Biochemistry (Moscow), Vol. 69, No. 8, 2004, pp. 909-917. Translated from Biokhimiya, Vol. 69, No. 8, 2004, pp. 1118-1128.

Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM03-297, June 27, 2004.

Study of Secondary Specificity of Enteropeptidase in Comparison with Trypsin

A. G. Mikhailova*, V. V. Likhareva, B. V. Vaskovsky, S. K. Garanin, L. V. Onoprienko, I. A. Prudchenko, L. D. Chikin, and L. D. Rumsh

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, Moscow 117997, Russia; fax: (7-095) 335-7103; E-mail: anna@enzyme.siobc.ras.ru

Received December 26, 2003
Revision received February 3, 2004

Abstract—A comparative study of secondary specificities of enteropeptidase and trypsin was performed using peptide substrates with general formula A-(Asp/Glu)$_n$-Lys(Arg)-↓-B, where n = 1-4. This was the first study to demonstrate that, similar to other serine proteases, enteropeptidase has an extended secondary binding site interacting with 6-7 amino acid residues surrounding the peptide bond to be hydrolyzed. However, in the case of typical enteropeptidase substrates containing four negatively charged Asp/Glu residues at positions P2-P5, electrostatic interaction between these residues and the secondary site Lys99 of the enteropeptidase light chain is the main factor that determines hydrolysis efficiency. The secondary specificity of enteropeptidase differs from the secondary specificity of trypsin. The chromophoric synthetic enteropeptidase substrate G_5DK-F(NO_2)G (k_{cat}/K_m = 2380 $mM^{-1}\cdot min^{-1}$) is more efficient than the fusion protein $PrAD_4K$-P26 (k_{cat}/K_m = 1260 $mM^{-1}\cdot min^{-1}$).

Key words: enteropeptidase, autolysis, trypsin, trypsinogen, peptide substrates, fusion proteins

Enteropeptidase (enterokinase, EC 3.4.21.9), a highly specific processing proteinase, activates trypsinogen, thereby initiating the cascade of reactions activating digestive enzymes [1]. When converting trypsinogen to trypsin, enteropeptidase highly efficiently catalyzes the hydrolysis of the polypeptide chain downstream of the N-terminal sequence tetra-aspartyl-lysine. This high specificity is the main characteristic feature of enteropeptidase that attracts the interest of enzymologists to this enzyme. Enteropeptidase is widely used as a tool for specific cleavage of various fusion proteins containing the -DDDDK- sequence (the enteropeptidase linker) inserted between the carrier and target proteins [2-7]. This enzyme is used not only in applied studies but also in fundamental works (specifically, in determination of highly specific proteolysis determinants). In structural aspect, enteropeptidase represents a typical trypsin-like serine proteinase: its light chain, corresponding to amino acid residues 801-1035, contains the entire set of amino acid residues required for the catalytic site formation and shares a high extent of homology with trypsin. However, in addition to the primary binding site S1 (Asp981), which determines the primary trypsin specificity of enteropeptidase, the light chain of the enzyme contains the secondary binding site S2 (Lys889), which coordinates the four aspartic acid residues at positions P2-P5 of the substrate [8], thereby imparting the characteristic high specificity to enteropeptidase. In 1998, we discovered unusual Ca^{2+}-dependent autolysis of the heavy enteropeptidase chain, corresponding to residues 118-800 and linked to the light chain through a disulfide bond [7, 9, 10]. As a result of autolysis, the efficiency of enteropeptidase substantially (by two orders of magnitude) decreased solely with respect to the natural substrate, trypsinogen. However, the enzyme activity with respect to artificial low- and high-molecular-weight substrates containing the linker sequence -DDDDK- was completely retained. When the autolysis sites (-QNMEK^{465}-TIFQ-, -NNYEK^{360}-INCN-, -NEWER^{384}-TQGS- and -GRRER^{420}-VGLL-) were determined [7, 10], a question arose as to how the hydrolysis of the polypeptide chain in the above-mentioned sequences agrees with presumable absolute specificity of enteropeptidase: in theory, its substrate should contain a Lys/Arg residue at position P1 and four Glu/Asp residues at positions P2-P5 [11].

The study of autolysis kinetics of the enteropeptidase heavy chain upon its incubation with EDTA showed that this process is a zero-order reaction, because the initial

* To whom correspondence should be addressed.

0006-2979/04/6908-0909 ©2004 MAIK "Nauka / Interperiodica"

reaction rate does not depend on the concentration of enteropeptidase, being equal to $4 \cdot 10^{-9}$ M/min (20°C, 10^{-3} M EDTA). Hence, this is apparently an intramolecular process [12]. It is believed that autolysis takes place simultaneously at the bonds located downstream of lysine (Lys360 and Lys465) and arginine (Arg384 and Arg422) residues; however, the final autolysis product is always the C-terminal half of the heavy chain (residues 466-800), linked though a disulfide bond to the light chain. We assumed that either autolysis is as a unique process for this enzyme, like trypsin activation, or the specificity of the enteropeptidase is broader than had been believed earlier and needs to be studied in more detail. An example of the first case is autolysis of proteinase of tobacco etch virus, which takes place at the site that does not correspond to the specificity of this enzyme with respect to any peptide or protein substrate [13]. However, a wide use of enteropeptidase in gene engineering studies on processing fusion proteins allowed the authors of [2-5, 14] to obtain data indicating that enteropeptidase preparations are able to hydrolyze some target peptides and proteins that do not contain the sequence -$(Asp)_4$Lys- required for enteropeptidase hydrolysis. Hydrolysis of these proteins and peptides at the bonds located downstream of arginine or lysine residues was originally ascribed to contamination of enteropeptidase preparations with some trypsin-like enzyme [3]. We noted that, in all cases of this unusual enteropeptidase hydrolysis of peptide bonds formed by Lys and Arg carbonyls (including autolysis discovered by us), at least one Glu or Asp residue is necessarily present at positions P2-P5 with respect to the peptide bonds hydrolyzed. In view of this, we assumed that enteropeptidase can hydrolyze substrates with a general formula:

$$A\text{-}Asp_m(Glu_m)\text{-}Xaa_n\text{-}Lys(Arg)\text{-}\downarrow\text{-}B, \qquad (1)$$

where Xaa is Asp, Glu, or any other amino acid; $m = 0$-1; $n = 0$-3; A and B are protective groups or various amino acid residues.

To corroborate a universal nature of the proteolysis catalyzed by enteropeptidase, we used some biologically active peptides comprised of seven to nine amino acid residues—human angiotensin II (DR-VYIHPF, AT) and cattle hemoglobin β-chain peptides (LTAEEK-A (Hb 2-8) and MLTAEEK-AA (Hb 1-9))—as first substrates of this type [12, 15]. At the next stage of work, we used synthetic tryptophanyl-containing peptides WDDR-G and WDDK-G [16]. Our first results showed that enteropeptidase is indeed able to hydrolyze peptides at the bonds formed by the carboxyl groups of Lys or Arg residues if the substrate contains less than four negatively charged amino acid residues at positions P2-P5. However, truncation of the linker sequence as a result of a decrease in the number of negatively charged amino acid residues deteriorates the binding of such substrates to the secondary site Lys889 of the enzyme light chain by an order of magnitude [12, 15, 16].

In this work, we continued our previous studies on this subject. To determine the effect of the length of polypeptide chain on the efficiency of enteropeptidase hydrolysis, we used other hemoglobin fragments containing the EEK sequence. In addition, a comparative analysis of the secondary specificities of enteropeptidase and trypsin with respect to peptide substrates of this type has been performed. We also tested new synthetic chromophoric substrates of enteropeptidase containing *p*-nitrophenylalanine ($F(NO_2)$) as P1′ residue. A change in absorbance at 310 nm, occurring as a result of hydrolysis of such peptide bonds, allows direct spectrophotometric monitoring of the initial hydrolysis rate both at pH 4-6 [17] and pH 7-8 [18].

MATERIALS AND METHODS

Materials. The reagents used in the study were from Merck and Serva (Germany), Sigma and Bio-Rad (USA), Reanal (Hungary), and Kriokhim (Russia). Crystalline trypsin was from the Plant of Medical Preparations (St. Petersburg Meat-Packing Plant, Russia). HPLC was performed on a Beckman System Gold chromatograph (USA). Spectrophotometric measurements were performed on a Gilford 2400-2 spectrophotometer (USA).

Synthesis of peptide substrates. All peptide fragments of cattle hemoglobin were obtained by standard methods of classic peptide chemistry in solution, using the principle of maximal protection. The peptide LTAEEKAAV (Hb (2-10)) was obtained by condensing two protected fragments—Boc-T(Bzl)A and E(OBzl)E(OBzl)K(Z)AAV-OBzl—by the DCC/HOBt method, with subsequent unblocking with saturated HBr solution in CH_3COOH. The protected fragments were obtained by stepwise elongation of the chain, starting with the C terminus, by the methods of mixed anhydrides and activated (*p*-nitrophenyl and pentachlorophenyl) esters with HOBt catalysis. Free peptide was isolated by chromatography on a column with Sephadex G-15 in 0.1 M acetic acid. In total, 55 mg of the peptide Hb (2-10) was obtained.

The peptide MLTAEEKA (Hb (1-8)) was obtained similarly to peptides Hb (2-10), Hb (2-8), and Hb (1-9) [12]. The peptides WDDRG and WDDKG were synthesized by the method of Mutt et al. [14].

Peptides $G_5DK\text{-}F(NO_2)G$ and $GD_4K\text{-}F(NO_2)G$ were synthesized by the solid-phase method on a Beckman 990B device (USA) using *p*-(Boc-Gly-oxymethyl)phenylacetamidomethyl polymer (0.53 mmol/g) obtained from Advanced Chemtech (USA). Peptide bonds were formed using the DCC/HOBt method with preliminary activation at 0°C for 10 min; reagents were used in 3.5-fold excess. After unblocking and subsequent

elution from resin by liquid HF in the presence of *p*-cresol (10 : 1 v/v), peptides were purified by gel filtration on Sephadex G-15 in 0.1 M acetic acid and then subjected to HPLC. In total, we obtained 107 mg of the first peptide and 39 mg of the second peptide. The homogeneity of each peptide was determined by HPLC (using Nucleosil C18 columns) and N-terminal amino acid analysis. The structure of the peptides was confirmed by NMR (Brucker 500 MHz, Germany), amino acid analysis (Biotronic LC 3000, Eppendorf-Biotronic, Germany), and MALDI mass spectrometry (VISION 2000, Thermo Bioanalysis Corp., UK). The degree of purity of the peptides was at least 95% (according to HPLC data).

Enteropeptidase was isolated from bovine duodenum and purified as described earlier [7]. The activity of enzyme preparations was determined using trypsinogen activation [7].

Kinetic measurements. *1. HPLC.* Each peptide (0.25-10 mM; or 0.01-0.5 mM in the case of GD_4K-F(NO_2)G and G_5DK-F(NO_2)G) was incubated with the enzyme in 0.1 M Tris-HCl buffer (pH 8.0; 37°C). Enteropeptidase concentration was 72.8 nM (AT, Hb (2-8), and Hb (1-8)), 33.3 nM (Hb (1-9)), 22.6 nM (WDDRG, WDDKG, and Hb (2-10)), and 1 nM (G_5DK-F(NO_2)G and GD_4K-F(NO_2)G). Trypsin concentrations used for the trypsin hydrolysis were 2.5 μM (in the case of Hb (2-8)), 30 nM (in the case of Hb (1-9)), 0.58 μM (in the case of G_5DK-F(NO_2)G), and 1 μM (in the case of GD_4K-F(NO_2)G). Rate constants of hydrolysis of peptides by enteropeptidase and trypsin were determined by incubating at least six concentrations of the corresponding substrate with the enzyme. When determining the k_{cat}/K_m ratio in the case of enteropeptidase hydrolysis of the substrates Hb (1-9) and Hb (2-10), low concentrations of substrates were used (10^{-5}-10^{-4} M; [S] << K_m). At certain time intervals, corresponding to at most 20% conversion of each substrate, six to eight aliquots (2-5 μl) were taken from incubation medium, diluted 10-25 times with 0.1% TFA, and stored at −70°C. The composition of samples was analyzed by HPLC in 0.1% TFA on columns Nucleosil 7/C8 (Macherey-Nagel, Germany; 4 × 250 mm; acetonitrile gradient 0-60%), Nova-Pac C8 (Waters, USA; 3.9 × 150 mm; acetonitrile gradient 0-40% during 35 min; elution rate 1 ml/min), and Luna C18 (Phenomenex, USA; 2 × 250 mm; acetonitrile gradient 0-60%; elution rate 0.3 ml/min). Initial reaction rate was calculated by the ratio between the areas of peaks of the substrate and one of the products. A correction for the difference in the molecular absorbance at 222 nm was introduced after exhaustive hydrolysis. To confirm structures of hydrolysis products, aliquots of fractions corresponding to individual peaks were analyzed by MALDI mass spectrometry.

2. Spectrophotometric method. Hydrolysis of the substrates GD_4K-F(NO_2)G (13 nM enteropeptidase or 1 μM trypsin) and G_5DK-F(NO_2)G (20 nM enteropeptidase or 0.58 μM trypsin) was monitored by a decrease in absorbance upon hydrolysis of the bond Lys–Phe(NO_2) ($\Delta\varepsilon_{310}$ = 550) at 37°C in a thermostatically controlled cuvette of the spectrophotometer. To determine the initial hydrolysis rate, we used at least six concentrations of each substrate within the range 0.1-0.5 mM in 10 mM Tris-HCl buffer (pH 7.5) containing 50 mM $CaCl_2$ and 0.1% *n*-octyl-β-D-glucopyranoside. In some cases, aliquots of incubation medium were taken as described in method 1 and simultaneously analyzed by HPLC.

Kinetic parameters (k_{cat} and K_m) of hydrolysis of substrates were calculated using the method of Eisenthal–Cornish-Bowden [19].

RESULTS

Peptides with a general formula A-$(Asp/Glu)_n$-Lys(Arg)-↓-B (2), containing one or two Asp or Glu residues at positions P2-P3 (n = 1-2) and Lys or Arg at position P1, were used as enteropeptidase substrates containing linkers of varying length. Hydrolysis of these peptides was monitored using HPLC, by recording a decrease in the area of the peak corresponding to the original peptide and increase in the area of the peak corresponding to hydrolysis products during incubation with enteropeptidase (Fig. 1). Kinetic constants of hydrolysis of the substrates tested are summarized in Table 1. We found that, indeed, enteropeptidase is able to hydrolyze peptides at the bonds formed by the carboxyl groups of Lys or Arg residues if the substrate contains less than four negatively charged amino acid residues at positions P2-P5. However, substrates containing a truncated linker sequence (with a decreased number of negatively charged amino acid residues) bound to the secondary site Lys889 of the light chain of the enzyme much less strongly (by an order of magnitude). The K_m values for all these substrates (more than 10^{-3} M) are greater by an order of magnitude than the corresponding parameters for typical synthetic substrates of the enzyme—fusion proteins and peptides containing the full-length enteropeptidase linker -DDDDK- (K_m ~ 10^{-4} M). Note that substrates differing in the number (one or two) and nature (aspartyl or glutamyl) of negatively charged amino acid residues interacted with enteropeptidase with almost equal efficiencies. The same applies to the substrates containing Arg or Lys at position P1.

The k_{cat} values for WDDRG and WDDKG, AT and Hb (2-8) were also similar and low (~30 min^{-1}). The total efficiency of hydrolysis of such peptides consisting of five, seven, or eight amino acid residues containing one or two negatively charged amino acid residues at positions P2-P3 of the substrates was almost equal and accounted for no more than 1% of the corresponding value obtained for a fusion protein containing the full-length linker -$(Asp)_4$-Lys- (Table 1). Our data have not revealed any difference

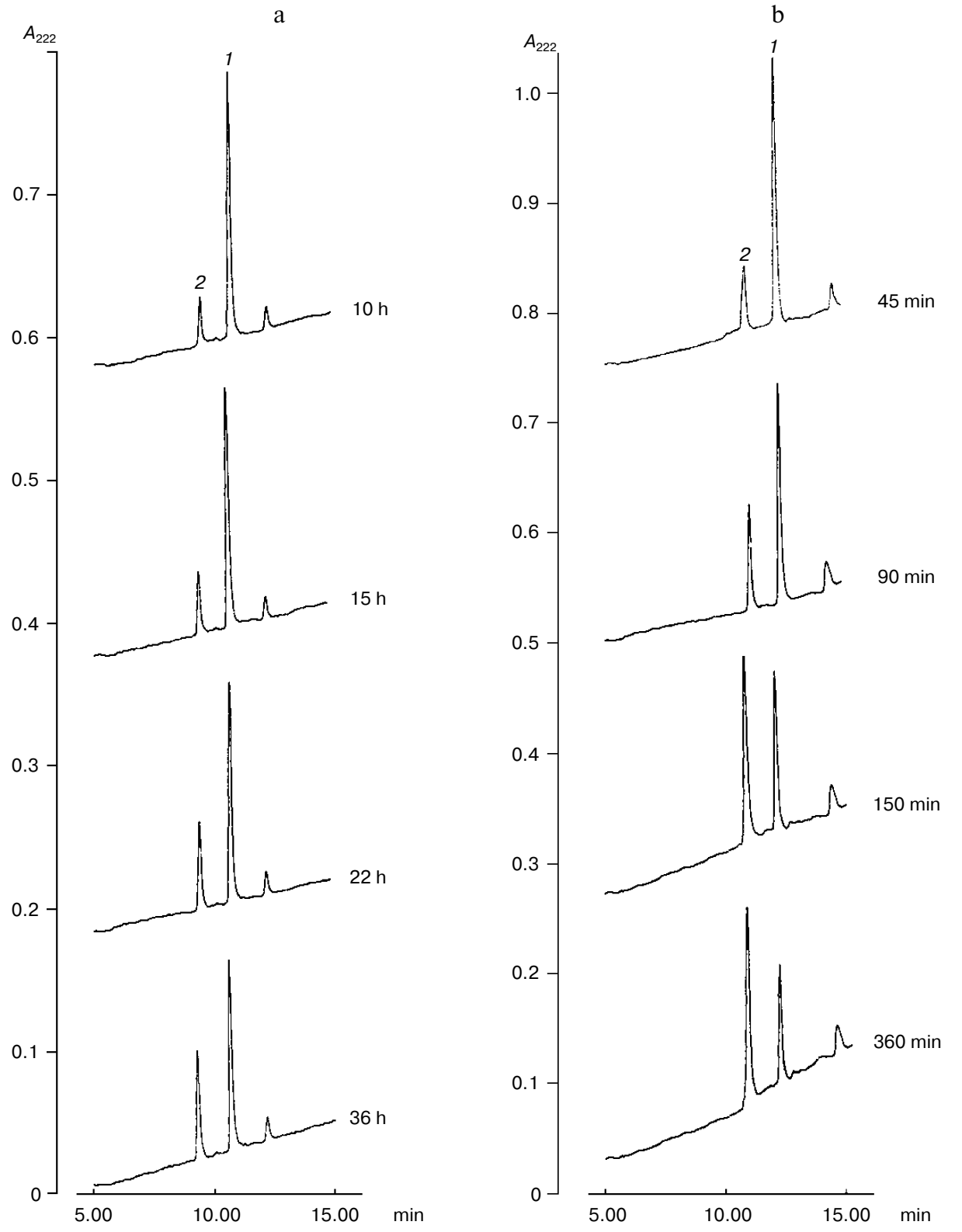

Fig. 1. Hydrolysis of hemoglobin β-chain peptides by enteropeptidase (pH 8.0, 37°C). Analysis of incubation medium by HPLC on a Nova-Pac C8 column (3.9 × 150 mm; acetonitrile gradient 0-40% in 0.1% TFA; elution rate 1 ml/min). a) Hb (2-8), [S] = 5.33 mM, [E] = 72.8 nM; *1*) peak of substrate LTAEEKA; *2*) peak of product LTAEEK. b) Hb (1-9), [S] = 6.00 mM, [E] = 33.3 nM; *1*) peak of substrate MLTAEEKAA; *2*) peak of product MLTAEEK.

Table 1. Kinetic constants of hydrolysis of substrates by enteropeptidase (EP) and trypsin (Tr) (pH 8.0, 37°C)

Substrate	K_m, mM		k_{cat}, min^{-1}		k_{cat}/K_m, $mM^{-1} \cdot min^{-1}$	
	EP	Tr	EP	Tr	EP	Tr
AT; **DR**-VYIHPF	3.3 ± 0.5	—	24.0 ± 2	—	7.3 ± 1.2	—
W**DDK**-G	1.6 ± 0.2	—	32.6 ± 4.3	—	19.9 ± 0.6	—
W**DDR**-G	2.1 ± 0.2	—	40.6 ± 2.9	—	19.5 ± 0.3	—
Hb (2-8); LTA**EEK**-A	4.2 ± 0.8	8.9 ± 1.8	29.4 ± 3.3	74.2 ± 14.8	7.0 ± 1.5	8.4 ± 1.8
Hb (1-9); MLTA**EEK**-AA	4.0 ± 1.0	1.7 ± 0.4	1510 ± 189	2900 ± 300	378 ± 105	1700 ± 400
Hb (1-8); MLTA**EEK**-A	—	—	—	—	7.0 ± 0.8	—
Hb (2-10); LTA**EEK**-AAV	—	—	—	—	400 ± 40	—
G_5**DK**-F(NO_2)G*	0.437 ± 0.044	1.95 ± 0.2	1040 ± 150	47.2 ± 5.0	2380 ± 600	24.2 ± 6.0
G**D_4K**-F(NO_2)G*	0.160 ± 0.016	10.0 ± 1.0	1070 ± 170	50 ± 5.0	6700 ± 1700	5.0 ± 1.5
G**D_4K**-Nfa* [15-17]	0.200	—	1000	—	5000	80**
						0.1***
PrA**D_4K**-P26**** [7]	0.125		157.0		1260	

* 50 mM Ca^{2+}.
** Calculated using the data of Grant et al. [21]; 10 mM Ca^{2+}.
*** Calculated using the data of Grant et al. [21]; 0.1 mM Ca^{2+}.
**** Fusion protein containing a modified protein A as a carrier and recoverin as a target protein.

in the efficiency of enteropeptidase hydrolysis of Lys- or Arg-containing substrates. However, trypsin hydrolyzed Arg-containing substrates five- to tenfold more effectively than Lys-containing substrates, which may be due to different types of binding of these positively charged amino acid residues in the S1 binding site [20]. Enteropeptidase contains the primary trypsin site S1, Asp981 (Asp189 in the standard chymotrypsin numbering system); however, its natural substrate, trypsinogen, is a Lys-type substrate. The effect of the substitution of this residue with Arg has not been studied before.

Thus, as we assumed, hydrolysis of type 2 substrates with $n < 4$ was much less efficient than hydrolysis of typical enteropeptidase substrates with $n = 4$. Unexpected results were obtained for the nonapeptide Hb (1-9). Both substrates—hemoglobin fragments Hb (1-9) and Hb (2-8)—contain identical amino acid sequences at positions P6-P1′. However, at equal K_m values (~10^{-3} M), elongation of the peptide Hb (2-8) LTAEEKA at the N- and C-termini by one amino acid residues resulted in a dramatic increase in hydrolysis efficiency: k_{cat} for Hb (1-9) MLTAEEKAA was 1510 min^{-1} (i.e., hydrolysis of this compound was only threefold less effective than hydrolysis of the fusion protein containing the full-length linker PrAD_4K-P26) (Table 1).

We also determined the constants of trypsin hydrolysis of cattle hemoglobin β-chain peptides Hb (2-8) and Hb (1-9) (Table 1). The efficiencies of enteropeptidase and trypsin hydrolysis of the peptide Hb (2-8) were similar. However, the transition from the heptapeptide Hb (2-8) to the nonapeptide Hb (1-9) caused a further increase in hydrolysis efficiency.

To determine the effect of the length of polypeptide chain in N- and C-terminal regions on the efficiency of enteropeptidase hydrolysis, we used other fragments of cattle hemoglobin β-chain, which contained the -EEK- sequence – MLTAEEKA (Hb (1-8)) and LTAEEKAAV (Hb (2-10)) (Table 1). Thus, two pairs of peptide substrates of enteropeptidase, containing identical amino acid sequences at positions P6-P1′ but yielding hydrolysis product differing in length by 1 or 2 amino acid residues only at the N or C terminus (specifically, MLTAEEK-AA and LTAEEK-AAV (I), and LTAEEK-A and MLTAEEK-A (II)) were obtained. The efficiency of hydrolysis of group I peptides (k_{cat}/K_m = 378 and 400 $mM^{-1} \cdot min^{-1}$, respectively) was by two orders of magnitude greater than that of group II peptides (k_{cat}/K_m = 7.0 $mM^{-1} \cdot min^{-1}$ in the case of both substrates). A similar consistent pattern was observed in the case of trypsin hydrolysis of peptides of group I and II (k_{cat}/K_m = 1700 and 8.4 $mM^{-1} \cdot min^{-1}$, respectively). Elongation of polypeptide chain at N-terminal region caused no increase in hydrolysis efficiency. Efficiencies of enteropeptidase and trypsin hydrolysis of substrates of this type are determined by the length of the leaving C-terminal group. Apparently, the interaction between the enzyme and the amino acid residue located at position P2′ of the substrate is critical for hydrolysis efficiency. Further elongation of substrates both at P3′

and P7 had no effect on hydrolysis efficiency. However, our data on enteropeptidase hydrolysis of angiotensin II are indicative of a key role of secondary interactions between the amino acid residues at positions P3-P5(6): kinetic constants obtained for this substrate, containing a fairly long leaving group (B = VYIHPF) but a shorter N-terminal region (A = H, n = 1), are low and practically equal to the corresponding parameters obtained for Hb (1-8), Hb (2-8), WDDKG and WDDRG (Table 1).

At the beginning of the study, the selection of substrates corresponding to formula (2) was random. However, later we selected several series of substrates with minimal differences, such as WDDKG and WDDRG, Hb (1-8) and Hb (2-8), Hb (1-9) and Hb (2-10). For a systematic study of the extended center of serine proteinases, only one series of substrates, differing only in one position from one another, is usually used. For this reason, taking into account our data that the leaving group of the substrate containing at least two amino acid residues is needed for efficient hydrolysis, at the next stage of the study we synthesized a series of type 2 peptides containing the sequence *p*-nitro-L-phenylalanine-glycine (B = $F(NO_2)G$) as a C-terminal outgoing group and a combination of Asp and Gly residues at positions P2-P5(6). At first, we determined the constants of enteropeptidase and trypsin hydrolysis for the peptides GD_4K-$F(NO_2)G$ and G_5DK-$F(NO_2)G$ (Table 1). The peptides whose $F(NO_2)$ residue forms the bond hydrolyzed are convenient tools that allow direct spectrophotometric monitoring of hydrolysis. For example, the substrates that are hydrolyzed at the bond Lys–$Phe(NO_2)$ are highly sensitive substrates of fungal aspartate proteinases at pH 5.0-6.5 [17]. Unfortunately, changes in absorbance at optimal pH range of serine proteinases (7.5-8.0) are much less pronounced. Nevertheless, the absorbance change ($\Delta\varepsilon_{310}$) in the case of peptides containing a number of Asp residues located upstream of the Lys residue was 550, which allows the spectrophotometric method to be successfully used for recording hydrolysis rate of these substrates. The kinetic constants of hydrolysis of the peptides GD_4K-$F(NO_2)G$ and G_5DK-$F(NO_2)G$ (Table 1), determined independently using two methods—spectrophotometrically and chromatographically (using HPLC, as in the case of all other substrates), were practically equal. The first of the two substrates, representative of typical synthetic peptide and protein substrates of enteropeptidase, was used as a control. The parameters of hydrolysis of the synthetic substrate β-naphthylamide glycyl-tetra-L-aspartyl-L-lysine (GD_4K-Nfa), which was first suggested by Grant and Tailor [21] for determination of human enteropeptidase, were practically the same (Table 1). Nevertheless, in our opinion, it is more correct to use the chromophoric peptide substrate for studying the secondary specificity of the

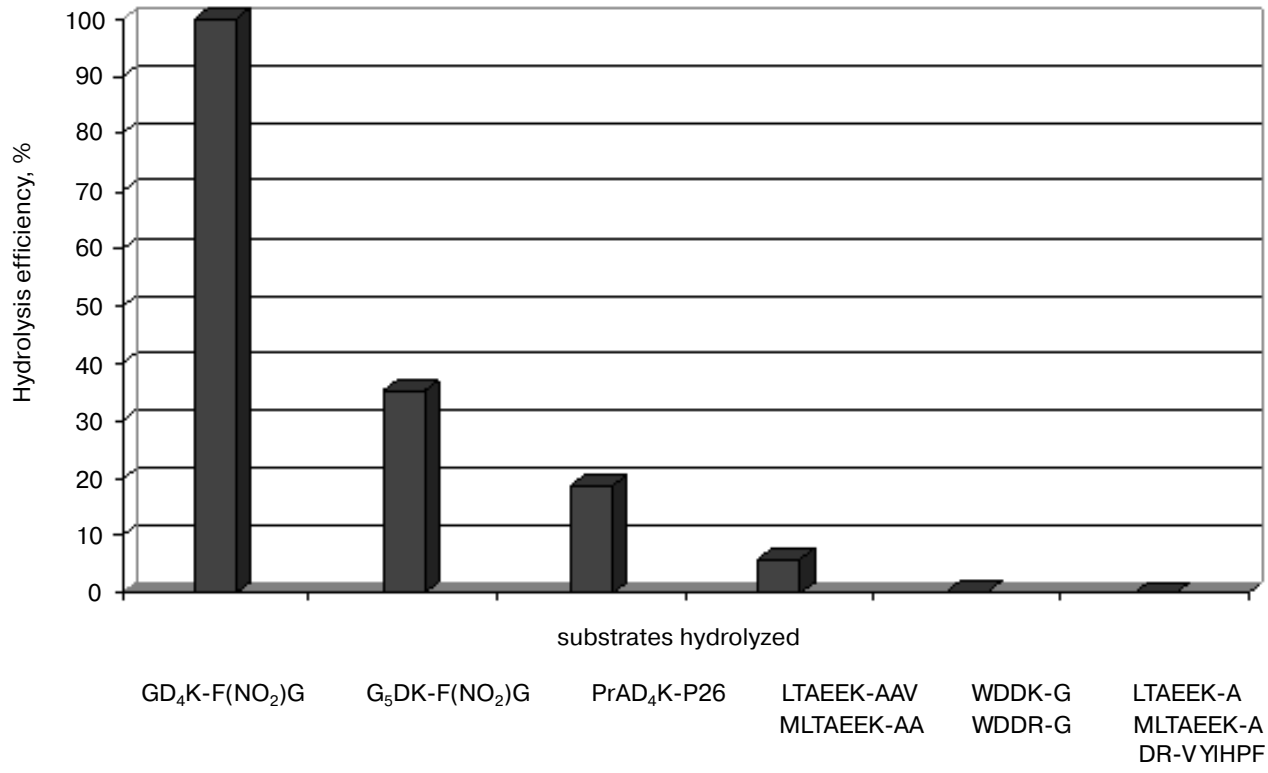

Fig. 2. Efficiency of hydrolysis of synthetic peptide and protein substrates by enteropeptidase. The k_{cat}/K_m value for $GD_4KF(NO_2)G$ was taken as 100%.

Table 2. Selectivity of hydrolysis of some type A-(Asp/Glu)$_n$-Lys(Arg)-↓-B substrates by enteropeptidase (EP) and trypsin (Tr): the k_{cat}/K_m ratio for enteropeptidase and trypsin hydrolysis of each substrate

Substrate	EP/Tr
Hb (2-8); LTA**EEK**-A	0.83
Hb (1-9); MLTA**EEK**-AA	0.22
G$_5$**DK**-F(NO$_2$)G	98
G**D$_4$K**-F(NO$_2$)G	1340

enzyme. Enteropeptidase, containing the primary trypsin site, can hydrolyze (yet less efficiently) the amide substrates of trypsin, such as *p*-nitroanilides of benzoyl-DL-arginine (K_m = 1.27·10^{-2} M; k_{cat} = 42 min^{-1} [22]) and pyroglutamyl-prolyl-L-arginine (K_m ~ 2·10^{-3} M [23]), not to mention the highly effective thioester Z-L-Lys-SBzl [8]. However, no authors have yet reported purely trypsin-type hydrolysis of peptides and proteins by enteropeptidase (i.e., when no one negatively changed amino acid group is contained at position P2-P5) (Mikhailova and Likhareva, unpublished data).

The second synthetic chromophore substrate of this series, G$_5$DK-F(NO$_2$)G, was unexpectedly found to be a highly effective substrate of enteropeptidase (even more effective than the fusion protein PrAD$_4$K-P26 (Table 1 and Fig. 2). The efficiency of hydrolysis of G$_5$DK-F(NO$_2$)G, evaluated by the k_{cat}/K_m ratio, was 2380 mM^{-1}·min^{-1}, i.e., only threefold lower than that of a typical substrate GD$_4$K-F(NO$_2$)G and more than sixfold greater that the effectives of hydrolysis of the peptides MLTAEEK-AA and LTAEEK-AAV (although the last two peptides contain two negatively charged amino acid residues) (Fig. 2). This substrate may be also used for spectrophotometric determination of the activity of enteropeptidase.

Conversely, the constants of trypsin hydrolysis of the substrates GD$_4$K-F(NO$_2$)G and G$_5$DK-F(NO$_2$)G were low (5 and 24.2 mM^{-1}·min^{-1}, respectively; Table 1). The adverse effect of four Asp residues at positions P2-P5 in trypsinogen on trypsin hydrolysis and the physical role of this phenomenon in preventing undesirable trypsin autoactivation have been long known [24]. Let us express the selectivity of hydrolysis of each substrate as the k_{cat}/K_m ratio of enteropeptidase and trypsin hydrolysis (Table 2). In the case of GD$_4$K-F(NO$_2$)G, a standard enteropeptidase substrate, this ratio is 1340. The substrate G$_5$DK-F(NO$_2$)G, containing Gly residues at positions P3-P6 and only one Asp residue at position P2, was found to be an almost specific enteropeptidase substrate: the selectivity of its hydrolysis was 98 (Table 2).

DISCUSSION

Secondary specificity of serine proteinases. The specificity of serine proteinases, such as chymotrypsin and trypsin, is largely determined by the P1–S1 interaction. In the case of trypsin, amino acid residues Asp189, Gly216, and Gly226 form a negatively charged S1-binding site, which is responsible for the specificity of binding by trypsin of substrates containing Arg and Lys at P1. However, the interaction between proteinase and substrate continues beyond the S1 site and often involves additional binding sites. Six or seven amino acid residues, located at both sides of the peptide chain hydrolyzed (i.e., at positions P3-P3′), interact with the enzyme. The secondary interactions do not play a key role in the determination of specificity of chymotrypsin and trypsin (although it is possible to speak of certain advantages and disadvantages). However, binding at these sites markedly increases catalytic efficiency of the enzyme. Good peptide substrates of serine proteinases are hydrolyzed with the k_{cat}/K_m ratio about 10^8-10^9 M^{-1}·min^{-1} [27]. Note that hydrolysis efficiency increases chiefly due to an increase in the catalytic component k_{cat} (≈6000 min^{-1}) rather than due to an increase in the affinity (K_m). It still remains obscure how such remote interactions contribute to acceleration of catalysis, although there are many hypotheses accounting for this phenomenon [25]. Importantly, the S1 and Sn–Sn′ binding sites in serine proteases function in coordination: k_{cat}/K_m increases with an increase in the peptide length only in the case of a proper P1/S1 interaction (for trypsin, for example, in the case of substrates containing Arg or Lys at position P1 [25]).

X-Ray analysis showed that the peptide substrate fragment P3-P3′ forms an antiparallel β-sheet with the secondary binding sites to the enzyme (S3-S3′), with side groups of neighboring amino acid residues of substrate being oppositely directed. Bulky side chains (e.g., in P1′ and P3′) compete for the same site on the enzyme surface, the specificities of S1′ and S3′ are the same, etc. The chemical nature of P2-P*n*- and P1′-P*n*′-specificities of trypsin has been intensively studied using both substrate series [26-28] and (in the case of P1′-P*n*′) acyl transfer to peptide nucleophiles [29-32].

The main chain of peptide substrate P1-P3 forms β-sheet with the main chain of residues 214-216 of the enzyme. This process is accompanied by formation of hydrogen bonds between the oxygen atom of the carbonyl of Ser214 and NH group of residue P1 of the substrate, NH of Trp215 and carbonyl of P3, and carbonyl of Gly216 and NH of P3. These interactions are a common feature of chymotrypsin-type proteinases; they are required for effective hydrolysis of the substrate. Our data on a low efficiency of enteropeptidase hydrolysis of substrate AT, which contains only the P2 residue (Table 1), show that an extended secondary site of the light

enteropeptidase chain has much in common with similar sites of trypsin and chymotrypsin [25].

The S2-S3 site of chymotrypsin and trypsin represents a shallow hydrophobic groove bounded by His57, Trp215, and Leu99. For this reason, the most preferable amino acid residues at position P2 in these enzymes are aliphatic apolar residues (e.g., Leu). However, Pro is the best residue at position P2 [27, 28]. The P2 position between residues Leu99 and Trp215 in chymotrypsin and trypsin is unfavorable for bulk aromatic residues [27]. Apparently, residue 99 is one of the key determinants of the secondary specificity of trypsin-type serine proteinases. For example, unlike trypsin, this position in kallikrein is occupied by Tyr; in this case, the Phe residue at position P2 of the substrate is the best, getting in a "sandwich" between Tyr99 and Trp215 [27]. Factor Xa also contains the Tyr residue at position 99, which is involved in the formation of a secondary aromatic site for ligand binding by this enzyme [33]. Special attention should be paid to enteropeptidase, containing a secondary positively charged site (Lys889 of the light chain) at position 99, which interacts with the negatively charged residues P2-P5. In addition, the enteropeptidase light chain contains Phe instead of Trp at position 215. According to X-ray data, its phenyl ring serves as a hydrophobic platform supporting the side group of the secondary site Lys99 [8]. The well-known experimental fact that negatively charged amino acid residues (at least two) at positions P2-P5 have an extremely adverse effect on the efficiency of trypsin hydrolysis of such substrates may be accounted for by negative electrostatic effect on the binding of Arg/Lys residues, located at position P1, to the S1 site. It is known that Ca^{2+} binding by Asp residues located at position P2-P3 decreases K_m by a factor of 3-4 but has no effect on k_{cat} [24]. However, even a single Asp or Glu both at P2 and P1′ does not facilitate effective hydrolysis of the corresponding peptides by trypsin [26].

The S3 site of trypsin and chymotrypsin does not display a pronounced specificity and may bind even D-amino acids; the side chain of the P3 residue is exposed from the active site [25]. However, the proline residue is unfavorable at this position because the hydrogen bond required for β-sheet formation cannot be formed in this case [27].

Similarly, the leaving C-terminal group of the substrate, P1′-Pn′, interacts with the S1′-Sn′ sites of the enzyme, increasing by three orders of magnitude the efficiency of hydrolysis of the substrates capable of realizing these interactions [30]. As a result of β-sheet formation, P1′ and P3′ become unidirectional, so that the S1′ and S3′ sites overlap. These two binding sites in trypsin and chymotrypsin form loops 40 (residues 34-41) and 60 (residues 58-68), which synergistically determine the S1′ preference of trypsin (both anionic and cationic) with respect to bulky hydrophobic aliphatic or aromatic residues (specifically, Met, Phe, and Leu) [32]. The preferred substrates of chymotrypsin are those containing Arg or Lys at position P1′, because they can electrostatically interact with Asp35 and Asp64 [30, 32]. Mutations at these secondary sites can lead to target enzymes with specified properties (for example, mutant rat trypsin K60E with high specificity hydrolyzed the Arg–Arg bond [29, 32]).

The P2′ residue is oppositely directed and interacts with loop 150. However, it is believed that the main factor that determines hydrolysis efficiency is the hydrogen bond formed between NH of the main chain of residue P2′ and the carbonyl oxygen of the Phe41 main chain in the enzyme [25, 29-31]. In view of his, the Pro residue will be extremely unfavorable at this position; acyl transfer to such nucleophilic agents as H-Ala-OH and H-Ala-OMe does not occur, contrary to the case of H-Ala-NH_2 and elongated peptides that are able to realize this hydrogen bond [31]. Apparently, this fact accounts for the differences (by two orders of magnitude) in the efficiency of hydrolysis of substrates of trypsin and enteropeptidase after their elongation by a single P2′ residue (Table 1). Interestingly, position 41 in the enteropeptidase light chain is occupied by Val instead of Phe. It should be noted that loops 40, 60, and 150 of this enzyme significantly differ from those in trypsin and chymotrypsin. Earlier, data on S′ specificity of enteropeptidase were practically absent, except for its high tolerance to the nature of the amino acid residue at position P1′ [6].

The data obtained in this work may be summarized as follows.

1. Using peptide substrates of enteropeptidase with the general formula A-(Asp/Glu)$_n$-Lys(Arg)-↓-B (n = 1-4), containing a linker of varying length, we showed that enteropeptidase, similar to other serine proteinases, has an extended secondary site interacting with six or seven amino acid residues, which are located at both sides of the hydrolyzed peptide bond (P3/4-P3′). However, this effect is manifested only in the case of substrates with a truncated linker, containing one or two Asp/Glu residues upstream of the Lys/Arg bond hydrolyzed. In the case of standard substrates of enteropeptidase, containing four negatively charged Asp/Glu residues at positions P2-P5, the factor that determines hydrolysis efficiency is electrostatic interaction between these residues and the secondary site Lys99 in the enteropeptidase light chain.

2. The secondary specificity of enteropeptidase differs from the secondary specificity of trypsin. The synthetic chromophore substrate G_5DK-F(NO_2)G is an effective substrate of enteropeptidase: k_{cat}/K_m for this substrate is 2380 $mM^{-1}\cdot min^{-1}$, which is greater than the corresponding value obtained for the fusion protein $PrAD_4K$-P26 (1260 $mM^{-1}\cdot min^{-1}$). Hydrolysis of this substrate is more than sixfold more effective than hydrolysis of peptides MLTAEEK-AA and LTAEEK-AAV (Fig. 2). Conversely, trypsin hydrolyzes the last two substrates more effectively: k_{cat}/K_m = 1700 $mM^{-1}\cdot min^{-1}$ in the case of MLTAEEK-AA and only 24.2 $mM^{-1}\cdot min^{-1}$ in the case

of G_5DK-F(NO_2)G. The data summarized in Table 2 show the absence of selectivity of hydrolysis of LTAEEK-A, an approximately fivefold preference of trypsin hydrolysis in the case of MLTAEEK-AA, and, by contrast, preference of enteropeptidase hydrolysis in the case of peptides G_5DK-F(NO_2)G and GD_4K-F(NO_2)G (almost by two and three orders of magnitude, respectively).

3. The ability of enteropeptidase to hydrolyze with high efficiency the peptide bond downstream of Lys/Arg residues preceded by 1-3 Asp/Glu residues hampers the use of this enzyme for targeted hydrolysis of fusion proteins that contain the enteropeptidase linker -DDDDK- [2-5]. It should be also taken into account that both the carrier protein and the target product may undergo degradation if they contain the sequences -(Asp/Glu)$_n$-Lys(Arg)-, where n = 1-3.

4. On the other hand, this property makes it possible to use enteropeptidase for sequencing natural proteins. Unlike trypsin, which hydrolyzes polypeptide molecule at the bonds located downstream of all Lys/Arg residues, enteropeptidase will produce large peptide fragments.

In conclusion, it should be noted that k_{cat}/K_m values determined for the most effective synthetic substrates of enteropeptidase, which contain four negatively charged amino acid residues at positions P2-P5 ((5-7)·10^6 $M^{-1}\cdot min^{-1}$) are nevertheless significantly lower than the standard values of this parameter determined for good peptide substrates of serine proteinases (~10^8-10^9 $M^{-1}\cdot min^{-1}$ [25]). Only in the case of its physiological substrate, trypsinogen, enteropeptidase realizes the entire reserve of hydrolysis efficiency: k_{cat}/K_m for trypsinogen activation is $3\cdot10^8$ $M^{-1}\cdot min^{-1}$ [10, 23]. As we showed earlier [7, 9, 10], this process requires the involvement of one more secondary substrate-binding site (S_{II}), which interacts solely with trypsinogen and is located in the region 118-465 in the heavy chain of the enzyme.

This study was supported by the Russian Foundation for Basic Research (project No. 02-04-48553).

REFERENCES

1. Kunitz, M. (1939) *J. Gen. Physiol.*, **22**, 429-446.
2. Sharma, A., Khoury-Christianson, A. M., White, S. P., Dhanjal, N. K., Huang, W., Paulhiac, C., Friedman, E. J., Manjula, B. N., and Kumar, R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 9337-9341.
3. Uegaki, K., Nemoto, N., Shimizu, M., Wada, T., Kyogoku, Y., and Kobayashi, Y. (1996) *FEBS Lett.*, **379**, 47-50.
4. Safi, W., Maiorano, J. N., and Davidson, W. S. (2001) *J. Lipid Res.*, **42**, 864-872.
5. Agnihotri, R., Crawford, H. C., Haro, H., Matrisian, L. M., Havrda, M. C., and Liaw, L. (2001) *J. Biol. Chem.*, **276**, 28261-28267.
6. Hosfield, T., and Lu, Q. (1999) *Analyt. Biochem.*, **269**, 10-16.
7. Mikhailova, A. G., and Rumsh, L. D. (2000) *Appl. Biochem. Biotechnol.*, **88**, 159-174.
8. Lu, D., Futterer, K., Korolev, S., Zheng, X., Tan, K., Waksman, G., and Sadler, J. E. (1999) *J. Mol. Biol.*, **292**, 361-373.
9. Mikhailova, A. G., and Rumsh, L. D. (1998) *Bioorg. Khim.*, **24**, 282-287.
10. Mikhailova, A. G., and Rumsh, L. D. (1999) *FEBS Lett.*, **442**, 226-230.
11. Lu, D., and Sadler, J. E. (1998) in *Handbook of Proteolytic Enzymes* (Barret, A. J., Rawlings, N. D., and Woessner, J. F., eds.) Academic Press, London, pp. 50-54.
12. Likhareva, V. V., Mikhailova, A. G., and Rumsh, L. D. (2002) *Vopr. Med. Khim.*, **48**, 561-569.
13. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001) *Protein Eng.*, **14**, 993-1000.
14. Mutt, V., Tatemoto, K., Carlquist, M., and Light, A. (1981) *Biosci. Rep.*, **1**, 651-659.
15. Likhareva, V. V., Vaskovsky, V. B., Shepel', N. E., Garanin, S. K., Mikhailova, A. G., and Rumsh, L. D. (2003) *Bioorg. Khim.*, **29**, 129-134.
16. Likhareva, V. V., Mikhailova, A. G., Vaskovsky, B. V., Garanin, S. K., and Rumsh, L. D. (2002) *Lett. Peptide Sci.*, **9**, 71-76.
17. Hofmann, Th., and Hodges, R. S. (1982) *Biochem. J.*, **203**, 603-610.
18. Mikhailova, A. G., Vorotyntseva, T. I., Bessmertnaya, L. Ya., and Antonov, V. K. (1984) *Biokhimiya*, **49**, 1483-1487.
19. Eisenthal, R., and Cornish-Bowden, A. (1974) *Biochem. J.*, **139**, 715-720.
20. Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R., and Rutter, W. J. (1985) *Science*, **222**, 291-297.
21. Grant, D. A. V., and Hermon-Taylor, J. (1979) *Biochim. Biophys. Acta*, **567**, 207-215.
22. Baratty, J., and Maroux, S. (1976) *Biochim. Biophys. Acta*, **452**, 488-496.
23. Lu, D., Yuan, X., Zheng, X., and Sadler, J. E. (1997) *J. Biol. Chem.*, **272**, 31293-31300.
24. Abita, J. P., Delaage, M., and Lazdunski, M. (1969) *Eur. J. Biochem.*, **8**, 314-324.
25. Hedstrom, L. (2002) *Chem. Rev.*, **102**, 4501-4523.
26. Bianchini, E. P., Louvain, V. B., Marque, P.-E., Juliano, M. A., Juliano, L., and Le Bonniec, B. F. (2002) *J. Biol. Chem.*, **277**, 20527-20534.
27. Fiedler, F. (1987) *Eur. J. Biochem.*, **163**, 303-312.
28. Pozsgay, M., Cs. Szabó, G., Bajusz, S., Simonsson, R., Gáspár, R., and Elodi, P. (1981) *Eur. J. Biochem.*, **115**, 497-502.
29. Kurth, T., Grahn, S., Thormann, M., Ullman, D., Hofmann, H.-J., Jakubke, H.-D., and Hedstrom, L. (1998) *Biochemistry*, **37**, 11434-11440.
30. Kurth, T., Ullman, D., Jakubke, H.-D., and Hedstrom, L. (1997) *Biochemistry*, **36**, 10098-10104.
31. Schellenberger, V., Turck, C. W., and Rutter, W. J. (1994) *Biochemistry*, **33**, 4251-4257.
32. Grahn, S., Kurth, T., Ullman, D., and Jakubke, H.-D. (1999) *Biochim. Biophys. Acta*, **1431**, 329-337.
33. Reyda, S., Sohn, Ch., Kiebe, G., Rall, K., Ullman, D., Jakubke, H.-D., and Stubbs, M. T. (2003) *J. Mol. Biol.*, **325**, 963-977.