2, and P3 position, using the results of the previous analysis to design subsequent libraries. By stepwise analysis of different positions, the residues surveyed in each consecutive round of peptide synthesis were restricted, thus decreasing the complexity and number of peptides required for sufficient coverage.

P1 Library I—The first P1-diverse library (see Table 1 for nomenclature) was constructed by synthesizing 19 separate pools of 361 ACC tripeptides, each with a different fixed P1 residue and containing a randomized mixture of 19 residues at the P2 and P3 positions. The fluorogenic TPP I enzyme activity assay was used to analyze each pool, thus yielding the relative substrate specificity of each P1 residue averaged over all P2 and P3 residues. Subsequently, after analysis of the P2 and P3 libraries (see below), a second more restricted P1 library was constructed. This library, designated P1 library II, consisted of 38 separate ACC tripeptides, where all 19 residues were surveyed in the P1 position in the context of two different P3-P2 sequences (Ala-Ala and Arg-Nle).

Peptide pools in the P1 library I were digested using TPP I, and the substrate specificity constant, kcat/KM, was calculated directly from the rates of reaction at low substrate concentrations (direct method) or by determination of the individual kinetic parameters kcat and Km at medium to high substrate concentrations (Michaelis-Menten method). Note that both the substrate specificity constants and kinetic parameters are apparent values, reflecting activity of all 361 peptides in each pool. For the majority of the P1 peptide pools, the results obtained from both methods yielded similar results (supplemental Table 1). TPP I digestion favored the large hydrophobic residues Phe, Leu, and Nle at P1 (Fig. 1). The individual values for peptide pools with the P1 residue fixed at His, Val, Ile, Thr, Gly, and Pro could not be derived because of their very slow rate of digestion above the ACC background signal and...
required further investigation using purified peptides (see below for analysis of P1 library II).

There were two drawbacks using the complex mixture for the evaluation of P1. First, some of the fixed P1 peptide pools had high background fluorescence because of the presence of free ACC, which could not be removed using standard peptide synthesis extraction methods. For example, Arg, Ile, Thr, and Val had low coupling efficiencies to the aniline group on ACC, probably because of the branched side-chain structures of these amino acids. In these syntheses, more uncoupled free ACC was cleaved from the peptidyl resin along with the ACC tripeptides, and its presence interfered with the analysis of weak substrates. Second, although the current method can rank the P1 position in a relative manner, the presence of 361 peptides in the pool prevents determination of true $k_{\text{cat}}$ and $K_M$ values.

To address these issues, peptides with diverse P1 residues and fixed P3-P2 sequences were individually synthesized and purified by HPLC. Two combinations were chosen for this series. One was based on a widely used TPP I substrate, Ala-Ala-Phe-AMC, and thus the P3-P2 sequence was fixed as Ala-Ala. The other contained the P3-P2 sequence Arg-Nle based on analysis of the P3 library (see below). Peptides (25–200 μM) were incubated with purified TPP I, and substrate specificity constants were calculated either using the Michaelis-Menten method or, when $K_M$ values exceeded 400 μM, the direct method. $k_{\text{cat}}$ and $K_M$ values were reported for individual peptides that had $K_M$ values below 400 μM (supplemental Table 2). Analysis of the individual peptides using the fluorometric TPP I assay revealed a similar rank order compared with that obtained from P1 library I. The hydrophobic residues Leu, Nle, and Phe were the most favorable residues at the P1 position (Fig. 2), followed by Tyr, Trp, Glu, Asp, Gln, and Ala. Interestingly, in cases when the kinetic constants $K_M$ and $K_{\text{cat}}$ could be determined, linear regression plots (not shown) indicated that the substrate specificity constant correlated well with substrate affinity and to a greater degree than with $k_{\text{cat}}$ ($k_{\text{cat}}/K_M$ versus $1/K_M$, $r^2 = 0.94$; $k_{\text{cat}}/K_M$ versus $k_{\text{cat}}$, $r^2 = 0.18$). This analysis also showed clearly that peptides containing P1 Asn, His, Lys, Arg, Ser, Val, Ile, Thr, Gly, and Pro are all poor substrates for TPP I, with specificity constants less than 1% of that of the best substrates. The influence of the P1 position on the substrate specificity was quite marked, with over a 10,000-fold difference in $k_{\text{cat}}/K_M$ between the optimal P1 residues (Leu, Nle, and Phe) and the poorest P1 residue (Pro) (see “Discussion” and Fig. 12).

There was excellent agreement on the effect of the P1 residue on substrate specificity when determined by using peptides with P3-P2 sequences being either Ala-Ala, Arg-Nle, or a mixture of all possible residues (Fig. 2). The noticeable exception was when Arg, Lys, or to a lesser extent His occupied the P1 position in the context of the Arg-Nle P3-P2 sequence. This suggests an unfavorable effect of placing positively charged residues at both P1 and P3 positions. Another noteworthy observation is that Ile and Val produce low substrate specificity in the P1 position for both peptide sets compared with other hydrophobic residues. This discrepancy may be explained in that Ile and Val differ in having a branched side chain at the β-carbon, which may sterically hinder substrate binding. Thr likewise has a β-branched structure and results in low substrate specificity when present in the P1 position.

P2 Library—The P2-diverse library was restricted at the P1 position to the following four optimum residues found when evaluating P1 library I: Leu, Nle, Phe, and Tyr. Here, for each of these fixed P1 residues, 19 pools of 19 ACC tripeptides were synthesized, each with a different fixed P2 position and containing a randomized mixture of 19 residues at the P3 position (see Table 1). These 76 peptide pools were evaluated by using the fluorescence assay to determine the relative contribution of each P2 residue in the context of four different P1 residues averaged over all P3 residues. Subsequently, after analysis of the P3 library and validation of an LC/MS/MS-based assay, we analyzed a subset of the peptide pools to determine the specificity constants for 589 individual peptides from 31 pools in the P2 library, allowing more thorough evaluation of potential P3-P2-P1 interactions (see the P3 library analysis below).

The results showed that the non-natural amino acid, Nle, which has an unbranched aliphatic side chain, was the most favorable residue overall at P2 for TPP I digestion. Of the natural amino acids, Pro and Ala followed by Phe, Tyr, and Trp were the most favorable residues (Fig. 3). In general, the P2 position did not show the high selectivity observed at P1, as the highest and lowest $k_{\text{cat}}/K_M$ values only differed by a factor of ~100 (supplemental Table 3, and also see Fig. 12).

P3 Library—The P3-diverse library was constructed by synthesizing 133 individual ACC tripeptides that surveyed all 19 residues in the P3 position in the context of the following seven different P2-P1 sequences: Nle-Leu, Nle-Nle, Pro-Nle, Ala-Phe, Nle-Phe, Pro-Phe, and Thr-Phe.

The specificity constant $k_{\text{cat}}/K_M$ was calculated after determining individual kinetic parameters using the Michaelis-Menten method. The results showed that the positively charged residues Arg, His, and Lys, as well as Tyr, were favored at the P3 position, and Pro was the least...
favorable residue (Fig. 4). These data also revealed that the relative effect of different P3 residues on the substrate specificity constant depended on the identity of the P1 residue. Arg clearly is the optimal P3 residue when the natural amino acids Phe or Leu occupy P1. In contrast, His, Tyr, and Arg are the optimal P3 residues when the non-natural amino acid Nle occupies P1 (Fig. 4 and see below).

Because one purpose of this study was to develop an improved assay for TPP I activity, we compared the new ACC substrates with Ala-Ala-Phe-ACC, which has similar kinetic properties to the widely used TPP I substrate Ala-Ala-Phe-AMC ($k_{cat}/K_M \approx 0.025 \mu M^{-1} s^{-1}$ for both substrates). Inspection of Fig. 4 indicates that the Ala-Ala-Phe sequence was not optimal, with a number of fluorogenic peptides having considerably higher substrate specificity constants ($k_{cat}/K_M \approx 0.1$ to $0.2 \mu M^{-1} s^{-1}$). Although these consist predominantly of peptides containing the non-natural amino acid Nle (e.g. P3-P2-P1 sequences Arg-Nle-Leu, Arg-Nle-Nle, Arg-Nle-Phe, His-Nle-Nle, and Tyr-Nle-Nle), a number of the more specific substrates contained natural P3-P2-P1 sequences (e.g. Arg-Pro-Phe, Arg-Ala-Phe, Lys-Pro-Phe, and Gln-Pro-Phe). Examination of the individual kinetic constants revealed that the higher substrate specificity constants for the more favorable substrates could largely be attributed to their higher affinity (lower $K_M$) for TPP I compared with that of the corresponding Ala-Ala-Phe substrate (Table 2 and supplemental Table 4).

Given that the use of Nle in the P2 position produced some of the most specific substrates, we also synthesized additional peptides with the optimal residues at P3 (His, Tyr, and Arg) and P1 (Nle and Phe) but containing norvaline (Nva) at P2. (Like Nle, Nva is a non-natural amino acid that contains an unbranched aliphatic side chain but is one methylene shorter than Nle.) Analysis revealed a trend where peptides that contain a P2 Nva residue have substrate specificity constants $\sim$30%
higher than that of the corresponding P2 Nle peptides, and this is primarily because of an effect on $K_{M}$ (Table 2 and supplemental Table 4).

**LC/MS/MS Analysis of Peptide Pools**—The results obtained from analysis of the individual peptides comprising the P1-Ii and P3 libraries suggested some interdependence of the P1, P2, and P3 positions on TPP I substrate specificity. To investigate this further without the considerable effort and expense required to synthesize additional individual peptides, we implemented an LC/MS/MS-based assay that allowed us to determine the substrate specificity constants of individual peptides within peptide pools. Briefly, TPP I was incubated for varying periods of time with peptide pools at concentrations of individual peptides well below their anticipated $K_{M}$. After stopping the reaction, the mixture was analyzed by LC/MS/MS. From the retention time, the precursor mass, and the mass of the product ions, all peptides within a pool could be distinguished (see “Experimental Procedures”). Control experiments indicated that the integrated peak areas of the selected product ions were proportional to the amount of peptide analyzed and that there was excellent run-to-run reproducibility (data not shown).

We validated the LC/MS/MS assay using the seven P2 library pools that contained all 133 peptides that had been analyzed as individual peptides (see above under “P3 Library”). Initially, time course experiments using the Xaa-Nle-Phe-ACC peptide pool indicated that substrate disappearance could be fit with a single exponential decay model (Fig. 5, top panel), as expected if the digestion was first order with respect to substrate concentration (see “Experimental Procedures”). Analysis of MS spectra indicated that disappearance of substrate was the result of hydrolysis of the P1–ACC bond and thus reflected TPP I activity. Given no appreciable reverse reaction and $K_{M} \gg [S]$, the substrate specificity constant can be calculated from the integrated Michaelis-Menten expression $k_{cat}/K_{M} = \ln([S]_{0}/[S]_{t})/(t \times [E]_{0})$ (see “Experimental Procedures”). We found that direct calculation of $k_{cat}/K_{M}$ from single time points, where $[S]_{t}/[S]_{0}$ ranged from 0.8 to 0.2, or curve fitting of the entire time course yielded similar results (data not shown). Thus, for the remaining pools, we analyzed 0-, 40-, and 480-min time points by LC/MS/MS, using the 40-min time point for "good" substrates and the 480 min time point for "poor" substrates. The results of this analysis agreed well with those obtained from the fluorometric assay of individual peptides, thus validating the approach (Fig. 5, bottom panel).

We analyzed a total of 589 peptides (31 pools of 19 peptides) from the P2 library using the LC/MS/MS assay and calculated their individual substrate specificity constants (values listed in supplemental Table 5 and graphed in supplemental Fig. 1). To simplify the presentation of this large amount of information, we normalized the $k_{cat}/K_{M}$ values for peptides with a given P2–P1 sequence to the corresponding P3 Arg peptide, and we then averaged these normalized values across all available P2 residues for peptides with the same P1 and P3 residues (Fig. 6). This revealed that for all natural amino acids surveyed at P1 (Leu, Phe, and Tyr), the optimal P3 residue clearly was Arg. In contrast, for peptides with a P1 Nle, the optimal P3 residues were His followed by Tyr and Arg.

These results were also evident from the fluorescence-based assay of the peptides composing the P3 library (see Fig. 4). In terms of trends for the suboptimal P3 residues, the only apparent differences were that P1 Nle

**TABLE 2**

Kinetic properties of representative fluorogenic TPP I substrates

| Substrate                  | $k_{cat}$ | $K_{M}$ | $k_{cat}/K_{M}$ |
|----------------------------|-----------|---------|-----------------|
| Ala-Ala-Phe-AMC            | 5.2       | 188     | 0.028           |
| Ala-Ala-Phe-ACC            | 5.3       | 233     | 0.022           |
| Arg-Nle-Leu-ACC            | 3.6       | 28      | 0.133           |
| Arg-Nle-Nle-ACC            | 4.2       | 37      | 0.120           |
| Arg-Nva-Nle-ACC            | 4.7       | 30      | 0.169           |
| Arg-Pro-Phe-ACC            | 8.2       | 54      | 0.156           |

$K_{M}$ values for 133 peptides derived from mass spectrometry analysis of P2 library peptide pools are plotted against the $k_{cat}$ values for 133 peptides derived from single time points, where $[S]_{t}/[S]_{0}$ ranged from 0.8 to 0.2, or curve fitting of the entire time course yielded similar results (data not shown). Thus, for the remaining pools, we analyzed 0-, 40-, and 480-min time points by LC/MS/MS, using the 40-min time point for "good" substrates and the 480 min time point for "poor" substrates. The results of this analysis agreed well with those obtained from the fluorometric assay of individual peptides, thus validating the approach (Fig. 5, bottom panel).

![FIGURE 5. Validation of the LC/MS/MS TPP I assay. Top panel, reaction time course. LC/MS/MS was used to determine disappearance of substrate during hydrolysis of the Xaa-Nle-Phe-ACC P2 library pool as described under “Experimental Procedures.” Representative peptides with the indicated P3 residue are shown. Bottom panel, comparison of the LC/MS/MS and fluorescence TPP I assays. $k_{cat}/K_{M}$ values for 133 peptides derived from mass spectrometry analysis of P2 library peptide pools are plotted against the $k_{cat}/K_{M}$ values of the individual peptides from the P3 library calculated using the direct fluorescence method.](image-url)

![FIGURE 6. Analysis of individual peptides within the P2 library using the LC/MS/MS assay. Substrate specificity constants of peptides within a given pool (fixed P1 and P2 residues) were first normalized to the corresponding value of the peptide containing P3 Arg within that pool. For peptides with a given P3 and P1 residue, the normalized value was averaged across all available P2 residues, and the mean value ± range is displayed.](image-url)
favored a P3 Phe and that P1 Tyr favored the negatively charged amino acids Glu and Asp at P3 (Fig. 6).

*Substrate Specificity at the P1’ and P2’ Position*—We used nonfluorogenic peptide libraries and the LC/MS/MS assay to explore further the substrate specificity of TPP I. We initially evaluated the P1’ position by synthesizing 19 individual tetrapeptides with diverse P1’ residues, with the P3-P2-P1 sequence fixed as Ala-Ala-Phe based on the sequence of the commonly used TPP I substrate Ala-Ala-Phe-AMC. The peptides had a free N terminus and a C-terminal amide. Equimolar amounts of these 19 peptides plus three fluorogenic substrates serving as internal controls (Ala-Ala-Phe-ACC, Asn-Nle-Nle-ACC, and Ala-Ala-Pro-ACC) were mixed, digested with TPP I for periods varying from 20 s to 8 h, and analyzed using the LC/MS/MS-based assay (see “Experimental Procedures”). The $k_{\text{cat}}/K_M$ values for each component were calculated by curve fitting as described above. The results obtained for the three internal standards agreed with the previous fluorescence-based analysis of the individual peptides, further validating this approach (data not shown).

The substrate specificity constants for the Ala-Ala-Phe-Xaa tetrapeptides are shown in Fig. 7 (solid bars). The most favored residues at P1’ were Tyr and Phe, followed by the moderately favorable residues Val, Trp, Ile, and Ala. Tetrapeptides with polar residues in the P1’ position were less favorable, and those with positively charged side chains were particularly unfavorable. Finally, like the P1 position, no hydrolysis was observed for peptides with a P1’ proline.

This analysis also allowed us to compare the Ala-Ala-Phe-Xaa tetrapeptides with the corresponding Ala-Ala-Phe fluorogenic peptides. Both the ACC and AMC coumarin derivatives were relatively poor substrates, having specificity constants similar to that of Ala-Ala-Phe-Asn and Ala-Ala-Phe-Leu, ~40-fold lower than that of the optimal tetrapeptide Ala-Ala-Phe-Tyr. This suggests that either the aromatic coumarin is unfavorable because of steric constraints or that the nonpeptidic amide bond is somewhat resistant to hydrolysis. Regardless, the identity of the P1’ residue had a significant effect, with the substrate specificity constant of the optimal Ala-Ala-Phe-Tyr being 500-fold higher than that of the second least specific substrate Ala-Ala-Phe-Arg.

We extended this analysis using a P2’-diverse library consisting of 19 pools, each containing 18 peptides. Synthesis was conducted using 19 separate reactions with a different fixed P2’ residue, an isokinetic mixture of 18 amino acids in the P1’ position, and the P3-P2-P1 sequence Ala-Ala-Phe (Table 1). (Edman sequencing and MS analysis revealed that Lys was not present at the P1’ position, suggesting that Lys coupling from the isokinetic mixture of amino acids failed. However, given that analysis of the tetrapeptides revealed that P1’ Lys was extremely unfavorable, the P1’ Lys-deficient pentapeptide pools were used for further characterization of TPP I substrate specificity.) After adding three fluorogenic peptides as internal standards (see above), each pool was digested by TPP I for 0, 4, 40, and 480 min and analyzed using quantitative LC/MS/MS (see “Experimental Procedures”).

The effect of peptide length on P1’ is shown in Fig. 7. In Fig. 7, the open bars represent pentapeptides containing the indicated P1’ residue averaged over all P2’ positions except Pro, with the range being indicated by the vertical lines. Comparison of the pentapeptides with the corresponding tetrapeptides (Fig. 7, solid bars) indicates that in general, the relative contribution of different P1’ residues to substrate specificity is similar for the different length peptides. Interestingly, both length and the identity of the P2’ position have a relatively small effect on the more favorable P1’ residues compared with the less favorable P1’ residues. In general, the improvement found by increasing the length of the peptide is modest, with the optimal pentapeptide tested (Ala-Ala-Phe-Trp-Val, $k_{\text{cat}}/K_M = 2.3 \mu M^{-1} s^{-1}$) having only an ~2-fold higher substrate specificity than the corresponding optimal tetrapeptide (Ala-Ala-Phe-Tyr, $k_{\text{cat}}/K_M = 0.94 \mu M^{-1} s^{-1}$) (supplemental Table 6).

The relative effects of different P2’ residues are more clearly demonstrated by Fig. 8. Here, peptides with the same P1’ residue are grouped, with each graph showing the relative effect of the P2’ position. In Fig. 8, solid lines indicate the substrate specificity constants for pentapeptides with different P1’ residues averaged across all P2’ residues except Pro, whereas the dashed lines in Fig. 8 indicate the values for the corresponding tetrapeptides. Note that the scales are different for different P1’ residues. Several general features are apparent from this analysis. First, the presence of Pro at the P2’ position markedly decreases hydrolysis of most peptides by TPP I (Fig. 8). Second, except for Pro, in general, the presence of a given P2’ residue has either a beneficial or neutral effect on substrate specificity (compare the dashed lines with individual solid bars in Fig. 8). Hydrophobic residues at P2’ tend to have the most beneficial effect, in particular Val, Leu, and Nle for all P1’ residues and Ile for the poorer P1’ residues. In contrast, hydrophilic residues and Gly at P2’ tend to have little effect or even decrease the substrate specificity constant compared with the corresponding P1’ tetrapeptides. However, compared with P1’, except for Pro, the overall effect of the P2’ residue on substrate specificity is relatively small (see “Discussion” and Fig. 12).

Use of Fluorogenic Peptides for Analysis of Enzyme Activity in Biological and Clinical Samples—When devising biochemical assays using fluorogenic peptides, one has to consider both the activity of the enzyme of interest as well as additional potentially interfering activities. For instance, TPP I activity would be overestimated if the fluorophore were liberated following cleavage by another tripeptidylpeptidase or by the combined actions of a dipeptidase and an aminopeptidase. Alternatively, TPP I activity could be underestimated if there were nonproductive consumptions of substrate resulting from a single cleavage by an aminopeptidase or by a dipeptidase. To test these possibilities, we incubated a series of different fluorogenic tripeptides, selected to represent a broad range of $k_{\text{cat}}/K_M$ values, with either highly purified recombinant human TPP I or biological samples (brain extracts from an unaffected human as well as CLN2(+/+) and CLN2(--/--) mice), and we measured the apparent enzymatic activity at pH 4.5. Fig. 9 demonstrates that the measurements obtained using human brain extracts and purified...
human TPP I showed excellent agreement. There was little activity found when using extracts from CLN2(+/−) mouse brain. Taken together, these results indicate that in brain the majority of the activity measured using all substrates tested was because of TPP I (Fig. 9, open squares). Interestingly, the correlation between activities obtained with different substrates in purified human enzyme and CLN2(+/−) mouse brain was considerably poorer than the correlation found using purified human enzyme and human brain (compare Fig. 9, open and filled circles). Additional experiments using mouse liver and human fibroblast extracts revealed a consistent trend in that there was significantly better correlation when comparing the relative hydrolysis of different substrates using different enzyme sources within a species than when comparing across species (data not shown). This suggests that the human and mouse TPP I have evolved to have slightly different substrate preferences.

We also measured the activity as a function of pH for 12 fluorogenic peptides by using human lymphoblast extracts. Use of purified human TPP I gave an activity profile with a single acidic pH optimum in the range pH 4–5 (Fig. 10 and data not shown). Ala-Ala-Phe-ACC and Ala-Ala-Phe-AMC had similar pH profiles to each other, with high activity in the neutral pH range (Fig. 10, upper panel). In contrast, the Arg-Ala-Phe-ACC peptide had extremely low activity in the neutral pH

**FIGURE 8. Analysis of the P2′ library.** Pools of peptides containing the same P2′ residue and a mixture of 18 P1′ residues were digested with purified human TPP I, analyzed by LC/MS/MS, and the substrate specificity constants calculated for each individual peptide. Values are grouped according to the P1′ residue. The dashed line in each graph shows the \( k_{cat}/K_M \) value for the P1′ tetrapeptide, and the solid line shows the \( k_{cat}/K_M \) values for the average of all P2′ residues except Pro for the 18 pentapeptides having the same indicated P1′ residue.
Substrate Specificity of Tripeptidyl-peptidase I

FIGURE 9. Hydrolysis of ACC tripeptides using different enzyme sources. Thirty-six ACC tripeptides were selected that covered a broad range of substrate specificity constants. Individual peptides (200 µm) were incubated with either purified human TPP I or the indicated brain extract. Correlation coefficients were as follows: purified human enzyme versus human brain, \( r^2 = 0.99 \); purified human enzyme versus CLN2(+/+) mouse brain, \( r^2 = 0.78 \).

FIGURE 10. Hydrolysis of fluorogenic substrates as a function of pH. Top panel: the substrates were incubated with a human lymphoblast extract (left axis) or purified human TPP I (right axis) at the indicated pH. Reactions contained 150 mM NaCl, 0.1% Triton X-100, and 100 mM of the following acid-sodium salt buffers: formate, pH 3.5; acetate, pH 4.0–5.5; phosphate, pH 6.0–7.0. Bottom panel: summary of relative hydrolysis of the indicated substrates incubated with human lymphoblast extracts at pH 4.5 and 7.0.

range (Fig. 10, upper panel), and its pH profile was representative of three other ACC peptides with P3-P2-P1 sequences Arg-Nle-Nle, Arg-Pro-Leu, and Arg-Pro-Nle (partial data shown in Fig. 10, lower panel). The other fluorogenic peptides had appreciable activity in the neutral pH range, but all were considerably lower than the Ala-Ala-Phe coumarin derivatives (Fig. 10, lower panel). Interestingly, some substrates such as Arg-Pro-Phe-ACC (Fig. 10, upper panel) and Lys-Pro-Phe-ACC (data not shown) had broader pH optima, with activity at pH 5 greater than or equal to the activity at pH 4.5. Finally, other peptides such as Arg-Nle-Leu-ACC (Fig. 10, upper panel) and Arg-Nva-Nle-ACC (data not shown) had pH optima at both 4.5 and at 6.0. When considered with the observation that only a single acidic pH optimum was found using purified TPP I, these data emphasize the multiplicity of activities present within biological specimens that can hydrolyze some of the fluorogenic substrates. Spillover of these activities into the acidic pH range as well as other activities (see below) is a serious concern when attempting to measure accurately the low levels of residual TPP I activity that can arise from hypomorphic mutations in the CLN2 gene.

We further explored the utility of the new fluorogenic substrates by analyzing different tissues from wild-type and CLN2(−/−) mice, comparing the new fluorogenic substrate Arg-Nle-Nle-ACC and the widely used substrate Ala-Ala-Phe-AMC. Although only low levels of activity were detected in CLN2(−/−) brain using both substrates, this was not the case for all tissues examined (Table 3). For instance, when comparing CLN2(−/−) and wild-type spleen tissues, ~20 and ~1% activity was detected using the Ala-Ala-Phe-AMC and Arg-Nle-Nle-ACC substrates, respectively. Kinetic measurements on CLN2(−/−) spleen revealed that the Ala-Ala-Phe-AMC substrate had nonlinear reaction progress curves (Fig. 11, top panel). The lag followed by an increase in hydrolytic activity might result from processive peptide hydrolysis (e.g. an aminopeptidase and/or a combination of aminopeptidase and dipetidase activities) or from proteolytic processing that could either activate or change the substrate specificity of another hydrolase. Regardless, no such activity was detected using the Arg-Nle-Nle-ACC substrate (Fig. 11, bottom panel). This clearly demonstrated the improved utility of the new fluorogenic substrate for the selective measurement of TPP I activity in biological specimens.

DISCUSSION

In this study, we conducted detailed substrate specificity measurements using synthetic peptide libraries and highly purified recombinant human TPP I. Our results confirm and extend earlier work that used partially purified enzyme preparations and commercially available peptides. For instance, Watanabe et al. (26) conducted a limited analysis of the P1 position using para-nitroaniline (pNA) peptides and partially purified TPP I from rat liver tritosomes. They found that Ala-Ala-Phe-pNA was ~60 times more specific than Ala-Ala-Ala-pNA, and no hydrolysis was detected for Ala-Ala-Pro-pNA. In another study, McDonald et al. (24) tested partially purified porcine TPP I using several fluorogenic Gly-Pro-Xaa tripeptides and found that the relative substrate specificity of the P1 position was Met > Leu > Ala > Arg. Considering that the overall geometry of Met is very similar to Nle, these earlier results are consistent with our findings that Leu, Phe, and Nle are optimal at P1 and that peptides with Pro at P1 are extremely poor substrates.

Recently, Oda and co-workers (36) reported an investigation of the substrate specificity of recombinant human TPP I using a number of hexapeptide derivatives with the unnatural amino acid p-nitrophenylalanine (Nph) at the P1’ position, which, after cleavage, undergoes a small change in absorbance at 300 nm. They found that the most effective substrate tested, Ala-Arg-Phe-Nph-Arg-Leu, had a substrate specificity constant that was 40 times higher than that of Ala-Ala-Phe-AMC. This is similar to our comparison of Ala-Ala-Phe-AMC and the optimal tetrapeptide Ala-Ala-Phe-Tyr (see “Results”). These investigators also examined the relative effect of five different residues at the P1 position in the context of Ala-Xaa-P1-Nph-Xaa-Leu, where Xaa represents a
randomized mixture of 19 amino acids, and they found that all had similar activities (100 to 67% relative hydrolysis rates for Phe, Nle, Tyr, Leu, and Trp), but they did not report any findings regarding other P1 substitutions. Although we find a greater range in differences among these hydrophobic P1 residues, possibly because of the greater sensitivity of the fluorescence and mass spectrometry-based assays, the overall results are not too dissimilar.

Oda and co-workers (36) also analyzed five different P2 residues in the context of the sequence Ala-P2-Phe-Nph-Arg-Leu and five different P3 residues in the context of the sequence Ala-P2-Phe-Nph-Arg-Leu. Their results are markedly different from our findings. For the P2 analysis, they report a 5-fold difference in the \( k_{cat}/K_M \) values of their peptides, with the relative order being Arg > Ala > Asp > Ser ~ His. In contrast, we find that the relative \( k_{cat}/K_M \) values of these residues varies by ~50-fold, with their optimal P2 substrate, Arg, being among our least favored P2 residues. Also, for their P3 analysis, Oda and co-workers (36) reported a 50-fold difference in the five \( k_{cat}/K_M \) values, with the relative order being Ala > Asp > His > Arg ~ Ser. In contrast, we find that one of their most disfavored P3 residues, Arg, is the optimal P3 residue in our study. It is possible that some of these discrepancies reflect the use of different reporter groups in the two studies (ACC versus Nph) and interdependencies between P1’ and other positions. Future studies that directly compare the different substrates under identical conditions are required to resolve these issues.

We synthesized and analyzed over 7,200 different fluorogenic and standard peptides (as individual peptides or within mixtures) for our systematic analysis of the P3, P2, P1, P1’, and P2’ residues. To summarize this large amount of data, we constructed a substrate specificity matrix that compares the relative substrate specificities of different residues at each position, with the optimal residue being assigned a score of 1 (Table 4). A plot of the relative \( k_{cat}/K_M \) values indicates that the P1 position is the most critical, followed by P1’, P2, and P3, with the P2’ position tolerating a wide range of substitutions other than Pro (Fig. 12). If we consider each position to be independent, we can assign a relative score to all possible pentapeptide sequences as the product of the relative substrate specificities at each position (see below).

The matrix may be useful as a first approximation to estimate the specificity of TPP I for different substrates. It should be noted, however, that analyses of the P2 and P2’ libraries and P1 library II revealed that there are peptides that exhibit synergistic or deleterious interactions between residues at different positions. For instance, the beneficial effects that we found for positively charged P3 residues may be related to the fact that these residues are located at the N terminus of the peptides. Thus, for P3 residues, the positive charge of the N terminus, coupled with the positive charge of the side chain, may enhance binding to a negatively charged subsite on TPP I. However, this may be complicated by charge repulsion or electrostatic steering effects that occur when other positively charged residues are present on the peptide substrate (see Fig. 2). Thus, a more sophisticated approach, including consideration of structural and electrostatic properties of both substrate and enzyme, may be required to predict TPP I substrate specificity.

A homology model of TPP I has been proposed based on the crystal structures of the related bacterial proteases (37). Unfortunately, the relatively low sequence identity between the mammalian and prokaryotic enzymes (25%) has limited the analysis to the prediction that the binding pocket is relatively open and that TPP I prefers substrates with bulky hydrophobic groups in the P1 position. No insights could be drawn regarding the N-terminal P3 residue or any of the P’ residues. Further analysis would benefit from detailed knowledge of the three-dimensional structure of TPP I. At that time, the actual substrate specificity constants for individual sequences listed in supplemental Tables 1–6 should provide useful training and testing sets for extended computational analysis.

It is worth attempting to relate our kinetic measurements to the function of TPP I in vivo. The lifetime of proteins in the lysosome varies considerably, from that of long lived resident lysosomal enzymes (hours to many days) to rapidly hydrolyzed substrates that enter the lysosome by the endocytic or autophagic pathways. Although it is difficult to predict what rates of hydrolysis would be biologically relevant, it seems likely that true substrates would be hydrolyzed in the seconds to hours.

### TABLE 3

Analysis of different mouse tissues using Ala-Ala-Phe-AMC and Arg-Nle-Nle-ACC substrates

| Substrate                | Genotype | Brain | Kidney | Liver | Heart | Spleen |
|--------------------------|----------|-------|--------|-------|-------|--------|
| Ala-Ala-Phe-AMC          | CLN2(+/+)| 11 ± 1| 277 ± 56| 331 ± 33| 25 ± 4| 3,314 ± 507|
|                          | CLN2(−/−)| 2,015 ± 36| 15,896 ± 743| 10,292 ± 617| 487 ± 15| 16,973 ± 66|
| Knock out/wild type      | 0.5%     | 1.7%  | 3%     | 5.2%  | 19.5% |
| Arg-Nle-Nle-ACC          | CLN2(+/+)| 15 ± 2| 371 ± 5 | 118 ± 3 | 28 ± 0| 621 ± 507|
|                          | CLN2(−/−)| 6,188 ± 161| 50,589 ± 3,028| 31,080 ± 3,373| 1,471 ± 52| 52,742 ± 322|
| Knock out/wild type      | 0.2%     | 0.7%  | 0.4%   | 1.9%  | 1.2%  |

*Data represent the mean ± S.D. of two animals for each genotype. Activity was measured for 60 min at 30 °C and at pH 4.5 by using substrate concentrations of either 200 μM for Ala-Ala-Phe-AMC or 100 μM for Arg-Nle-Nle-ACC. Similar results were found when activity was measured at pH 4.0 or 5.0 (data not shown).
Substrate Specificity of Tripeptidyl-peptidase I

### TABLE 4

| Substrate Specificity Matrix | P3<sup>a</sup> | P2<sup>a</sup> | P1<sup>a</sup> | P1<sup>b</sup> | P2<sup>c</sup> |
|-----------------------------|---------------|---------------|---------------|---------------|---------------|
| Arg                        | 1.00          | Nle           | 1.00          | Phe           | 1.00          |
| His                        | 8.63E-01      | Pro           | 5.19E-01      | Phe           | 9.01E-01      |
| Tyr                        | 4.74E-01      | Ala           | 5.10E-01      | Nle           | 8.85E-01      |
| Lys                        | 3.51E-01      | Phe           | 2.89E-01      | Tyr           | 2.14E-01      |
| Glu                        | 2.70E-01      | Tyr           | 2.72E-01      | Trp           | 1.04E-01      |
| Phe                        | 1.71E-01      | Trp           | 1.94E-01      | Ala           | 4.66E-02      |
| Ile                        | 1.36E-01      | Ile           | 1.80E-01      | Asp           | 2.31E-02      |
| Ala                        | 1.14E-01      | Val           | 1.79E-01      | Gln           | 1.91E-02      |
| Ser                        | 1.08E-01      | Thr           | 1.33E-01      | Ala           | 1.52E-02      |
| Glu                        | 9.12E-02      | Glu           | 1.32E-01      | Asn           | 5.54E-03      |
| Thr                        | 9.10E-02      | Leu           | 1.00E-01      | Arg           | 3.15E-03      |
| Val                        | 8.60E-02      | Ser           | 9.55E-02      | Lys           | 2.79E-03      |
| Asn                        | 8.09E-02      | Gly           | 6.78E-02      | Ser           | 2.30E-03      |
| Nle                        | 6.67E-02      | Gln           | 6.48E-02      | His           | 2.00E-03      |
| Trp                        | 5.84E-02      | Lys           | 2.55E-02      | Val           | 1.48E-03      |
| Leu                        | 5.31E-02      | Asn            | 2.04E-02      | Ile           | 6.97E-04      |
| Gly                        | 4.29E-02      | Asp            | 1.52E-02      | Thr           | 6.47E-04      |
| Asp                        | 3.06E-02      | Arg            | 9.40E-03      | Gly           | 1.02E-04      |
| Pro                        | 4.87E-03      | His            | 8.56E-03      | Pro           | 1.35E-05      |

<sup>a</sup> The substrate specificity of each peptide in 31 Xaa-P2-P1-ACC pools that were analyzed using the LC/MS/MS assay was normalized to the corresponding Arg-P2-P1-ACC peptide within the pool. Substrate specificity constants that were below the limits of detection were censored. The relative substrate specificity listed represents the mean normalized value for the indicated P3 residue.

<sup>b</sup> Data from the fluorescence analysis of the P2 library (direct method) were used. For each of the four fixed P1 positions, data were normalized using the value of the P2 Nle peptide pool. The four values for each P2 residue were then averaged.

<sup>c</sup> Substrate specificities determined for P1 library II (supplemental Table 2) were normalized to the peptide with the highest substrate specificity constant and the concentration of enzyme by the expression $\frac{k_{cat}}{K_m} = \frac{n_{max}}{n_{max}^{optimal}}$. 

![FIGURE 12. Distribution of relative substrate specificities at different positions](image)

The plot shows the log value for each residue calculated as described in Table 4.

![FIGURE 13. Histogram of the relative scores for TPP I digestion of pentapeptides.](image)

The optimal P3-P2-P1-P1-P2<sup>b</sup> sequence Arg-Nle-Leu-Phe-Val has assigned a score of 1, and the poorest sequence Pro-His-Pro-Pro-Pro has a score of $1.8 \times 10^{-7}$. The sequences for all peptides are distributed as a log normal distribution having a mean value of $6.2 \times 10^{-7}$. The data were transformed into log units, and a frequency histogram was constructed using an interval (bin) size of 0.1. Scores corresponding to the indicated half-lives (dotted lines) were calculated assuming a concentration of 500 nM TPP I in the lysosome and a value of $30 \mu M^{-1} s^{-1}$ for the optimum substrate. Although Nle has been used in generation of the histogram, and can be taken as an analog for Met, the plot does not reflect data from the NCLs.

One goal of this study was to provide information with which to interpret future studies directed at identifying substrates that accumulate in LINC. Accumulation of peptides and proteins is likely to reflect a number of processes, including intrinsic susceptibility to TPP I (which will include both primary and higher order substrate structure considerations), accessibility/susceptibility to other lysosomal proteases that could potentially compensate for the defective TPP I, and the rate at which unhydrolyzed substrates aggregate to form protease-resistant storage bodies.

Although little is known regarding the substrates of TPP I in vivo, mitochondrial ATP synthase subunit c is known to accumulate in LINC (17, 39). The specific pathophysiological processes that both lead to and result from this accumulation are not well understood, and this hydrophobic membrane protein has been identified as a prominent component in the storage material in many other NCLs (19). If subunit c is directly degraded by TPP I, the results presented here indicate that it would represent only an average substrate, with a predicted relative substrate specificity constant of $1.8 \times 10^{-5}$ based on its N-terminal sequence D11TA (Table 4; $1.8 \times 10^{-5} = 0.031 \times 0.18 \times 0.023 \times 0.26 \times 0.56$). Given the assumptions mentioned above, subunit c would be predicted to have a half-life of ~40 min in the lysosome (Fig. 13).
Substrate Specificity of Tripeptidyl-peptidase I

1. Goebel, H. H., Mole, S. E., and Lake, B. D. (eds) (1999) The Neuronal Cardin Lipoprotein (Batten Disease): Biomedical and Health Research, IOS Press, Amsterdam
2. Palmer, D. N., Barns, G., Hushands, D. R., and Jolly, R. D. (1986) J. Biol. Chem. 261, 1773–1777
3. Sleat, D. E., Donnelly, R. J., Lackland, H., Liu, C. G., Sohar, I., Pullarkat, R. K., and Lobel, P. (1997) Science 277, 1802–1805
4. Sohar, I., Sleat, D. E., Jadot, M., and Lobel, P. (1999) J. Neurochem. 73, 700–711
5. Rawlings, N. D., and Barrett, A. J. (1999) Biochim. Biophys. Acta 1429, 496–500
6. Vines, D. J., and Warburton, M. J. (1999) FEBS Lett. 443, 131–135
7. Lin, L., Sohar, I., Lackland, H., and Lobel, P. (2001) J. Biol. Chem. 276, 2249–2255
8. Comellas-Bigler, M., Fuentes-Prior, P., Maskos, K., Huber, R., Oyama, H., Uchida, K., Dunn, B. M., Oda, K., and Bode, W. (2002) Structure (Camb) 10, 865–876
9. Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., Dunn, B. M., and Oda, K. (2001) Nat. Struct. Biol. 8, 442–446
10. Ezaki, J., Takeda-Ezaki, M., Oda, K., and Kominami, E. (2000) Biochem. Biophys. Res. Commun. 268, 904–908
11. Golabek, A. A., Wujek, P., Walus, M., Bieler, S., Soto, C., Wisniewski, K. E., and Kida, E. (2004) J. Biol. Chem. 279, 31058–31067
12. Berry-Kravis, E., Sleat, D. E., Sohar, I., Meyer, P., Donnelly, R., and Lobel, P. (2000) Ann. Neurol. 47, 254–257
13. Sohar, I., Lin, L., and Lobel, P. (2000) Clin. Chem. 46, 1005–1008
14. Junaid, M. A., Brooks, S. S., and Pullarkat, B. K. (2001) Eur. J. Paediatr. Neurol. 5, Suppl. A, 63–68
15. Kleijer, W. J., van Diggelen, O. P., Kreulmans, J. L., Losekoot, M., Garrirten, Y. H., Stroink, H., Majoer-Krakauer, D., Franken, P. F., Eurlings, M. C., Taschner, P. E., Los, F. J., and Gaaljaard, R. J. (2001) Prenat. Diagn. 21, 99–101
16. Lukacs, Z., Santavuori, P., Keil, A., Steinfeld, R., and Kohlschutter, A. (2003) Clin. Chem. 49, 509–511
17. Palmer, D. N., Fearnley, I. M., Walker, J. E., Hall, N. A., Lake, B. D., Wolfe, L. S., Haltia, M., Martinus, R. D., and Jolly, R. D. (1992) Am. J. Med. Genet. 42, 561–567
18. Palmer, D. N., Martinus, R. D., Cooper, S. M., Midwinter, G. G., Reid, J. C., and Jolly, R. D. (1989) J. Biol. Chem. 264, 5736–5740
19. Palmer, D. N., Jolly, R. D., van Mil, H. C., Tyynela, J., and Westlake, V. J. (1997) Neuropeptides 25, 109–119
20. Fearnley, I. M., Walker, J. E., Martinus, R. D., Jolly, R. D., Kirkland, K. B., Shaw, G. J., and Palmer, D. N. (1990) Biochem. J. 268, 751–758
21. Tyynela, J., Suopanki, J., Santavuori, P., Baumann, M., and Haltia, M. (1997) J. Neuropathol. Exp. Neurol. 56, 369–375
22. Herva, R., Tyynela, J., Hirvasniemi, A., Syrjakallio-Ylitalo, M., and Haltia, M. (2000) Brain Pathol. 10, 215–222
23. Sohar, I., Sleat, D. E., and Lobel, P. (2003) in Handbook of Proteolytic Enzymes, (Barrett, A. J., Rawlings, N. D., and Woessner, J. F., eds) 2nd Ed., pp. 1893–1896, Elsevier Academic Press, London
24. McDonald, J. K., Hosington, A. R., and Eisenhauer, D. A. (1985) Biochem. Biophys. Res. Commun. 126, 63–71
25. Doebber, T. W., Divor, A. R., and Ellis, S. (1978) Anal. Biochem. 86, 561–567
26. Doebber, T. W., Divor, A. R., and Ellis, S. (1978) Anal. Biochem. 86, 561–567
27. Watanabe, Y., Kuyama, Y., and Fujimoto, Y. (1992) Biochem. Biophys. Res. Comm. 179, 1749–1755
28. Warburton, M. J., and Bernardini, F. (2001) FEBS Lett. 500, 145–148
29. Sleat, D. E., Sohar, I., Lin, R. M., and Lobel, P. (2001) Eur. J. Paediatr. Neurol. 5, Suppl. A, 57–62
30. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 7754–7759
31. Maly, D. J., Leonetti, F., Backes, B. J., Dauber, D. S., Harris, J. L., Craik, C. S., and Ellman, J. A. (2002) J. Org. Chem. 67, 910–915
32. Ostresh, J. M., Winkle, J. H., Hamashuchi, T. V., and Houghten, R. A. (1994) Biopolymers 34, 1681–1689
33. McGehee, G. M., Bickett, D. M., Wiseman, J. S., Green, M., and Berman, J. (1995) Methods Enzymol. 248, 35–46
34. Kaiser, E., Colless, R. L., Bossinger, C. D., and Cook, P. I. (1970) Anal. Biochem. 30, 595–598
35. Tian, Y., Ramesh, C. V., Ma, X., Naqui, S., Patel, T., Cenizal, T., Tiscione, M., Diaz, K., Crea, T., Arnold, E., Arnold, G. F., and Taylor, J. W. (2002) J. Pept. Res. 59, 346–276
36. Sohar, I., Wiseman, I. A., El-Banna, M., Kim, K. H., Mao, Q., Price, S., Macaulay, S. L., Sidman, R. L., Shen, M. M., Zhao, Q., Passini, M. A., Davidson, B. L., Steward, G. R., and Lobel, P. (2004) J. Neurosci. 24, 9117–9126
37. Oyama, H., Fujisawa, T., Suzuki, T., Dunn, B. M., Wlodawer, A., and Oda, K. (2005) J. Biochem. (Tokyo) 138, 127–134
38. Wlodawer, A., Durell, S. R., Li, M., Oyama, H., Oda, K., and Dunn, B. M. (2003) BMC Struct. Biol. 3, 9
39. Holtzman, E. (1989) Lysosomes, pp. 73–74, Plenum Publishing Corp., New York
40. Ezaki, J., Takeda-Ezaki, M., and Kominami, E. (2000) J. Biochem. (Tokyo) 128, 509–516
41. Mitchison, H. M., Bernard, D. J., Greene, N. D., Cooper, J. D., Junaid, M. A., Pullarkat, R. K., de Vos, N., Breuning, M. H., Owens, J. W., Mobley, W. C., Gardiner, R. M., Lake,
Substrate Specificity of Tripeptidyl-peptidase I

B. D., Taschner, P. E., and Nussbaum, R. L. (1999) Neurobiol. Dis. 6, 321–334
41. Sleat, D. E., Sohar, I., Pullarkat, P. S., Lobel, P., and Pullarkat, R. K. (1998) Biochem. J. 334, 547–551
42. Kopan, S., Sivasubramaniam, U., and Warburton, M. J. (2004) Biochem. Biophys. Res. Commun. 319, 58–65
43. Warburton, M. J., and Bernardini, F. (2002) Neurosci. Lett. 331, 99–102
44. Bernardini, F., and Warburton, M. J. (2002) Biochem. J. 366, 521–529
45. Junaid, M. A., Sklower Brooks, S., Wisniewski, K. E., and Pullarkat, R. K. (1999) Clin. Chim. Acta 281, 169–176
46. Lin, L., and Lobel, P. (2001) Hum. Mutat. 18, 165
47. Sleat, D. E., Gin, R. M., Sohar, I., Wisniewski, K., Sklower-Brooks, S., Pullarkat, R. K., Palmer, D. N., Lerner, T. J., Boustany, R. M., Uldall, P., Siakotos, A. N., Donnelly, R. J., and Lobel, P. (1999) Am. J. Hum. Genet. 64, 1511–1523
48. Schechter, I., and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162