THE PROTEIN SOCIETY

Human enteropeptidase light chain: Bioengineering of recombinants and kinetic investigations of structure and function

Eliot T. Smith and David A. Johnson*

Department of Biomedical Sciences, James H. Quillen College of Medicine, East Tennessee State University, Johnson City, Tennessee

Received 17 January 2013; Revised 15 February 2013; Accepted 16 February 2013
DOI: 10.1002/pro.2239
Published online 25 February 2013 proteinscience.org

Abstract: The serine protease enteropeptidase exhibits a high level of substrate specificity for the cleavage sequence DDDDK~ X, making this enzyme a useful tool for the separation of recombinant protein fusion domains. In an effort to improve the utility of enteropeptidase for processing fusion proteins and to better understand its structure and function, two substitution variants of human enteropeptidase, designated R96Q and Y174R, were created and produced as active (>92%) enzymes secreted by *Pichia pastoris* with yields in excess of 1.7 mg/Liter. The Y174R variant showed improved specificities for substrates containing the sequences DDDDK ($k_{cat}/K_M = 6.83 \times 10^6\ M^{-1}\ sec^{-1}$) and DDDDR ($k_{cat}/K_M = 1.89 \times 10^7\ M^{-1}\ sec^{-1}$) relative to all other enteropeptidase variants reported to date. BPTI inhibition of Y174R was significantly decreased. Kinetic data demonstrate the important contribution of the positively charged residue 96 to extended substrate specificity in human enteropeptidase. Modeling shows the importance of the charge–charge interactions in the extended substrate binding pocket.

Keywords: enteropeptidase; enterokinase; recombinant protein; kinetics; extended peptide substrates; *Pichia pastoris*

Introduction

Enteropeptidase (EC 3.4.21.9), originally named enterokinase, is a trypsin-like serine protease that activates pancreatic digestive enzymes in the duodenum.[1–7] The proteolytically active light chain has high selectivity for the sequence DDDDK (D_4K).[8,9] Physiologically, enteropeptidase removes the N-terminal pro-peptide of trypsinogen by cleaving the sequence $D_4K\sim$ IVGG to activate trypsin. Genetic deficiencies of enteropeptidase have been reported in two families with symptoms of intestinal malabsorption, diarrhea, and growth failure, revealing the importance of human enteropeptidase (hEP) in the digestive process.[10,11] Inhibitors of enteropeptidase have been designed and tested in a murine model to evaluate their efficacy as anti-obesity drugs.[12]

Although some nonspecific cleavages have been reported for this enzyme,[6,13–16] the high selectivity

Abbreviations: bEP, bovine enteropeptidase; bEP_L, bovine enteropeptidase light chain; BPTI, basic pancreatic trypsin inhibitor; CCP, cleavage control protein; DTNB, 5,5′-dithiobis-(2-nitrobenzoic acid); EP_L, enteropeptidase light chain; GD_4K-na, Gly-Asp_4-Lys-napthylamide; GD_4K-pNA, Gly-Asp_4-Lys-p-nitroanilide; GD_4R-pNA, Gly-Asp_4-Arg-p-nitroanilide; hEP, human enteropeptidase; hEP_L, human enteropeptidase light chain; MUGB, 4-methylumbelliferyl-p-guanidinobenzoate; $rbEP_L$, recombinant bovine enteropeptidase light chain; $rhEP_L$, recombinant human enteropeptidase light chain; rhEPLsc, supercharged recombinant human enteropeptidase light chain C112S; STI, soybean trypsin inhibitor; Z-Lys-SBzl, N-α-carbobenzyloxy-L-lysine thiobenzyl ester.

Additional Supporting Information may be found in the online version of this article.

The authors have no competing interests affecting the objectivity or integrity of the publication.

Grant sponsor: NIH; Grant number: R15HL091770.

*Correspondence to: David A. Johnson, Box 70582, Johnson City, TN 37614. E-mail: davidj@etsu.edu

of enteropeptidase light chain (EP_L) for the sequence $D_4K\sim$ X makes it useful as a biotechnological tool for site-specific separation of recombinant fusion protein domains.[17,18] Recombinant bovine enteropeptidase light chain ($rbEP_L$) has been expressed in *P. pastoris*,[19,20] *Saccharomyces cerevisiae*,[21] and *Escherichia coli*.[18,22,23] The monomeric $rbEP_L$ from *E. coli* proved superior to the native heterodimeric bEP at cleaving fusion proteins.[18] Human EP_L has greater catalytic efficiency compared to bovine EP_L for $D_4R\sim$ X and $D_4K\sim$ X substrates[24,25] and recombinant hEP light chain ($rhEP_L$) was shown to have a 10-fold higher specificity constant (k_{cat}/K_M) for the $D_4K\sim$ X sequence than bEP_L.[26]

The x-ray crystal structure (PDB code: 1EKB) for bEP_L in complex with the inhibitor VD_4K-chloromethane was used in conjunction with $rbEP_L$ variants from a baculovirus expression system to reveal the critical role of K99 (chymotrypsinogen numbering system) in the S2-S4 (Schechter-Berger system)[27] extended binding site that interacts with Asp P2-P4 substrate residues.[27,28] A recent study used the human-to-bovine substitutions R96K and K219Q to shift the catalytic properties of $rhEP_L$ to more closely parallel bEP_L, demonstrating additive contributions of positive residues R96 and K219 to improved extended specificity.[29] They constructed *in silico* models of hEP_L based on the bEP_L x-ray structure[28] and proposed that R96 may shield the P2 residue from the solvent better than K96 and that K219 may form salt bridges with Asp-P3 and Asp-P5.[29] Others created, in *E. coli*, a Y174R variant of bEP_L that showed an approximate four-fold improvement in specificity toward the $GD_4K\sim$ IVGG substrate as a result of the proposed salt bridges by R174 with Asp-P3 and Asp-P4.[22] Recently, a supercharged $rhEP_L$ ($rhEP_L$sc) was created in *E. coli* to improve enzyme solubility and refolding yields[30,31] and an x-ray crystal structure of this mutant was obtained (PDB code: 4DGJ).[28,30,31]

In an effort to improve the utility of $rhEP_L$ for processing fusion proteins and to better understand the structure and function of hEP_L, two $rhEP_L$ variants were created and produced as active enzymes secreted by *P. pastoris*. Kinetic analyses employed synthetic D_4K and D_4R p-nitroanilde (D_4K-pNA and D_4R-pNA) substrates and the N-α-carbobenzyloxy-L-lysine-thiobenzyl ester (Z-Lys-SBzl) substrate for comparison with other reports. Potential Asn-linked glycosylation sites were mutated to Gln to yield non-glycosylated proteins. In one variant, arginine was replaced with glutamine at position 96 to create R96Q; furthermore, the kinetic properties of this variant in comparison with others suggests that the human R96 residue influences extended substrate affinity through charge-dependent interaction with Asp-P2. In the second variant (Y174R), tyrosine at residue 174 was replaced with arginine, which has been shown to improve the specificity in $rbEP_L$.[22] The human Y174R enzyme shows the highest specificity for the Asp_4 motif of any EP_L variant reported to date.

Results

Protein engineering and molecular cloning

The amino acid sequences for the variants R96Q and Y174R were designed based on the sequence of native hEP_L (Fig. 1). As a result of the tendency of *P. pastoris* to hyperglycosylate secreted proteins,[32–36] the conservative amino acid substitutions N75Q, N113Q, N135Q, and N175Q were used to disrupt four potential Asn-linked glycosylation sites in both R96Q and Y174R (Fig. 1). The Furin-like protease Kexin (Kex2) of the Golgi apparatus in *P. pastoris* can cleave dibasic sequences such as KR∼ X.[37] The substitution R98Q was incorporated into both R96Q and Y174R variants to eliminate a potential internal Kexin cleavage site (Fig. 1), supported by data that an R98A substitution does not reduce the activity of bEP_L.[28]

DNA coding for the human R96Q and Y174R mutants was commercially codon-optimized for expression in *P. pastoris,* synthesized, and cloned into the pPICzα A plasmid for secreted expression. Plasmids encoding R96Q or Y174R were used to transform the X-33 strain of *P. pastoris* and the most productive clones were identified in a screening protocol. During secretion, the Kexin protease removes α-mating factor via cleavage of the sequence KR∼ IVGG, resulting in secretion of active R96Q or

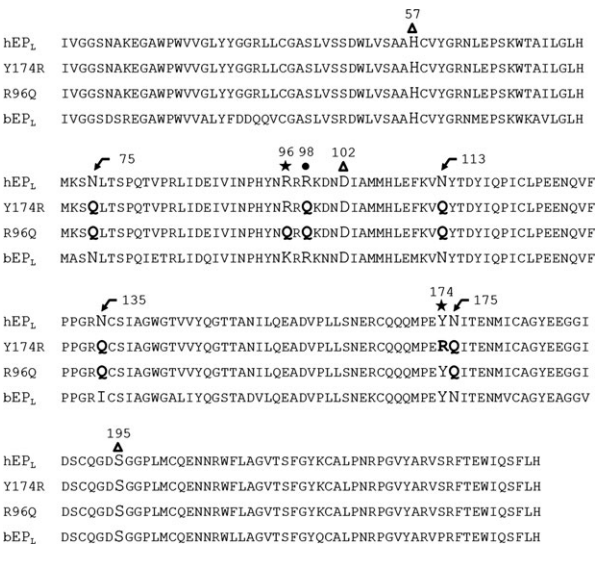

Figure 1. Sequence alignment of human and bovine EP_L with human R96Q and Y174R variants. All engineered substitutions appear in bold and residues 96 and 174 are labeled with stars. Asn-to-Gln substitutions (bent arrows) disrupt four potential Asn-linked glycosylation sites. The substitution R98Q (dot) eliminates one possible Kexin cleavage site. Serine protease active site residues (triangles) are also indicated.

Table I. *Kinetic Parameters of rhEP$_L$ Variants*

	K_M (M)	k_{cat} (sec^{-1})	k_{cat}/K_M (M^{-1} sec^{-1})
Z-Lys-SBzl			
Tag*off^{TM}	$1.90 \times 10^{-4} \pm 1.90 \times 10^{-5}$	(336 ± 11)*	$(1.77 \times 10^6 \pm 1.86 \times 10^5)$*
R96Q	$1.65 \times 10^{-4} \pm 1.05 \times 10^{-5}$	354 ± 11	$2.15 \times 10^6 \pm 8.20 \times 10^4$
Y174R	$1.72 \times 10^{-4} \pm 1.20 \times 10^{-5}$	319 ± 11	$1.86 \times 10^6 \pm 1.43 \times 10^5$
GD_4K-pNA			
Tag*off^{TM}	$2.98 \times 10^{-5} \pm 3.98 \times 10^{-6}$	(135 ± 4)*	$(4.53 \times 10^6 \pm 3.62 \times 10^5)$*
R96Q	$2.76 \times 10^{-4} \pm 3.11 \times 10^{-5}$	148 ± 7.4	$5.37 \times 10^5 \pm 3.40 \times 10^4$
Y174R	$1.79 \times 10^{-5} \pm 2.07 \times 10^{-6}$	122 ± 3.5	$6.83 \times 10^6 \pm 6.59 \times 10^5$
GD_4R-pNA			
Tag*off^{TM}	$6.29 \times 10^{-6} \pm 7.89 \times 10^{-7}$	(45 ± 2)*	$(7.20 \times 10^6 \pm 1.34 \times 10^6)$*
R96Q	$1.80 \times 10^{-5} \pm 2.23 \times 10^{-6}$	42 ± 2.4	$2.33 \times 10^6 \pm 3.93 \times 10^5$
Y174R	$2.58 \times 10^{-6} \pm 5.36 \times 10^{-7}$	49 ± 1.6	$1.89 \times 10^7 \pm 3.86 \times 10^6$

Values are expressed as mean ± S.D. of three independent replicates.
Assays conducted in 100 mM Tris, 0.02 mM $CaCl_2$, 0.01% Triton X-100, 10% DMSO, pH = 8.0.
* Active site concentrations of Tag*off^{TM} could not be determined by titration because of the limited supply. The parenthetical k_{cat} values are taken from the average of the k_{cat} for R96Q and Y174.

Y174R variants. Approximately 1.7 mg/L of active R96Q and 2.2 mg/L of active Y174R, as estimated by Z-Lys-SBzl activity, were expressed in 5 L batches of bioreactor fermentation. Untransformed X-33 *P. pastoris* do not secrete enzymatic activity for the Z-Lys-SBzl substrate.

Purification

Both enzymes were recovered from culture media by affinity chromatography using the reversible, Kunitz-type protease inhibitor, soybean trypsin inhibitor (STI). The R96Q and Y174R variants were purified 1352-fold and 969-fold, respectively, from fermentation media (Supporting Information Table SI). Although the total protein recoveries of purified R96Q and Y174R were similar, the total and specific activities were unexpectedly low for Y174R (Supporting Information Table SI). Assays during purification were in the absence of $CaCl_2$ or Triton X-100, which were subsequently found to enhance the activity of Y174R consistent with a report on solubility problems with recombinant hEP.[30] Subsequent kinetic analyses including 0.02 mM $CaCl_2$ and 0.01% Triton X-100 show that both variants have similar K_M and k_{cat} values for Z-Lys-SBzl (Table I), suggesting that Y174R is less soluble than R96Q. The single-step STI affinity purification yielded highly purified enzymes [Fig. 2(A)] and the A280-based concentration estimates were confirmed to within 8% by active site titration.[38]

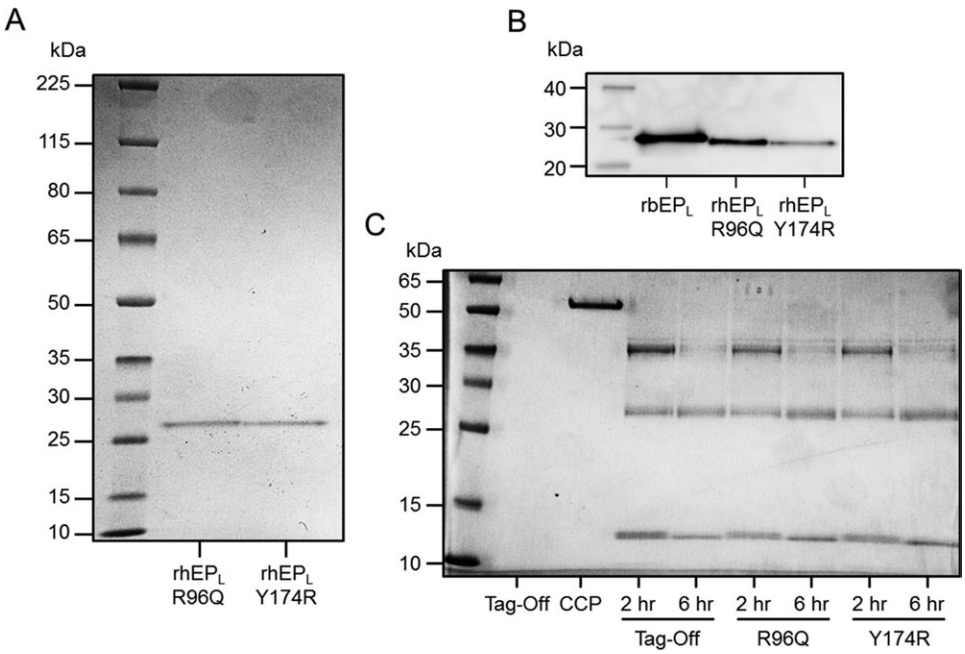

Figure 2. Molecular characterization of purified enzymes. (A) SDS-PAGE shows high purity and correct mass (~ 26 kDa) for nonglycoslyated enteropeptidase variants R96Q and Y174R. (B) Monoclonal antibody against bEP$_L$ also labels human variants R96Q and Y174R by Western blot. (C) Cleavage assay shows that R96Q and Y174R variants cleave the CCP like Tag*off^{TM}.

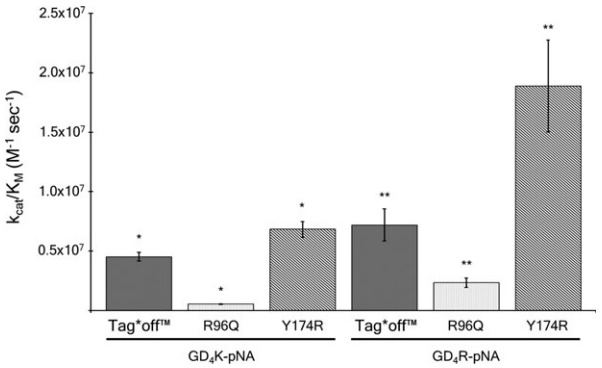

Figure 3. Specificity constants of Tag*off™, R96Q, and Y174R for the extended substrates GD_4K-pNA and GD_4R-pNA. All values are mean average ± standard deviation, as listed in Table I. $P < 0.005$ for all * and ** comparisons.

Identity confirmation and stability analyses

The purified recombinant enzymes were analyzed by SDS-PAGE [Fig. 2(A)], confirming high levels of purity at the predicted molecular weight of 26 kDa for non-glycosylated $rhEP_L$. Western blot analysis [Fig. 2(B)] further confirms the identities of the $rhEP_L$ variants using a monoclonal antibody against bEP_L with nonglycosylated $rbEP_L$ as a positive control. Both substitution variants selectively cut an enteropeptidase cleavage control protein (CCP; EMD Millipore) with the same cleavage pattern as observed for Tag*off™ positive control enzyme [Fig. 2(C)]. Tag*off™ is a nonglycosylated recombinant hEP_L (EMD Millipore).

Over the course of five days, Tag*off™, R96Q, and Y174R were evaluated for stability at 37°C in 100 m*M* Tris, 0.02 m*M* $CaCl_2$, 0.01% Triton X-100, pH = 8.0. Tag*off™ and R96Q had half-lives longer than one week (16 days and 9 days, respectively); however, the Y174R substitution reduced the half-life to 48 hours (Supporting Information Fig. SI). Residue R174 may provide a target for EP_L autolysis. All of the enzymes showed long-term stability in low pH storage buffers at +4°C and −20°C.

Kinetic analyses

The Michaelis Constant K_M, the catalytic rate constant k_{cat}, and the specificity constant k_{cat}/K_M were evaluated for Tag*off™, R96Q, and Y174R using the synthetic peptide substrates Z-Lys-SBzl, GD_4K-pNA, and GD_4R-pNA (Table I). These assays were conducted in 100 m*M* Tris, 0.02 m*M* $CaCl_2$, 0.01% Triton X-100, 10% DMSO, pH = 8.0. $CaCl_2$ enhances EP_L activity at the 0.02 m*M* concentration.[26] NaCl was excluded because it has been shown to reduce EP_L activity for GD_4K-na.[26] As a result of the limited supply of Tag*off™, the stock concentration of active Tag*off™ could not be determined by active site titration; consequently, all Tag*off™ k_{cat} values reported here are estimated based on the averages of the values for the R96Q and Y174R variants, which are very similar (Table I); furthermore, the estimate reported here for Tag*off™ with GD4K-pNA is similar to the k_{cat} reported elsewhere for hEP_L with the substrate GD_4K-na.[26]

For the single residue substrate Z-Lys-SBzl, no statistically significant differences were observed among the K_M values; likewise, the k_{cat} value for R96Q is only 1.1 times that for Y174R ($P < 0.05$) (Table I). The comparison of k_{cat}/K_M ratios reveals that R96Q cleaves Z-Lys-SBzl about 1.2 times as efficiently as either Tag*off™ ($P < 0.05$, estimated) or Y174R ($P < 0.05$) (Table I). The Y174R variant and Tag*off™ do not have statistically different k_{cat}/K_M values.

For the substrate GD_4K-pNA, the enzymes behave very differently. The Y174R variant shows the smallest K_M for this substrate (Table I). The K_M for Tag*off™ is 1.6 times larger than that for Y174R ($P < 0.01$) (Table I). With the weakest affinity for GD_4K-pNA, R96Q has a K_M that is 9.3-fold greater than Tag*off™ ($P < 0.0005$) and 15.5-fold greater than the Y174R variant ($P < 0.0005$) (Table I). The k_{cat} of R96Q was 1.2-fold greater than that of Y174R ($P < 0.005$) (Table I). The single amino acid substitution Y174R improved the catalytic efficiency for the GD_4K-pNA substrate by 1.5-fold compared to Tag*off™ ($P < 0.01$, estimated) (Table I, Fig. 3). The specificity of R96Q for D_4K-pNA showed only 12% of the value for Tag*off™ ($P < 0.0001$, estimated) and only 8% of the Y174R result ($P < 0.0001$) (Table I, Fig. 3). Compared to 100 m*M* Tris-HCl, pH = 8.0, the addition of 0.02 m*M* $CaCl_2$ and 0.01% Triton X-100 improved the specificity constants of R96Q and Y174R by six times and 20 times, respectively, for GD_4K-pNA.

K_M values for GD_4R-pNA were lowest for Y174R, followed by Tag*off™ at 2.4 times Y174R ($P < 0.005$) and then R96Q at 6.9 times Y174R ($P < 0.001$) (Table I). As with the other two synthetic substrates, the k_{cat} values of R96Q and Y174R are very similar. The k_{cat}/K_M specificity constant of Y174R for the GD_4R-pNA substrate is 2.6 times that of Tag*off™ ($P < 0.005$, estimated) and 8 fold that of R96Q ($P < 0.005$) (Table I, Fig. 3). The k_{cat}/K_M for Tag*off™ is three times that of R96Q ($P < 0.005$, estimated) (Table I, Fig. 3). Compared to 100 m*M* Tris-HCl, pH = 8.0, the addition of 0.02 m*M* CaCl2 and 0.01% Triton X-100 improved the specificity constants of R96Q and Y174R for GD_4R-pNA by 6.8 and 28 times, respectively.

Comparison of the relative efficiencies of each enzyme for the extended peptide synthetic substrates reveals that the GD_4R-pNA sequence is preferred over GD_4K-pNA by all variants, especially for Y174R, by 2.8 times ($P < 0.01$) (Table I, Fig. 3). The Tag*off™ enzyme only showed 1.3 times greater selectivity for GD_4R-pNA ($P < 0.005$) compared with GD_4K-pNA, although the K_M was reduced by 42% (Table I, Fig. 3). While R96Q had the lowest

Table II. *Inhibition of rhEP$_L$ by BPTI*

	K_i (M)
Tag*off^{TM}	$4.61 \times 10^{-8} \pm 1.66 \times 10^{-8}$
R96Q	$6.32 \times 10^{-8} \pm 1.84 \times 10^{-9}$
Y174R	$2.28 \times 10^{-7} \pm 7.02 \times 10^{-9}$

Values are expressed as mean ± SD of three independent replicates.
Assays conducted in 300 μM Z-Lys-SBzl, 100 mM Tris, 0.02 mM $CaCl_2$, 0.01% Triton X-100, 10% DMSO, 1 mM DTNB, pH = 8.0.

specificity constants with either of the extended peptide substrates, its preference for GD_4R-pNA was 4.3 times that of GD_4K-pNA ($P < 0.005$) (Table I, Fig. 3).

The substitution Y174R significantly decreases the binding affinity for the inhibitor BPTI (Table II). The inhibition equilibrium constant, K_i, for Y174R is four times larger ($P < 0.0001$) than for Tag*off^{TM} and 3.6 times larger ($P < 0.0001$) than for R96Q (Table II). The R96Q substitution did not significantly change the K_i from that of Tag*off^{TM} (Table II).

Discussion

The extended substrate binding pocket of EP_L can be thought of as a five-walled cage where residues 96, 99, 174, and 219 form walls 1–4 (Fig. 4), with the active site (S195) and the S1 specificity pocket

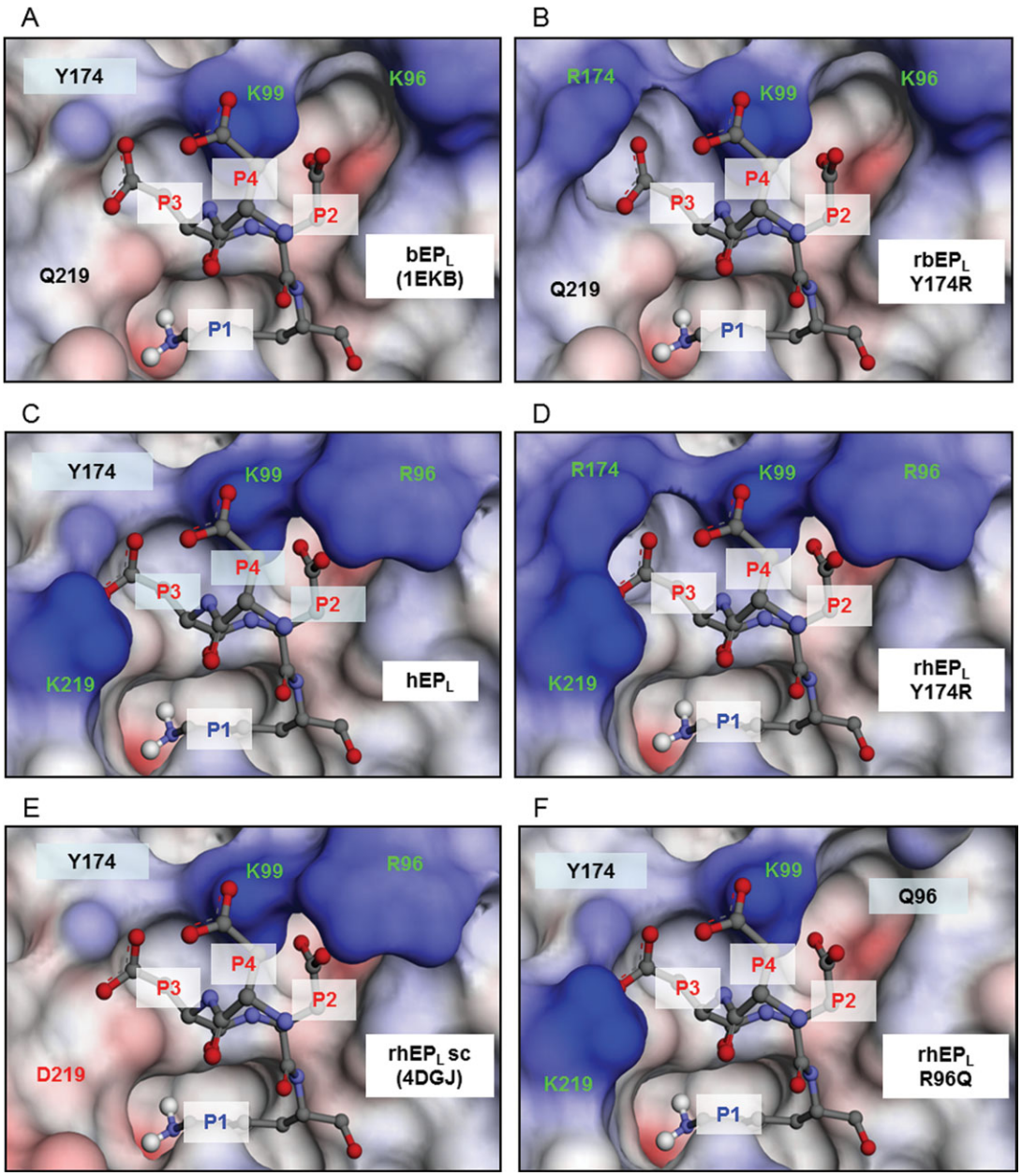

Figure 4. Proposed electrostatic models of EP_L binding pockets with bound D_3K substrate. Positively charged (blue) residues (green labels) improve binding affinity for Asp P2-P4. Negative charges are red. (A) Bovine EP_L (PDB Code: 1EKB). (B) Bovine EP_L with Y174R substitution. (C) Human EP_L (by the substitution of PDB model 4DGJ). (D) Human EP_L with Y174R substitution. (E) Supercharged human EP_L (PDB Code: 4DGJ). (F) Human EP_L with R96Q substitution.

(D189) forming the floor. The open fifth wall and the open ceiling of the cage allow the substrate to enter. In all variants, "wall 2" (K99) and the "floor" are required for D_4K specificity.[28] Added specificity comes from the properties of the other walls. In bEP_L, "wall 1" is K96, "wall 3" is Y174, and "wall 4" is Q219 [Fig. 4(A)]. The positive charge of K96 has been shown to improve specificity;[28] however, Y174 contributes only a weak hydrogen bond between the tyrosine hydroxyl group and the carboxyl side chain of Asp-P3.[28] In the study of the *E. coli* variant $rbEP_L$ Y174R [Fig. 4(B)], the presence of the ε-guanidino group of R174 on "wall 3" of the cage was shown to improve extended substrate specificity and was proposed to form a salt bridge with Asp-P3.[22] According to the x-ray crystal structure of bEP_L [Fig. 4(A)], Q219 side chain does not interact with the substrate in any way, extending away from the binding site instead.

The improved specificity of hEP_L over bEP_L can be attributed in part to differences in two of the cage walls [Fig. 4(A and C)] and in part to species-specific differences in the stability of the active site and specificity pocket.[29] First, amino acid substitution by others has shown that R96 stabilizes Asp-P2 better than K96.[29] Second, another substitution by the same group has also shown that K219 improves the substrate specificity of $rhEP_L$ compared to Q219, especially in conjunction with R96.[29] The positively charged K219 side chain amino group may by attracted toward the binding site in the presence of Asp-P3 in hEP_L[29] to form "wall 4" [Fig. 4(C, D, and F)]; whereas Q219 and D219 side chains extend into the solvent in the crystal structures of bEP_L[28] [Fig. 4(A and B)] and $rhEP_Lsc$[34] [Fig. 4(E)].

The human Y174R variant presented here combines the improved specificity of the bovine Y174R substitution[22] with the positive influences of residues R96 and K219[29] in the human enzyme. The result is the novel human Y174R [Fig. 4(D)] with improved extended substrate specificity relative to previously reported rEP_L variants. This Y174R enzyme may prove advantageous for the separation of recombinant protein fusion constructs. Although Y174R has only a slightly better specificity than Tag*off^{TM} for the D_4K sequence, Y174R displays a very high specificity constant for the D_4R sequence (Table I, Fig. 3). This reduces nonspecific cleavage by Y174R in the presence of excess fusion protein if the D_4R linker is employed. The shorter enzyme half-life of Y174R may be because of the autolysis at R174. Although poor stability is typically undesirable in enzymes, the limited half-life of Y174R under catalytic conditions (pH = 8.0, 37°C) may actually contribute to a reduction of nonspecific substrate cleavage relative to native EP_L.

The R96Q variant generated during this project reveals that the absence of positive charge at position 96 reduces the catalytic efficiency by 94% relative to Tag*off^{TM} for D_4K and by 68% relative to Tag*off^{TM} for D_4R (Table I, Fig. 3). Others have shown that the K96A substitution in $rbEP_L$ reduces its specificity by approximately 80% of native bEP_L for GD_4K-na.[28] It has also been previously shown that the human-to-bovine substitution $rhEP_L$ R96K reduces its specificity by only about 6% relative to $rhEP_L$.[29] These data suggest that the positive charge of residue 96 contributes significantly to the substrate specificities of all EP_L variants, despite crystallographic evidence suggesting that K96 in bEP_L does not coordinate with Asp-P2.[28]

Examination of an x-ray crystal structure of the β-trypsin + BPTI complex (PDB Code: 3OTJ)[39] aligned with the structure of EP_L reveals that BPTI residue Arg39 lies in close proximity to EP_L residue 174. As a result, an electrostatic repulsion likely occurs, accounting for the approximately 5-fold increase in the K_i of Y174R for BPTI.

Nonglycosylated rhEP variants Y174R and R96Q were easily expressed into the growth media of *P. pastoris* as properly folded active enzymes. Both enzymes were highly purified by chromatography on immobilized STI. The Y174R variant displayed increased specificity for the GD_4R-pNA substrate relative to a commercial sample of rhEP (Tag*off^{TM}). The potential utility of a D_4R linker between fusion proteins[24,25] is thus amplified by the use of Y174R. The kinetic properties of R96Q demonstrate the importance of R96 to the binding of extended peptide substrates.

Materials and Methods

The hEP_L amino acid sequence was taken from NCBI Ref. Seq. NP_002763.2. DNAs were codon-optimized and synthesized by different companies; R96Q by GenScript and Y174R by DNA 2.0. The recombinant human EP_L named Tag*off^{TM} and a CCP were products of EMD Millipore.

Molecular cloning

The DNA sequences for R96Q and Y174R were cloned into the pPICzα A plasmid (Life Technologies). The linearized vectors were used to transform X-33 strain *P. pastoris* (Life Technologies) by electroporation. Transformants were selected on YPDS agar plates with 100 μg/mL Zeocin and 100 μg/mL Ampicillin. Selected colonies were grown in 12-well microtiter plates as 2 mL cultures in synthetic minimal medium (SMM) (10 g/L MSG, 2.5 g/L $(NH_4)_2SO_4$, 100 m*M* KH_2PO_4 at pH = 5.0, 13.4 g/L YNB, and 0.4 mg/L biotin) containing 0.5% methanol.[40] Cultures were grown for four days at room temperature with daily addition of 0.5% methanol. Culture supernatants were tested for enteropeptidase activity using the synthetic substrate

Z-Lys-SBzl and activities were normalized to wet cell pellet weights for selection of the best expressers.

Fermentation

Cultures were grown using the strategy in the Life Technologies *Pichia* Fermentation Process Guidelines, with some modification. Basal salts medium contained 26.7 mL/L H_3PO_4, 18.2 g/L K_2SO_4, 14.9 g/L $MgSO_4$, 4.13 g/L KOH, 40 g/L glycerol and was modified by adding 10 g/L MSG and 2.5 g/L $(NH_4)_2SO_4$.[40] $CaSO_4$ was omitted to reduce the formation of insoluble calcium phosphates. Basal salts medium was supplemented with PTM4 trace elements.[41] Antifoam 204 (0.5 mL/L) was used to prevent foaming. Fermentation occurred at a pH of 5.0 and pure O_2 was used to maintain the dissolved O_2 content (dO_2) at 35% by regulating the impeller speed. Culture density and protein accumulation were monitored daily.

Purification

Culture supernatants, adjusted to pH = 8.0, were centrifuged and filtered through glass fiber prefilters and 0.45 μm PE filters. Enzymes were affinity purified at pH = 8.0 on an agarose column with immobilized STI. The wash buffer contained 50 m*M* Tris, 0.5 *M* NaCl, pH = 8.0 while 50 m*M* acetic acid, 0.5 *M* NaCl, pH = 3.5 was used for protein elution. Active fractions (Z-Lys-SBzl) were pooled and concentrated by ultrafiltration on a 10K MWCO membrane (Millipore) and stored in 100 m*M* glycine, pH = 3.0. Active site titrations of R96Q and Y174R were performed with 4-methylumbelliferyl-p-guanidinobenzoate (MUGB).[38]

SDS-PAGE and western blotting

SDS-PAGE was performed with 12% [Fig. 2(A and B)] or 4-12% [Fig. 2(C)] NuPAGE gels using MOPS SDS running buffer (Life Technologies) and stained with NuBlu Coomassie stain (NuSep). Western blot on PVDF was blocked for 1 hour in TBST + 1% BSA at 20°C, then reacted with 0.1 μg/mL of a mouse monoclonal antibody against bEP_L (GenScript) in TBST. After washing with TBST, the blot was then incubated for 1 hour with HRP-conjugated goat anti-mouse antibody (0.08 μg/mL) in TBST (Pierce), washed in TBST, and visualized using Super Signal West Pico ECL (Pierce). Images were captured and processed using a GBox (Syngene).

Cleavage of CCP

CCP was incubated with Tag*off™, R96Q, or Y174R enzyme at 37°C in 20 m*M* Tris-HCl, 50 m*M* NaCl, 2 m*M* CaCl2, pH = 7.4 (provided with Tag*off™). At 2 and 6 hours, samples were collected and frozen (−20°C) prior to SDS-PAGE analysis.

Stability assay

Enzymes were incubated for five days at 37°C in 20 m*M* Tris-HCl, 50 m*M* NaCl, 2 m*M* CaCl2, pH = 8.0 and daily samples were stored at −20°C prior to assays in 300 μ*M* Z-Lys-SBzl, 100 m*M* Tris, 0.02 m*M* CaCl2, 0.01% Triton X-100, 10% DMSO, 1 m*M* 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB), pH = 8.0.

Kinetic assays

All enzyme activity assays were conducted in 100 μL format on a 96-well microtiter plate reader at 25 ± 0.5°C. All substrate stocks were freshly dissolved in DMSO. The assay conditions were 100 m*M* Tris, 0.02% $CaCl_2$, 0.01% Triton X-100, 10% DMSO, pH = 8.0. Assays with 0.05–1.0 m*M* Z-Lys-SBzl (Sigma or Bachem) also included 1 m*M* DTNB.[42,43] The extended peptide substrates GD_4K-pNA and GD_4R-pNA (LifeTein) were used at concentrations between 0.02 and 1.0 m*M* and 5 and 100 μ*M*, respectively. Reactions were initiated by adding enzyme to a 2 n*M* final concentration into premixed buffer and substrate. Absorbances at 410 nm were followed kinetically for 10 minutes. Product extinction coefficients were 13,600 M^{-1} cm^{-1} for the thiobenzyl assays and 8800 M^{-1} cm^{-1} for p-nitroaniline. The kinetic parameters K_M and k_{cat} were determined by hyperbolic regression analysis of data using Hyper32 (J. Easterby, ©2003).

BPTI inhibition assays were conducted in the same format and with 100 m*M* Tris, 0.02% $CaCl_2$, 0.01% Triton X-100, 10% DMSO, pH = 8.0, containing 1 m*M* DTNB and using 300–600 μ*M* Z-Lys-SBzl as the substrate. Here, 2 n*M* enzyme concentrations were premixed with 50–600 n*M* BPTI and incubated at room temperature for 15 minutes prior to the addition of Z-Lys-SBzl to initiate the reactions. Dixon plots were generated from the reaction velocity data to determine the K_i values.[44]

Molecular modeling

Modeling was performed in Discovery Studio 3.1 (Accelrys), using bEP_L (PDB Code: 1EKB) to create the proposed model of $rbEP_L$ Y174R [Fig. 4(B)] by single amino acid substitution. This side chain was repositioned by manual bond rotation, compared to the previously proposed position of R174,[22] and its geometry was cleaned using the software's Dreiding-like force field. The structure of $rhEP_L$sc (PDB Code: 4DGJ) provided the template for modeling all human EP_L variants in complex with the D_3K substrate (taken from 1EKB). The key residues R96, R174, and K219 were repositioned, as above, in comparison with the previous reports[22,29] to reflect possible interactions with the D_3K substrate [Fig. 4(C–F)]. Residue Q96 [Fig. 4(F)] positioning could not be predicted based on currently available data.

Acknowledgment

We thank Dr. Michelle Duffourc of the ETSU Molecular Biology Core Facility for her helpful advice and for DNA sequencing.

References

1. Moss S, Lobley RW, Holmes R (1972) Enterokinase in human duodenal juice following secretin and pancreozymin and its relationship to bile salts and trypsin. Gut 13:851.
2. Lobley RW, Moss S, Holmes R (1973) Proceedings: brush-border localization of human enterokinase. Gut 14:817.
3. Lobley RW, Franks R, Holmes R (1977) Subcellular localization of enterokinase in human duodenal mucosa. Clin Sci Mol Med 53:551–562.
4. Hermon-Taylor J, Perrin J, Grant DA, Appleyard A, Bubel M, Magee AI (1977) Immunofluorescent localisation of enterokinase in human small intestine. Gut 18: 259–265.
5. Liepnieks JJ, Light A (1979) The preparation and properties of bovine enterokinase. J Biol Chem 254: 1677–1683.
6. Light A, Savithri HS, Liepnieks JJ (1980) Specificity of bovine enterokinase toward protein substrates. Anal Biochem 106:199–206.
7. Light A, Janska H (1989) Enterokinase (enteropeptidase): comparative aspects. Trends Biochem Sci 14: 110–112.
8. LaVallie ER, Rehemtulla A, Racie LA, DiBlasio EA, Ferenz C, Grant KL, Light A, McCoy JM (1993) Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase. J Biol Chem 268:23311–23317.
9. Lu D, Yuan X, Zheng X, Sadler JE (1997) Bovine proenteropeptidase is activated by trypsin, and the specificity of enteropeptidase depends on the heavy chain. J Biol Chem 272:31293–31300.
10. Haworth JC, Gourley B, Hadorn B, Sumida C (1971) Malabsorption and growth failure due to intestinal enterokinase deficiency. J Pediatr 78:481–490.
11. Hadorn B, Haworth JC, Gourley B, Prasad A, Troesch V (1975) Intestinal enterokinase deficiency. Occurrence in two sibs and age dependency of clinical expression. Arch Dis Child 50:277–282.
12. Braud S, Ciufolini MA, Harosh I (2012) Enteropeptidase: a gene associated with a starvation human phenotype and a novel target for obesity treatment. PLoS One 7:e49612.
13. Chen Z, Han S, Cao Z, Wu Y, Zhuo R, Li W (2012) Fusion expression and purification of four disulfide-rich peptides reveals enterokinase secondary cleavage sites in animal toxins. Peptides 39C:145–151.
14. Liew OW, Ching Chong JP, Yandle TG, Brennan SO (2005) Preparation of recombinant thioredoxin fused N-terminal proCNP: analysis of enterokinase cleavage products reveals new enterokinase cleavage sites. Protein Expr Purif 41:332–340.
15. Likhareva VV, Mikhailova AG, Rumsh LD (2002) Hydrolysis by enteropeptidase of nonspecific (model) peptide sequences and possible physiological role of this phenomenon. Vopr Med Khim 48:561–569.
16. Shahravan SH, Qu X, Chan IS, Shin JA (2008) Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. Protein Expr Purif 59:314–319.
17. LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Biotechnology 11: 187–193.
18. Collins-Racie LA, McColgan JM, Grant KL, DiBlasio-Smith EA, McCoy JM, LaVallie ER (1995) Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. Biotechnology 13:982–987.
19. Vozza LA, Wittwer L, Higgins DR, Purcell TJ, Bergseid M, Collins-Racie LA, LaVallie ER, Hoeffler JP (1996) Production of a recombinant bovine enterokinase catalytic subunit in the methylotrophic yeast Pichia pastoris. Biotechnology 14:77–81.
20. Peng L, Zhong X, Ou J, Zheng S, Liao J, Wang L, Xu A (2004) High-level secretory production of recombinant bovine enterokinase light chain by Pichia pastoris. J Biotechnol 108:185–192.
21. Choi SI, Song HW, Moon JW, Seong BL (2001) Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. Biotechnol Bioeng 75: 718–724.
22. Chun H, Joo K, Lee J, Shin HC (2011) Design and efficient production of bovine enterokinase light chain with higher specificity in *E. coli*. Biotechnol Lett 33: 1227–1232.
23. Huang L, Ruan H, Gu W, Xu Z, Cen P, Fan L (2007) Functional expression and purification of bovine enterokinase light chain in recombinant *Escherichia coli*. Prep Biochem Biotechnol 37:205–217.
24. Mikhailova AG, Likhareva VV, Teich N, Rumsh LD (2007) The ways of realization of high specificity and efficiency of enteropeptidase. Protein Pept Lett 14: 227–232.
25. Gasparian ME, Bychkov ML, Dolgikh DA, Kirpichnikov MP (2011) Strategy for improvement of enteropeptidase efficiency in tag removal processes. Protein Expr Purif 79:191–196.
26. Gasparian ME, Ostapchenko VG, Dolgikh DA, Kirpichnikov MP (2006) Biochemical characterization of human enteropeptidase light chain. Biochemistry 71: 113–119.
27. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27:157–162.
28. Lu D, Fütterer K, Korolev S, Zheng X, Tan K, Waksman G, Sadler JE (1999) Crystal structure of enteropeptidase light chain complexed with an analog of the trypsinogen activation peptide. J Mol Biol 292:361–373.
29. Ostapchenko VG, Gasparian ME, Kosinsky YA, Efremov RG, Dolgikh DA, Kirpichnikov MP (2012) Dissecting structural basis of the unique substrate selectivity of human enteropeptidase catalytic subunit. J Biomol Struct Dyn 30:62–73.
30. Simeonov P, Berger-Hoffmann R, Hoffmann R, Sträter N, Zuchner T (2011) Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. Protein Eng Des Sel 24:261–268.
31. Simeonov P, Zahn M, Sträter N, Zuchner T (2012) Crystal structure of a supercharged variant of the human enteropeptidase light chain. Proteins 80:1907–1910.
32. Lockhart BE, Vencill JR, Felix CM, Johnson DA (2005) Recombinant human mast-cell chymase: an improved procedure for expression in *Pichia pastoris* and purification of the highly active enzyme. Biotech Appl Biochem 41:89–95.
33. Niles AL, Maffitt M, Haak-Frendscho M, Wheeless CJ, Johnson DA (1998) Recombinant human mast cell tryptase beta: stable expression in *Pichia pastoris* and

purification of fully active enzyme. Biotechnol Appl Biochem 28:125–131.
34. Pepeliaev S, Krahulec J, Cerný Z, Jílková J, Tlustá M, Dostálová J (2011) High level expression of human enteropeptidase light chain in Pichia pastoris. J Biotechnol 156:67–75.
35. van Oort E, de Heer PG, van Leeuwen WA, Derksen NI, Müller M, Huveneers S, Aalberse RC, van Ree R (2002) Maturation of Pichia pastoris-derived recombinant pro-Der p 1 induced by deglycosylation and by the natural cysteine protease Der p 1 from house dust mite. Eur J Biochem 269:671–679.
36. van Oort E, Lerouge P, de Heer PG, Séveno M, Coquet L, Modderman PW, Faye L, Aalberse RC, van Ree R (2004) Substitution of Pichia pastoris-derived recombinant proteins with mannose containing O- and N-linked glycans decreases specificity of diagnostic tests. Int Arch Allergy Immunol 135:187–195.
37. Daly R, Hearn MT (2005) Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. J Mol Recognit 18:119–138.
38. Jameson GW, Roberts DV, Adams RW, Kyle WS, Elmore DT (1973) Determination of the operational molarity of solutions of bovine alpha-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. Biochem J 131:107–117.
39. Kawamura K, Yamada T, Kurihara K, Tamada T, Kuroki R, Tanaka I, Takahashi H, Niimura N (2011) X-ray and neutron protein crystallographic analysis of the trypsin-BPTI complex. Acta Crystallogr D67: 140–148.
40. Boettner M, Prinz B, Holz C, Stahl U, Lang C (2002) High-throughput screening for expression of heterologous proteins in the yeast Pichia pastoris. J Biotechnol 99:51–62.
41. Stratton J, Chiruvolu V, Meagher M (1998) High cell-density fermentation. Methods Mol Biol 103: 107–120.
42. Johnson DA (2006) Human mast cell proteases: activity assays using thiobenzyl ester substrates. Methods Mol Biol 315:193–202.
43. Green GD, Shaw E (1979) Thiobenzyl benzyloxycarbonyl-L-lysinate, substrate for a sensitive colorimetric assay for trypsin-like enzymes. Anal Biochem 93: 223–226.
44. Dixon M (1953) The determination of enzyme inhibitor constants. Biochem J 55:170–171.