


Review

# Pseudotrypsin: A Little-Known Trypsin Proteoform

Zdeněk Perutka  and Marek Šebela \* 

Department of Protein Biochemistry and Proteomics, Centre of the Region Haná for Biotechnological and Agricultural Research, Faculty of Science, Palacký University, Šlechtitelů 27, 783 71 Olomouc, Czech Republic; Perutka.Zdenek@seznam.cz

\* Correspondence: marek.sebela@upol.cz; Tel.: +420-585-634-927

Academic Editor: Paolo Iadarola

Received: 11 September 2018; Accepted: 9 October 2018; Published: 14 October 2018



**Abstract:** Trypsin is the protease of choice for protein sample digestion in proteomics. The most typical active forms are the single-chain  $\beta$ -trypsin and the two-chain  $\alpha$ -trypsin, which is produced by a limited autolysis of  $\beta$ -trypsin. An additional intra-chain split leads to pseudotrypsin ( $\psi$ -trypsin) with three chains interconnected by disulfide bonds, which can be isolated from the autolyzate by ion-exchange chromatography. Based on experimental data with artificial substrates, peptides, and protein standards,  $\psi$ -trypsin shows altered kinetic properties, thermodynamic stability and cleavage site preference (and partly also cleavage specificity) compared to the above-mentioned proteoforms. In our laboratory, we have analyzed the performance of bovine  $\psi$ -trypsin in the digestion of protein samples with a different complexity. It cleaves predominantly at the characteristic trypsin cleavage sites. However, in a comparison with common tryptic digestion, non-specific cleavages occur more frequently (mostly after the aromatic residues of Tyr and Phe) and more missed cleavages are generated. Because of the preferential cleavages after the basic residues and more developed side specificity, which is not expected to occur for the major trypsin forms (but may appear anyway because of their autolysis),  $\psi$ -trypsin produces valuable information, which is complementary in part to data based on a strictly specific trypsin digestion and thus can be unnoticed following common proteomics protocols.

**Keywords:** autolysis; chain; cleavage; digestion; peptide; proteoform; pseudotrypsin; specificity; trypsin

## 1. Cleavage Specificity of Trypsin

Trypsin, a serine protease, is commonly used as an important enzymatic reagent in biochemistry and biology. It is almost indispensable especially for the digestion of protein samples to peptides in bottom-up proteomics [1]. Apart from this application, trypsin is a tool in working with cell cultures. During trypsinization, surface adhesion proteins are degraded, which allows adherent cells to be detached from each other and the walls of plastic containers or plates in which they are being cultured. In industry, interestingly, trypsin is applied to hydrolyze allergenic proteins for the production of hypoallergenic milk [2]. In proteomics, sample digests are typically analyzed for protein identification by nanoflow liquid chromatography coupled to tandem mass spectrometry (nLC-MS/MS). MS-based data on peptides are searched against amino acid sequence databases, which benefits from the relative stringent cleavage specificity of trypsin as the search algorithms incorporate the cleavage rule as a filtering criterion. According to a study with complex samples from 2004, the enzyme cleaves peptide bonds in proteins at pH 8.3 exclusively at the carboxyl end of arginine and lysine residues [3]. This is in agreement with the canonical trypsin cleavage rule postulated a long time ago [4] even though it was built on results obtained at that time only with a limited amount of substrates [5]. Thus, it has long been accepted that trypsin does not cleave before proline and its activity is suppressed either

if a cysteine appears next to Arg/Lys or the basic residue is N- or C-terminally adjacent to an acidic residue. This old rule was questioned when a large data set of 14.5 million MS/MS spectra of peptides from *Shewanella oneidensis* was processed to statistically evaluate the cleavage sites [5]. Interestingly, numerous cleavages before proline were found. Their number was even higher than that referring to the cleavages before cysteine.

An average length of tryptic peptides is 14 amino acids. This number has been deduced from an in silico digestion of human proteins in the UniProt database [1]. Because of this reasonable size as well as the presence of a positive charge at the C-terminal Arg or Lys, which enhances the ionization process in the positive ionization mode, tryptic peptides are highly amenable to mass spectrometric measurements. In fact, there are at least two defined positive charges in tryptic peptides (at both N- and C-termini), which is favorable for a good fragmentation in MS/MS analyses [1,2].

## 2. Nonspecific and Missed Cleavage Sites

In addition to the cleavages after Arg or Lys, proteomics studies have often reported the formation of semitryptic and nonspecific peptides during the digestion process involving trypsin [1,6]. The semitryptic cleavage assumes that one of the cleavage sites is tryptic, but the other site may be at any residue. A minor chymotrypsin or chymotrypsin-like activity yields nonspecific cleavages C-terminal to phenylalanine, tyrosine, tryptophan, or leucine residues [7]. This can result either from the presence of a chymotrypsin contamination, which is variable in trypsin preparations supplied by different vendors [1], or pseudotrypsin ( $\psi$ -trypsin), a product of trypsin autolysis, which possesses such an activity in addition to the characteristic trypsin properties [8]. The presence of nonspecific peptides in tryptic digests (excluding C-terminal peptides) is also elucidated by a secondary non-enzymatic cleavage between Asp and Pro residues yielding peptides with an N-terminal proline. This is because of the lability of the respective bond, which is easily hydrolyzed in solution as well as broken in the gas phase [3].

When the protein substrate is not cleaved to a completion, missed cleavages occur, which make the assignment of experimental data to amino acid sequence databases less specific and straightforward. Missed cleavage sites were investigated for example by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using human cell line or *Mycobacterium* proteins separated by two-dimensional gel electrophoresis and digested in-gel by a porcine trypsin [9]. The analysis showed that about 90 % of the detected peptides with missed cleavage sites could be attributed to the following sequence motifs: (1) Arg or Lys with a neighboring proline at the C-terminal side, (2) two successive basic residues (Arg-Arg, Arg-Lys, Lys-Arg, Lys-Lys), and (3) Arg or Lys with an aspartic acid or glutamic acid residue at either N-terminal or C-terminal side. When processing peptide MS- and MS/MS-based data by database searches (during which theoretical peptide sequences are generated according to the selected input cleavage rules), the user may, among others, adjust a maximum number of missed cleavage sites. Usually, a setting of 0, 1, or 2 missed cleavages for peptides is recommended. The presence of missed cleavages in tryptic peptides represents a challenge in quantitative MS-based proteomics, which uses peptides as surrogates for their parent proteins [10]. Under optimal conditions, peptides should be stoichiometric with the parent protein to enable accurate quantitation. However, if a protein is digested into multiple overlapping peptides, the specific signal is attenuated and in consequence the quantification becomes underestimated.

There are also side reactions of trypsin known, for example its transpeptidase activity, which yields the addition of a single amino acid (Arg or Lys) to peptides in the reaction mixture. There are also dipeptides added or two longer peptides are combined together. The additions have been described for both N- and C-termini of peptides [7]. However, the incidence of the modified peptides is much lower compared to their unmodified counterparts.

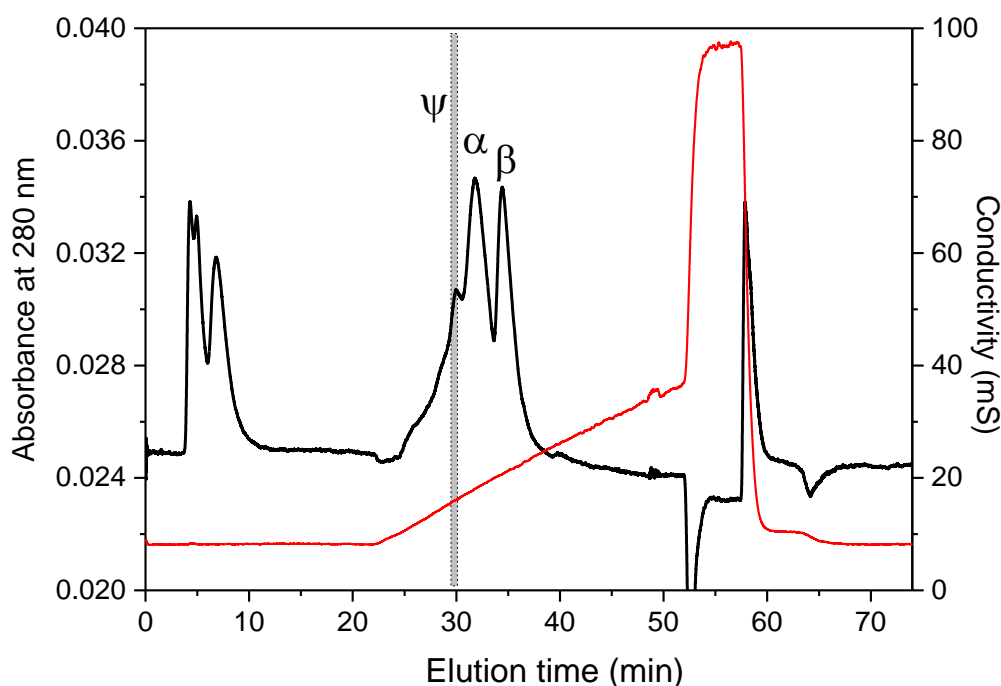
### 3. Trypsin Forms

Wilhelm Kühne, who discovered trypsin and coined its name in 1876, noticed that the enzyme was produced as an inactive zymogen (trypsinogen) in pancreatic cells [11]. In the 1930s, Moses Kunitz and John H. Northrop elaborated procedures to isolate both trypsinogen and trypsin in crystalline forms, which largely contributed to the development of enzymology [12]. Trypsinogen is stable at acidic pH (2–4). In a neutral or alkaline solution, it is activated by a limited proteolysis catalyzed by either duodenal enteropeptidase (enterokinase) or trypsin itself [4]. Thus, in the latter case, an autocatalytic process occurs. The complete amino acid sequence of bovine trypsinogen was deduced from peptide sequencing experiments by several independent groups in the 1960s [13,14] with later corrections at ambiguous positions (some Asn/Asp and Gln/Glu were not distinguished in early studies because of limitations of the sequencing methods used at that time). In addition to the dominant cationic trypsinogen, a minor anionic trypsinogen is produced in bovine pancreas [15]. The UniProt accession numbers are P00760 (TRY1\_BOVIN) and Q29463 (TRY2\_BOVIN), respectively; the corresponding mature trypsin sequences share 72% identity.

The cationic trypsinogen sequence spans the length of 229 amino acids (Figure 1). Altogether, there are six disulfide bonds in the molecule connecting the following cysteine residues: 13 and 143, 31 and 47, 115 and 216, 122 and 189, 154 and 168 plus 179 and 203 [14]. By the removal of the N-terminal hexapeptide VDDDDK, the single chain  $\beta$ -trypsin is formed as the predominant product of trypsinogen activation. Trypsin autolysis with a cleavage between Lys-131 and Ser-132 (numbered according to the trypsinogen convention) produces  $\alpha$ -trypsin, which has two chains connected by disulfide bonds) [4]. A further cleavage at the bond Lys-176–Asp-177 yields  $\psi$ -trypsin with three discrete chains [16]. Another known active trypsin forms include for example the two chain  $\delta$ -trypsin [17–19] and  $\gamma$ -trypsin [20] with chain splits at the bonds Arg-105–Val-106 and Lys-155–Ser-156, respectively. On the contrary, a cleavage between Lys-49 and Ser-50 has been shown to inactivate the enzyme [18]. More degraded autolysis products are considered inactive [4]. Bovine trypsin is characterized by a molar absorption coefficient  $\epsilon_{280}$  of  $40,000 \text{ mol}\cdot\text{L}^{-1}\cdot\text{cm}^{-1}$  [21] resulting from the presence of 4 Trp, 10 Tyr, and 12 Cys residues (oxidized to disulfides in the protein) in the sequence. A certain degree of variability in these numbers can be found for trypsin enzymes from different mammalian or other sources of origin [4]. All mentioned variants are often called “trypsin isoforms” in the literature but this is not satisfactory. According to the International Union of Pure and Applied Chemistry (IUPAC), the term isoform refers only to genetic differences and not to a variation at the protein level. Hence, the real isoforms are e.g., the cationic and anionic trypsin. To solve this terminological confusion, the term “proteoform” has recently been introduced, which covers all molecular forms encoded by a single gene, including changes due to genetic variations, alternatively spliced transcripts and posttranslational modifications [22].



the given order of elution times) at pH 7.1 was 45 min long. We have recently replaced the HEMA-BIO 1000 SB column by a Uno S12 column (15 × 68 mm), which allowed reducing the separation time, at the expense of resolution, but still yielded a pure  $\psi$ -trypsin (Figure 2). In this case, however, the use of a gradient elution was necessary—the buffer B contained 1 M NaCl (Perutka et al., unpublished results). In the 1970s, a French group, which studied kinetic properties of  $\psi$ -trypsin at that time, developed a convenient purification method based on affinity chromatography [26,27]. The yield was 15–20 % of the initial fully active trypsin. To prepare the affinity column, a trypsin inhibitor from egg-white chicken ovomucoid was attached to aminoethyl-cellulose using glutaraldehyde as a coupling reagent. The equilibration buffer for separation runs was 0.1 M Tris-HCl, containing 50 mM CaCl<sub>2</sub>, pH 7.1.  $\psi$ -Trypsin appeared already in the flow-through fraction (just after the void volume). Thus, it was necessary to include a size-exclusion chromatographic step prior to the affinity separation in order to remove low-molecular-weight contaminants such as peptides. In contrast, the proteoforms  $\alpha$  and  $\beta$  remained bound and could be eluted (as an unresolved mixture) by applying an acidic elution buffer of pH 2.3 [27]. This low affinity of  $\psi$ -trypsin is in agreement with results on the formation of its complex with pancreatic trypsin inhibitor, where the corresponding dissociation constant was increased by five orders of magnitude compared to that for a mixture of  $\alpha$ - and  $\beta$ -trypsin [28].



**Figure 2.** Ion-exchange chromatography of trypsin autolyzate. The depicted separation was performed on a Uno S12 column (15 × 68 mm; Bio-Rad, Hercules, CA, USA) using a linear gradient of NaCl (0–1 M) in 100 mM Tris-HCl, pH 7.1, containing 10 mM CaCl<sub>2</sub>. The gray box indicates a time window, in which the respective  $\psi$ -trypsin fraction was collected. The black line shows absorbance at 280 nm, the red line refers to conductivity of the eluate.

## 5. Molecular Properties and Structure of Trypsin and Pseudotrypsin

Accurate experimental molecular mass values of bovine  $\psi$ -trypsin were first determined by electrospray ionization (ESI)-MS [29,30]. For  $\beta$ -trypsin,  $\alpha$ -trypsin, and  $\psi$ -trypsin, the following average numbers (relative monoisotopic molecular masses) were obtained in the given order: 23296, 23310, and 23325 [29] or 23294, 23312 and 23328 [30]. These numbers are in accordance with sequence-based calculated values of 23293, 23311, and 23329, respectively [30], which reflect a relative mass difference of 18 resulting from each consecutive autolytic cleavage. For  $\psi$ -trypsin, the most recent data are  $23332 \pm 4$  (MALDI-MS) and  $23330 \pm 0.1$  (ESI-MS) [8]. Thus, based on an accurate molecular mass

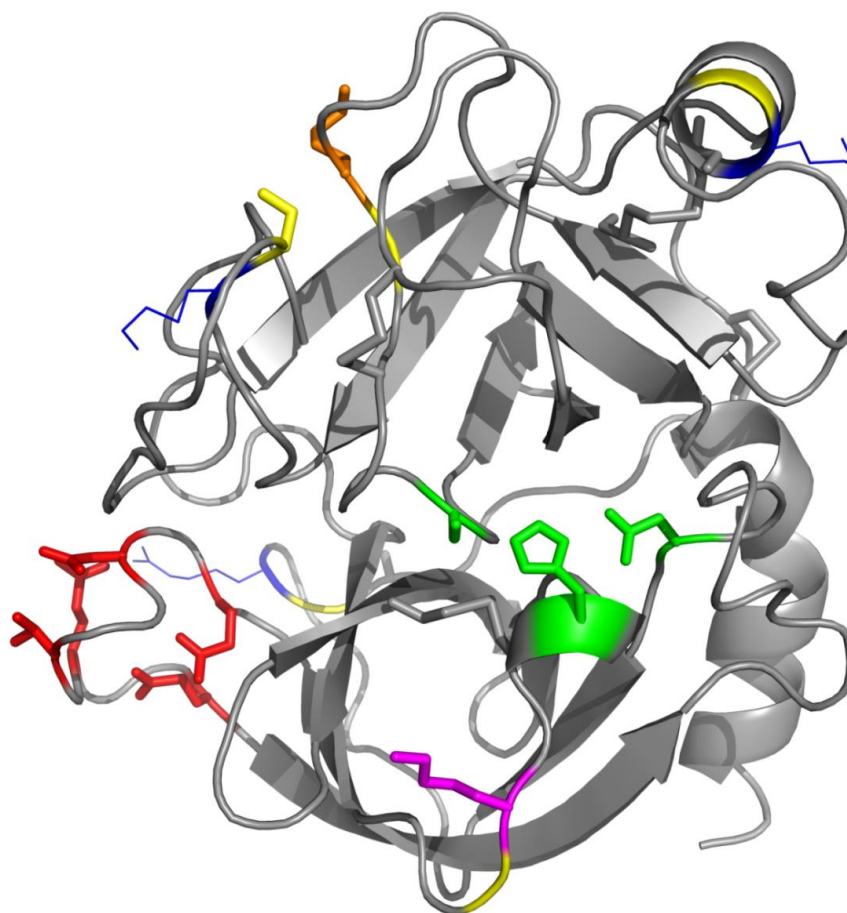


determination, the purity of  $\psi$ -trypsin preparations can easily be evaluated, also to rule out the presence of a chymotrypsin contamination coming from the original trypsin material.

Isoelectric points of proteins can be estimated from their amino acid sequences for example using the software tool ProtParam (<https://web.expasy.org/cgi-bin/protparam/>). A theoretical *pI* value of the cationic bovine trypsin (Uniprot accession number P00760, positions 24–246) is 8.69. Similarly, for porcine trypsin (Uniprot accession number P00761, positions 9–231), the result is 8.26. Interestingly, higher experimental values of 10.0/10.5 [31,32] were published for the bovine enzyme and 10.2/10.8 for the porcine enzyme [33,34]. For  $\psi$ -trypsin, no experimental data on *pI* are available to our knowledge, but similar values could be expected. The anionic bovine trypsin (Uniprot accession number Q29463, positions 24–247) has a theoretical *pI* value of 4.90, which agrees well with an experimental result [35].

Chymotrypsin and trypsin were among first proteins with experimentally determined spatial structures. The crystal structure of bovine chymotrypsin appeared already in 1967 [36] and was refined later on [37]. At that time, a similarity in the three-dimensional folding of trypsin and chymotrypsin could be assumed because of the amino acid sequence homology and matching positions of disulfide bonds. The crystal structures of bovine  $\beta$ -trypsin and trypsinogen were solved in the 1970s [38–40]. Figure 3 shows a view of the trypsin structure (PDB accession code 1AQ7 [41]) visualized using PyMol 1.3. Trypsin is a globular protein. Its overall fold comprises two six-stranded Greek-key  $\beta$ -barrels [42]. The active site with the catalytic triad of amino acids is located between the two barrels. His-57 and Asp-102 belong to the N-terminal barrel whereas Ser-195 originates from the C-terminal barrel (this numbering is according to the chymotrypsinogen convention, see in Walsh and Neurath) [13]. Helices represent only minor secondary structure components, for example at the C-terminus. The enzyme contains a calcium ion, which is important for activity. Its coordination chemistry involves several residues from the calcium-binding loop [38]. The calcium ion interacts with the side-chain oxygens (in the trypsinogen numbering) of Glu-58, Glu-65 (this one via a coordinated water molecule), and Glu-68 plus the carbonyl oxygens of Asn-60 and Val-63. No crystal structure of  $\psi$ -trypsin has been solved up to now. As this proteoform contains two chain splits (between: 1. Lys-131 and Ser-132, 2. Lys 176 and Asp-177; according to the trypsinogen numbering), the whole molecule is loosened somehow in comparison to that of  $\beta$ -trypsin. The bond Lys-176–Asp-177 is located close to the anionic binding site (i.e., specificity site: Asp-177, which is the position 189 when expressed in the chymotrypsinogen numbering). Upon the autolytic splitting, the binding site arrangement is disconnected. In consequence, the affinity for polypeptide trypsin substrates is lowered and the cleavage specificity broadened [8,24]. This structural alteration does not prevent from binding of a pancreatic trypsin inhibitor; only the dissociation constant of the enzyme-inhibitor complex is increased [24,28].  $\psi$ -Trypsin still keeps a certain level of specificity, which is based on hydrophobic interactions, as confirmed using synthetic ester substrates [27].

After trypsinogen activation, the new N-terminal residue (Ile-16, in chymotrypsinogen numbering) inserts into a cleft, where it establishes an ion pair (via  $\alpha$ -amino group) with Asp-194 next to the catalytic serine. This results in a conformational rearrangement. The amino group of Gly-193 moves into a position, which completes the oxyanion hole at the active site [2]. The hole is formed by the trypsin amide hydrogens of Gly-193 and Ser-195 in favor of stabilization of the developing negative charge on the carbonyl oxygen atom of the cleaved substrates.



**Figure 3.** The crystal structure of bovine  $\beta$ -trypsin. This overall view was made in PyMOL 1.3 using a pdb formatted file (accession number 1AQ7 [41]) downloaded from the PDB database (<http://www.rcsb.org/pdb>). Trypsin fold comprises two  $\beta$ -barrels. The active site with the catalytic triad of amino acids is located between them (in the center of this figure) and highlighted in green. Red color on the bottom left shows the calcium-binding residues. The KD bond, which is cleaved to generate  $\psi$ -trypsin is expressed as an orange side chain (Lys-176) at a yellow backbone segment (Asp-177). Blue-line side chains at other yellow segments indicate the presence of the basic residues, where the  $\beta$ -trypsin polypeptide chain is cleaved to produce  $\alpha$ -trypsin (Lys-131, above the calcium site and not far away from Lys-176 in this projection),  $\gamma$ -trypsin (Lys-155, top right, in the helix), and  $\delta$ -trypsin (Arg-105, behind the calcium site). Finally, magenta color at the bottom indicates Lys-49, which is disconnected from Ser-50 to yield the inactive autolytic form.

## 6. Pseudotrypsin Activity with Artificial Substrates (Enzyme Kinetics)

Compared to  $\alpha$ -trypsin and  $\beta$ -trypsin, the overall structural change resulting from the additional intrachain split in  $\psi$ -trypsin yields differences in the activity and specificity. This was evaluated already in the 1960s and 1970s by measuring kinetic parameters for low-molecular-weight substrates (Table 1). Smith and Shaw [16] recognized during the chromatographic purification that  $\psi$ -trypsin did not show any measurable activity with  $N^\alpha$ -benzoyl-D,L-arginine-4-nitroanilide (Bz-Arg-pNA) as a substrate under experimental conditions optimized for  $\alpha$ -trypsin (i.e., no amidase activity). Nevertheless, they could demonstrate a hydrolytic activity of  $\psi$ -trypsin by detecting a stoichiometric incorporation of [ $^{14}\text{C}$ ] diisopropyl fluorophosphate at the active site, which is typical for serine proteases or esterases in general. However, the rate of  $^{14}\text{C}$  incorporation was very slow, which indicated a decreased reactivity of  $\psi$ -trypsin compared to  $\alpha$ -trypsin. Similarly, a slow conversion of the active site titrant *p*-nitrophenyl-*p'*-guanidinobenzoate (NPGb) was observed. Comparative kinetic data measured with several artificial ester substrates are shown in Table 1. For example, the affinity of  $\psi$ -trypsin for the

cationic  $N^\alpha$ -benzoyl-L-arginine ethyl ester (Bz-Arg-OEt) represented by the determined Michaelis constant value is lower by three orders of magnitude than that of  $\alpha$ -trypsin [16]. Such a big difference in affinity was observed also for carboxybenzyl-L-lysine esters or  $N^\alpha$ -*p*-tosyl-L-arginine methyl ester [27]. Not only the affinity, but also the activity (catalytic constant) is often largely different. A typical chymotrypsin substrate  $N^\alpha$ -acetyl-L-tyrosine ethyl ester (Ac-Tyr-OEt) is hydrolyzed by  $\alpha$ -trypsin almost 3000 times faster than by  $\psi$ -trypsin (Table 1). At the same time, the respective  $K_m$  values are comparable as the binding of this neutral substrate is not affected so much by the disconnection of Asp-177 (the specificity site) in  $\psi$ -trypsin [16]. As a consequence of the reduced affinity and activity towards cationic substrates, the efficiency constant values  $k_{cat}/K_m$  are decreased by many orders of magnitude (up to 6) for  $\psi$ -trypsin [16,27].

**Table 1.** Kinetic data for bovine  $\alpha$ -trypsin and  $\psi$ -trypsin with artificial substrates. This table is adapted from Smith and Shaw [16] and completed by results with non-fractionated trypsin [43]. All experiments were performed at pH 8.0 and 25 °C.

Enzyme	Substrate <sup>a</sup>	$k_{cat}$ (s <sup>-1</sup> )	$K_m$ (mol·L <sup>-1</sup> )	$K_i$ (mol·L <sup>-1</sup> )
$\alpha$ -Trypsin	Bz-Arg-OEt (0.05 M CaCl <sub>2</sub> )	24	$2.5 \times 10^{-6}$	-
$\psi$ -Trypsin	Bz-Arg-OEt (0.001 M CaCl <sub>2</sub> )	0.19	$1.1 \times 10^{-2}$	-
$\psi$ -Trypsin	Bz-Arg-OEt (0.05 M CaCl <sub>2</sub> , benzamidine)	-	-	$3.7 \times 10^{-2}$
Trypsin (non-fractionated)	Bz-Arg-OEt (0.025 M CaCl <sub>2</sub> )	14.6	$4.3 \times 10^{-6}$	-
$\alpha$ -Trypsin	Ac-Tyr-OEt (0.05 M CaCl <sub>2</sub> , 10% 2-propanol)	57	$6.2 \times 10^{-2}$	-
$\psi$ -Trypsin	Ac-Tyr-OEt (0.001 M CaCl <sub>2</sub> , 10% 2-propanol)	0.02	$3.7 \times 10^{-2}$	-
Trypsin (non-fractionated)	Ac-Tyr-OEt (0.05 M CaCl <sub>2</sub> , 5% dioxane)	14.5	$4.2 \times 10^{-2}$	-
Trypsin (non-fractionated)	Bz-Arg-pNA (benzamidine) <sup>b</sup>	-	-	$1.8 \times 10^{-5}$

<sup>a</sup> Further information on the composition of the reaction mixture is provided in parentheses; <sup>b</sup> This reaction was performed at pH 8.15 and 15 °C.

Benzamidine has been shown a potent competitive inhibitor of trypsin. It is approximately of the same size as the side chain of lysine and arginine and contains both a positively charged group and hydrophobic moiety in its structure. The  $K_i$  value of benzamidine for the reaction of trypsin with Bz-Arg-pNA is  $1.8 \times 10^{-5}$  mol·L<sup>-1</sup> [4]. In contrast, the binding of the inhibitor to  $\psi$ -trypsin is characterized by an increased  $K_i$  value  $3.7 \times 10^{-2}$  mol·L<sup>-1</sup>, which is in agreement with the  $K_m$  value for the neutral (and thus non-specific) substrate Ac-Tyr-OEt (Table 1). The irreversible trypsin inhibitor TLCK does not inactivate  $\psi$ -trypsin at all [16]. Further evidence of the reduced  $\psi$ -trypsin affinity to positively charged ligands was observed with a basic pancreatic trypsin inhibitor. The second-order rate constant of the association is decreased by a factor of 16 when  $\psi$ -trypsin is used instead of  $\alpha$ - and  $\beta$ -trypsin and the dissociation constant is increased by a factor of  $1.5 \times 10^5$  from  $6 \times 10^{-14}$  mol·L<sup>-1</sup> (a quasi-irreversible binding in the case of trypsin) to  $9 \times 10^{-9}$  mol·L<sup>-1</sup>. This value is similar to that of chymotrypsin, which also associates with this inhibitor but lacks the trypsin specificity site [28]. When the disulfide bond Cys-179–Cys-203 in trypsin is selectively reduced and the emerged cysteines subsequently carboxymethylated, the resulting enzyme derivative still binds the inhibitor efficiently. Interestingly, the observed kinetic parameters of the association and dissociation correspond to those determined for  $\psi$ -trypsin [28].

The reaction mechanism of trypsin is illustrated in Scheme 1 (based on that in [27]) Acylation and deacylation rate constants ( $k_2$  and  $k_3$ , respectively) were measured with the active site titrant NPGb and pure bovine  $\alpha$ -,  $\beta$ -, and  $\psi$ -trypsin preparations [27]. Whereas the acylation rate was found to be 1000 slower for  $\psi$ -trypsin than for the other proteoforms at optimum pH, the deacylation rates were rather comparable. A similar difference in the acylation rate was observed with *p*-acetoxypheylguanidine *p*-toluenesulfonate. This compound belongs to “inverse substrates” as it contains the specific cationic center within the leaving group instead of the acyl moiety [44]. Measurements with different





which confirmed the previous data obtained with glucagon [24]. Interestingly,  $\psi$ -trypsin provided 1.5-fold higher number of peptides containing missed cleavage (Arg and Lys) sites, which probably results from the lower substrate binding ability of this proteoform [8].

## 8. The Use of Pseudotrypsin for Protein Identification in Proteomics

At the present time,  $\psi$ -trypsin is not commonly used in biochemistry and proteomics as it is not commercially available and its purification takes a few days with a low yield of pure protein at the end. From 200 mg of bovine trypsin as a starting material [8], only milligrams of the final product could be obtained. The applicability of  $\psi$ -trypsin has a potential in protein identification experiments involving ESI- or MALDI-MS/MS as it generates more peptides with missed cleavage sites than native trypsin and also nonspecific peptides terminated mainly by Phe and Tyr residues (resembling partially the chymotrypsin mode of action). In consequence, higher sequence coverage values and increased number of matched peptides can be achieved. Interestingly, pseudotrypsin has been hypothesized to cause the cleavage of a Cys–Gly bond in NDRG1 protein [47]. Another possibility of application resides in studying posttranslational modifications. Such a modification may under a common routine interfere with trypsin digestion, for example when a phosphorylation occurs close to an arginine or lysine, and would then require selecting of another protease [48]. On the other hand, for its broader cleavage specificity,  $\psi$ -trypsin is not recommendable for mass spectrometry-based protein quantification experiments [49] because of the possible distribution of the same predicted canonical cleavage site in more peptides (multiple cleavage products due to missed cleavages or additional cleavage sites).

A comparative in-gel digestion of a gel fraction of rat urine proteins resulted in 22 identifications after the use of  $\psi$ -trypsin. The same number was reached with a commercial non-fractionated trypsin, but only 17 were identified in both cases. Hence a simple combination of the two digestions provided about 20 % more identification [8]. The total number of the matched peptides was 233 and 199, respectively. The numbers of the cleaved Arg and Lys sites were comparable (around 130); in the case of Phe and Tyr site cleavages, two times more peptides were produced by  $\psi$ -trypsin. Recently, this proteoform was applied to analyze nuclear proteins after a DNase treatment of barley nuclei and sodium dodecyl sulfate polyacrylamide gel electrophoresis of the released protein material (Perutka et al., unpublished results). Peptides from the digests were analyzed by MALDI and ESI MS/MS and the results compared with parallel tryptic digestions. The identified nonspecific peptides in the  $\psi$ -tryptic digests represented 15–20 % of the total peptide number (compared to 7 % for a standard trypsin). In agreement with previous reports [8,24], peptides with C-terminal Tyr and Phe residues (and also Leu) were found in a significant percentage representation but were only minor to those resulting from the characteristic trypsin cleavage after the basic Arg and Lys residues (Figure 4). The higher number of Arg-ending over Lys-ending peptides in the MALDI-TOF/TOF MS/MS results (but not in the ESI-based results; not shown) probably reflected the fact that Arg-peptides provide more intense signals in MALDI-TOF MS [50], and thus they are preferably selected for the subsequent data-dependent fragmentation. Database searches allowed identifying novel proteins which had not previously been recognized based on standard tryptic peptides and deposited in the database UNcleProt [51]. They accounted for around 10 % of all identifications. The digestion performance of  $\psi$ -trypsin was compared relatively to that of a commercial trypsin using MALDI-TOF MS-based quantification of bovine serum albumin peptides (Perutka et al., unpublished results). Tryptic digestion was performed in a buffer made of H<sub>2</sub><sup>18</sup>O [52]. During proteolysis, labeled standards were generated by incorporating <sup>18</sup>O isotope into the carbonyl group of the nascent peptides. The  $\psi$ -tryptic digest was made in a buffered H<sub>2</sub><sup>16</sup>O and then mixed equivolumetrically with the labeled standards. The ratios of non-labeled versus labeled peptides were calculated from the areas of isotopically resolved peaks in MALDI-TOF MS spectra. As a result, the observed overall overnight digestion performance of  $\psi$ -trypsin was found lower by around 20 %.

P1 amino acid	G	A	P	V	L	I	M	F	W	Y	S	T	C	N	Q	K	H	R	D	E
side chain	Nonpolar							Aromatic			Polar				Basic		Acidic			
$\Psi$ -trypsin	2.0%	0.9%	0.3%	0.5%	2.2%	0.1%	0.9%	2.1%	0.4%	4.2%	0.7%	0.6%	0.0%	1.1%	0.4%	32.7%	0.4%	49.2%	1.2%	0.1%
$\alpha/\beta$ trypsin	0.8%	0.2%	0.0%	0.2%	1.1%	0.0%	0.0%	2.4%	0.0%	2.1%	0.0%	0.1%	0.0%	0.7%	0.0%	31.3%	0.5%	60.5%	0.1%	0.0%

**Figure 4.** C-terminal amino acids (P1 cleavage site) of peptides in  $\psi$ -tryptic and tryptic digests. This table summarizes results obtained after in-gel digestions of barley nuclear proteins. The percentage values were averaged from data for almost 40 distinct protein fraction samples. MALDI TOF/TOF MS/MS allowed to identify 1199 and 1238 peptides, respectively. PEAKS Studio 8 software (Bioinformatics Solutions, Waterloo, ON, Canada) was used to process the MGF-formatted files by searching against a *Hordeum vulgare* protein sequence database downloaded from the National Center for Biotechnology Information, USA, at 1% peptide false discovery rate (Perutka, Z. et al.; unpublished results). Peptide sequences containing C-termini of the identified proteins were not included in the calculation.

## 9. Concluding Remarks

Early studies identified  $\psi$ -trypsin as a proteoform resulting from trypsin autolysis. Amino acid analyses of its polypeptide chains revealed the existence of an additional split between Lys-176 and Asp-177 compared to the primary structure of  $\alpha$ -trypsin. The enzyme can be obtained from a trypsin autolyzate by ion exchange chromatography. Purification protocols that are available utilize one of the characteristic features of  $\psi$ -trypsin: in contrast to  $\alpha$ - or  $\beta$ -trypsin, it is not modified by the trypsin inactivator TLCK and shows a minimum retention on cation exchangers such as Sulfoethyl-Sephadex.

Enzyme kinetics studies with synthetic amino acid esters demonstrated a largely decreased affinity and activity towards cationic substrates. This has been elucidated by the presence of the characteristic chain split, which disconnects the specificity site (Asp-177) from the active site (Ser-183, His-46, Asp-90; all indicated according to the trypsinogen numbering convention). The reaction mechanism of  $\psi$ -trypsin with specific substrates differs from that of  $\alpha$ - or  $\beta$ -trypsin in the rate limiting step. No crystal structure of  $\psi$ -trypsin has been solved yet, but the anticipated existence of the modified ('neutral') active site and the ability to hydrolyze tyrosine-derived ester substrates lead to cleavages characteristic of chymotrypsin. So far, only a few studies have focused on the digestion of peptides and proteins by  $\psi$ -trypsin. These studies have confirmed its preferential action on Arg and Lys residues but also at Phe and Tyr residues (which is typical for chymotrypsin) as minor cleavage sites. The produced peptides contain frequent missed cleavages because of the absent specificity and reduced affinity towards the cationic sites. However, overnight digestions of protein samples provide enough peptides for identification experiments involving nLC-MALDI or nLC-ESI MS/MS. Subsequently, combining data from tryptic and  $\psi$ -tryptic digestions has been shown to be advantageous in order to increase the number of matched peptides and sequence coverage values. Furthermore, peptides with missed cleavage sites could be beneficial for studying posttranslational modifications of proteins.

$\psi$ -Trypsin should not be used as an equivalent substitute of trypsin or chymotrypsin and, in fact, there is no need to do it. However, as it makes preferential cleavages after basic Arg and Lys residues and has a more developed side specificity for aromatic and Leu residues, which is not expected to occur for pure major trypsin forms  $\alpha$  and  $\beta$  (but may appear anyway because of their autolysis), it produces valuable complementary information. The unavailability of any commercial material represents a big obstacle for the application of  $\psi$ -trypsin in common proteomics research. On the other hand, a single-step affinity chromatographic method has already been introduced [27], which could make the preparation of the enzyme easier (after a revision and transformation with the use of modern chromatographic materials).

**Author Contributions:** Marek Šebela came with the idea, wrote the Sections 1–5 and finally edited the whole text, Zdeněk Perutka wrote the Sections 6–9 and provided some of his unpublished results.

**Funding:** This work was supported by grant no. LO1204 (National Program of Sustainability I) from the Ministry of Education, Youth and Sports, Czech Republic.

**Conflicts of Interest:** The authors declare that there is no conflict of interests.

## Abbreviations

Ac-Tyr-OEt	$N^{\alpha}$ -acetyl-L-tyrosine ethyl ester
Bz-Arg-OEt	$N^{\alpha}$ -benzoyl-L-arginine ethyl ester
Bz-Arg-pNA	$N^{\alpha}$ -benzoyl-D,L-arginine-4-nitroanilide
ESI	electrospray ionization
IUPAC	International Union of Pure and Applied Chemistry
MALDI	matrix-assisted laser desorption/ionization
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NPGb	<i>p</i> -nitrophenyl- <i>p'</i> -guanidinobenzoate
nLC	nanoflow liquid chromatography
PDB	Protein Data Bank
SE-Sephadex	Sulfoethyl-Sephadex
TLCK	$N^{\alpha}$ - <i>p</i> -tosyl-L-lysine-chloromethyl ketone
TOF	time-of-flight
TPCK	<i>N</i> - <i>p</i> -tosyl-L-phenylalanine chloromethyl ketone

## References

1. Burkhart, J.M.; Schumbrutzki, C.; Wortelkamp, S.; Sickman, A.; Zahedi, R.P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteom.* **2012**, *75*, 1454–1462. [[CrossRef](#)] [[PubMed](#)]
2. Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin. The leading protease in proteomics. *Mass Spectrom. Rev.* **2013**, *32*, 453–465. [[CrossRef](#)] [[PubMed](#)]
3. Olsen, J.V.; Ong, S.E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteom.* **2004**, *3*, 608–614. [[CrossRef](#)] [[PubMed](#)]
4. Keil, B. Trypsin. *Enzymes* **1971**, *3*, 249–275. [[CrossRef](#)]
5. Rodriguez, J.; Gupta, N.; Smith, R.D.; Pevzner, P.A. Does trypsin cut before proline? *J. Proteom. Res.* **2008**, *7*, 300–305. [[CrossRef](#)] [[PubMed](#)]
6. Picotti, P.; Aebersold, R.; Domon, B. The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteom.* **2007**, *6*, 1589–1598. [[CrossRef](#)] [[PubMed](#)]
7. Schaefer, H.; Chamrad, D.C.; Marcus, K.; Reidegeld, K.A.; Blüggel, M.; Meyer, H.E. Tryptic transpeptidation products observed in proteome analysis by liquid chromatography-tandem mass spectrometry. *Proteomics* **2005**, *5*, 846–852. [[CrossRef](#)] [[PubMed](#)]
8. Dyčka, F.; Franc, V.; Fryčák, P.; Raus, M.; Řehulka, P.; Lenobel, R.; Allmaier, G.; Marchetti-Deschmann, M.; Šebela, M. Evaluation of pseudotrypsin cleavage specificity towards proteins by MALDI-TOF mass spectrometry. *Protein Pept. Lett.* **2015**, *22*, 1123–1132. [[CrossRef](#)] [[PubMed](#)]
9. Thiede, B.; Lamer, S.; Mattow, J.; Siejak, F.; Dimmler, C.; Rudel, T.; Jungblut, P.R. Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 496–502. [[CrossRef](#)]
10. Lawless, C.; Hubbard, S.J. Prediction of missed proteolytic cleavages for the selection of surrogate peptides for quantitative proteomics. *OMICS* **2012**, *16*, 449–456. [[CrossRef](#)] [[PubMed](#)]
11. Kühne, W. *Über das Trypsin (Enzym des Pankreas)*. *Verhandlungen des Naturhistorisch-medizinischen Vereins zu Heidelberg*; Carl Winter's Universitätsbuchhandlung: Heidelberg, Germany, 1877; Volume 1, pp. 194–198, reprinted in *FEBS Lett.* **1976**, *62*, E8–E12. [[CrossRef](#)]
12. Kunitz, M.; Northrop, J.H. Isolation from beef pancreas of crystalline trypsinogen, trypsin, a trypsin inhibitor, and an inhibitor-trypsin compound. *J. Gen. Physiol.* **1936**, *19*, 991–1007. [[CrossRef](#)] [[PubMed](#)]

13. Walsh, K.A.; Neurath, H. Trypsinogen and chymotrypsinogen as homologous proteins. *Proc. Natl. Acad. Sci. USA* **1964**, *52*, 884–889. [[CrossRef](#)] [[PubMed](#)]
14. Mikeš, O.; Holeyšovský, V.; Tomášek, V.; Šorm, F. Covalent structure of bovine trypsinogen. The position of remaining amides. *Biochem. Biophys. Res. Commun.* **1966**, *24*, 346–352. [[CrossRef](#)]
15. Puigserver, A.; Desnuelle, P. Identification of an anionic trypsinogen in bovine pancreas. *Biochim. Biophys. Acta* **1971**, *236*, 499–502. [[CrossRef](#)]
16. Smith, R.L.; Shaw, E. Pseudotrypsin. A modified bovine trypsin produced by limited autodigestion. *J. Biol. Chem.* **1969**, *244*, 4704–4712. [[PubMed](#)]
17. Maroux, S.; Rivery, M.; Desnuelle, P. An autolyzed and still active form of bovine trypsin. *Biochim. Biophys. Acta* **1967**, *140*, 377–380. [[CrossRef](#)]
18. Maroux, S.; Desnuelle, P. On some autolyzed derivatives of bovine trypsin. *Biochim. Biophys. Acta* **1969**, *181*, 59–72. [[CrossRef](#)]
19. Kumazaki, T.; Ishi, S. Characterization of active derivatives produced by acetamidation and selective autolysis of bovine trypsin. *J. Biochem.* **1979**, *85*, 581–590. [[CrossRef](#)] [[PubMed](#)]
20. Lacerda, C.D.; Teixeira, A.E.; de Oliveira, J.S.; Silva, S.F.; Vasconcelos, A.V.B.; Gouveia, D.G.; da Silva, A.R.; Santoro, M.M.; dos Mares-Guia, M.L.; Santos, A.M.C. Gamma trypsin: Purification and physicochemical characterization of a novel bovine trypsin isoform. *Int. J. Biol. Macromol.* **2014**, *70*, 179–186. [[CrossRef](#)] [[PubMed](#)]
21. Walsh, K.A.; Wilcox, P.E. Serine proteases. *Methods Enzymol.* **1970**, *19*, 31–41. [[CrossRef](#)]
22. Smith, L.M.; Kelleher, N.L. The Consortium for Top Down Proteomics. Proteoform: A single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187. [[CrossRef](#)] [[PubMed](#)]
23. Schroeder, D.D.; Shaw, E. Chromatography of trypsin and its derivatives. Characterization of a new active form of bovine trypsin. *J. Biol. Chem.* **1968**, *243*, 2943–2949. [[PubMed](#)]
24. Keil-Dlouhá, V.V.; Zylber, N.; Imhoff, J.M.; Tong, N.T.; Keil, B. Proteolytic activity of pseudotrypsin. *FEBS Lett.* **1971**, *16*, 291–295. [[CrossRef](#)]
25. Santos, A.M.C.; Oliveira, J.S.D.; Bittar, E.R.; Silva, A.L.D.; Guia, M.L.D.M.; Bemquerer, M.P.; Santoro, M.M. Improved purification process of  $\beta$ - and  $\alpha$ -trypsin isoforms by ion-exchange chromatography. *Braz. Arch. Biol. Technol.* **2008**, *51*, 511–521. [[CrossRef](#)]
26. Foucault, G.; Kellershohn, N.; Seydoux, F.; Yon, J.; Parquet, C.; Arrio, B. Comparative study of some conformational properties of  $\alpha$ ,  $\beta$  and  $\psi$  trypsins. *Biochimie* **1974**, *56*, 1343–1350. [[CrossRef](#)]
27. Foucault, G.; Seydoux, F.; Yon, J. Comparative kinetic properties of  $\alpha$ ,  $\beta$  and  $\psi$  forms of trypsin. *Eur. J. Biochem.* **1974**, *47*, 295–302. [[CrossRef](#)] [[PubMed](#)]
28. Vincent, J.P.; Lazdunski, M. Trypsin-pancreatic trypsin Inhibitor association. Dynamics of the interaction and role of disulfide bridges. *Biochemistry* **1972**, *11*, 2967–2977. [[CrossRef](#)] [[PubMed](#)]
29. Chowdhury, S.K.; Chait, B.T. Analysis of mixtures of closely related forms of bovine trypsin by electrospray ionization mass spectrometry: Use of charge state distributions to resolve ions of the different forms. *Biochem. Biophys. Res. Commun.* **1990**, *173*, 927–931. [[CrossRef](#)]
30. Ashton, D.S.; Ashcroft, A.E.; Beddell, C.R.; Cooper, D.J.; Green, B.N.; Oliver, R.W.A. On the analysis of bovine trypsin by electrospray-mass spectrometry. *Biochem. Biophys. Res. Commun.* **1994**, *199*, 694–698. [[CrossRef](#)] [[PubMed](#)]
31. Cunningham, L.W. Molecular-kinetic properties of crystalline diisopropyl phosphoryl trypsin. *J. Biol. Chem.* **1954**, *211*, 13–19. [[PubMed](#)]
32. Günther, A.R.; Santoro, M.M.; Rogana, E. pH titration of native and unfolded  $\beta$ -trypsin: Evaluation of the  $\Delta\Delta G^0$  titration and the carboxyl pK values. *Braz. J. Med. Biol. Res.* **1997**, *30*, 1281–1286. [[CrossRef](#)] [[PubMed](#)]
33. Buck, F.F.; Vithayathil, A.J.; Bier, M.; Nord, F.F. On the mechanism of enzyme action. LXXIII. Studies on trypsins from beef, sheep and pig pancreas. *Arch. Biochem. Biophys.* **1962**, *97*, 417–424. [[CrossRef](#)]
34. Travis, J.; Liener, I.E. The crystallization and partial characterization of porcine trypsin. *J. Biol. Chem.* **1965**, *240*, 1962–1966. [[PubMed](#)]
35. Ianucci, N.B.; Albanesi, G.J.; Marani, M.M.; Fernández Lahore, H.M.; Cascone, O.; Camperi, S.A. Isolation of trypsin from bovine pancreas using immobilized benzamidine and peptide CTPR ligands in expanded beds. *Sep. Sci. Technol.* **2007**, *40*, 3277–3287. [[CrossRef](#)]
36. Matthews, B.W.; Sigler, P.B.; Henderson, R.; Blow, D.M. The three-dimensional structure of tosyl- $\alpha$ -chymotrypsin. *Nature* **1967**, *214*, 652–656. [[CrossRef](#)] [[PubMed](#)]



37. Birktoft, J.J.; Blow, D.M. Structure of crystalline  $\alpha$ -chymotrypsin: V. The atomic structure of tosyl- $\alpha$ -chymotrypsin at 2 Å resolution. *J. Mol. Biol.* **1972**, *68*, 187–240. [[CrossRef](#)]
38. Bode, W.; Schwager, P. The refined crystal structure of bovine  $\beta$ -trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. *J. Mol. Biol.* **1975**, *98*, 693–717. [[CrossRef](#)]
39. Bode, W.; Fehlhhammer, H.; Huber, R. Crystal structure of bovine trypsinogen at 1.8 Å resolution. I. Data collection, application of Patterson search techniques and preliminary structural interpretation. *J. Mol. Biol.* **1976**, *106*, 325–335. [[CrossRef](#)]
40. Fehlhhammer, H.; Bode, W.; Huber, R. Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J. Mol. Biol.* **1977**, *111*, 415–438. [[CrossRef](#)]
41. Sandler, B.; Murakami, M.; Clardy, J. Atomic structure of the trypsin-aeruginosin 98-B complex. *J. Am. Chem. Soc.* **1998**, *120*, 595–596. [[CrossRef](#)]
42. Page, M.J.; Di Cera, E. Combinatorial enzyme design probes allostery and cooperativity in the trypsin fold. *J. Mol. Biol.* **2010**, *399*, 306–319. [[CrossRef](#)] [[PubMed](#)]
43. Inagami, T.; Sturtevant, J.M. Nonspecific catalyses by  $\alpha$ -chymotrypsin and trypsin. *J. Biol. Chem.* **1960**, *235*, 1019–1023. [[PubMed](#)]
44. Nakano, M.; Tanizawa, K.; Nozawa, M.; Kanaoka, Y. Efficient tryptic hydrolysis of aryl esters with a cationic center in the leaving group. Further characterization of “inverse substrates”. *Chem. Pharm. Bull.* **1980**, *28*, 2212–2216. [[CrossRef](#)]
45. Seydoux, F.; Yon, J. On the specificity of tryptic catalysis. *Biochem. Biophys. Res. Commun.* **1971**, *44*, 745–751. [[CrossRef](#)]
46. Keil-Dlouhá, V.; Zylber, N.; Tong, N.T.; Keil, B. Cleavage of glucagon by  $\alpha$ - and  $\beta$ -trypsin. *FEBS Lett.* **1971**, *16*, 287–290. [[CrossRef](#)]
47. Ghalayini, M.K.; Dong, Q.; Richardson, D.R.; Assinder, S.J. Proteolytic cleavage and truncation of NDRG1 in human prostate cancer cells, but not normal prostate epithelial cells. *Biosci. Rep.* **2013**, *33*, e00042. [[CrossRef](#)] [[PubMed](#)]
48. Kirkpatrick, D.S.; Gerber, S.A.; Gygi, S.P. The absolute quantification strategy: A general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005**, *35*, 265–273. [[CrossRef](#)] [[PubMed](#)]
49. Nigam, A.; Subramanian, M.; Rajanna, P.K. Non-specific digestion artifacts of bovine trypsin exemplified with surrogate peptides for endogenous protein quantitation. *Chromatographia* **2018**, *81*, 57–64. [[CrossRef](#)]
50. Krause, E.; Wenschuh, H.; Jungblut, P.R. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal. Chem.* **1999**, *71*, 4160–4165. [[CrossRef](#)] [[PubMed](#)]
51. Blavet, N.; Uřinová, J.; Jeřábková, H.; Chamrad, I.; Vrana, H.; Lenobel, R.; Beinhauer, J.; Šebela, M.; Doležel, J.; Petrovská, B. UNcleProt (Universal Nuclear Protein database of barley): The first nuclear protein database that distinguishes proteins from different phases of the cell cycle. *Nucleus* **2016**, *8*, 70–80. [[CrossRef](#)] [[PubMed](#)]
52. Havliř, J.; Thomas, H.; Šebela, M.; Shevchenko, A. Fast-response proteomics by accelerated in-gel digestion of proteins. *Anal. Chem.* **2003**, *75*, 1300–1306. [[CrossRef](#)] [[PubMed](#)]

