



Non-Canonical Amino Acids in Analyses of Protease Structure and Function

Peter Goettig ¹,*⁽), Nikolaj G. Koch ^{2,3,†,‡} and Nediljko Budisa ^{3,4}

- ¹ Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Paracelsus Medical University, Strubergasse 21, 5020 Salzburg, Austria
- ² Biocatalysis Group, Technische Universität Berlin, 10623 Berlin, Germany; nikolaj.koch@unibas.ch
- ³ Bioanalytics Group, Institute of Biotechnology, Technische Universität Berlin, 10623 Berlin, Germany; nediljko.budisa@umanitoba.ca
- ⁴ Department of Chemistry, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
- * Correspondence: peter.goettig@pmu.ac.at; Tel.: +43-662-2420-80604
- ⁺ Current address: Department of Chemistry, Institute of Physical Chemistry, University of Basel, 4058 Basel, Switzerland.
- [‡] Current address: Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland.

Abstract: All known organisms encode 20 canonical amino acids by base triplets in the genetic code. The cellular translational machinery produces proteins consisting mainly of these amino acids. Several hundred natural amino acids serve important functions in metabolism, as scaffold molecules, and in signal transduction. New side chains are generated mainly by post-translational modifications, while others have altered backbones, such as the β - or γ -amino acids, or they undergo stereochemical inversion, e.g., in the case of D-amino acids. In addition, the number of non-canonical amino acids has further increased by chemical syntheses. Since many of these non-canonical amino acids confer resistance to proteolytic degradation, they are potential protease inhibitors and tools for specificity profiling studies in substrate optimization and enzyme inhibition. Other applications include in vitro and in vivo studies of enzyme kinetics, molecular interactions and bioimaging, to name a few. Amino acids with bio-orthogonal labels are particularly attractive, enabling various cross-link and click reactions for structure-functional studies. Here, we cover the latest developments in protease research with non-canonical amino acids, which opens up a great potential, e.g., for novel prodrugs activated by proteases or for other pharmaceutical compounds, some of which have already reached the clinical trial stage.

Keywords: bioorthogonal tags; cross-linking; click chemistry; genetic code expansion; non-canonical amino acids; inhibitory potency; pharmaceutical compound; protease specificity; proteinogenic amino acid; unnatural amino acid

1. Introduction

All natural proteins contain the 20 canonical or standard L-amino acids (cAAs) present in all living organisms and viruses. In addition, archaea, bacteria, and eukaryotes possess selenocysteine (Sec) as the 21st proteinogenic amino acid [1]. Another proteinogenic amino acid is pyrrolysine (Pyl), which is present in some archaea and bacteria (Figure 1) [2,3]. Remarkably, Pyl is genetically encoded and read from UAG stop codons (*amber*) in frame. A natural orthogonal pair of aminoacyl-tRNA synthetase (PylRS) and tRNA^{Pyl} enables the insertion of Pyl into proteins. Similarly, in-frame reading of UGA stop codons (*opal*) allows to incorporate Sec into proteins. However, the mechanism is more complicated, as seryl-tRNA synthetase generates seryl-tRNA^{Sec}, followed by conversion to selenocysteyltRNA^{Sec}. Moreover, Sec incorporation requires a specific elongation factor at the ribosome and recognition elements in the mRNA [4]. The molecular properties of many proteinogenic amino acids are often altered by post-translational modifications (PTM), such as



Citation: Goettig, P.; Koch, N.G.; Budisa, N. Non-Canonical Amino Acids in Analyses of Protease Structure and Function. *Int. J. Mol. Sci.* 2023, *24*, 14035. https://doi.org/ 10.3390/ijms241814035

Academic Editor: Istvan Simon

Received: 14 July 2023 Revised: 18 August 2023 Accepted: 20 August 2023 Published: 13 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). disulfide formation, glycosylation, phosphorylation, hydroxylation, methylation, lipidation or decarboxylation, to name the most important ones. About 500,000 PTM sites have been catalogued in mammalian proteomes alone, while archaea require many PTMs to survive in extreme environments, and bacteria share some PTMs with eukaryotes [5–7].

However, the range of natural amino acids from biological sources is much larger than the 20 canonical or 22 proteinogenic ones, as up to 900 non-canonical amino acids (ncAAs) are known to date. However, there is good evidence that some amino acids are of extraterrestial abiotic origin. For example, after the fall of the Murchison meteorite in Australia in 1969, it was soon found to contain Gly, Ala, Val, Pro, Glu and ncAAs, such as 2-methylalanine and sarcosine [8]. More than 100 of these primordial amino acids have been discovered in meteorites to date, which is not surprising, since the intergalactic space is abundant in all necessary precursor molecules. Most chiral amino acids from meteorites occur as racemic mixtures, while some samples exhibited an excess of the L-form or, less frequently, of the D-form [9]. Recently, a new family of hydroxyamino acids unknown on Earth, e.g., β -aminomethyl succinic acid, was discovered in the Murchison meteorite [10].

In living cells, some non-proteinogenic amino acids are essential in the metabolism, e.g., ornithine (Orn), citrulline (Cit) and argininosuccinate (ArgSA) in the urea cycle [11]. Others function as neuronal messengers similar to GABA, such as β -amino acids and D-amino acids, which are derived from canonical amino acids, since corresponding racemases exist in mammalian tissues (Figure 1) [12]. *D*-serine (*D*-Ser) acts as co-agonist of the N-methyl-*D*-aspartate glutamate or NMDA receptors, whereas *D*-Asp is a major regulator of adult neurogenesis and a crucial component in the development of endocrine function [13]. However, *D*-Asp is increasingly found during aging in long-lived proteins, since L\alpha-Asp spontaneously racemizes via a succinimidyl intermediate to L β -, D α - and D β -isomers [14].

β-amino acids are also present in biomolecules, e.g., as a building block of the natural anticancer drug taxol (paclitaxel) or, like β-Ala, as a component of vitamin B5 (pantothenate), which is a precursor of coenzyme A [15,16]. In addition, natural ncAAs, such as dimethylarginine (ADMA) Alg and homoarginine (hArg), participate in the physiological nitric oxide processes, involving atherosclerosis and atherogenesis. In nature, there are several other "homo-amino acids", such as homoserine, homocysteine (hCys) or homophenylalanine (hPhe), which are one methylene group (CH₂) longer than their canonical counterparts (Figure 1) [17,18]. S-adenosylmethionine (SAM), S-adenosyl homocysteine (SAH), hCys, 3-sulfinoalanine (3SA) and taurine (Tau) participate in the sulfur metabolism and the transfer of methyl groups, whereby hCys and Tau play a role as neurotransmitters [19,20]. Cyanobacteria produce some ncAAs, such as β-methylamino-*L*-alanine (BMAA), 2,4-diaminobutyric acid (Dab) and N-2-aminoethyl-glycine (Aeg), which have neurodegenerative potential for humans [21]. In the following, L-amino acids are not explicitly designated, in contrast to D-amino acids (Figure 1, Table 1; for details of nomenclature, see Appendix A).

Regarding the terminology, a clear distinction can be made between proteinogenic, nonproteinogenic, natural non-canonical (ncAA) and non-natural amino acids (nnAA), which corresponds to the more frequently used term unnatural amino acids (uAA). Unnatural or synthetic amino acids are often designated ncAAs as well (Figure 2, Table 2, nomenclature details in Appendix B). Nevertheless, it is necessary to use these terms separately, to retrieve the relevant publications in PUBMED. Accordingly, the appropriate terminology is used in this review to distinguish these amino acids, preferentially using ncAA for natural ones and uAA for synthetic ones, which are often referred to as ncAAs or, more rarely, as "unusual amino acids" in the literature as well. With respect to methods, and in general considerations, the term ncAA comprises natural and unnatural amino acids. Protease designations, such as A01.001 for the aspartic protease pepsin, follow the nomenclature of the MEROPS database (https://www.ebi.ac.uk/merops/, accessed on 19 August 2023) [22]. Associated with MEROPS, and published about a decade ago, the most comprehensive and systematic work on proteases is the Handbook of Proteolytic Enzymes, with descriptions of proteolytic mechanisms [23,24]. Our main goal is to present the highly significant and widely applied usage of ncAAs in basic and commercial research of proteases. Thus, we provide a



comprehensive overview, spanning the earliest beginnings to the most recent developments, with respect to methods and numerous examples of individual proteases.

Figure 1. Natural non-canonical amino acids. More details on these molecules, such as occurrence, biological functions, and usage, are given in Table 1. Details on nomenclature (IUPAC) and CAS numbers (Chemical Abstract Services) are given in Appendix A.

AA	Standard Name	Type and Occurrence	Functions and Usage	
Sec Pyl	selenocysteine pyrrolysine	all organisms (ao) archaea and bacteria	redox processes methyltransferase catalysis	
Hyp pSer pThr	hydroxyproline phosphoserine phosphothreonine	vertebrates (ve) eukaryotes (eu) eu	PTM, collagen stability PTM, signaling/cancer PTM, signaling/cancer	
Orn Cit ArgSA SAM SAH 3SA hCys/Hcy Tau Sar 2AA Kyn Aad hSer β-Ala	ornithine citrulline argininosuccinic acid S-adenosylmethionine S-adenosylhomocysteine 3-sulfinoalanine homocysteine taurine sarcosine anthranilic acid kynurenic acid aminoadipic acid homoserine β -alanine	a0 a0	urea cycle urea cycle Urea cycle DNA methylation (eukaryotes) sulfur metabolism sulfur metabolism sulfur metabolism glycine biosynthesis tryptophan biosynthesis tryptophan degradation lysine biosynthesis methionine metabolism vitamin B5 component	
D-Ser D-Asp GABA DOPA	D-serine D-aspartic acid γ-aminobutyric acid 3-hydroxytyrosine	ve ve ve ve	neurotransmitter, coagonist NMDA receptor neurotransmitter, agonist NMDA receptors neurotransmitter, GABA receptors neurotransmitter, dopamine precursor	
Abu/hAla ADMA Aeg Aib Alg Aze Can Dab Dha hArg Met(o) Nle/α-Ahx Nva/Ape Pip Sem/SeMet Tem/TeMet	α -aminobutyric acid dimethylarginine N-2-aminoethylglycine α -aminoisobutyric acid allylglycine azetidine-2-carboxylic acid canavanine 2,4-diaminobutyric acid dehydroalanine homoarginine methionine sulfoxide norleucine norvaline pipecolic acid selenomethionine	ve ve cyanobacteria ao plants, <i>Convallaria majalis</i> plants, <i>Leguminosae</i> cyanobacteria bacteria, <i>Lactococcus lactis</i> ao bacteria bacteria bacteria ao Met analog in proteins Met analog in proteins	metabolite nitric oxide synthase regulator toxin metabolite rare metabolite toxin toxin toxin lantibiotics bacterial growth inhibitor aging proteins immunity regulator X-ray crystallography X-ray crystallography	

Table 1. Natural non-canonical amino acids, occurrence and functions.



Figure 2. Unnatural non-canonical amino acids. Further details on these molecules, such as experimental applications, are given in Table 2. Details on nomenclature (IUPAC) and CAS numbers (Chemical Abstract Services) are given in Appendix B.

AA	Standard Name	Characteristics and Usage	
Aca	trans-4-aminocyclohexylalanine	substrate, inhibitor	
ACCA	cis-3-aminomethylcyclobutane carboxylic acid	inhibitor	
ε-Ahx	aminocaproic acid/6-aminohexaonic acid	substrate, inhibitor, ABP	
Ama	trans-4-aminomethylcyclohexylalanine	substrate, inhibitor	
Amf	4-aminomethylphenylalanine	substrate, inhibitor	
Anb	5-amino-2-nitrobenzoic acid	substrate, inhibitor, ABP	
Cha	cyclohexylalanine	substrate, inhibitor	
Chg	cyclohexylglycine	inhibitor	
Сра	3-(2-cyano-4-pyridyl)alanine	inhibitor	
CycL	cyclo-leucine/cyclo-Leu	inhibitor	
DfeGly	difluoroethylgycine	inhibitor, ABP	
DFM	difluoromethionine	protein modification	
DfpGly	difluoropropylgycine	inhibitor	
Dnp-Lys	N(6)-(2,4-dinitrophenyl)lysine	substrate (fluorophore)	
4F-Phe	4-fluorophenylalanine (¹⁹ F)	NMR	
5F-Trp	5-fluorotryptophan (¹⁹ F)	NMR	
F(gua)	4-guanidinophenylalanine	substrate, inhibitor, ABP	
hCha	homocyclohexylalanine	substrate, inhibitor	
Igl	2-indanylglycine	substrate, ABP	
MfeGly	monofluoroethylgycine	inhibitor	
MLeu	α-methyl-leucine	inhibitor	
1-Nal	1-naphthylalanine	inhibitor	
2-Nal	2-naphthylalanine	inhibitor	
Oic	octahydroindole-2-carboxylic acid	substrate, inhibitor, ABP	
Phg	phenylglycine	substrate, inhibitor	
Phi	perhydroindol-2-carboxylic acid	inhibitor	
Sep	selenotryptophans (selenolo[3,2-b]pyrrole/[2,3-b]pyrrole)	X-ray cystallography	
TfeGly	trifluoroethylgycine	inhibitor, ABP	
Thfg	tetrahydrofuranylglycine	inhibitor	
Thi	3-(2-thienyl)alanine/β-thienylalanine or 3-(3-thienyl)alanine	substrate, inhibitor	
Thp	thioproline	inhibitor	
Tic	1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid	substrate, inhibitor	
Tle	<i>tert</i> -leucine/ <i>tert</i> -butyl-glycine	substrate, inhibitor, ABP	
CN-Phe	4-cvanophenylalanine	fluorophore	
3CY	3-chlorotyrosine	protein modification	
3NY	3-nitrotyrosine	protein modification, fluorophore	
Вра	4-benzovlphenylalanine	photo-crosslinker	
Dap	2.3-diaminopropionic acid	ABP for Ser/Cvs proteases	
DiZPK	$3-(3-\text{methyl}-3H-\text{diazirine}-3-\text{vl})-\text{propamino}(\text{carbonvl}-N\varepsilon-\text{lysine})$	protein modification/photo-crosslinker	
Hco	7-(hydroxy-coumarin-4-yl)-ethylglycine	protein modification/fluorophore	
NPY	nitropiperonyltyrosine	caging and decaging	
PABK	Nε-4-azidobenzyloxycarbonyllysine	click reactant, caging and decaging	
Aha	azidohomoalanine	click reaction, 1,3-dipolar cycloaddition	
AzF	4-azidophenylalanine	1,3-dipolar cycloaddition	
Hpg	homopropargylglycine	1,3-dipolar cycloaddition	
Plk	N-propargyllysine	1,3-dipolar cycloaddition	
Sac	S-allvlcvsteine	photo-click reaction	

 Table 2. Unnatural non-canonical amino acids, usage, and function.

2. Methods for the Incorporation of ncAAs in Peptides and Proteins

Non-canonical amino acids (ncAAs) can be introduced into peptides and recombinant proteins using various in vivo and in vitro methods or by specific chemical and enzymatic modifications. As these methods are very numerous, we recommend interested readers to study the referenced literature of this review to learn all the necessary details. These insertions or modifications can be made to the protein backbone or on the side chains of amino acids, providing useful alterations to the properties of the protein. In nature, alterations of the backbone of proteins usually occur after the translation process through post-translational modifications. On the other hand, when we want to introduce changes to protein backbones that are not naturally occurring, we can use proline analogs to modify them in a living organism [25]. In vitro ribosomal translation systems can be used to introduce other modifications, such as D-, β - or γ -amino acids, into target polypeptides. Additionally, specific engineered systems can incorporate α -hydroxy acids into proteins in response to in-frame *amber* codons [26]. Overall, the incorporation of ncAAs into proteins provides a powerful tool for protein engineering, enzyme catalysis, biomedicine and biotechnology [27].

2.1. Chemical Modification of Standard Amino Acids

Historically, chemical modification of natural amino acid residues was the first method to introduce unnatural side chains into peptides and proteins for analytical and preparative applications, as Cys alkylation of the enzyme glyoxalase with iodoacetamide dates back to 1933 [28,29]. Proteases can be repurposed as ligases to synthesise peptides with unnatural side chains. For example, enzymatic synthesis of corresponding peptides is a different route using proteases as ligases, like the cysteine protease papain for the generation of the tripeptide ethyl ester Ala-Aib-Ala-OEt containing α -aminoisobutyric acid [30]. Various proteases have been widely used as ligases to synthesize polypeptides and protein, especially with ncAAs [31].

Site-specific chemical modifications of residues involved in catalysis were used in several functional investigations of proteases. Mutation of the catalytic Lys145 to Cys rendered the *Escherichia coli* signal peptidase inactive, whereas the reaction of this Cys with 2-bromoethylamine and 3-bromopropylamine restored the activity of the resulting γ -thia-lysine and its homolog [32]. Ser-to-Cys mutations of subtilisin from *Bacillus lentus* were used to explore the S2, S1 and S1' pocket specificity via chemical modification to Cys-SR, with a variable length of R and increasing charges [33]. Several investigated variants remained active, but positive and negative charges, as well as a branched side chain at position 222, reduced the catalytic efficiency 100-fold. Essentially, chemical modification of Cys62 (S2) to methanethiosulfonate, similar to Cys156/166 (S1) and Cys217 (S1'), reduced the catalytic efficiency stepwise down to 10% of the wild-type enzyme by increasing the negative charge in each position [34,35].

2.2. Substitution of Specific cAAs by ncAAs in Auxotrophy-Based Methods and Its Relevance for Structural Biology

As early as 1957, in a ground-brea_{ki}ng experiment with the Met-auxotrophic *E. coli* ML304d strain, Cowie and Cowen showed that basically all methionine residues in newly synthesized proteins could be replaced by selenomethionine (SeMet) [36]. In 1962, already about 30 ncAAs, such as norleucine (Nle, α -Ahx), fluoro-Phe, Aza-Trp and methyl-Trp, were incorporated into proteins of corresponding auxotrophic bacteria [37]. Typically, the method worked only for Met, Leu, Ile, Phe, Tyr, Trp, Arg and Lys analogs. Since the early 1990s, ncAAs have been incorporated into target proteins by expression in *E. coli*, mostly for structural studies with X-ray crystallography with multi-wavelength anomalous dispersion (MAD) [38,39]. Auxotrophic bacteria were employed, which cannot synthesize a particular amino acid, such as methionine. The medium was then supplemented with 19 standard amino acids and selenomethionine (SeMet, Sem) or, in rare cases, telluro-methionine (TeMet, Tem, Figure 1) [40,41]. The method was extended, e.g., by incorporating selenium

containing Trp (Sep), which exhibited an anomalous signal of the selenium sites in a crystallized model protein (Figure 2) [42].

Since about 1985, undulator beamlines have been available at synchrotrons, which allowed to adjust distinct X-ray wavelengths for anomalous scattering of transition and heavy metals, to obtain phase information for structure determination [43]. In the following years, anomalously scattering atoms, either soaked in protein crystals or incorporated in proteins themselves, were utilized to collect X-ray diffraction data sets with phase information [44]. Crystals of the natural ncAA selenolanthionine allowed to determine the anomalous scattering factors f' and f'' of selenium at the K edge with polarized synchrotron radiation [45,46]. Moreover, detectors were improved, in particular charged coupled devices (CCD) [47]. A complex of selenobiotin and streptavidin was the basis for structure determination with the multi-wavelength anomalous diffraction method (MAD) [48]. Already in 1989, Tem was incorporated into fungal proteins, and the first incorporation of Tem in *E. coli* dihydrofolate reductase was reported in 1994 [49,50]. Incorporation of Sem to 100% and Tem to 75% was achieved for Staphylococcus aureus pyrrolidone carboxyl peptidase, with unchanged enzymatic properties [40]. The application of the method of reassigned sense codons for auxotrophic E. coli allowed to incorporate methionine analogs, such as Sem and Tem, at high levels [41]. Subsequently, Tem incorporation in model proteins facilitated their structure determination both with the MAD and with the multiple isomorphous replacement (MIR) approach [51]. An excellent overview of biological research with selenium and tellurium analogs of Met and Cys, such as Sem, Tem, Sec and Tec (TeCys), was provided by Musiol and Moroder [52]. Based on the incorporation of Sem, the complex of MMP3/TIMP1 is a fine example of structure determination with the MAD method (Figure 3A) [53].

Meanwhile, single-wavelength anomalous dispersion (SAD) became the dominant method for de novo protein structure determination; since the year 2007, more deposited PDB coordinates were obtained by SAD than by MAD [54,55]. In addition to this development, the quantitative incorporation of Sem into proteins expressed in yeast, insect and mammalian cells was established for structural biology [56–58]. Although many atoms can be utilized, even sulfur, phosphorus and other light atoms, Sem-SAD is still the most common approach. Also, the extremely intense radiation of X-ray free-electron lasers (XFEL) is suitable for SAD phasing with protein microcrystals [59]. This method is constantly being improved, e.g., with the SAD-SMAR algorithm, which facilitates to determine the anomalous scattering substructure in a highly efficient manner [60]. Currently, about 30% of all crystal structures are solved by experimental phasing methods; the rest only requires molecular replacement with a related protein model [61].

As the genetic information of an organism can be encoded with only 30–40 sense codons, there is potential to recode more than 20 sense codons with ncAAs [62]. Tirrell and co-workers have proposed an alternative approach to stop codon suppression (SCS) by incorporating ncAAs at rare sense codons [63]. In 2014, Budisa and Bohlke attempted to exploit the degeneracy of the genetic code by freeing rare sense codons from their original coding roles and reprogramming them to code for specific ncAAs permanently [64]. However, with methods for reassigning sense codons still in their infancy, the field is currently dominated by stop codon suppression technologies, which have been increasingly improved and expanded [65–67]. In an unusual approach of the reassigned sense codon, cell-free protein synthesis with an *E. coli* 30S ribosomal extract was applied, in order to replace all natural Asp and Phe positions with 2-naphthylalanine (2-Nal) or 4-phenylazophenylalanine in the HIV protease [68]. To this end, the tRNAs for Asp and Phe had to be inactivated by corresponding antisense mRNA, resulting in an acceptable translation efficiency.

Special cases of modified cAAs and ncAAs are isotope-labeled ones, which are widely employed in nuclear magnetic resonance spectroscopy (NMR spectroscopy) and mass spectrometry (MS). Hereby, hydrogen, carbon, nitrogen and many other atom nuclei are replaced by deuterium (²H/D), ¹³C, ¹⁵N, ¹⁹F, ³¹P, etc. to solve structures of small and large biomolecules, investigate their dynamics, diversify structures and discover new drugs [69–72]. In an exemplary study, the reassigned sense codon method was utilized to incorporate 5-¹⁹F-Trp (an isotope-labeled 5F-Trp) into thrombin, to measure the dynamics of the protease by NMR, as well as enzyme kinetics, which was further supported by X-ray crystallography (Figure 3B) [73]. In addition, NMR with labeled AAs is useful in solving membrane protein structures, e.g., in the solid state, as well as in the emerging field of metabolomics [74,75]. Isotope labeling of proteins can be utilized in combination with NMR and MS or in MS alone, as for proteomics [76,77]. Numerous overview articles describe other promising applications of auxotrophy-based methods [78,79].



Figure 3. Non-canonical amino acids in structure-function studies. (**A**) Human MMP-3 in complex with its natural inhibitor TIMP-1. Two selenomethionine residues (Sem, with orange spheres) were sufficient to determine the structure by MAD phasing at 2.8 Å resolution (PDB code 1UEA). The catalytic Zn²⁺ (black sphere) is coordinated by three His, a Glu and the N-terminus of TIMP-1. (**B**) Thrombin in which all Trp residues were replaced by 5-¹⁹F-Trp via the auxotrophy-based reassigned sense codon method (PDB 6V64). ¹⁹F-Trp allowed to measure molecular dynamic processes by NMR, with the fluorine atoms depicted as green spheres. The covalently bound chloromethyl ketone inhibitor PPACK, shown as a ball-and-stick model, contains the modified Arg-cmk, occupying the S1 pocket, a Pro in the S2 pocket and a *D*-Phe binding to the S4 pocket formed largely by F-5-Trp215. The residues of the catalytic triad His57, Asp102 and Ser195 are shown as sticks, as is Asp189 at the bottom of the S1 pocket, which confers the specificity for basic substrates.

2.3. Site-Directed Insertion of ncAAs Using Orthogonal Pairs

In order to incorporate a non-canonical amino acid (ncAA) into a protein sequence, a distinct set of aminoacyl-tRNA synthetase (aaRS) and tRNA molecules is needed. These molecules must be orthogonal, meaning they must not interact with the endogenous aaRS-tRNA pairs, and should be capable of reading stop codons in the target mRNA sequence to allow the accurate incorporation of the desired ncAA [80]. Typically, the selection of molecular machineries capable of incorporating ncAAs is performed by generating gene libraries of the incorporating synthetase [81]. Two Archaean systems, the *Methanocaldococcus jannaschii* tyrosine pair (MjTyrRS:tRNA^{Tyr}) and the *Methanosarcina* pyrrolysine pair (PyIRS:tRNA^{Pyl}), have been used in *Escherichia coli* [82,83]. These systems enable the sitespecific incorporation of diverse ncAAs into proteins in vivo, providing a powerful tool for creating novel proteins with unique properties [84]. In the contemporary literature, there are many excellent reviews that cover these topics in detail [85]. Therefore, interested readers are encouraged to consult these reviews for further information [86,87].

Over the past two decades, our research group and others have discovered that pyrrolysyl-tRNA synthetase (PylRS) is an especially versatile enzyme that can incorporate diverse amino acid substrates into proteins. As already mentioned, the natural substrate of PylRS is pyrrolysine (Pyl), a rare proteinogenic amino acid found in few proteomes, mainly in methanogenic microorganisms (Figure 1, Table 1) [88–90]. Pyl is a lysine analog with a 4-methyl-pyrroline-5-carboxylate ring attached to its side chain. Pyl is recognized and charged onto tRNA^{Pyl} by PylRS, which is a class II aminoacyl-tRNA synthetase, with the cognate tRNA anticodon CUA, creating a natural PylRS:tRNA^{Pyl} orthogonal pair that fulfills all requirements for a natural orthogonal translation system (OTS) [91]. This OTS is capable of in-frame stop codon suppression and subsequent site-specific incorporation of an ncAA into a recombinant target protein. Overall, PylRS provides a powerful tool for the incorporation of ncAAs into proteins, enabling the creation of novel proteins with unique properties [92].

Expanding the genetic code with orthogonal pairs has enabled the site-specific incorporation of numerous ncAAs into proteins, offering great potential for mimicking post-translational modifications (PTMs) in a precisely controlled manner [93]. In contrast, traditional methods of protein modification often result in non-specific labeling, leading to a complex and heterogeneous protein mixture. This micro-heterogeneity is a critical issue when aiming to achieve specific PTMs with desired chemical modifications, both natural and unnatural. Therefore, there is a need for selective methods that enable the absolute control of the position of the reactive handle within a protein [94]. Over the past few years, it has been possible to incorporate various ncAAs into proteins in a site-specific manner using different hosts such as bacteria, yeasts, mammalian cells and even multicellular organisms like *Caenorhabditis elegans* and *Drosophila melanogaster* [95,96].

As the stop codon suppression method is very versatile and already available from biological systems, the pyrrolysine tRNA synthetase-tRNA pair is widely utilized, due to its high promiscuity regarding the acceptance of ncAAs through reassignment of the *amber* stop codon UGA [97]. To incorporate natural, post-translationally modified AAs, such as phospohorylated serine (pSer), in a site-specific manner, this method was utilized for the *amber* and the Leu codon CUA with orthogonal translation systems in *E. coli* (Figure 1) [98,99]. Apart from *E. coli*, organisms like yeast, *Leishmania* and even mammalian cell cultures were successfully employed for the incorporation of hundreds of unnatural ncAAs [92,100]. This method could be expanded by native chemical ligation, in order to incorporate specific groups in proteins for FRET measurements [101].

Engineering elongation factors and the ribosome itself is another route to generating proteins with ncAAs that has been reviewed by Cui and colleagues [102]. Based on early efforts to create orthogonal mRNA-ribosome pairs, alternative Shine–Dalgarno sequences were introduced in the mRNA and the corresponding recognition elements in the 16S rRNA [103,104]. An *E. coli* 23S rRNA with two mutations in the peptidyltransferase center allowed to incorporate *D*-Phe and *D*-Met into bacterial dihydrofolate reductase (DHFR) and insect luciferase by cell-free expression [105]. Similarly, various β -amino acids could be incorporated into DHFR, e.g., β , β -dimethyl- β -alanine [106]. Recently, it was demonstrated that whole *E. coli* cells can be labeled with fluorescent ncAAs, such as 4-cyanophenylalanine (CN-Phe) and two oxazole derivatives, which was facilitated by modified ribosomes and a PylRS-tRNA^{Pyl} orthogonal pair [107]. The novel PURE translation system produces affinity-tagged polypeptides, whereby so-called hyper-accurate ribosomes with mutations in the S12 protein significantly enhanced yields of ncAA containing peptides, exemplified for α -methyl-Cys [108]. Nevertheless, the genetic code itself can be modified by using quadruplets instead of triplets of DNA and RNA bases, as well as by unusual new base molecules [109].

3. Protease Substrates, Inhibitors and Activity-Based Probes with ncAAs

In general, many ncAAs confer stability to peptides, due to resistance against most proteases from the three kingdoms of life [110–112]. Frequently, novel ncAAs are tested as antimicrobial peptides against bacteria and fungi. For example, the peptide ILLKKLLKKI, which consisted entirely of D-amino acids, killed *Mycobacterium tuberculosis* as effectively as the non-peptidic drug rifampicin [113–115]. Similarly, the protease stability of the antimicrobial peptide KRLFKKLLKYLRKF was significantly increased by incorporation of *D*-Lys, *D*-Arg, 2,4-diaminobutyric acid (Dab), 2,3-diaminopropionic acid, hArg, 4-aminobutyric acid and β -2-thienylalanine (Thi) (Figures 1 and 2) [116].

Numerous examples of peptidic chromogenic and fluorogenic protease substrates with ncAAs are described in the literature. Among them are natural ncAAs and synthetic uAAs, such as Nle, Orn, cyclohexylalanine (Cha) and many others, which can be most favored by a protease in distinct positions according to the Schechter–Berger nomenclature [117]. Small differences in substrate specificity of proteases are revealed by using ncAAs at certain positions of substrate analogs and activity-based probes, which allow to study different protease activities in vitro and in living cells for bioimaging [118]. So-called hybrid combinatorial substrate libraries (HyCoSuL) are suited for the investigation of various classes of proteases with cAAs and ncAAs in tetrapeptide-ACCs (7-amino-4-carbamoylmethylcoumarin) comprising P4–P1 residues [119]. Numerous activity-based probes (ABPs) have been developed to detect active proteases in vitro and in vivo. A comprehensive overview of the general topic of ABPs directed against proteases is given by Vizovišek and coworkers [120]. The first study using an ABP employed a biotinylated fluorophosphonate or FP-biotin, which was apparently able to detect several serine hydrolases in human tissues [121]. Early studies of caspases detected them in apoptotic cells with compounds like FAM-Val-Ala-Asp-fluoromethyl ketone [122]. Since some ncAA side chains bind more strongly and selectively to the specificity pockets of proteases, their systematic screening can result not only in optimal substrates and inhibitors, but also in highly specific ABPs, which facilitates the monitoring of protease activity in test tubes, body fluids or tissue samples and even in living cells. In general, ncAAs can stabilize peptides and proteins against proteolytic degradation and help in the discovery of corresponding inhibitors [123,124]. Usually, the classical and straightforward way to create small-molecule inhibitors is through synthesis with organic chemistry methods [125]. Many inhibitors are so-called peptidomimetics, which often contain non-cleavable pseudo-peptide bonds, although they are often combined with natural ncAA and uAA residues, which were already considered in early computational studies [126,127]. In the following, only a few enzyme kinetic parameters are mentioned, such as catalytic efficiency or the specificity constant (k_{cat}/K_M), preferentially inhibition constants (K_i) instead of IC₅₀ values. Equilibrium association and dissociation constants (K_a, K_d) are also mentioned for clarity and comparability [128–130].

3.1. Serine Proteases

3.1.1. Digestive Trypsin-like Serine Proteases

The classical digestive serine proteases secreted from the human pancreas are trypsin (cationic trypsin, S01.127), chymotrypsin B (S01.152) and elastase-1 (S01.153) [131]. As the early enzymatic studies employed the bovine and porcine proteases, many experiments refer to bovine α -chymotrypsin (CTRA, S01.001), bovine trypsin (TRYB, S01.151) and porcine elastase-1 (CELA1 as well as for the human enzyme). Nevertheless, their specificity for the P1 residue in substrates remains characteristic and defines many proteases as trypsin-like (Arg/Lys), chymotrypsin-like (Phe/Tyr/Leu) and elastase-like (Ala/Val). These serine proteases and other types, such as subtilisin-like ones and α/β -hydrolases, exhibit a catalytic Ser, which is activated as a nucleophile by the two triad residues His and Asp in various structural variations [24].

Some of the work performed from the 1960s to the 1980s on chymotrypsin substrates, containing about a dozen ncAAs, was summarized in the studies of the Jakubke and Laskowski groups [132,133]. In addition, a quantitative structure–activity relationship (QSAR) was derived for the substrate fragments, i.e., essentially, the individual residues and overall log (k_{cat}/K_M) values of the substrates resulted from additive contributions of the fragments [133]. Based on these data, CTRA aldehyde inhibitors were synthesized with modified Phe residues in the P1 position, of which Tos-Phe-Ala-Thr-Phe(p-NO₂)-CHO had a K_i of 11 nM [134]. A previous investigation on the chymotrypsin-like *Streptomyces griseus* protease B (SGPB, S01.262) with the uncommonly favored P1-Glu analyzed the interaction with the 55-residue-long third domain of the turkey ovomucoid inhibitor (OMTKY3, I01.003) using a homologous series of the aliphatic P1-side chains Gly, Ala, Abu, Ape (Nva), α -Ahx (Nle) and Ahp (Figure 1) [135]. In a more comprehensive work on OMTKY3, some ncAAs, such as Abu, Ape, α -Ahx and Hse, were incorporated into the P6-P3' reactive-site loop of the inhibitor, which was assayed with CTRA and other serine proteases [132]. These studies established the sequence-to-reactivity algorithms (SRAs), that combine structural data, free energy and kinetic constants into a predictive tool [136].

Bovine pancreatic trypsin inhibitor (BPTI, aprotinin, I02.001) inhibits trypsin and chymotrypsin in a substrate-like manner, with the segment Pro13-Ile19 binding the S3–S4' specificity sites. An enzyme kinetic and structural study of bovine β -trypsin and BPTI with the P1-Lys15 mutations to Abu, DfeGly and TfeGly demonstrated that ethyl-Gly largely reduced inhibition, whereas the fluoroethyl-Gly side chains, in particular completely fluorinated uAAs in BPTI, enhanced the inhibitory effect towards α -chymotrypsin [137].

Increasing fluorination was correlated with lower K_i, e.g., K15Abu had a K_i of 1.37 μ M, K15TfeGly (440 nM), MfeGly (340 nM), CHF₂-bearing DfeGly (68 nM), DfpGly (70 nM) and PfpGly with a CF₂-CF₃ group (277 nM), see Figure 1 [138]. By thioester-mediated chemical ligation, an analogue of BPTI was generated, which yielded the modified disulfide N α -ethanethiol-Gly38-Cys14 [139]. Compared to the reaction of unmodified BPTI with chymotrypsin, the equilibrium association constant K_a dropped from 1.7×10^7 M⁻¹ s⁻¹ about 20-fold, which was explained by structural perturbations, as corroborated by 2D ¹H-NMR TOCSY experiments. Using a cell-free expression system, a highly potent variant of the trypsin inhibitor ecotin with allylglycine (Alg) was obtained that allowed to cleave at the P1 reactive site upon iodine treatment and the subsequent release of active trypsin [140].

Arguably, the most comprehensibly analyzed model inhibitor is the sunflower trypsin inhibitor (SFTI-1, I12.002). A series of SFTI variants with β -/ γ -amino acids and N-substituted β -Ala residues in the P1 position (residue 5) was synthesized. In addition, the N-substituted β -alanines β Nhlys and β Nhphe were introduced and exhibited inhibition against bovine CTRA and TRYB in vitro. The two analogs with [b3hLys5]SFTI-1 and [b3hPhe5]SFTI-1 displayed activity comparable to monocyclic SFTI-1 for TRYB and [Phe5] for CTRA with K_a = 1.0×10^{10} M⁻¹ and K_a = 2.0×10^9 M⁻¹, respectively, whereby the K_i is about 500 pM [141]. Monocyclic linear SFTI-1 was engineered in the P1 position with Phe or N-benzylglycine (Nphe) and in the P1' position with Hse or [N-(2-hydroxyethyl)-glycine] (Nhse), in order to efficiently inhibit CTRA, resulting in a best K_i of about 9 nM [142].

Spumigins are linear analogs of tetrapeptidic protease inhibitors from the cyanobacterium *Nodularia spumigena*, found in the Baltic Sea [143]. They consist of an N-terminal capping group, which is a Tyr derivative with an OH group instead of the NH₂ group, followed by hPhe or 4-Me-hPhe, then Pro or 3-Me-Pro, and as C-terminal residue either Arg, arginal or the corresponding arginol with a CH₂-OH moiety. It can be estimated that the variant spumigin E, which exhibits a P1 arginal, inhibits TRYB with an IC₅₀ of about 200 nM. Other cyanobacteria such as Dolichospermum planctonicum produce a cluster of so-called radiosumins, which essentially are dipeptides comprising two unusual ncAAs with various double bonds [144]. Radiosumin C consists of two L-(2-amino-3-(4-amino-2cyclohexen-1-ylidene)-propionic acid (Aayp) as P1 and P1' residues, with two acetyl caps and a free C-terminal carboxylate, whereas radiosumin D differs slightly by a methylated 4-amino group of the P1 residue (L-Amyp) and a D-Aayp in P1' position. Radiosumin C inhibited the three human isoforms trypsin-1 (S01.127), trypsin-2 (S01.258) and trypsin-3 (mesotrypsin, S01.174) with IC₅₀ values of 1.7, >7.2 and 2.0 µM, respectively. Extracts from the marine sponge Theonella swinhoei, found near Madagascar, contained the three cyclotheonellazoles A-C and the related oriamide, whereby compound A inhibited CTRA with an with IC₅₀ of 620 pM and CELA1 with an IC₅₀ of 34 pM [145]. A capped N-terminal Gly-Ala is attached to a ring formed by amino-Ala-Nle-Ile, then an inserted carbonyl, a complex amino acid with a Tyr side chain, but a thiazole group and an ethylene group in the "backbone", an amino acid with HO₃S-CH₂- as side chain that closes the ring. Another complex cyclic inhibitor of trypsin-like proteases, lyngbyastatin 4, was discovered in the marine cyanobacterium Lyngbya confervoides. [146]. Lyngbyastatin 4 features several ncAAs, such as N-me-Tyr, 3-amino-6-hydroxy-2-piperidone (Ahp), Abu and hTyr; it selectively inhibited chymotrypsin (IC₅₀ = 300 nM) and elastase (IC₅₀ = 30 nM), whereas the activity of trypsin, thrombin and plasmin was not affected.

One of the most uncommon types of uAAs comprises side chains with nucleic bases from DNA or RNA, respectively, which have been successfully incorporated into proteins [147]. For example, MCoTI-II from the squash *Momordica cochinchinensis* is a macrocyclic inhibitor of trypsin, comprising a so-called cysteine knot and a P1-Lys, which was substituted with Guanine-Ala as an isoster of Arg, however, it did not improve the picomolar inhibitory potency of the natural form [148].

3.1.2. Thrombin and Other Blood Coagulation Factors

Based on its crucial role in blood coagulation, thrombin (S01.217) and its major substrate fibrinogen are one of the most studied proteolytic systems [149]. Prothrombin, also known as coagulation factor II, possesses a so-called Gla domain, like several other coagulation factors. The Gla domain contains 10 γ -carboxy glutamates (Gla), which bind Ca²⁺ with high affinity, facilitating the proper function of these coagulation factors. One of the most common chromogenic substrates of the trypsin-like thrombin in enzymatic assays is D-Phe-Pip-Arg-pNA (S-2238, Pip is pipecolic acid), which resembles many substratelike inhibitors, such as the irreversible *D*-Phe-Pro-Arg-chloromethyl ketone (PPACK) or D-Phe-Pro-D-Arg-D-Thr-CONH₂, inhibiting in the nanomolar range [150]. Measurements of the coagulation reaction for thrombin with fibrinogen by thrombin time (TT) and the thrombospondin time (APTT) confirmed that the inhibitor *D*-Phe-Pro-Orn-(N δ -SO₂NH₂)-OMe significantly prolonged the clotting time [151]. For a combined X-ray crystallographic and NMR study, 5-19F-Trp was incorporated into the zymogen prethrombin-2 and mature thrombin by expression in minimal medium, resulting in additional evidence for the conformational selection model of enzyme kinetics with a shifting E*-E equilibrium (Figure 3B) [73,131].

As Pro is the preferred P2-residue in thrombin substrates and inhibitors, a large set of P1-Arg-boronate inhibitors was synthesized with Pro analogues, which contained perhydroindole (Phi), azabicyclo[2.2.2]octane (Abo) and azabicyclo[2.2.1]heptane (Abh) or the Pro mimick N-cyclopropylhexyl glycine in P2 position [152]. The corresponding compounds inhibited thrombin and plasmin with IC_{50} values in the low nanomolar and

picomolar range. The same P2 screening strategy was applied to neutrophil elastase (HNE, see section below), which prefers small hydrophobic P1 residues, yielding inhibitors with low nanomolar K_i values [153]. Further examples of peptidic thrombin inhibitors with nanomolar potency exhibit the basic structure Z-D-Phe-Pro-boro-Mpg/Irg, whereby Phe could be replaced by -CH-Phe₂ side- chains and the boro-moieties by phosphonates, respectively [154]. The highly potent thrombin inhibitor hirudin (I14.001) exhibits a K_i value in the low femtomolar range, depending on the source of 1 to 22 fM. Hirudin-derived heptadeca peptides, in which the critical residue Ile59, i.e., position 11, was replaced with tert-butylalanine (Tba), decreased the K_i from 840 to 390 pM [155]. Similarly, several engineered variants of hirudin, especially the so-called hirunorms IV and V, inhibited thrombin in the picomolar range, due to the presence of Aib, β -Ala, Cha, Nal, D-Ala, D-Glu and hPhe (Figures 1 and 2) [156]. The crystal structure of human α -thrombin in complex with hirunorm V confirmed that the N-terminus of the inhibitor binds via a reversed backbone as follows: Chg1 (S2), Val2 (S1) and 2-Nal (S4) and Thr4 (S3) (Figure 4B) [157]. Including ncAAs in the natural fragment 1-47 of hirudin by the mutations Val1tBug (tertbutylglycine, Tle) and Tyr3-2-Nal decreased the K_d (K_i) for the fast thrombin form from 41 nM to 1.1 nM [158]. Furthermore, the tBug1-Arg2-2-Nal variant replaced the original Val-Ser-Tyr and inhibited the fast (Na⁺ bound) and the slow (Na⁺ free) form of thrombin up to 7000-fold more effectively than the natural hirudin fragment, with K_d values of 15.4 pM (fast) and 220 pM (slow), respectively [159].

In addition, the established engineering of hirudin facilitated incorporation of uAAs like 7-azatryptophan and 3-nitrotyrosine (3NY), which reduced the binding affinity for thrombin, but the products were useful as fluorescent probes [160,161]. Rationally designed bivalent thrombin inhibitors targeted the fibrinogen exosite of thrombin and blocked the active site with a non-scissile bond at P1-P1', based on arginyl ketomethylene isosteres, whereby Xaa as the P1' residue was Gly, Ala or Pro [162]. Hereby, the most potent compound was *D*-Cha-Pro-Arg ψ [CO-CH₂-S]-Gly₅-Asp-Tyr-Glu-Pro-Ile-Pro-Glu-Glu-Tyr-Cha-*D*-Glu-OH, with a K_i value of 350 fM.

Crystal structures of bovine thrombin and the inhibitors NAPAP, 4-TAPAP and MQAP revealed the distinct binding modes of these peptidomimetics, which occupy the S4, S2 and S1 subsites with either naphtyl or tosyl sulfonamide-Gly-(4-amino)Phe/Arg and a piperidine or piperidine-2-carboxylic acid (pipecolic acid, Pip), see Figure 4C [163]. The inhibition constants for human thrombin were roughly 6 nM (NAPAP), 170 nM (4-TAPAP) and 20 nM (MQPA), while the related compound N-a2-naphthyl-sulfonyl-3-amidino-Phe-4-methylpiperidide (L-NAPAMP) improved the K_i to 2.5 nM [164]. Analogs of the thrombin inhibitor NAPAP, in which the P2-Gly was replaced by uAAs, such as 1-aminomethylglycine (Amg), maintained the potency with a K_i of 10 nM and good selectivity among other trypsin-like proteases, and 4-amidino-Phe in P1 was introduced for alkylation extensions [165]. Additional carboxylic groups reduced the systemic and hepato-biliary clearance of similar NAPAP analogs in rats, which may promote pharmacological applications [166]. A study of extended thrombin inhibitor constructs resulted in compounds such as dansyl-Arg-D-Pip (P4–P1), β-Gly-Gly-Gly-D-amino-pentanoic acid-(P1'-P4')-hirudin-55-65 (exosite I) or μ -amino dodecanoic acid in the P1'–P4' stretch, which all inhibited in the lower picomolar range [167,168]. A prime side screening with sarcosine (Sar), D/L-Ala, D/L-3-aminoisobutyric acid, N-methyl-a-Ala and Cha improved the inhibitory potency of the resulting Bbs-Arg-(D-Pip)-[linker]-DYEPIPEEA-Cha-D-Glu, with a phenyl ring and ζ-amino heptanoic acid as the linker main chain [169,170]. The compound inhibited thrombin with a K_i of 230 fM, relatively close to the potency of hirudin with a K_i of at least 22 fM. Another synthetic compound, D-Arg-Oic-Pro-D-Ala-(p-Me)F-NH2 (FM19), showed promising effects in platelet inhibition of acute coronary syndromes associated with diabetes, despite being a low-affinity inhibitor of thrombin (Figure 2) [171]. The X-ray crystallographic structure corroborated a reverse binding mode, where the *D*-Arg residue of *FM19* binds to the S1 pocket.

Aeruginosins are natural small-molecule inhibitors of the coagulation factors thrombin and factor VIIa, which contain a D-amino acid in the P3 position and inhibited in the nanomolar range. However, a D-P3 residue usually occupies the S4 subsite of trypsin-like serine proteases, whereby the strongest inhibition of thrombin was observed for a hybrid aeruginosin with a corresponding "P3" *D*-3R-chloroleucine with an IC₅₀ of 1.6 nM [172]. Sub-picomolar inhibition was found for the 32-residue-long tsetse fly thrombin inhibitor (TS), which was screened for sulfo-Tyr (sY) variants, yielding (N-Me)L32-sY9/sY12-TTI with a K_i of 60 fM [173].

A study on inhibitors of the urokinase-type plasminogen activator (uPA, S01.231) showed that the murine counterpart of upain-1, the cyclic mupain-1 (CPAYSRYLDC) with L-4-guanidino-cyclohexyl-Ala instead of P1-Arg6, inhibited murine uPA 10-fold better with a K_i of 45 nM [174]. A similar screening of mupain-1 was based on a phage-display peptide library with P1-Arg6 analogs, such as 4-guanidino-Phe or F(gua), which improved the K_i to about 185 nM for the variant position, supported by the Asp9Ala mutation (Figure 4D) [175]. Activated protein C (APC), thrombin and factor Xa were screened for the P4–P1 specificity with about 50 AAs, of which the unnatural ones often resembled the natural ones [176]. Based on these results, fluorescent substrates of the type of Ac-Lys-Dab(Z)-Igl-Arg-ACC reached catalytic efficiencies of 71,000 up to 8,700,000 $M^{-1} s^{-1}$, whereby Dab(Z) and Igl are benzyloxycarbonyl-L-2,4-diaminobutyric acid and L-2-indanyl-glycine. Subsequently, the corresponding ABPs allowed to detect and to quantify their target protease from the coagulation cascade in human plasma. A similar approach resulted in the first fluorescent ABP that is selective for factor XIa over other coagulation proteases, namely, Bodipy-PEG4-Tyr(2,6-Cl₂-Z)-Nle-Glu(Bzl)-ArgP(OPh)₂ [177]. As blood coagulation factors and related proteases are favored targets for ABPs, the factor VII activating protease could be detected in human blood plasma by Cy5- ε -Ahx-Pro-D-Tyr-Lys-Arg-P(OPh)₂, with the fluorescent cyanine5 group (Cy5).

A de novo design study was based on a DNA library with a random sequence of ten residues encoding at least three uAAs cyclized by two Cys residues. These ncAAs included homopropargylglycine (Hpg), canavanine (Can), 4-Bromo-Phe, a Lys analog with a double bond between C γ and C δ , tBu-Gly (Tle) and a Pro with a sulfur in γ -position. K_i values for thrombin inhibition ranged from 6 to 35 nM for the three best cyclic inhibitors, which comprised a linker and a His₆ tag [178].

3.1.3. Kallikrein-Related Peptidases, Cathepsins, Neutrophil Serine Proteases and Tryptases

Kallikrein 1 (KLK1, formerly hK1, S01.160), the "true tissue kallikrein", has a dual chymotryptic/tryptic specificity for P1-Tyr/Arg [164,179]. Surprisingly, the screening for substrates with basic ncAAs in the P1- and P3-positions resulted in KLK1 inhibitors of the nanomolar range, exhibiting *trans*-4-aminomethylcyclohexylalanine (Ama) in Abz-F-Ama-S-R-Q-EDDnp, whereas variations of basic and hydrophobic uAAs in P3 resulted in phenylacetyl-Aca-Ser-Arg-NH₂ as the best inhibitor (Aca is trans-4-aminocyclohexylalanine, Figure 2) [180,181]. Kallikrein-related peptidase KLK2 exclusively prefers basic P1 residues, as corroborated by the best fluorogenic substrate Ac-Orn-Phe-Arg-AMC, with the shorter Lys homolog ornithine (Orn) in the P3 position [182]. The chymotryptic PSA or kallikreinrelated peptidase 3 (KLK3, S01.162) is a major target of biochemical and pharmaceutical research, because it is a critical factor in prostate cancer. Screening of P2 and P3 libraries with ncAAs, such as Cha, Hse, Met(o), Nle, Nva, Phg and Nal, yielded peptidomimetic aldehyde inhibitors like Z-SSK-X-L-al and boronic compounds like Z-S-(N-Me)SKLL, with a K_i value of about 200 nM [183]. Similar to tissue kallikrein KLK1, plasma kallikrein (KLKB1, S01.212) regulates the blood pressure by releasing bradykinin from kininogens. Moreover, it is involved in the contact activation system by activating factor XII, which subsequently initiates the intrinsic blood coagulation pathway, and it is a suitable target in diabetesrelated pathologies [184]. A complex cyclic peptide with ncAA substitutions, such as hArg, Aze, Ala(ψ CH₂NH) and three tribromomethylbenzene-linked Cys residues, inhibited KLKB1 with a K_i of 360 pM and it had positive effects in rodent diabetes model systems.

Human cathepsins comprise two serine proteases (CatA, S10.002 and CatG, S01.133), of which cathepsin G is more important due to its role in the neutrophil leukocytes' mediated

immune response [185]. A study on the human chymotrypsin-like CatG may serve as a major example for the usage of ncAAs in chromogenic substrates. As aliphatic and aromatic side chains such as Leu, Phe and His are preferred in P1 by CatG, a tetrapeptide library with 5-amino-2-nitrobenzoic (Anb) as chromophore was generated, with the general formula Ac-Phe-Val-Thr-Xaa-Anb [186]. The substrate with 4-guanidino-Phe F(gua) in P1 had a catalytic efficiency (k_{cat}/K_M) ten times higher than the second best 4-amino-Phe, which was still preferred over P1-Tyr and Phe. The already mentioned sunflower trypsin inhibitor (SFTI-1) is one of the most frequently engineered model systems that was adapted to many serine proteases, such as chymotrypsin and CatG, which were strongly inhibited by P1-Arg5 variants with para-substituted Phe, e.g., F, NO₂, NH₂, Me and Gua, according to the association constant K_a [187,188]. Matriptase-1 (S01.302) and Matriptase-2 (S01.308) are cancer- and iron homeostasis-related type II trans-membrane serine proteases with a strong preference for P1-Arg. A systematic screening of uAAs in positions P4–P2 of inhibitors with a ketobenzothiazole serine trap moiety in the P1' position resulted in a compound with a K_i of 4 nM for matriptase-1 and 400 pM for matriptase-2 [189].

The optimization process for fluorogenic FRET or IQF substrates in the P1' to P3' positions in a comparative study of the serine proteases cathepsin G (CatG) and human neutrophil elastase (HNE, elastase-2), proteinase 3 (PR3) and "non-structural protein 4" (NSP4) found the most potent substrates exhibiting Nva in P1' and 3-Pal or Met(o) in P3' [190]. Previously, it was observed that the most efficient ABP for CatG contained a P1-F(gua), whereas HNE reacted only with a P1-Leu probe [191]. As HNE levels rise in blood upon microtrauma, plasma and in vitro stability was tested for decapeptides with ethylglycine, 4,4-difluoroethylglycine (DfeGly) and 4,4,4-trifluoroethylglycine (TfeGly) in positions 4, 6 and 7 (Figure 2) [192,193]. A study with the same setup demonstrated that peptides with these ncAAs in the P2, P1' and P2' positions were better substrates for chymotrypsin and pepsin [194]. A large hybrid combinatorial study substrate library (Hy-CoSuL) with 102 uAAs was employed to explore the S1–S4 pockets of HNE, resulting in the ideal substrate Ac-Nle(OBzl)-Met(o)₂-Oic-Abu-ACC and the corresponding ultrasensitive ABP *PK101* [195]. The corresponding ABP, namely, biotin-PEG4-Nle(O-Bzl)-Met(o)₂-Oic-Abu-PO₃Ph₂, was crystallized in complex with HNE and revealed details of the interaction with the S4–S1 pockets, e.g., exosite binding beyond S4 [196]. FR901277 is a natural bicyclic inhibitor of elastases isolated from Streptomyces resistomicificus, with an IC_{50} of 180 nM for HNE and 260 nM for porcine elastase (Figure 4E) [197]. The sequence can be described as Thr-Orn-Thr-dehydroxy-Thr-AA4-Phe-AA6-Val, whereby AA4 and AA6 are complex ring-containing amino acids, with additional unusual ester and amide bonds formed from the Thr-OH to the Val8 carboxylate and from the Orn2 ϵ NH₂ to a carboxylate of AA6. A high-resolution X-ray structure confirmed that the compound binds from the S4 to S2' pockets of porcine elastase, with dehydroxy-Thr as the P1 residue. Focusing on NSP4, the optimized substrate Ac-hCha-F(gua)-Oic-Arg-ACC was obtained, as well as the highly specific ABP biotin- ϵ -Ahx-hCha-F(gua)-Oic-ArgP(OPh)₂ (Figure 2) [198]. Based on a Hy-CoSuL, highly selective ABPs for cellular assays were developed, in order to target the four neutrophil serine proteases HNE, CatG, PR3 and NSP4 [199]. For another trypsin-like serine protease from leukocytes-granzyme B from natural killer cells-optimal substrates were generated, which contained ncAAs in positions P5 to P2 such as Tic, Oic, Hyp, Tyr(Bzl) and others, which subsequently allowed to synthesize specific ABPs [200].

Human mast cells secrete both the trypsin-like α -tryptase (S01.143) and the more abundant β -tryptase (S01.015) as part of distinct immune responses, whereby mainly β -tryptase forms tetramers upon activation [201]. As tryptases are involved in inflammation and allergy, inhibitors could be valuable pharmaceutical compounds, such as cyclotheonamides from the marine sponge species *Ircina* [202]. The cyclotheonamides E and E4 inhibit human tryptase, with IC₅₀ values around 5 nM, and the E5 variant inhibits at 85 nM. Both contain the unusual ncAA with an amino-methyl side chain and a Tyr derivative with a CH=CH group inserted between the C α and the carbonyl C.



Figure 4. Trypsin-like proteases with various inhibitors. (A) Bovine trypsin in complex with a variant of bovine pancreatic trypsin inhibitor (BPTI, PDB 4Y0Z). The protease with the catalytic Ser195 (dark red patch) is depicted as molecular surface and the inhibitor as a secondary structure or with side chains as ball-and-stick models (insert). Abu15 binds to the S1 pocket, and the other subsites are occupied by P4 to P2' of the inhibitor. (B) The complex of human α -thrombin with the engineered hirunorm V (green ball and sticks) essentially shows reversed binding of Chg1 (S2 subsite), Val2 (S1), 2-Nal (S4) and Thr4 (S3). The electrostatic potential at the molecular surface of thrombin is contoured for -5.0 to +5.0 K_BT/e , whereby the two positively charged exosites are important for protein interactions (PDB 5GDS). (C) Active bovine ε -thrombin in complex with the inhibitor 4-TAPAP (tosyl-4-p-amidino-Phe-piperidine), which binds to the S4, S1 and S2 pockets. The catalytic triad residues Asp189 and Trp215 are displayed as stick models (PDB 1ETT). (D) Complex of uPA with the cyclic mupain-1 (CPAYSRYLDC) variant and 4-guanidino-Phe6, shown as cyan ball-and-stick model that binds to the S1 pocket (PDB 4X1R). The cyclic inhibitor is stabilized by the intramolecular disulfide Cys1-Cys10. (E) Porcine elastase (CELA1) and human neutrophil elastase (HNE) are inhibited by the natural bicylic inhibitor FR901277. A crystal structure revealed the peptidic linkage of the Orn2 side chain to the nCAA AA6 and the ester bond formed by Thr3-OH and the C-terminal carboxylate of Val8. Another remarkable feature is the P1 residue dehydroxy-threonine, a homolog of dehydroalanine (Dha), with the -CH=CH₂ side chain (PDB 1QR3). (F) A potent inhibitor of the trypsin-like Zika NS2-NS3 protease was formed by cyclization via condensation of an N-terminal Cys with Cpa. The crystal structure of the bZiPro construct revealed the cleaved inhibitor bound to the active site, with the novel uAA ("Cys/Cpa") occupying the S4 subsite (PDB 6JPW).

3.1.4. Viral Serine Proteases

Flavivirueses carry a single-stranded RNA and cause various diseases, such as Dengue fever, Yellow fever or European tick-borne encephalitis, which causes up to 80,000 deaths per year worldwide [203]. Therefore, the crucial major protease flavivirin (S07.001) from the trypsin-related flaviviral serine proteases is a target for drug development. In a comprehensive study of the dengue virus (DENV) NS2B-NS3 protease, fluorescent substrates of the type Bz–Xaa-Arg–AMC enhanced the catalytic efficiency with Xaa = Amf (4-aminomethylphenylalanine), Ama and Aca compared to Arg [204]. With respect to the canonical inhibitor sequence Bz-Arg-Lys-Nle-NH₂ for dengue virus serotype 2 (DENV-2) protease, a substitution of the Bz cap with a more complex residue and a P1'-Phg decreased the K_i to 400 nM [205]. An optimization of DENV protease inhibitors in the P3 and P2 positions yielded trifluoromethyl-Bz-4-amidino-Phe-Lys-Phg-OH with an IC₅₀ of 210 nM [206]. Systematic screening for cap (P3) and P1'- Phg derivatives around the P2-Arg-P1-Lys core of DENV-2 and West Nile virus (WNV) proteases resulted in compounds with thiophene or thiazole caps and P1'-4-hydroxy-phenylglycine ethers, such as Phg-O-4-CF₃-Bz, which reached K_i values of 12 nM (DENV) and 39 nM (WNV), whereby virus replication was also inhibited [207]. Substrates with ncAAs were analyzed for the NS2B-NS3 proteases of DENV, WNV and Zika virus (ZIKV) and allowed to synthesize corresponding ABPs [208]. As D-amino acids are favored over L-enantiomers in P4, the specific ZIKV NS2B-NS3 protease substrate Ac-D-Arg-Lys-Orn-Arg-ACC was generated as well as the corresponding phosphonate inhibitor/ABP biotin-ε-Ahx-D-Arg-Lys-Orn-Arg-P(OPh)₂ [209]. Apart from Arg, $Arg(Me)_2$ was the most specific P1 residue for all these flaviviral proteases, whereas P2-Orn and hLys (DENV) were preferred, and D-Arg or D-Leu (DENV) were favored in P4. Derived from a substrate of the Zika virus protease, a peptide with an N-terminal Cys and a C-terminal 3-(2-cyano-4-pyridyl)alanine (Cpa) cyclized spontaneously to a dihydrothiazole and yielded an inhibitor of proteolytic stability and high affinity for the NS2B-NS3 bZiPro construct with a K_i of about 140 nM. However, the linear form of the inhibitor was observed in a complex bZiPro crystal structure, with the novel uAA "Cys/Cpa" binding to the S4 pocket and beyond (Figure 4F) [210].

The related hepatitis C virus (HCV) comprises the NS2 cysteine protease CPro-1 (C18.001), which directly precedes the NS3 serine protease (hepacivirin, CPro-2, S29.001), with a Zn^{2+} contributing to autoproteolysis of the viral polyprotein [211]. The P1 specificity was investigated for the NS3 domain with and without the NS4 activator peptide, using substrates that contained both cAA and ncAA residues like hCys, Abu, and Alg, nearly reaching the catalytic efficiency of P1-Cys substrates [212]. In a similar study, substrates with ncAAs derived from the viral NS5A/B protein showed no improvement but generated the H₂N-EDVLC-Tic-Nle-SY-OH inhibitor (Tic: 1,2,3,4-tetrahydroiso-quinoline-3carboxylate) with a K_i of 340 nM [213]. A purely computational QSAR screen for ncAAs in hexapeptides, docked to the S6-S1 subsites of hepacivirin, predicted inhibitors with subpicomolar Ki values, namely, Ac-Asp-D-Gla-Trc-Asp-Asp-Fab/Cyo, which exhibited 4carboxy-Trp (Trc), α -keto- δ , δ -difluoro-aminobutyric acid (Fab) or α -ketocysteine (Cyo) [214]. Extensive screening for the P3 to P2' positions of peptidomimetic inhibitors with Chg, Nva or Phg, yielded a best K_i of 12 nM, and an X-ray structure of an NS3/4a protease complex gave further insights into details of inhibitor binding [215]. Several inhibitors with low nanomolar potency consist of uAAs, such as Pro derivatives with isochinoline ethers as substituents at the C γ or a residue with a vinyl-cyclopropyl side chain, which covalently links to the non-catalytic Cys159, achieving an IC_{50} of 2 nM [216]. Proline derivatives fused to C60 fullerenes at the C β and C γ atoms with a carboxylate substitution at the C δ atom inhibited the HCV NS3/4A serine protease with an IC_{50} of 150 nM, making them promising virostatic lead compounds [217]. After successful clinical phase III trials, the low nanomolar inhibitor of hepacivirin, Voxilaprevir, was approved by the US Food and Drug Administration (FDA) for treatment of patients [218]. It features a Tle residue and three cyclopropane groups, among them an uAA with a difluoromethylcyclopropyl moiety, including the $C\alpha$ atom.

3.1.5. Subtilisin-like Serine Proteases and α/β -Hydrolases

An early screening study with ncAA containing inhibitors of the P1-Arg-specific furin (S08.071) and the subtilisin-like proprotein convertase-1 (PC1, S08.072) utilized the sequence D-Tyr-Arg-Ser-Lys-Arg-Xaa-Val-Gln-Lys-Asp, whereby Xaa in the P1' position was various ncAAs [219]. Inhibition of both proteases occurred with γ -Abu, β -Cha and β -Ala as Xaa, which resulted in a nanomolar K_i for furin with the β -Ala derivative. Essentially, this project was extended by using, e.g., Tle, Sar, MLeu, Aib, D-Tic, and L-Tic demonstrating that inhibition of furin was more efficient with P1' Tle (*tert*-butylglycine) (Figures 1 and 2) [220]. A more potent inhibitor of furin was 4-guanidinomethyl-phenylacteyl-Arg-Tle-Arg-4amidinobenzylamide, with a K_i value of 5.5 pM [221]. Substrate-derived compounds with aminooxy-acetic acid (Aoaa) or 8-amino-3,6-dioxa-octanoic acid (Adoa) in the P1 and P1' positions inhibited the subtilisin kexin isoenzyme I (S08.063) from the proprotein convertase family in the low micromolar range [222]. Semisynthetic variants of the inhibitor eglin C (I13.001) obtained by native chemical ligation, phage display and stop codon suppression for the incorporation of kynurenic acid (Kyn) were captured by subtilisin BPN' (S08.034) for selection [223]. SufA (S08.138) from the subtilisin-like serine protease family hydrolyzes human fibrinogen during infections with the gram-positive bacterium Finegoldia magna, which motivated substrate screening of uAA P1-Arg analogs that led to the synthesis of the inhibitor Cbz-6-AmNphthP(OPh)₂ with micromolar potency [224].

Prolyl oligopeptidase (POP, prolyl endopeptidase, S09.001) belongs to the α/β -hydrolase family and exhibits a large propeller domain, which forms a compartment that harbors the active site. It cleaves after P1-Pro, whereby a marked preference for Pro in positions P4, P3, P1' and P3' was observed [225]. Several POP keto heterocyclic inhibitors have been developed based on analogs of the central P1-Pro, such as thioproline (Thp), Phi, Abo and Abh, resulting in IC₅₀ values in the low nanomolar range (Figure 2) [226]. The synthetic inhibitor benzylcarboxylate-Pro-2-formylpiperidine was crystallized in a complex with human POP containing an unnatural imino acid derivative (Figure 5A) [227].

Quite unusually, penicillin-binding proteins (PBPs) with peptidase activity can cleave peptidoglycans at positions with D-amino acids, resulting in biofilm formation. Among them are PBP4, PBP4a and PBP5 from *E. coli* and *Bacillus* subtilisin (*D*-Ala-*D*-Ala peptidases S13.001/2 and S11.001/8), respectively, which also cleave at *D*-Asn, *D*-His, *D*-Trp and various other ncAAs [228]. General inhibitions of PBP protease activity have been achieved by incorporating ncAAs into pharmaceutical peptides for better efficacy by increasing intracellular concentrations [69,229]. Lactoferrin (S60.001) is first and foremost an iron transport protein that, surprisingly, can function as a protease with trypsin-like specificity, cleaving substrates such as Z-Phe-Arg-AMC [230].

3.2. ATP-Dependent Proteases

Self-compartmentalizing proteases perform important tasks in the degradation of misfolded proteins in all organisms and in the immune system of higher organisms, with the 26S proteasome as the most prominent example of the N-terminal nucleophile (Ntn) hydrolases with a catalytic Thr1 [231]. In addition, the ATP-dependent ClpXP-complex exists in many bacteria and human mitochondria, whereby ClpP (caseinolytic protease P, S14.003) exhibits a chymotrypsin-like P1 specificty, whereas ClpX is an AAA-ATPase/unfoldase. Based on a HyCoSuL profiling, the ideal substrate Ac-Phe(3,4-Cl₂)-hArg-Leu-ACC for *Staphylococcus aureus*, *E. coli* and human ClpP protease was obtained [232]. A related study screened many P3 to P1 residues, including uAAs, and established that 2,5-dichlorobenzoyl-Trp-Leuboronate inhibits ClpP of *S. aureus* only in the nanomolar range [233]. In an indirect manner, natural cyclopeptidic inhibitors of the ClpC1 unfoldase from *Mycobacterium tuberculosis* can effect a completely uncontrolled protease activity of ClpP that kills the mycobacterial cells [234]. These protease activity-stimulating compounds, e.g., ilamycins, rufomycins, and cyclomarins from marine *Streptomycetes* contain unusual ncAAs, such as nitro-Tyr, various N ϵ 1-substituted Trp or isopentanol and butene side chains.



Figure 5. Compartmentalized serine and threonine proteases. **(A)** Prolyloligopeptidase active site with the aldehyde inhibitor Cbz-Pro-formyl-piperidine, an imino acid (PDB code 2XDW). The covalently bound inhibitor occupies the S3 to S1 subsites (yellow ball-and-stick model), and the left panel displays the propeller domain (red) and the peptidase domain (orange). The catalytic triad (Ser554, Asp641, H680) is shown as sticks in the right panel. **(B)** The eukaryotic 20S core of the proteasome consists of two rings of seven α subunits and two rings of seven β subunits, with the active β 1 (PGPH, caspase-like), β 2 (green trypsin-like) and β 5 (yellow, chymotrypsin-like) subunits (left panel). The synthetic inhibitor (salmon ball-and-stick model) occupies specificity pockets S1 to S4, with the ncAA aminomethyl-Phe mimicking a P1-Arg or Lys. A covalent bond is formed between the vinyl sulfone and the catalytic Thr1 (PDB code 4INR). Such compounds enhance the effects of anti-cancer drugs, e.g., bortemozib, which mainly target the β 5 active sites (right panel).

Lon or PIM1 is a highly conserved serine protease (S16.002) present in all domains of life. It consists of a substrate-binding domain, a serine protease domain, and an ATPase domain on a single polypeptide chain, which forms hexamers and mainly degrades misfolded proteins [235]. A HyCoSuL screening for *E. coli* Lon specific substrates resulted in the potent covalent boronate inhibitor Pyz-hArg-NptGly-Leu-B(OH)₂ with a k_{inact}/K_{I} of 3400 s⁻¹ × M⁻¹, which resembles the proteasome inhibitor bortemozib, but confers higher selectivity for Lon over the proteasome (NptGly is neopentylglycine) [236].

The eukaryotic proteasome with the active subunits β 1, β 2 and β 5 (T01.010/11/12) and its bacterial counterparts such as HslVU (ClpQ, T01.006) in E. coli are the most widespread and important representatives of this protease class. Although the ubiquitin-26S proteasome machinery is crucial for protein degradation, immuno- and thymo-proteasomes fulfill additional tasks [237]. Dysregulated proteasomes are associated with cancer, which makes them suitable targets for pharmaceutical inhibitors that can selectively bind to the active sites of the $\beta1$ (PGPH, caspase-like specificity for Asp and Glu), $\beta2$ (trypsin-like) and β 5 (chymotrypsin-like) subunits. A study on internally quenched fluorescent (IQF) substrates with both ncAAs and uAAs in the P1' and P2' positions for the 20S proteasome found an increased specificity for the chymotrypsin-like subunit in comparison to shorter tetrapeptides [238]. Both fluorescent substrates and ABPs for distinct protease subunits of the 20S core particle in human cell lines contained various ncAAs in the P1 position and facilitated to monitor the activity and localization of proteasomes [239]. However, the human proteasome is an important target in cancer therapy because inhibition of its chymotryptic activity results in a highly selective cytotoxicity to cancer cells [240]. Resembling chymotryptic substrate stretches, new CBZ-AzBzH-Ile-Leu-boronate-Phe/Leu (AzBzH, (S)-6-Azido-2-(benzyl-oxy-carbonylamino)-hexanoic acid peptidomimetics were synthesized that had lower IC₅₀ values than the pharmaceutical drugs bortezomib and carfilzomib, which contains an hPhe. Nevertheless, inhibitors of the trypsin-like proteasome activity serve as enhancers or sensitizers of bortezomib and carfilzomib in the treatment of cancer, and thus the highly potent inhibitor α -azido-Phe-Leu-Leu-4-aminomethyl-Phe-methyl vinyl sulfone was developed and structurally characterized as *Saccharomyces cerevisiae* 20S complex (Figure 5B) [241]. Although they are not ATP-dependent and are classified as a mixed protease family (P01), the octameric β -aminopeptidases BapA from *Sphingosinicella xenopeptidilytica* and DmpA from *Ochrobactrum anthropi* belong to the Ntn hydrolases with an $\alpha\beta\beta\alpha$ architecture like the proteasome [242]. Both BapA and DmpA cleave ncAAs with an N-terminal β -amino group, i.e., β 2AAs with a side chain R linked to the C α and β 3-AAs with the side chain R at the C β [243]. BapA and DmpA hydrolyzed dipeptides such as H- β 2hPhe- β 2hAla-OH, which was accompanied by a diastereomeric conversion of β 2hAla, but they also catalyzed the ligation of oligopeptides, such as H-[β 2hPhe]3- β 2hAla-OH.

3.3. Cysteine Proteases

Most cysteine protease contain a catalytic triad or dyad, with a nucleophilic Cys and an activating, stabilizing His [24]. As already mentioned, human cathepsins comprise serine proteases, as well as aspartic proteases (CatD, CatE) and 11 cysteine proteases [185]. Cathepsin B (CatB, C01.060) is related to the prototypical papain (C01.001) and is a major drug target, as it is involved in Alzheimer's disease and various cancer processes. Therefore, a recent study on CatB was performed with a library containing many ncAAs, in order to obtain optimal substrates and fluorescent ABPs [244]. Surprisingly, the most preferred P1 residues of CatB were exclusively uAAs, such as Lys(2Cl-Z), Cys(Bzl), Cys(MeBzl), Cys(Me)Bzl and Nle(OBzl), as well as hSer(Bzl) in P2, Phg in P3 and hCha in P4. Based on these results, ABPs for cathepsins B, L, V and S were obtained and tested in cancer cell lines. An emerging field in this context is the investigation of protease activity in cells and organisms or tumor tissue extracts with the HyCoSuL approach, which was applied with ABPs for cathepsins B, S and L [245]. Cathepsin C or dipeptidyl peptidase I (CatC, DDPI, C01.070) is a lysosomal papain-like cysteine protease and an important activator of other proteases with functions in the immune system and inflammation. A study of human, bovine and malarial CatC with a natural P1-Arg and P2-Met preference revealed that 4-benzoylphenylalanine (Bpa) and Nle(OBzl) are essentially more favorable in the P1 position, and Hse and Abu in P2, whereas Pip is the preferred P2 residue in malarial CatC [246]. Screening of P1 and P2 specificity in papain-like proteases with basic uAAs led to the synthesis of better substrates with respect to P2-Phe/Arg-P1-Arg for cysteine protease B (CPB) and cruzaine from Leishmania mexicana and Trypanosoma cruzi parasites, respectively, as well as for cathepsin L (CatL) and the eponymous papain from Carica papaya [247,248]. By contrast, CatB, a rather unspecific protease with a preference for P1-Gly, turned over ε -Ahx-Leu-Thr(OBzl)-AMC at a high rate, like papain, which favored a P1-Cys(SBzl) [249]. The cytosolic self-compartmentalizing bleomycin hydrolase (C01.084) is relatively unspecific but a target for cancer therapy, for which the substrates Lys(2-Cl-Cbz)-ACC and Cys(Bn) and the corresponding vinylsulfones ABPs have been developed [250].

The C2 family of cysteine proteases comprises the calpains, e.g., calpain-I (C02.001), whereby a Ca²⁺ regulates the activity of the multisubunit calpastatins (I27). In a comparative study of calpain-1/2 (μ /m) and cathepsins B, K and L, epoxide compounds with low nanomolar potency were developed, whereby the two best inhibitors, *WR13* and *WR18*, exhibited Alg in the P2-position and γ -cyano- α -aminobutyric acid [251]. In order to assess the role of the rotamers for distinct residues at the calpain–calpastatin interface, more than 100 uAAs, such as Amt (α -methyl-tryptophan) and 4-methyl-Phe, including the natural ncAA Sec and others, were modeled using the ROSETTA3 software [252].

Caspases are dimeric cysteine proteases (family C14), which often cleave after the consensus sequence DXXD (P4-P1). Typically, fluoromethyl ketones are the classical covalent caspase inhibitors, e.g., Cbz-Val-Asp-FMK, which was used as a starting model for the incorporation of 2-aminobenzoic acid (anthranilic acid, 2AA) derivatives, yielding Cbz-Val-2AA-Asp-FMK, which inhibited caspases 1, 3, 6, 7, 8 and 9, with IC₅₀ values in the low nanomolar range [253]. A comparative study of caspase-3 and 7 inhibitors, as well as ABPs, demonstrated that DEVD was a discriminating sequence [254]. However, one of the most potent inhibitors of Casp3, with an IC₅₀ of 23 nM, was Ac-3Pal-Asp-D- β hLeuhLeu-Asp-AOMK (acyloxymethyl ketone, 3Pal is 3-pyridyl-alanine), consisting altogether of three ncAAs. This compound was the basis for the synthesis of carboxyfluorescein (FAM) and biotin-tagged ABPs (Figure 6A). As the Ac-LETD-AOMK inhibitor is very potent, but does not discriminate between caspases 3, 6, 7, 8 and 9, a HyCoSuL screening revealed the compound β -Ala-Leu-Glu-Hyp-Asp-AOMK, which was highly specific for Casp8 and Casp9 [255]. Similarly, a comparative HyCoSuL study of Casp-3, 6, 7, 8, 9 and 10 for the P4–P2 positions found very selective substrates for caspases 8, 9 and 10, such as Ac-D-hPhe-Aad-Thr(Bzl)-Asp-ACC, Ac-Oic-Tle-His-Asp-ACC and Ac-Nle(OBzl)-Dab-Nle-Asp-ACC [256]. Internally quenched fluorescence (IQF) substrates with the fluorophore ACC and N(6)-(2,4-dinitrophenyl)lysine (Dnp-Lys) as the quencher exhibited an enhanced signal in specificity profiling of caspases 3, 7, 8 and 9, as well as of HNE, legumain and MMPs 2 and 9 [257]. MALT1 is the only human paracaspase (C14.026) with a unique preference for P1-Arg. In order to obtain potent ABPs, libraries with fluorogenic substrates were screened for the P5-P2' positions, resulting in an idealized recognition sequence $F(gua)-LVSR\downarrow GT/Abu$ and, consequently, in the ABP F(gua)-LVSR-AOMK with an acyloxymethyl ketone warhead [258]. A related study improved corresponding substrates and ABPs with respect to their selectivity for MALT1 by reducing the cross-reactivity with CatB, whereby a P2-Pip was highly critical [259]. Legumain or asparagine endopeptidase (AEP, family C13.004) has a similar fold as caspases, but it is active as a monomer [260]. Employing aza-peptidyl Michael acceptor and expoxide warheads, a cyclopropyl-uAA-Pro compound was discovered that inhibited legumain, with an IC₅₀ of 4 nM, and Cy5-linked counterparts are highly selective ABPs for legumain.

Based on a HyCoSuL P2 screen with many ncAAs, the substrate specificity profile for two ubiquitin-specific peptidases (DUBs), MERS papain-like protease (PlPro, C16.011) and human UCH-L3 (C12.003) was determined, which allowed to generate ubiquitinbased substrates and ABPs containing suitable ncAAs positioned in the C-terminal Ub motif [261]. Autophagins such as ATG4 proteases (family C54) process Atg8 proteins for their conjugation to the membrane component phosphatidylethanolamine, which results in autophagosome formation, as well as for their subsequent deconjugation by SENPs (family C48). Using a positional substrate scanning library (PSSCL) for Atg4 of Trypanosoma crucei of the type Ac-Xaa-Xaa-Cha-Gly-ACC allowed to distinguish it from human Atg4, SENP1, SENP2 and UCH-L3 [262]. The parasite *Plasmodium falcipare* possesses the papain-like falcipains 2 and 3 (C01.046/063) as key targets for anti-malarial drugs [263]. In a study that included the related trypanosomal rhodesain (C01.072) inhibitors with a central Michael system (CO-CH₂=CH₂-CO) and mostly Chg and Phg in the P2 position, the most potent compounds had K_i values around 400 nM. A screening for about 20 cinnamic acid variants as C-terminal caps was undertaken for a hPhe-Leu dipeptide core with 4-aminoquinoline, a component of chloroquine (CQ) as an N-terminal cap [264]. Some of these compounds showed potency against the proteases in the lower μ M range and against the CQ-resistant P. falciparum strain W2.



Figure 6. Cysteine proteases in complex with ncAA-containing compounds. **(A)** Similar to other caspases, human Casp-3 is active as a symmetrical dimer. Ac-3Pal-Asp-*D*- β hLeu-hLeu-Asp-acyloxymethyl ketone (AOMK), depicted as ball and sticks, inhibits Casp-3 with an IC₅₀ of 23 nM and binds covalently to the catalytic Cys163 near the dyad residue His121, occupying the S5 to S1 specificity pockets (PDB 4JJE). **(B)** The viral SARS-CoV-2 MPro is a cysteine protease with a chymotrypsin-like fold. MPro is also active as a dimer, which is stabilized by the so-called domain III (with transparent surface), and the protease consists of the two half domains I and II. Covalent binding to the catalytic Cys145 was established by the efficient activity-based probe Biotin-Peg₄-Abu-Tle-Leu-Gln-vinyl sulfone [265], shown as a spherical model. (PDB 6Y2E). **(C)** *GM4* is a tricyclic compound with His-P1 and comprises a γ -amino acid cis-3-aminocyclobutane carboxylic acid in the P1' position that inhibits MPro at a K_d value of 5.2 nM. A crystal structure revealed the binding mode from the S4 to S1' pockets (PDB 7Z2S) The catalytic dyad is displayed as yellow patches on the molecular surface of MPro. The N-terminus of the dimer counterpart molecule (orange surface) contributes mainly with Ser1 to the S1 and S3 subsite, and vice versa.

Not surprisingly, the SARS-CoV-2 corona virus during the COVID 19 pandemic has stimulated many academic and pharmaceutical studies to explore and develop peptidomimetic inhibitors of the main protease MPro of SARS-CoV-2. This cysteine hydrolase is also termed 3-chymotrypsin-like protease (3CLPro, C30.007), exhibiting a preference for P2-Leu and P1-Gln residues as an active dimer [266]. A recent approach with variations in the P2 and P3 positions, comprising mostly ncAAs, yielded 18 synthetic tripeptidyl MPro inhibitors with the aldehyde warhead β -(S-2-oxopyrrolidin-3-yl)-alaninal at the P1 position, whereby a P3 O-tBu-Thr conferred the highest potency [267,268]. The best compound reached nanomolar values for IC_{50} in vitro, as well as for cellular and antiviral EC_{50} , and crystallographic data corroborated the mode of binding and inhibition. Proline derivatives merged with C60 fullerenes at the C β and C γ atoms and substituted with a carboxylate at the C δ atom inhibited MPro with IC₅₀ values around 1 μ M, while a malonic acid fullerene fusion molecule exhibited an IC_{50} of 200 nM [269]. Also, acyloxymethyl ketones as ABPs for the main cysteine-type protease of SARS-CoV-2 were developed, with the P4–P1 sequence Abu-Tle-Leu-Gln/His, which was additionally modified with P2-Cha, P1-Cit and GlnMe₂ [270]. Based on a comprehensive substrate profiling, the most favored Ac-Abu-Tle-Leu-Gln-ACC and the irreversible SARS-CoV-2 MPro inhibitor Ac-Abu-D-Tyr-Leu-Gln-vinyl sulfone were synthesized, as well as corresponding ABPs with a PEG linker and a biotin moiety (Figure 6B) [271]. Moreover, the crystal structure of an ABP-MPro complex was determined, and active MPro was visualized in nasopharyngeal epithelial cells of patients with COVID-19 infection. A structure-based study combined with virtual screening in the AutoDockFR (AutoDock4, v4.2.6) software was successful, followed by a molecular dynamics simulation with GROMACS 2018.6. Eventually, chloro-acetamidebased compounds were obtained with electrophilic warheads that had low micromolar

affinities [265]. The flexible in vitro translation (FIT) system allows to reprogram the genetic code with so-called flexizymes, which are flexible tRNA acylation ribozymes that facilitate the preparation of acyl-tRNA libraries with all types of ncAAs and the ability to read liberated codons for the synthesis of unusual peptides [272]. Based on this method, cyclic γ 2,4-peptides with up to 16 residues and different uAAs were prepared, to find potent SARS-CoV-2 MPro inhibitors [273]. In particular, His-P1 and the cyclic γ -amino acids cis-3-aminocyclobutane carboxylic acid (γ 1) and (1*R*,3*S*)-3-aminocyclopentane carboxylic acid (γ 2) in the P1' position enhanced the potency of these inhibitors with low nanomolar IC₅₀ values and a best K_d below 1 nM. The binding mode of the cyclobutane-containing inhibitor *GM4* with a K_d of 5.2 nM was analyzed in more detail in a high-resolution crystal structure (Figure 6C).

A less-studied viral protease is the dimeric hepatitis C peptidase 2 or NS2/3 (CPro-1, C18.001), which exhibits a unique fold with two composite His-Glu/-Cys catalytic triads [274]. In addition to inhibition by tosyl-phenylalanyl chloromethyl ketone, a very comlex uAA compound was an effective inhibitor, comprising an expoxide core, a carboxyamide and larger aromatic ring systems with an amino group. Due to the covalent, irreversible inhibition, only an EC₅₀ value of 96 μ M for autoprocessing was measured [275].

3.4. Aspartic Proteases

Typical aspartic proteases such as the prototypic pepsin (A01.001) possess a pair of catalytic Asp residues and nearby stabilizing residues such as Ser, Thr and Tyr [276]. Pepsin was discovered in the early 19th century. It was among the first proteins to be crystallized and was the first one with significant X-ray diffraction. Since the mid-1970s, the natural pepsin inhibitor pepstatin from various Actinomyces bacteria has been characterized, with the sequence isovaleryl-Val-Val-Sta-Ala-Sta, containing the ncAA statin, which is essentially a Leu with a CH₂-OH insertion between the C α and the carbonyl C [277]. A complex structure of human pepsin with pepstatin was determined by X-ray crystallography by the group of Michael James (Figure 7A) [278]. The synthetic derivative Ac-Val-Sta-Ala-Sta inhibited porcine pepsin with a K_i of 1 nM, which is close to that of pepstatin at about 100 pM. Beta secretase 1 (BACE1, A01.004) is a membrane-anchored pepsin-like aspartic protease, which prefers hydrophobic recognition stretches and cleaves the amyloid precursor protein (APP) as a critical step in the development of Alzheimer's disease. In the search for optimal substrates for in vitro assays, a profiling study found that the dodecapeptide Ile-Ser-Glu-Ile-Thi-Thi↓Nva-Ala-Glu-Phe-Arg-His-NH₂ was the best substrate, comprising β -2-thienylalanine (Thi) and norvaline (Nva) (Figure 2) [279]. The transmembrane multidomain γ -secretase, with its aspartic protease subunit presentiin-1 (A22.001), was efficiently inhibited by peptidomimetics comprising hydroxyethyl urea side chains at the P1' and/or P3' positions, whereas amyloid beta ($A\beta$) production was significantly reduced [280]. Moreover, inhibition of γ -secretase-dependent Notch signaling promoted the differentiation of neuroblastoma cells, with implications for tumorigenesis and malignancy.

Polypeptidic inhibitors of the aspartic peptidase renin (A01.007) with antihypertensive potential were derived from the hexapeptide pepstatin. Also, its shorter derivative norstatin in the P1 position conferred nanomolar inhibition in compounds with P3 1-Nal [281]. Renin inhibitors, derived from angiotensinogen residues 8–13, were synthesized with two uAAs, e.g., Phe(OMe) in the P3 and P2 position, Sta and two pseudo-dipeptides in P1-P1' and P2'-P3' [282]. In addition, when incorporated into Boc-Tyr(OMe)-His-AHNA-OEt, a renin inhibitor with β -hydroxy- γ -amino acids, such as 4-amino-3-hydroxynonanoic acid (AHNA), turned out to be a potent compound with an IC₅₀ of 13 nM, and the related compound Boc-Tyr(OMe)-MeLeu-AHPPA- ε -Ahx-Iaa had an IC₅₀ of about 1 nM [283,284]. Due to drug resistance, new inhibitor strategies are constantly developed against viral proteases involved in diseases like AIDS and hepatitis [285]. Early studies on inhibitors of aspartic HIV protease (retropepsin, A02.001) showed that tetrahydrofuranyl glycine (Thfg) increases the binding capacity compared with P2-Asn of the compound *Ro31–8959* [286]. Based on these previous studies, P2-Thfg and P3-pyrazine amide residues yielded corresponding inhibitors with IC_{50} values of 70 pM [287]. In similar studies, allophenylnorstatine (Apns or 3-amino-2-hydroxy-4-phenylbutyric acid) was used as a lead building block, with aromatic side chain variations mostly at the P3 and P2' positions of the HIV protease, resulting in a K_i of 500 pM and acceptable pharmacokinetics [288]. This study was continued, mentioning the peptide Ser-Phe-Asp-Apns-Pro-Ile-Val-NH₂ (P4 to P3') as a potent inhibitor with a K_i of 5 nM [289]. Following these design principles and including other uAAs, HIV protease inhibitors with picomolar potency and antiviral activity were obtained [290]. Extensive structural, functional, and physiological studies were performed for kynostatin (KNI-272), a rationally designed inhibitor, which inhibits HIV-1 protease with a K_i of 5.5 pM [291]. It comprises the three ncAAs methylthioalanine, Apns and thioproline (Thp), as well as the capping groups 5-isoquinolyloxyacetyl (iQoa) and NH-tert-butyl (tBu), thereby mimicking the P3 to P2' residues of a substrate. In combined X-ray and neutron diffraction experiments, high-resolution structures of the HIV-1 protease iQoa-Mta-Apns-Thp-NH-tBu complex were obtained that showed fine details of the hydrogen atom interactions in the active site (Figure 7B) [292]. Kynostatin reached clinical trial phases I/II, but the effect on HIV-infected patients was unsatisfactory [293]. Ahmpatinin iBu (Ac-Val-Val-Sta-Ala-Ahmppa) isolated from a Streptomyces strain significantly inhibited HIV-1 protease with an IC_{50} below 2 nM, with Ahmppa having a methyl-Tyr side chain instead of the Leu equivalent in statin [294]. In addition to preventing the critical dimerization of the two HIV protease subunits by simple lipopeptides with palmitic acid, e.g., Pam-Tyr-Glu-Leu-OH, several tetrapeptides, such as Ac-Ile-Ser-(2Nal)-Glu-Leu-OH, inhibited HIV-1 protease in the nanomolar range [295]. Also, replacing the uAA L-counterpart with Ac-D-4F-Phe in the P3 position decreased the IC₅₀ about 100-fold to 3 nM [296]. Interestingly, HIV protease inhibitors that comprise very large uAAs with fullerene moieties in the side chain can be quite potent, such as Fmoc-Phe(4-aza-C60)-Hyp-Hyp-Lys-OH, with a K_i of 76 nM [297].

Beyond academic research, ncAAs play an important role in the field of antiviral agents and their medical applications. Usually, they are incorporated into peptides and peptidomimetics that inhibit HIV and HCV proteases (hepatitis C virus peptidase, C18.001), such as saquinavir, lopinavir, atazanavir, grazoprevir and glecaprevir [298]. Among these ncAAs are (S)-decahydroisoquinoline-3-carboxylic acid amide, (S)-Piperazine-2-carboxylic acid, Tle, (2S,4R)-4-Hydroxy-2-pipecolic acid, 3,3-diphenyl-alanine or (R)-5,5-Dimethyl-1,3-thiazolidine-4-carboxylic acid. Also, pentapeptides such as iQoa-Mta-Apns-Dmt-Nbu, in which the critical P1 residue is Apns, inhibit HIV protease with a K_i of 2.3 pM [299]. The inhibitor Ac-Phg-Tle-Apns-NH-iBu inhibited HTLV-1 protease (human T-cell leukemia virus-1 protease, A0.2012) in the nanomolar range and, similarly, Apns was crucial in the P1 position of several potent peptidomimetic compounds directed against the aspartic proteases of HIV, HTLV-I and malarial plasmepsin II (A01.023) from Plasmodium falcipare and *vivax* [300,301]. In case of the related plasmepsin V (A01.075), which is an integral membrane type I protein, the natural Arg derivative cananavine (Can) in P3 and residues like Chg in P2 conferred inhibitory potency in the low nanomolar or even picomolar range. The crystal structure of compound WEHI-842 in a complex with the soluble protease domains confirmed the canonical binding (Figure 7C) [302]. Canavanine is translationally active and can replace Arg residues in cell-free protein biosynthesis (Figure 1) [303].



Figure 7. Aspartic proteases in complex with ncAA inhibitors. (**A**) Human pepsin in complex with the bacterial inhibitor pepstatin from *Actinomycetes* ($K_i = 100 \text{ pM}$), comprising an isovaleryl acetyl cap (P4-iVa) and two statin residues (P1 and P2' -Sta), as observed in an X-ray structure (PDB 1PSO). The inhibitor is shown as a ball-and-stick model, and the catalytic Asp32 and Asp215 are indicated as red patches in the lower panel. Statin is a β -hydroxy- γ -amino acid with a side chain in the γ -position that corresponds to the one of Leu. (**B**) HIV protease or retropepsin is active as a homodimer with both Asp25 of chain A and B as catalytic residues. Kynostatin (*KNI-272*, ball and sticks) inhibits HIV protease with a K_i of 5.5 pM, but it was not effective as a pharmaceutical compound. The tripeptidic core consists of the three ncAAs methylthioalanine, allophenylnorstatine (Apns) and thioproline (Thp), with two capping groups (iQoa and tBu) (PDB 1HPX). A neutron diffraction study showed the positions of hydrogen atoms (white spheres) and of deuterated water molecules (lower panel, PDB 2ZYE). (**C**) The malarial plasmepsin V in complex with the highly potent inhibitor WEHI-842 (ball and sticks). This inhibitor occupies the S4 to S2' pockets with its peptidic core Can-Val-Sta and the two capping groups carboxybenzyl (Cbz) and phenetylamine (Pea) (PDB 4ZL4).

3.5. Metalloproteases

Matrix metalloproteinases (MMPs, family M10) constitute a large and important group of Zn²⁺-dependent enzymes that remodel and degrade the extracellular matrix (ECM), in particular various collagens, with additional functions in cell signaling and cancer [304]. They are regulated by the polypeptidic TIMP inhibitors (family I35), as structurally corroborated by the MMP3-TIMP1 complex, which is an example for using SeMet in structure solution (Figure 3A) [53]. MMPs are favored targets of numerous academic and commercial projects that aimed at the discovery of potent pharmaceutical compounds, comprising biosynthesized compounds and in silico screening of inhibitory compounds with ncAAs for several MMPs [305-307]. Peptidic compounds such as HS-(CH₂- D-Leu)-Phe-Ala-NH₂ efficiently inhibit MMP-1 and MMP-8 (fibroblast collagenase, M10.001 and neutrophil collagenase, M10.002), reaching the low nanomolar range [308]. An early example of screening with ncAAs using a positional scanning synthetic library (PSSCL) was the search for gelatinase A (MMP-2) and B (MMP-9) tetrapeptide inhibitors, e.g., His- ε -Ahx- β Ala-His, which were efficient in cell cultures, but the in vitro IC_{50} remained in the micromolar range [309]. Among small-molecule inhibitors of metalloproteinases, hydroxamates with variations of the P1' residue inhibited gelatinase A (MMP-2), gelatinase B (MMP-9), stromelysin-I (MMP-3) and neutrophil collagenase (MMP-8) in the picomolar range, whereby in the P1' position the following ncAAs conferred high potency: 3-Bal, 1-Nal, 2-Nal, 8-Qal and Tle (Figure 2) [310]. For inhibitor studies, it is favorable to crystallize complexes with the peptidase domain of MMPs, instead of the full-length multidomain protease, which often comprises propeller domains and additional insertions, such as fibronectin in the case of the gelatinases (Figure 8A). A small reverse hydroxamate inhibitor with a P1'-Leu and P2'-Tle was crystallized with the catalytic domain of gelatinase B/MMP-9 (Figure 8B) [311]. Inhibitor fingerprints of MMPs 2, 3, 7, 8, 9, 13 and 14 were obtained from a library of 1400 peptide hydroxamates by systematic variations with cAAs and ncAAs in the P1', P2' and P3' positions [312]. Using the stop codon suppression method, DOPA (3-hydroxy-Tyr) and HqAla (8-hydroxyquinolin-3-yl-Ala) were incorporated into N-TIMP-2 at positions Y36 and S69, respectively, which increased the specific inhibition around 1 nM towards the catalytic domains of MMP-2 and MMP-9, but not for MMP-14 [313]. Highly specific and efficient substrates with the fluorescent groups MCA (7-Methoxycoumarin-4-acetic acid) and DAP ((E)-1-(4-chlorophenyl)-3-[4-(dimethyl-amino)-phenyl]prop-2-en-1-one) were synthesized for MMP-9, comprising the sequence MCA-Pro-Leu-(4-iodo)Phe-DAP-Ala-Arg-NH₂, which occupies the S4 to S4' subsites (Figure 8C) [314]. Recently, a peptidomimetic inhibitor called TP0556351 with an IC_{50} of 200 pM was generated for gelatinase A/MMP-2 (Figure 8D) [315]. It exhibits a β -homo-Pro in the P5 position, followed by (N-Me)Glu-Leu-Dab-Glu and Asp as a Zn²⁺binding group and a (CH₂)₄-aryl-oxyphenyl as an S1' occupying moiety. Very efficient substrates for MMP-11 (stromelysin-3) and MMP-14 (membrane-type 1 MMP) were generated by placing a S-para-methoxybenzyl-Cys (MeOBn-C) in the P1' position, resulting in nearly 10- and 40-fold higher catalytic efficiency than the P1'-Leu substrates [316]. Several unusual nncAAs were investigated in this study, such as pPhe, an hPhe extended by a CH_2 group, or amino acids with n-pentane or n-hexane side chains. A mutational study of the S1' pocket shaping residue Gln215 in MMP-11 showed that the turnover of fluorogenic heptapeptides was increased at least 25-fold when the P1' residue Leu was replaced by Cys-(OMe)benzyl (MeOBn-C) [317]. In agreement with these approaches, the optimized FRET substrate Ac-GRRRK(Dabcyl)-GGAAN-(MeOBn-C)-RMGG-fluorescein was found to exhibit 25-fold better affinity for MMP-11 compared to previously synthesized substrates [318]. In addition to a similarly high affinity for MMP-14, this FRET substrate easily penetrated cell membranes and showed high fluorescence signals in cancer cell lines. Based on previously solved complex structures of MMP-13 (M10.013) with peptides and TIMP-inhibitors, a molecular modeling approach led to the synthesis of "TIMP peptidomimetics" [319]. The most potent one inhibited MMP-13, with an IC_{50} of 21 nM, and consisted of three uAAs: an N-terminal azido-bearing residue derived from Lys, a large residue that had a F₃C-O-Biphenyl-CH₂-side chain and an ether derived from Ser with a propionyl group, followed by Ser-Ala. The azide and triple bond allowed to cyclize according to the Cu⁺-catalyzed 1,3-dipolar cycloaddition. In a substrate and inhibitor design study, TACE (TNF- α -converting enzyme or ADAM17, M12.217) preferred Phg and hPhe at the P1' position, as well as basic AAs at the P2' position, including Cit and methionine sulfoxide Met(o) [320]. According to these substrate preferences, reverse hydroxamate TACE inhibitors with phenethyl and 5-methyl-thiophene-methyl side chains at P1' or nitro-Arg at P2' were generated. Even the stereochemistry of a single P1' residue has a marked effect on inhibitory potency, as demonstrated for a phosphonamidodithionate with a Glu in the R configuration targeting the Zn²⁺-dependent glutamate carboxypeptidase G from Pseudomonas [321].

Aminopeptidase N or alanine aminopeptidase (APN, M01.001) is a membrane-anchored Zn^{2+} metalloproteinase, which preferentially cleaves P1-Ala and exhibits an unusual specificity for P1'-Trp/Phe [322]. Substrate profiling with kinetic analysis and inhibitor assays showed that the most efficient H₂N-X-ACC substrates and H₂N-X-PO₃H₂ inhibitors had Nle and hCha P1 residues for both porcine and human APN [323]. APN is regulated by the extracellular adapter protein 14-3-3 binding, whereby phosphorylated peptides, such as APN 36–73 with pSer43 and pThr63, could mimic this specific interaction, which facilitated crystallographic structure determination [324]. The bacterial counterpart from *Neisseria meningitis* (NmAPN) is a potential drug target and was thus analyzed for its specificity in the S1 and S1' binding sites,

including uAAs [325]. The best P1 residue was hArg, and the phosphinate inhibitor hPhe-P-(CH₂)Tyr exhibited a K_i of 54 nM. Similarly, oat (Avena sativa) aminopeptidase accepted basic and aromatic residues in P1-ACC substrates, whereby hCha, hTyr, hPhe and hArg were preferred over the best cAA Arg [326]. Bestatin is a natural metalloproteinase inhibitor consisting of an α -hydroxy- β -amino acid with a phenyl ring in P1 and Leu as a P1' residue, whose potency was increased five-fold by a Tyr(OMe) in P1 with a K_i of 43 nM for the PfA-M1 protease (M01.029) of *Plasmodium falciparum* (Figure 9A–C) [327]. A study of the malariarelated amino-peptidases PfM1AAP and PfM17LAP revealed a similar substrate specificity, with the M17 proteases preferring bulky hydrophobic side chains, such as hCha, hPhe and Igl (2-indanyl-glycine) [328]. Methionyl aminopeptidases (MetAP) belong to the M24 family, of which human MetAP1 and MetAP2 as well as *E. coli* MetAP1 were investigated [329]. All three metAPs exhibited an exceptional specificity for Met, which matched hLeu and Nle in the case of both MetAP1s, while corresponding phosphonates inhibited in the low micromolar range. A comparative study on the APN activity of rat, pig and human kidney cell lysates with P1-ACC substrates from HyCoSuL substrates with variations in the P1 position identified hAPN, ERAP1/2 (endoplasmic reticulum aminopeptidases 1 and 2, M01.018/24), IRAP (insulin-regulated aminopeptidase, M01.011), MetAP1 (M24.017) and LTA4H (leukotriene A4 hydrolase, M01.004) [330]. Since ERAP1 and ERAP2 generate a fraction of the MHC-I peptides downstream of the proteasome and contribute to autoimmune diseases, cancer, spondylitis and psoriasis, efficient inhibitors are desired as new lead compounds [331]. Several modified dipeptides, such as hPhe-(PO_2H)-Leu-CONH₂, inhibit ERAP1/2, with an IC₅₀ in the low nM to pM range [332]. One variant appears as γ -amino acid with a phosphinic acid group in the β -position and two side chains, a combination of NO₂-hPhe (S1) and Phe (S1').



Figure 8. Metalloproteinase (MMP) complexes with ncAA compounds shown as ball-and-stick models. (**A**) Full-length MMP-2 (gelatinase A) comprises a peptidase domain (cat) with the catalytic Zn^{2+} (grey sphere), three fibronectin repeats (FN), a β 4-propeller domain and structural Ca²⁺ ions (green spheres), resembling MMP-9 (PDB 1CK7). Like other MMPs, it contains the active site motif HEXXHXXXXH. (**B**) The catalytic domain of MMP-9 (gelatinase B) with a reverse hydroxamate inhibitor. Leu and *tert*-leucine (Tle) occupy the S1' and S2' subsites, respectively (PDB 1GKC). (**C**) MMP-9-cat bound to the fluorogenic substrate *Probe 8* with the sequence MCA-Pro-Leu-Gly-4-iodo-Phe-Dab(dnp)-Ala-Arg-NH₂. Dab is diaminobutyric acid, N-substituted with dnp being 2,4-dinitrophenyl (PDB 4JIJ). (**D**) MMP-2-Cat displayed in complex with the reversed backbone inhibitor *TP0556351*, which binds the S5 to S1' subsites with the sequence β -hPro-nme-Glu-Asp-Dab-Glu-Asp-Ape(5)-NH₂. The β -homo-Pro carries a carboxyamide, nme-Glu is N-methylated, and the aryloxyphenyl group Ape(5) is connected to the Zn²⁺-binding Asp by a (CH₂)₄-linker (PDB 7XJO).

A series of α -mercapto acyl-dipeptide inhibitors targeted angiotensin-converting enzyme (ACE, preliminarily designated XM02.001 with two catalytic domains M02.001 and M02.004), which is overall rather unspecific, and neprilyisin (NEP, M13.001) [333]. The most potent compounds reached IC₅₀ values of 7 nM (ACE) and 1.2 nM, respectively, and were

tested in rats. Systematic variation of the organic residues around the core dipeptide CycL-Tyr-OR containing cyclo-leucine yielded a more potent compound than captopril [334]. Compounds with the proline-derived uAAs Phi and Abo, as present in perindopril and perindoprilate, were investigated as inhibitors of ACE [335]. Efficient inhibitors of ACE and dipeptidyl peptidase 3 (DPP3, M49.001) were generated by replacing the individual P1 residues of the cleavage site in the heptapeptide angiotensin with the cyclic ncAA *cis*-3-aminomethylcyclobutane carboxylic acid (ACCA), whereby the cyclobutane served as a bridging moiety in the backbone [336]. These protease-stable angiotensin variants had a significant effect on cancer cell cultures in reducing their proliferation. More than 70 years ago, studies on antimicrobial peptides demonstrated that Cbz-Gly-Xaa-COOH inhibited mammalian pancreatic carboxypeptidase (M14.001) as well, whereby Xaa can be Thi, 1/2-Nal or *p*-tolyl-*DL*-alanine[337,338].

Eukaryotes such as the fungi *Saccharomyces cerevisia* and *Neurospora crassa* possess a Zn^{2+} -dependent mitochondrial processing peptidase (MPP, M16.003), whose activity was studied with synthetic substrates containing P2-Arg and ncAAs in this position [339]. In addition, aminoethoxy acetic acid insertion repeats at the non-prime side enhanced the substrate turnover; however, Cit, N ω -nitro-Arg and others were disfavored.

A metalloprotease that should not be confused with PSA (KLK3) is the prostatespecific membrane antigen (PSMA), the multidomain glutamate carboxypeptidase II (GCPII, M28.010). It is anchored in the membrane and exhibits an unusual specificity for P1-Glu↓P1'-Glu scissile bonds. In order to develop a new inhibitor-directed enzyme prodrug therapy (IDEPT), a multifunctional agent was prepared using the strain-promoted alkyne-azide cycloaddition (SPAAC) of azidophenylalanine (AzF) incorporated into the suicide enzyme yCDtriple and the dibenzylcyclooctyne moiety of the PSMA-targeting agent DBCO-PEG₄-AH₂-TG97 [340]. This click reaction product efficiently killed PSMA-positive PCA by converting the prodrug 5-fluorocytosine into the cytotoxic 5-fluorouracil.



Figure 9. Malarial aminopeptidase PfA-M1 in complex with a bestatin-derived inhibitor. (**A**) PfA-M1 comprises four domains, whereby the inhibitor (yellow sphere model) is located in the center of the catalytic domain cat(II). (**B**) The catalytic Zn²⁺ (grey sphere) is coordinated by His496, His500 and Glu519. Glu497 activates the catalytic water molecule and belongs to the metalloprotease motif HEXXH, which is also present in MMPs, whereas the other side chains bind substrates. (**C**) The inhibitor consists of a modified statin (Sta) with a para-methoxy-Phe side chain instead of a Leu. Its amino group is bound by Glu319 and Glu463 as well as by Tyr580. Though the Sta derivative binds the S1 pocket, Leu is preferred in the S1' subsite.

Botulinum toxin or bontoxilysin from *Clostridium botulinum* is one of the most toxic biomolecules known and comprises a peptidase domain (endopep, bontoxilysin residues 118-450, M27.002) [341]. Derived from the endopep BoNT/C neuronal transmitter substrate SNAP-25, a single exchange of P1-Lys to Orn resulted in a 200-fold better substrate [342]. Correspondingly, incorporation of homoserine (hSer) and ornithine (Orn) instead of Ser and Lys at distinct positions of a substrate increased the sensitivity of a bontoxilysin mass spectrometry assay for clinical samples about 170-fold [343].

4. Modified Proteases

4.1. Incorporation of ncAAs into Proteases for Structure-Functional Studies

Since it is much more sophisticated to incorporate ncAAs into large polypeptides and proteins than in small molecules, the preparation of corresponding altered proteases is still in its infancy. Employing subtilisin from Bacillus lentus as model protein, regioisomers such as iso-His were generated via a S156C mutant, which did not significantly alter the catalytic properties [344]. By contrast, the site-specific incorporation of 3NY and 3-chlorotyrosine (3CY) instead of Phe205 near the catalytic center of the subtilisin-like protease KP-43 (S08,123) increased the activity at pH 7 to some extent [345]. Due to its unique recognition sequence ENLYFQ \downarrow G/S, tobacco etch virus protease (TEV, C04.004) is one of most frequently used enzymes in tailoring recombinant proteins. TEV protease was modified with caged tyrosine analogues utilizing a bioorthogonal Pyl-tRNA synthetase/tRNA pair, which facilitated optical control of proteolytic activity [346]. The caged uAA nitropiperonyltyrosine (NPY) replaced Tyr178 in a mutant that abolished the activity upon reconstitution by UV light treatment. Based on an efficient in vitro translation system employing UAG/amber codon suppression, the structural role of Asp-25 and Asp-125 in HIV protease was investigated by introducing *threo*- β -methylaspartate and β , β -dimethylaspartate in a site-specific manner, resulting in a specific activity of 0 to 45%, which was explained by perturbation in the S1 and S1' pockets [347,348]. Similarly, 7-(hydroxy-coumarin-4-yl)- ethylglycine (Hco) was incorporated in the NS2B-NSB3 protease of West Nile virus by cell-free protein synthesis via *amber* codon suppression. This unnatural variant of the protease allowed for sensitive measurements of fluorescence upon the binding of inhibitors and small ligand molecules (Figure 2) [349]. In addition, the alkaline protease from *Pseudomonas aeruginosa* (aeruginolysin, AprA, M10.056), a metzincin of the metalloendoproteases, was also modified by replacing the highly conserved Met214 with the uAA difluoromethionine (DFM), which had little effect on the structure and function of the protease [350].

The uAA Nε-4-azidobenzyloxycarbonyllysine (PABK) was used for bioorthogonal ligation by CUAAC. PABK is both a photo-affinity reagent and a chemical decaging compound via the strain-promoted alkyne-azide cycloaddition (SPAAC), enabling intracellular protein activation (Figure 2) [351]. This approach was applied to a model organism, namely, zebrafish embryos, expressing the exogenous blotched snakehead virus protease VP4, possessing a catalytic Ser/Lys dyad in which the lysine was replaced by PABK [352]. Upon decaging by 2-(diphenylphosphanyl)-benzamide, the activated protease cleaved a cytosolic substrate in developing zebrafish, resulting in a clear signal in the nucleus.

In a ground-breaking approach, genetically encoded 2,3-diaminopropionic acid (Dap) was incorporated into the model proteases TEV (C04.004), HtrA-2 (S0.278) and intramembrane RHBDL3 (formerly RHBDL4, rhomboid-like 3 protease, S54.006), replacing the catalytic Cys or Ser (Figure 2) [353]. After photoactivation, the proteases captured both known and novel substrates in complex mixtures in vitro and in mammalian cells by forming stable amide bonds instead of the otherwise transient acyl intermediates, permitting their purification and identification.

4.2. Cross-Linking with ncAAs

Site-specific cross-linking is a highly desired approach to investigating protein interactions in structural proteomics and in living cells [354,355]. For more in-depth details on this topic, see the highly topical review of Aydin and Conin, since our focus is on examples of proteases [356]. Thus, keratinase KerBL (subtilisin Carlsberg, S08.001) from *Bacillus licheniformis* was internally cross-linked by the reaction of the uAAs bromopropyl/butyl-Tyr with Cys mutants, which stabilized flexible regions [357]. Besides having a melting temperature that was more than 10 °C higher for the variants N159C/Y260BprY and N159C/Y260BbtY, their keratinolytic activity nearly doubled, and the catalytic efficiency for the substrate suc-AAPF-pNA did not change significantly. A recent study used ncAA mutagenesis in the heavy chain complementarity-determining region 3 (CDR-H3) of antibody fragments directed against human rhinovirus 14 (HRV14) 3C protease (family C30) [358]. Covalent links of Lys, His or Cys residues in the 3C protease were formed mainly with Tyr106 Fab mutants, having tyrosine analogues or aliphatic chains with bromine or iodine.

The photo-crosslinker DiZPK (3-(3-methyl-3H-diazirine-3-yl)-propamino-carbonyl-N ϵ -Lys), an analog of pyrrolysine (Pyl), was incorporated into the *E. coli* protease DegP (HtrA, S01.273) using the *amber* stop codon suppression method, using a specific DiZPK-tRNA and a DiZPK-tRNA synthetase (Figure 2) [359]. This approach enabled the cross-linking of periplasmatic interaction partners with DegP, a membrane-anchored elastase-like protease that is crucial for *E. coli* to survive acidic stress [360]. Moreover, the oligomerization state of DegP was analyzed by photo-crosslinking based on the uAAs DiZPK- or Bpa, confirming the formation of dimers and trimers in living cells [361]. Similarly, photo-activatable AzF was incorporated at positions 463 and 466 in the transmembrane segment of BACE1, which trimerized in human cell cultures upon UV irradiation, showing the Cu²⁺-sensing ability of this APP cleaving secretase together with fluorescent labels [362].

4.3. Click Reactions with Proteases and Their Inhibitors

The site-specific incorporation of ncAAs into proteins for click reactions with the classical Cu(I) catalyzed click reaction of alkynes and azides (CuAAC) is well established, e.g., for labeling recombinant proteins with biotin or fluorophores (Figure 10A) [363]. Other applications concern additional covalent linkers in polypeptides that increase the stability against proteolytic degradation, as demonstrated for triazole-stapled BCL9 α -helical peptides, which were generated by the classical Cu⁺-catalyzed click reaction of alkynes and an azide (CuAAC) [364]. A specific example of enhanced resistance to the ubiquitin (Ub)-specific protease 1 (USP1, C19.019) was observed for the poly-Ub chains that were generated by the CuAAC reaction of engineered human Ub with Gly76Aha (azidohomoalanine) and propargyl-derivatized Lys27/29 (Plk) [365].

Using double mutants of the Zika virus NS2B-NS3 protease, such as N48AzF/R169AzF, the Gd(III) ion chelator propargyl-DO3A was covalently linked with AzF side chains by the 1,3 dipolar cycloaddition [366,367]. Similar covalent linking via disulfide formation enabled individually labeled constructs to measure double electron–electron resonance (DEER/EPR), which eventually confirmed the closed conformation of the ligand-free and inhibited protease. Furthermore, the cysteine protease inhibitor cystatin B and other proteins in synaptosomes from rat cerebral cortex could incorporate Hpg within two hours, enabling a click reaction with biotin-azide and the subsequent high affinity binding to streptavidin-linked dynabeads for further analysis [368].

As a proof of the principle, it was demonstrated that AzF could be incorporated at basically any position of the yeast multisubunit 26S proteasome, by using the stop codon suppression method and facilitating further click reactions (Figure 2) [369]. Accordingly, reconstituted 26S proteasomes were covalently linked at these distinct AzF positions to dibenzocyclooctyne-linked fluorophores, allowing for the performance of FRET- and anisotropy-based assays for substrate interactions and concomitant conformational changes of the proteasome [370]. In addition, the proteasome-Pup (a ubiquitin equivalent) system of *Actinobacteria* was studied with pupylated proteins conjugated to synthetic peptides and uAAs employing the classical Cu(I)-catalyzed 1,3-dipolar cycloaddition of an alkyne group in Pup and an azide group in the peptide, resulting in the so-called Pup-click [371].

Recently, it was shown that the site-specific incorporation of numerous ncAAs into model proteins can be expanded to a library of aliphatic and particularly small ncAAs with short side chains [89]. Starting with the tRNA^{Pyl} containing the anti-*amber* stop codon and specific PyIRS from various archaean organisms, even the relatively short ncAA S-allylcysteine (Sac) has been incorporated, enabling photoclick reactions with Cys-SH groups and many related ncAAs, as well as azido and alkyne group containing ncAAs, such as AzF (Figures 2 and 10B) [372]. Further progress was made by engineering psychrophilic PyIRS orthologs from cold dwelling archaea to develop "cold" orthogonal translation systems (OTS) as an alternative to the commonly used meso- and thermophilic OTSs. Apparently, the psychrophilic homolog of PyIRS of *Methanococcoides burtonii* has high catalytic efficiency and promiscuity, which allows to incorporate many different ncAAs, with higher yields than previously reported [373].



Figure 10. Click reactions that take place under mild conditions in water or even in biological systems. (**A**) In the Cu(I)-catalyzed 1,3-dipolar azide-alkyne cycloaddition, methionine analogs, such as azidohomoalanine (Aha) and homopropargylglycine (Hpg), form stable triazoles by Cu⁺ catalysis in aqueous solution; R represents arbitrary polypeptides. (**B**) The thiol–ene coupling is a photoreaction of a thiol from Cys and an olefin, e.g., S-allylcysteine (Sac), leading to the formation of double thioether links, which are more stable than a disulfide bridge.

5. Conclusions and Outlook

Regarding the early studies with ncAA containing substrates and inhibitors of the prototypic serine proteases trypsin and chymotrypsin, their focus was on analyses of the mechanisms (see Section 3.1.1). This focus shifted to pharmaceutical applications to some extent, when blood coagulation factors like thrombin or the cancer-related KLKs CatA and CatG were investigated (Sections 3.1.2 and 3.1.3). In addition, more natural ncAAs were discovered in many cyclic natural small-molecule inhibitors of trypsin-like proteases. Medical and pharmaceutical applications were dominating for the infectious disease-promoting flaviviral serine proteases from Zika, Dengue and West Nile viruses (Section 3.1.4). Other serine proteases were not studied in detail, while the only relevant threonine proteases in the proteasome were intensively investigated with ncAA substrates, inhibitors and ABPs, due to their role in the immune system and in cancer (Section 3.2).

Among the cysteine proteases, the 11 cathepsins are involved in Alzheimer's disease and various types of cancer, whereas the caspases are most important for apoptotic processes (Section 3.3) They are often tested with the same HyCoSul libraries as their serine protease counterparts, based on similar specificity and mechanisms as in serine proteases. Nowadays, the top target is the SARS-CoV-2 MPro, as its inhibition seems most relevant in times of the Corona pandemic. After ncAA-based inhibitor studies of the prototypic pepsin, especially comparable inhibitors of the HIV protease were very successful with respect to clinical trials and pharmaceutical drug production (Section 3.4). It should be mentioned that ncAA-based HIV inhibitors are often highly effective against hepatitis C protease and the HTLV protease, whereas compounds directed against the malaria-related plasmepsins have not reached their full potential. The human MMPs are major targets of small inhibitors with ncAAs, due to their roles in cancer, also the ADAM, ACE and TACE proteases are still interesting targets (Section 3.5).

Currently, modified proteases are quite heterogeneous, but caging and decaging ncAAs are often incorporated into them, which facilitates activation by photo-reactions (Section 4.1). Again, mainly photo-reactive residues are employed for site-specific crosslinking of various proteases with diverse interacting proteins (Section 4.2). Based on the stop codon suppression method, AzF is the preferred residue for the CuAAC click reaction with an alkyne as counterpart and for photo-reactions (Section 4.3). Interestingly, targets from the ubquitin-proteasome are at the forefront of recent research. However, new reaction types, such as the photo-catalyzed thiol–ene reaction, have become available, which offers the advantage that Cys residues can be the reaction partners.

With the development of new methods in the life sciences, new applications for protease research with ncAAs can be found in the literature. For example, in surfaceenhanced Raman spectroscopy (SERS) experiments, trypsin activity could be monitored by an increasing signal intensity upon cleavage and solubilization of a Phe and 4-cyano-Phe (CN-F)-containing peptide, which was bound as a monolayer to a gold solubilization surface [374]. Site-specific incorporation of the ncAA acridonylalanine (Acd) in the LexA repressor protease (S24.001) allowed to monitor the interactions in the LexA:RecA complex [375]. Increasingly, ncAAs facilitate the synthesis of protease-activated prodrugs for pharmaceutical applications, for which Poreba provides a comprehensive review, including information on several clinical trials [376].

Future applications of unnatural amino acids in the life sciences seem limitless, especially in the field of proteases. The increased resistance of biomolecules with uAAs to proteolysis gives them an advantage over conventional peptides and proteins. For example, so called α -PNAs, which contain nucleobase amino acid residues with thymine or cytosine attached to the side chain, are very stable in human serum and offer new pharmacological antisense strategies [377]. Information from peptide-based drug discovery studies of caspases, MMPs and granzymes combining natural and unnatural amino acid substrates enables the prediction of protease-specific cleavage sites, as by the PROSPERous web server (https://prosperous.erc.monash.edu/, accessed on 19 August 2023), facilitating and accelerating future protease studies [378–380]. A relatively new and expanding area of research is the uAA-dependent optical control of nucleic acid processing, cell signaling, protein–protein interactions and enzymatic activity [381]. Very promising areas for applications of uAAs are medical and pharmaceutical research, especially the domain of vaccine design [382–384].

However, the nearly thirty year old proposal of Sidney Benner that the genetic code should be expanded by adding new, unnatural codons has not been consequently investigated, or it has at least not reached the level of successful protein expression, but corresponding approaches will certainly contribute to upcoming explorations [385]. Recent attempts towards a new genetic code with novel bases and 256 quadruplets might open the door for a completely artificial biology including synthetic living cells [109,386]. Nevertheless, the realization of a hypothetical organism operating with such an expanded genetic code faces practical challenges for two primary reasons. First, if the number of encoded amino acids in such a code is too small (e.g., 20 or 30), the code will become so degenerate that mutations (as the driving force of evolution) will become impossible. On the other hand, an excessively large repertoire, for instance around 100 amino acids, could impose such substantial metabolic burdens that the viability of such an "alien organism" is brought into question. Therefore, the use of quadruplets will most likely be limited to the elucidation of academic issues. Nature has probably chosen a code with triplets to create a balance between these two extreme scenarios [387].

However, within the context of triplet codes, significant potential exists for integrating new amino acids into the genetic code. In 2007, as demonstrated by Szostak and collaborators, in vitro ribosome-mediated translation showcased the insertion of 50 diverse non-canonical amino acids into peptide sequences through the native translation apparatus [388]. This strategy enables approximately 70% of the codons to be reassigned, capitalizing on the substrate tolerance of aminoacyl-tRNA synthetases, a feature that underscores the adaptive promiscuity of these enzymes [389,390]. Dieter Soll and coworkers estimated that 30 to 40 sense codons are sufficient to encode the genetic information of an organism, and a large number of sense codons (>20) may be available for recoding with non-canonical amino acids [391].

The expansion of the genetic code with synthetic nucleotides, nucleobases and nucleic acid backbones is another advanced area of research within synthetic biology that aims to expand the existing genetic code by incorporating new types of nucleic acids into DNA and RNA molecules. This expansion goes beyond the four naturally occurring nucleotides (adenine, cytosine, guanine and thymine/uracil) and involves the addition of chemically modified or fully synthetic nucleotides to the genetic code [392].

The future trajectory of this field lies in the integration of orthogonal translation with synthetic metabolism [393]. This approach seeks to minimize the costly external addition of non-canonical amino acids (or nucleobases) by engineering a synthetic metabolism that is compatible with these unique unnatural substrates. Technically, reprogrammed and orthogonalized translation is mainly encoded by translation components on plasmids. The next step is to use genome-editing methods to genomically integrate all components of reprogrammed translation with noncanonical amino acids. The use of these ncAAs or UAAs can be further "fortified" by getting cells "addicted" to them through adaptive laboratory evolution, as recently shown for fluorinated tryptophans and thienopyrrole [394–396]. This comprehensive strategy envisions the emergence of synthetic cells or "orthogonal life," providing promising platforms for novel protease activities and corresponding inhibitors. Similarly, synthetic cells may promote the generation of ncAA containing biomolecules that confer distinct new functions as molecular switches, catalysts, nanoprobes, metabolic regulators or carriers of activatable drugs to distinct tissues, which all seems immensely attractive. Ultimately, this holistic approach has the potential to usher in a new era of synthetic biology characterized by engineered life forms and functional biomolecules [397].

Author Contributions: Conceptualization, P.G. and N.B.; writing—original draft preparation, P.G.; writing—review and editing, N.B. and N.G.K.; visualization, P.G. and N.G.K.; project administration, P.G.; funding acquisition, P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Austrian Science Fund FWF as the lead agency, grant number I 3877-B21, and the Deutsche Forschungsgemeinschaft (DFG) in the D-A-CH program. This research was supported in part by the Canada Research Chairs Program (N.B., Grant No. 950-231971) and the NSERC Discovery Grant (N.B., RGPIN-05669-2020).

Acknowledgments: We are grateful to the Austrian Science Fund FWF as the lead agency and the participating Deutsche Forschungsgemeinschaft for supporting the grant I 3877-B21 of the D-A-CH program (P.G. and N.G.K.). Moreover, we are thankful for advice by Hans Brandstetter (University of Salzburg) and discussions with Jon Harris (Queensland University of Technology, Brisbane). Also, we would like to thank Daniela Schuster (Paracelsus Medizinische Universität Salzburg) for her kind support regarding the continuation of the abovementioned FWF project.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. IUPAC names and CAS or Nikkaji numbers (*) of natural non-canonical amino acids.

AA	IUPAC Name	CAS Number
Sec	(2 <i>R</i>)-2-amino-3-selanylpropanoic acid	10236-58-5
Pyl	N°-[[(2 <i>K</i> ,3 <i>K</i>)-3-methyl-3,4-dihydro-2 <i>H</i> -pyrrol-2-yl]carbonyl]-L-lysine	448235-52-7
Hvn	(2 <i>S</i> ,3 <i>S</i>)-3-hydroxypyrrolidine-2-carboxylic acid	4298-08-2
nyp	(2 <i>S</i> , <i>4R</i>)-4-hydroxypyrrolidine-2-carboxylic acid	51-35-4
pSer	(2S)-2-amino-3-phosphonooxypropanoic acid	407-41-0
pInr	(25,3R)-2-amino-3-phosphonooxybutanoic acid	27530-80-9
Orn	(2S)-2,5-diaminopentanoic acid	70-26-8
Cit	(2S)-2-amino-5-(carbamoylamino)pentanoic acid	372-75-8
ArgSA	(2S)-2-[[N'-[(4S)-4-amino-4-carboxybutyl]carbamimidoyl]amino]butanedioic acid	2387-71-5
SAM	(2 <i>S</i>)-2-amino-4-[[(2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i> ,5 <i>R</i>)-5-(6-aminopurin-9-yl)-3,4-dihydroxyoxolan-2-	29908-03-0
	yl]methyl-methylsulfonio]butanoate	
SAH	(25)-2-amino-4-[[(25,35,4K,5K)-5-(6-aminopurin-9-yi)-3,4-dihydroxyoxolan-2-	979-92-0
26 1	(2P) 2 amino 2 hydroxy 2 overrenone 1 cultinate	1115 (5 7
55A bCus/Hay	(2K)-2-amino-5-nydroxy-5-oxopropane-1-summate	1115-65-7
Tou	(25)-2-amino-4-sunanyibutanoic aciu	107 25 7
lau Sar	2 (mothylamino)acotic acid	107-33-7
	2-aminobenzoic acid	107-77-1
Kyn	4-oxo-1H-quipolipe-2-carboxylic acid	492-27-3
Aad	(2S)-2-aminohevanedioic acid	1118-90-7
hSer	(25)-2-amino-4-hydroxybutanoic acid	672-15-1
ß-Ala	3-aminopropanoic acid	107-95-9
Dear	(2P) 2 amino 2 hydrogramonomic soid	212.94 5
D-Ser	(2R) 2 aminohutanodiois asid	512-04-5 1783 96 6
CABA	A-aminobutaneticic acid	56-12-2
DOPA	(2S)-2-amino-3-(3 4-dihydroxyphenyl)propanoic acid	59-92-7
		33.727
Abu/hAla	(2S)-2-aminobutanoic acid	1492-24-6
ADMA	(25)-5-(diaminomethylideneamino)-2-(dimethylamino)pentanoic acid	30313-93-6
Aeg	2-amino 2 methylaronanoia acid	24123-14-0 62 57 7
Alo	(2S) 2 aminopont 4 opois acid	16338 48 0
Arg	(2S)-2-antihopent-4-enoic acid	2133_34_8
Can	(2S)-2-2mino-4-(diaminomethylideneamino)ovyhutanoic acid	5/3-38-/
Dab	(25)-2 4-diaminohutanoic acid	1758-80-1
Dha	2-aminoprop-2-enoic acid	1948-56-7
hArg	(2S)-2-amino-6-(diaminomethylideneamino)hexanoic acid	156-86-5
Met(o)	(2S)-2-amino-4-methylsulfinylbutanoic acid	3226-65-1
Nle/α -Ahx	(2S)-2-aminohexanoic acid	327-57-1
Nva/Ape	(2S)-2-aminopentanoic acid	6600-40-4
Pip	(2S)-piperidine-2-carboxylic acid	3105-95-1
Sem/SeMet	(2S)-2-amino-4-methylselanylbutanoic acid	3211-76-5
Tem/TeMet	(2(2S)-2-amino-4-methyltellanylbutanoic acid	J663.934H *

36 of 52

Appendix B

Table A2. IUPAC names and CAS or Nikkaji numbers (*) of unnatural non-canonical amino acids.

AA	IUPAC Name	CAS Number
Aca	(2S)-2-amino-3-(4-aminocyclohexyl)propanoic acid	J1.610.797B *
ACCA	cis-(15,35)-3-(aminomethyl)cyclobutane-1-carboxylic acid	1400744-20-8 (HCl)
ε-Ahx	6-aminohexaonic acid	60-32-2
Ama	(2S)-2-amino-3-(4-aminomethylcyclohexyl)propanoic acid	-
Amf	(2S)-2-amino-3-[4-(aminomethyl)phenyl]propanoic acid	1991-96-4
Anb	5-amino-2-nitrobenzoic acid	13280-60-9
Cha	(2S)-2-amino-3-cyclohexylpropanoic acid	27527-05-5
Chg	(2S)-2-amino-2-cyclohexylacetic acid	14328-51-9
Сра	(2S)-2-amino-3-(2-cyanopyridin-4-yl)propanoic acid	169949-53-5
CycL	1-aminocyclopentane-1-carboxylic acid	52-52-8
DfeGly	(2S)-2-amino-4,4-difluoro butanoic acid	J786.797B *
DFM	(2S)-2-amino-4-(difluoromethylsulfanyl)butanoic acid	J1.943.889I *
DfpGly	(2S)-2-amino-4,4-difluoro pentanoic acid	-
Dnp-Lys	(2S)-2-amino-6-(2,4-dinitroanilino)hexanoic acid	14401-10-6
4F-Phe	(2S)-2-amino-3-(4-fluorophenyl)propanoic acid (¹⁹ F)	51-65-0
5F-Trp	(2S)-2-amino-3-(5-fluoro-1 <i>H</i> -indol-3-yl)propanoic acid (¹⁹ F)	16626-02-1
F(gua)	(2S)-2-amino-3-[4-(diaminomethylideneamino)phenyl]propanoic acid	2776-36-5
hCha	(2S)-2-amino-4-cyclohexylbutanoic acid	116622-38-9
Igl	(2S)-2-amino-2-(2,3-dihydro-1 <i>H</i> -inden-2-yl)acetic acid	155239-51-3
MfeGlv	(2S)-2-amino-4-fluorobutanoic acid	I1.434.018A *
MLeu	methyl (2S)-2-amino-2.4-dimethylpentanoic acid	90104-02-2
1-Nal	(2S)-2-amino-3-naphthalen-1-vlpropanoic acid	55516-54-6
2-Nal	(2S)-2-amino-3-naphthalen-2-vlpropanoic acid	6960-34-5
Oic	(2S.3aS.7aR)-2.3.3a.4.5.6.7.7a-octahydro-1H-indole-2-carboxylic acid	I532.164F *
Phg	(2S)-2-amino-2-phenylacetic acid	2935-35-5
Phi	(2S.3aS.7aS)-2.3.3a,4.5.6.7.7a-octahydro-1 <i>H</i> -indole-2-carboxylic acid	80875-98-5
0	(2S)-2-amino-3-(4H-selenopheno[3,2-b]pyrrol-6-yl)propanoic acid	
Sep	(selenolo[3,2-b]pyrrole)	J1.076.917E *
TfeGlv	(2S)-2-amino-4.4.4-trifluorobutanoic acid	15960-05-1
Thfg	(2S)-2-amino-2-(oxolan-3-vl)acetic acid	I1.827.275I *
	(2S)-2-amino-3-thiophen-2-vlpropanoic	
Thi	acid/(2S)-2-amino-3-thiophen-3-vlpropanoic acid	22574-47-6/3685-51-6
Thp	(4R)-1.3-thiazolidine-4-carboxylic acid	34592-47-7
Tic	1.2.3.4-tetrahydroisoguinoline-3-carboxylic acid	67123-97-1
Tle	(2S)-2-amino-3.3-dimethylbutanoic acid	20859-02-3
		1(2420 20.0
CN-Phe	(25)-2-amino-3-(4-cyanophenyi)propanoic acid	167479-78-9
3CY	(25)-2-amino-3-(3-chloro-4-hydroxyphenyl)propanoic acid	7423-93-0
3IN Y	(25)-2-amino-3-(4-hydroxy-3-nitrophenyl)propanoic acid	621-44-3
Вра	(25)-2-amino-3-(4-benzoyIphenyI)propanoic acid	104504-45-2
Dap	(25)-2,3-diaminopropanoic acid	4033-39-0
DIZPK	(25)-2-amino-6-[5-(5-methylalazirin-5-yi)propylearbamoylaminojnexanoic acid	1337883-32-3
HCO	(2K)-2-amino-4- $(/-nydroxy-2-oxochromen-4-yi)$ butanoic acid	905442-42-4
	(25)-2-amino-3- $(4-(1-(6-n)tropenzo]u)]$	207727-80-4
radk	(25)-2-ammo-o-(4-azido-denzyioxycardonyiamino)nexanoic acid	2004713-47-3
Aha	(2S)-2-amino-4-azidobutanoic acid	120042-14-0
AzF	(2S)-2-amino-3-(4-azidophenyl)propanoic acid	33173-53-4
Hpg	(2S)-2-aminohex-5-ynoic acid	98841-36-2
Plk	(2S)-2-amino-6-{[(prop-2-yn-1-yloxy)carbonyl]amino}hexanoic acid	1215204-46-8
Sac	(2R)-2-amino-3-prop-2-enylsulfanylpropanoic acid	21593-77-1

References

- Rother, M.; Quitzke, V. Selenoprotein synthesis and regulation in Archaea. *Biochim. Biophys. Acta Gen. Subj.* 2018, 1862, 2451–2462. [CrossRef]
- Meng, K.; Chung, C.Z.; Söll, D.; Krahn, N. Unconventional genetic code systems in archaea. *Front. Microbiol.* 2022, 13, 1007832. [CrossRef]
- 3. Zhang, Y.; Gladyshev, V.N. High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Res.* **2007**, *35*, 4952–4963. [CrossRef]
- 4. Böck, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: The 21st amino acid. *Mol. Microbiol.* **1991**, *5*, 515–520. [CrossRef]
- 5. Gong, P.; Lei, P.; Wang, S.; Zeng, A.; Lou, H. Post-Translational Modifications Aid Archaeal Survival. *Biomolecules* **2020**, *10*, 584. [CrossRef]
- Macek, B.; Forchhammer, K.; Hardouin, J.; Weber-Ban, E.; Grangeasse, C.; Mijakovic, I. Protein post-translational modifications in bacteria. *Nat. Rev. Microbiol.* 2019, 17, 651–664. [CrossRef]
- Prus, G.; Hoegl, A.; Weinert, B.T.; Choudhary, C. Analysis and Interpretation of Protein Post-Translational Modification Site Stoichiometry. *Trends Biochem. Sci.* 2019, 44, 943–960. [CrossRef]
- 8. Kvenvolden, K.; Lawless, J.; Pering, K.; Peterson, E.; Flores, J.; Ponnamperuma, C.; Kaplan, I.R.; Moore, C. Evidence for Extraterrestrial Amino-acids and Hydrocarbons in the Murchison Meteorite. *Nature* **1970**, *228*, 923–926. [CrossRef]
- 9. Elsila, J.E.; Aponte, J.C.; Blackmond, D.G.; Burton, A.S.; Dworkin, J.P.; Glavin, D.P. Meteoritic Amino Acids: Diversity in Compositions Reflects Parent Body Histories. *ACS Cent. Sci.* 2016, *2*, 370–379. [CrossRef]
- 10. Koga, T.; Naraoka, H. A new family of extraterrestrial amino acids in the Murchison meteorite. Sci. Rep. 2017, 7, 636. [CrossRef]
- 11. Zou, H.; Li, L.; Zhang, T.; Shi, M.; Zhang, N.; Huang, J.; Xian, M. Biosynthesis and biotechnological application of non-canonical amino acids: Complex and unclear. *Biotechnol. Adv.* 2018, *36*, 1917–1927. [CrossRef]
- 12. Bastings, J.J.A.J.; van Eijk, H.M.; Olde Damink, S.W.; Rensen, S.S. d-amino Acids in Health and Disease: A Focus on Cancer. *Nutrients* **2019**, *11*, 2205. [CrossRef]
- 13. Genchi, G. An overview on d-amino acids. Amino Acids 2017, 49, 1521–1533. [CrossRef]
- 14. Fujii, N.; Takata, T.; Fujii, N.; Aki, K.; Sakaue, H. D-Amino acids in protein: The mirror of life as a molecular index of aging. *Biochim. Biophys. Acta Proteom.* **2018**, *1866*, 840–847. [CrossRef]
- Kudo, F.; Miyanaga, A.; Eguchi, T. Biosynthesis of natural products containing β-amino acids. *Nat. Prod. Rep.* 2014, *31*, 1056–1073. [CrossRef]
- 16. von Nussbaum, F.; Spiteller, P. β-Amino Acids in Nature. In *Highlights in Bioorganic Chemistry: Methods and Applications;* Schmuck, C., Wennemers, H., Eds.; Wiley-VCH Verlag: Weinheim, Germany, 2004; Volume 1, pp. 63–89.
- 17. Hedges, J.B.; Ryan, K.S. Biosynthetic Pathways to Nonproteinogenic α-Amino Acids. Chem. Rev. 2020, 120, 3161–3209. [CrossRef]
- 18. Papageorgiou, N.; Androulakis, E.; Papaioannou, S.; Antoniades, C.; Tousoulis, D. Homoarginine in the shadow of asymmetric dimethylarginine: From nitric oxide to cardiovascular disease. *Amino Acids* **2015**, *47*, 1741–1750. [CrossRef]
- 19. Hrncic, D.; Rasic-Markovic, A.; Macut, D.; Mladenovic, D.; Susic, V.; Djuric, D.; Stanojlovic, O. Sulfur–Containing Amino Acids in Seizures: Current State of the Art. *Curr. Med. Chem.* **2018**, *25*, 378–390. [CrossRef]
- 20. Stipanuk, M.H. Metabolism of Sulfur-Containing Amino Acids: How the Body Copes with Excess Methionine, Cysteine, and Sulfide. J. Nutr. 2020, 150, 2494S–2505S. [CrossRef]
- 21. Manolidi, K.; Triantis, T.M.; Kaloudis, T.; Hiskia, A. Neurotoxin BMAA and its isomeric amino acids in cyanobacteria and cyanobacteria-based food supplements. *J. Hazard. Mater.* **2019**, *365*, 346–365. [CrossRef]
- 22. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef] [PubMed]
- 23. Rawlings, N.D.; Salvesen, G. *Handbook of Proteolytic Enzymes*, 3rd ed.; Academic Press: London, UK; Oxford, UK; Boston, MA, USA; New York, NY, USA; San Diego, CA, USA, 2013; pp. 1–4094.
- 24. Elsässer, B.; Goettig, P. Mechanisms of Proteolytic Enzymes and Their Inhibition in QM/MM Studies. *Int. J. Mol. Sci.* 2021, 22, 3232. [CrossRef] [PubMed]
- Kubyshkin, V.; Davis, R.; Budisa, N. Biochemistry of fluoroprolines: The prospect of making fluorine a bioelement. *Beilstein J. Org. Chem.* 2021, 17, 439–460. [CrossRef] [PubMed]
- Li, Y.-M.; Yang, M.-Y.; Huang, Y.-C.; Li, Y.-T.; Chen, P.R.; Liu, L. Ligation of Expressed Protein α-Hydrazides via Genetic Incorporation of an α-Hydroxy Acid. ACS Chem. Biol. 2012, 7, 1015–1022. [CrossRef] [PubMed]
- 27. Agostini, F.; Völler, J.-S.; Koksch, B.; Acevedo-Rocha, C.G.; Kubyshkin, V.; Budisa, N. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 9680–9703. [CrossRef]
- Chalker, J.M.; Bernardes, G.J.L.; Lin, Y.A.; Davis, B.G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. Asian J.* 2009, 4, 630–640. [CrossRef]
- 29. Dickens, F. Interaction of halogenacetates and SH compounds: The reaction of halogenacetic acids with glutathione and cysteine. The mechanism of iodoacetate poisoning of glyoxalase. *Biochem. J.* **1933**, 27, 1141–1151. [CrossRef]
- Tsuchiya, K.; Numata, K. Chemoenzymatic synthesis of polypeptides containing the unnatural amino acid 2-aminoisobutyric acid. Chem. Commun. 2017, 53, 7318–7321. [CrossRef]

- 31. Goettig, P. Reversed Proteolysis—Proteases as Peptide Ligases. Catalysts 2021, 11, 33. [CrossRef]
- Paetzel, M.; Strynadka, N.C.J.; Tschantz, W.R.; Casareno, R.; Bullinger, P.R.; Dalbey, R.E. Use of Site-directed Chemical Modification to Study an Essential Lysine in *Escherichia coli* Leader Peptidase. J. Biol. Chem. 1997, 272, 9994–10003. [CrossRef]
- DeSantis, G.; Berglund, P.; Stabile, M.R.; Gold, M.; Jones, J.B. Site-Directed Mutagenesis Combined with Chemical Modification as a Strategy for Altering the Specificity of the S1 and S1' Pockets of Subtilisin Bacillus lentus. *Biochemistry* 1998, 37, 5968–5973. [CrossRef] [PubMed]
- Davis, B.G.; Khumtaveeporn, K.; Bott, R.R.; Jones, J.B. Altering the specificity of subtilisin Bacillus lentus through the introduction of positive charge at single amino acid sites. *Bioorg. Med. Chem.* 1999, 7, 2303–2311. [CrossRef] [PubMed]
- 35. Davis, B.G.; Shang, X.; DeSantis, G.; Bott, R.R.; Jones, J.B. The controlled introduction of multiple negative charge at single amino acid sites in subtilisin Bacillus lentus. *Bioorg. Med. Chem.* **1999**, *7*, 2293–2301. [CrossRef] [PubMed]
- 36. Cowie, D.B.; Cohen, G.N. Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulfur. *Biochim. Biophys. Acta* **1957**, *26*, 252–261. [CrossRef]
- 37. Richmond, H.M. The Effect of Amino Acid Analogues on Growth and Protein Synthesis in Microorganisms. *Bacteriol. Rev.* **1962**, 26, 398–420. [CrossRef]
- Boles, J.O.; Cisneros, R.J.; Weir, M.S.; Odom, J.D.; Villafranca, J.E.; Dunlap, R.B. Purification and characterization of selenomethionyl thymidylate synthase from *Escherichia coli*: Comparison with the wild-type enzyme. *Biochemistry* 1991, 30, 11073–11080. [CrossRef]
- Yang, W.; Hendrickson, W.A.; Crouch, R.J.; Satow, Y. Structure of Ribonuclease H Phased at 2 Å Resolution by MAD Analysis of the Selenomethionyl Protein. *Science* 1990, 249, 1398–1405. [CrossRef]
- 40. Boles, J.O.; Yu, H.N.; Patti, J.M. The biosynthetic incorporation of selenomethionine and telluromethionine into pyrrolidone carboxyl peptidase (PYRase) from *S. aureus. SAAS Bull. Biochem. Biotechnol.* **1997**, *10*, 13–17. [PubMed]
- Budisa, N.; Steipe, B.; Demange, P.; Eckerskorn, C.; Kellermann, J.; Huber, R. High-level Biosynthetic Substitution of Methionine in Proteins by its Analogs 2-Aminohexanoic Acid, Selenomethionine, Telluromethionine and Ethionine in *Escherichia coli*. *Eur. J. Biochem.* 1995, 230, 788–796. [CrossRef]
- 42. Bae, J.H.; Alefelder, S.; Kaiser, J.T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. Incorporation of β-selenolo[3,2-b]pyrrolylalanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **2001**, 309, 925–936. [CrossRef]
- 43. Hendrickson, W.A.; Smith, J.L.; Sheriff, S. Direct phase determination based on anomalous scattering. *Methods Enzymol.* **1985**, 115, 41–55. [CrossRef]
- 44. Hendrickson, W.A.; Horton, J.R.; LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J.* **1990**, *9*, 1665–1672. [CrossRef]
- Both, E.B.; Shao, S.; Xiang, J.; Jókai, Z.; Yin, H.; Liu, Y.; Magyar, A.; Dernovics, M. Selenolanthionine is the major water-soluble selenium compound in the selenium tolerant plant Cardamine violifolia. *Biochim. Biophys. Acta Gen. Subj.* 2018, 1862, 2354–2362. [CrossRef]
- Templeton, L.K.; Templeton, D.H. Biaxial tensors for anomalous scattering of X-rays in selenolanthionine. *Acta Crystallogr. Sect. A* 1988, 44, 1045–1051. [CrossRef]
- 47. Ferrer, J.L.; Roth, M.; Antoniadis, A. Data compression for diffraction patterns. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1998**, *54*, 184–199. [CrossRef] [PubMed]
- Hendrickson, W.A.; Pähler, A.; Smith, J.L.; Satow, Y.; Merritt, E.A.; Phizackerley, R.P. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl. Acad. Sci. USA* 1989, *86*, 2190–2194. [CrossRef] [PubMed]
- 49. Boles, J.O.; Lewinski, K.; Kunkle, M.; Odom, J.D.; Dunlap, R.B.; Lebioda, L.; Hatada, M. Bio–incorporation of telluromethionine into buried residues of dihydrofolate reductase. *Nat. Struct. Biol.* **1994**, *1*, 283–284. [CrossRef] [PubMed]
- 50. Ramadan, S.E.; Razak, A.A.; Ragab, A.M.; El-Meleigy, M. Incorporation of tellurium into amino acids and proteins in a telluriumtolerant fungi. *Biol. Trace Elem. Res.* **1989**, *20*, 225–232. [CrossRef]
- Budisa, N.; Karnbrock, W.; Steinbacher, S.; Humm, A.; Prade, L.; Neuefeind, T.; Moroder, L.; Huber, R. Bioincorporation of telluromethionine into proteins: A promising new approach for X-ray structure analysis of proteins. *J. Mol. Biol.* 1997, 270, 616–623. [CrossRef] [PubMed]
- 52. Moroder, L.; Musiol, H.-J. Amino acid chalcogen analogues as tools in peptide and protein research. *J. Pept. Sci.* 2020, 26, e3232. [CrossRef]
- 53. Gomis-Rüth, F.-X.; Maskos, K.; Betz, M.; Bergner, A.; Huber, R.; Suzuki, K.; Yoshida, N.; Nagase, H.; Brew, K.; Bourenkov, G.P.; et al. Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* 1997, 389, 77–81. [CrossRef] [PubMed]
- Liu, Q.; Hendrickson, W.A. Contemporary Use of Anomalous Diffraction in Biomolecular Structure Analysis. In *Protein Crystallography: Methods and Protocols*; Wlodawer, A., Dauter, Z., Jaskolski, M., Eds.; Springer: New York, NY, USA, 2017; pp. 377–399.
- 55. Rose, J.P.; Wang, B.-C. SAD phasing: History, current impact and future opportunities. *Arch. Biochem. Biophys.* **2016**, 602, 80–94. [CrossRef] [PubMed]
- 56. Barton, W.A.; Tzvetkova-Robev, D.; Erdjument-Bromage, H.; Tempst, P.; Nikolov, D.B. Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Sci.* 2006, *15*, 2008–2013. [CrossRef] [PubMed]

- 57. Ouerdane, L.; Mester, Z. Production and Characterization of Fully Selenomethionine-Labeled Saccharomyces cerevisiae. *J. Agric. Food Chem.* **2008**, *56*, 11792–11799. [CrossRef] [PubMed]
- 58. Wenzel, S.; Imasaki, T.; Takagi, Y. A practical method for efficient and optimal production of Seleno-methionine-labeled recombinant protein complexes in the insect cells. *Protein Sci.* **2019**, *28*, 808–822. [CrossRef]
- Brewster, A.S.; Bhowmick, A.; Bolotovsky, R.; Mendez, D.; Zwart, P.H.; Sauter, N.K. SAD phasing of XFEL data depends critically on the error model. *Acta Crystallogr. Sect. D* 2019, 75, 959–968. [CrossRef]
- 60. Rius, J.; Torrelles, X. Extending the novel |[rho]|-based phasing algorithm to the solution of anomalous scattering substructures from SAD data of protein crystals. *Acta Crystallogr. Sect. A* 2022, *78*, 473–481. [CrossRef]
- 61. Barbarin-Bocahu, I.; Graille, M. The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: A case-study report. *Acta Crystallogr. Sect. D* 2022, *78*, 517–531. [CrossRef]
- Hoesl, M.G.; Budisa, N. In vivo Incorporation of Multiple Noncanonical Amino Acids into Proteins. *Angew. Chem. Int. Ed. Engl.* 2011, 50, 2896–2902. [CrossRef]
- 63. Kwon, I.; Kirshenbaum, K.; Tirrell, D.A. Breaking the Degeneracy of the Genetic Code. J. Am. Chem. Soc. 2003, 125, 7512–7513. [CrossRef]
- 64. Bohlke, N.; Budisa, N. Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: Rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol. Lett.* **2014**, *351*, 133–144. [CrossRef] [PubMed]
- 65. Budisa, N. Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire. *Angew. Chem. Int. Ed. Engl.* 2004, 43, 6426–6463. [CrossRef] [PubMed]
- 66. Budisa, N.; Minks, C.; Alefelder, S.; Wenger, W.; Dong, F.; Moroder, L.; Huber, R. Toward the experimental codon reassignment in vivo: Protein building with an expanded amino acid repertoire. *FASEB J.* **1999**, *13*, 41–51. [CrossRef] [PubMed]
- 67. Xue, H.; Wong, J.T. Future of the Genetic Code. Life 2017, 7, 10. [CrossRef] [PubMed]
- Kanda, T.; Takai, K.; Hohsaka, T.; Sisido, M.; Takaku, H. Sense Codon-Dependent Introduction of Unnatural Amino Acids into Multiple Sites of a Protein. *Biochem. Biophys. Res. Commun.* 2000, 270, 1136–1139. [CrossRef]
- Baumann, T.; Nickling, J.H.; Bartholomae, M.; Buivydas, A.; Kuipers, O.P.; Budisa, N. Prospects of in vivo Incorporation of Non-canonical Amino Acids for the Chemical Diversification of Antimicrobial Peptides. *Front. Microbiol.* 2017, 8, 124. [CrossRef]
- Foster, M.P.; McElroy, C.A.; Amero, C.D. Solution NMR of Large Molecules and Assemblies. *Biochemistry* 2007, 46, 331–340. [CrossRef]
- 71. Merkel, L.; Budisa, N. Organic fluorine as a polypeptide building element: In vivo expression of fluorinated peptides, proteins and proteomes. *Org. Biomol. Chem.* **2012**, *10*, 7241–7261. [CrossRef]
- 72. Sugiki, T.; Furuita, K.; Fujiwara, T.; Kojima, C. Current NMR Techniques for Structure-Based Drug Discovery. *Molecules* **2018**, 23, 148. [CrossRef]
- 73. Ruben, E.A.; Gandhi, P.S.; Chen, Z.; Koester, S.K.; DeKoster, G.T.; Frieden, C.; Di Cera, E. 19F NMR reveals the conformational properties of free thrombin and its zymogen precursor prethrombin-2. *J. Biol. Chem.* **2020**, *295*, 8227–8235. [CrossRef]
- Gowda, G.A.N.; Shanaiah, N.; Raftery, D. Isotope enhanced approaches in metabolomics. *Adv. Exp. Med. Biol.* 2012, 992, 147–164. [CrossRef] [PubMed]
- 75. Verardi, R.; Traaseth, N.J.; Masterson, L.R.; Vostrikov, V.V.; Veglia, G. Isotope Labeling for Solution and Solid-State NMR Spectroscopy of Membrane Proteins BT—Isotope labeling in Biomolecular NMR. In *Advances in Experimental Medicine and Biology*; Atreya, H.S., Ed.; Springer: Dordrecht, The Netherlands, 2012; Volume 992, pp. 35–62.
- Marshall, D.D.; Powers, R. Beyond the paradigm: Combining mass spectrometry and nuclear magnetic resonance for metabolomics. Prog. Nucl. Magn. Reson. Spectrosc. 2017, 100, 1–16. [CrossRef] [PubMed]
- 77. Tian, X.; Permentier, H.P.; Bischoff, R. Chemical isotope labeling for quantitative proteomics. *Mass Spectrom. Rev.* 2023, 42, 546–576. [CrossRef] [PubMed]
- Fang, K.Y.; Lieblich, S.A.; Tirrell, D.A. Incorporation of Non-Canonical Amino Acids into Proteins by Global Reassignment of Sense Codons. *Methods Mol. Biol.* 2018, 1798, 173–186. [CrossRef]
- 79. Hoesl, M.G.; Acevedo-Rocha, C.G.; Nehring, S.; Royter, M.; Wolschner, C.; Wiltschi, B.; Budisa, N.; Antranikian, G. Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem* **2011**, *3*, 213–221. [CrossRef]
- Hoesl, M.G.; Budisa, N. Recent advances in genetic code engineering in *Escherichia coli*. *Curr. Opin. Biotechnol.* 2012, 23, 751–757. [CrossRef]
- 81. Wang, L.; Schultz, P.G. Expanding the genetic code. Chem. Commun. 2002, 1, 1–11. [CrossRef]
- Wan, W.; Tharp, J.M.; Liu, W.R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta Proteins Proteom.* 2014, 1844, 1059–1070. [CrossRef]
- Wang, L.; Brock, A.; Herberich, B.; Schultz, P.G. Expanding the Genetic Code of *Escherichia coli*. Science 2001, 292, 498–500. [CrossRef]
- Baumann, T.; Exner, M.; Budisa, N. Orthogonal Protein Translation Using Pyrrolysyl-tRNA Synthetases for Single- and Multiple-Noncanonical Amino Acid Mutagenesis BT. In *Synthetic Biology—Metabolic Engineering*; Zhao, H., Zeng, A.-P., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 1–19.
- 85. Wang, L. Genetically encoding new bioreactivity. New Biotechnol. 2017, 38, 16–25. [CrossRef]
- 86. de la Torre, D.; Chin, J.W. Reprogramming the genetic code. Nat. Rev. Genet. 2021, 22, 169–184. [CrossRef] [PubMed]

- 87. Ros, E.; Torres, A.G.; Ribas de Pouplana, L. Learning from Nature to Expand the Genetic Code. *Trends Biotechnol.* **2021**, *39*, 460–473. [CrossRef] [PubMed]
- Exner, M.P.; Köhling, S.; Rivollier, J.; Gosling, S.; Srivastava, P.; Palyancheva, Z.I.; Herdewijn, P.; Heck, M.-P.; Rademann, J.; Budisa, N. Incorporation of Amino Acids with Long-Chain Terminal Olefins into Proteins. *Molecules* 2016, *21*, 287. [CrossRef]
- 89. Koch, N.G.; Goettig, P.; Rappsilber, J.; Budisa, N. Engineering Pyrrolysyl-tRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains. *Int. J. Mol. Sci.* 2021, 22, 11194. [CrossRef]
- 90. Tseng, H.-W.; Baumann, T.; Sun, H.; Wang, Y.-S.; Ignatova, Z.; Budisa, N. Expanding the Scope of Orthogonal Translation with Pyrrolysyl-tRNA Synthetases Dedicated to Aromatic Amino Acids. *Molecules* **2020**, *25*, 4418. [CrossRef] [PubMed]
- 91. Jiang, R.; Krzycki, J.A. PylSn and the Homologous N-terminal Domain of Pyrrolysyl-tRNA Synthetase Bind the tRNA That Is Essential for the Genetic Encoding of Pyrrolysine. *J. Biol. Chem.* **2012**, *287*, 32738–32746. [CrossRef]
- 92. Dumas, A.; Lercher, L.; Spicer, C.D.; Davis, B.G. Designing logical codon reassignment—Expanding the chemistry in biology. *Chem. Sci.* **2015**, *6*, 50–69. [CrossRef]
- 93. Reinkemeier, C.D.; Lemke, E.A. Synthetic biomolecular condensates to engineer eukaryotic cells. *Curr. Opin. Chem. Biol.* 2021, 64, 174–181. [CrossRef]
- Merkel, L.; Beckmann, H.S.G.; Wittmann, V.; Budisa, N. Efficient N-Terminal Glycoconjugation of Proteins by the N-End Rule. ChemBioChem 2008, 9, 1220–1224. [CrossRef]
- 95. Arranz-Gibert, P.; Vanderschuren, K.; Isaacs, F.J. Next-generation genetic code expansion. *Curr. Opin. Chem. Biol.* **2018**, *46*, 203–211. [CrossRef]
- 96. Brown, W.; Liu, J.; Deiters, A. Genetic Code Expansion in Animals. ACS Chem. Biol. 2018, 13, 2375–2386. [CrossRef]
- 97. Brabham, R.; Fascione, M.A. Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *ChemBioChem* 2017, 18, 1973–1983. [CrossRef]
- Pirman, N.L.; Barber, K.W.; Aerni, H.R.; Ma, N.J.; Haimovich, A.D.; Rogulina, S.; Isaacs, F.J.; Rinehart, J. A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation. *Nat. Commun.* 2015, *6*, 8130. [CrossRef] [PubMed]
- Rogerson, D.T.; Sachdeva, A.; Wang, K.; Haq, T.; Kazlauskaite, A.; Hancock, S.M.; Huguenin-Dezot, N.; Muqit, M.M.K.; Fry, A.M.; Bayliss, R.; et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* 2015, *11*, 496–503. [CrossRef] [PubMed]
- Wiltschi, B. Incorporation of non-canonical amino acids into proteins in yeast. *Fungal Genet. Biol.* 2016, 89, 137–156. [CrossRef]
 [PubMed]
- 101. Wissner, R.F.; Batjargal, S.; Fadzen, C.M.; Petersson, E.J. Labeling Proteins with Fluorophore/Thioamide Förster Resonant Energy Transfer Pairs by Combining Unnatural Amino Acid Mutagenesis and Native Chemical Ligation. J. Am. Chem. Soc. 2013, 135, 6529–6540. [CrossRef] [PubMed]
- Cui, Z.; Johnston, W.A.; Alexandrov, K. Cell-Free Approach for Non-canonical Amino Acids Incorporation Into Polypeptides. Front. Bioeng. Biotechnol. 2020, 8, 1031. [CrossRef]
- 103. Hui, A.; de Boer, H.A. Specialized ribosome system: Preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4762–4766. [CrossRef]
- 104. Rackham, O.; Chin, J.W. Cellular logic with orthogonal ribosomes. J. Am. Chem. Soc. 2005, 127, 17584–17585. [CrossRef]
- Dedkova, L.M.; Fahmi, N.E.; Golovine, S.Y.; Hecht, S.M. Construction of modified ribosomes for incorporation of D-amino acids into proteins. *Biochemistry* 2006, 45, 15541–15551. [CrossRef]
- 106. Maini, R.; Nguyen, D.T.; Chen, S.; Dedkova, L.M.; Chowdhury, S.R.; Alcala-Torano, R.; Hecht, S.M. Incorporation of β-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorg. Med. Chem.* 2013, 21, 1088–1096. [CrossRef] [PubMed]
- Chen, S.; Ji, X.; Gao, M.; Dedkova, L.M.; Hecht, S.M. In Cellulo Synthesis of Proteins Containing a Fluorescent Oxazole Amino Acid. J. Am. Chem. Soc. 2019, 141, 5597–5601. [CrossRef]
- Shakya, B.; Joyner, O.G.; Hartman, M.C.T. Hyperaccurate Ribosomes for Improved Genetic Code Reprogramming. ACS Synth. Biol. 2022, 11, 2193–2201. [CrossRef] [PubMed]
- 109. Mukai, T.; Lajoie, M.J.; Englert, M.; Söll, D. Rewriting the Genetic Code. Annu. Rev. Microbiol. 2017, 71, 557–577. [CrossRef]
- 110. Fiacco, S.V.; Kelderhouse, L.E.; Hardy, A.; Peleg, Y.; Hu, B.; Ornelas, A.; Yang, P.; Gammon, S.T.; Howell, S.M.; Wang, P.; et al. Directed Evolution of Scanning Unnatural-Protease-Resistant (SUPR) Peptides for in vivo Applications. *ChemBioChem* 2016, 17, 1643–1651. [CrossRef]
- Otvos Jr, L.; Bokonyi, K.; Varga, I.; Ertl, H.C.J.; Hoffmann, R.; Bulet, P.; Otvos, B.I.; Wade, J.D.; McManus, A.M.; Craik, D.J. Insect peptides with improved protease-resistance protect mice against bacterial infection. *Protein Sci.* 2000, *9*, 742–749. [CrossRef] [PubMed]
- 112. Yin, L.M.; Lee, S.; Mak, J.S.W.; Helmy, A.S.; Deber, C.M. Differential binding of L- vs. D-isomers of cationic antimicrobial peptides to the biofilm exopolysaccharide alginate. *Protein Pept. Lett.* **2013**, *20*, 843–847. [CrossRef]
- 113. Khara, J.S.; Priestman, M.; Uhía, I.; Hamilton, M.S.; Krishnan, N.; Wang, Y.; Yang, Y.Y.; Langford, P.R.; Newton, S.M.; Robertson, B.D.; et al. Unnatural amino acid analogues of membrane-active helical peptides with anti-mycobacterial activity and improved stability. J. Antimicrob. Chemother. 2016, 71, 2181–2191. [CrossRef] [PubMed]

- 114. Oh, J.E.; Lee, K.H. Synthesis of novel unnatural amino acid as a building block and its incorporation into an antimicrobial peptide. *Bioorg. Med. Chem.* **1999**, *7*, 2985–2990. [CrossRef]
- 115. Sarkar, T.; Chetia, M.; Chatterjee, S. Antimicrobial Peptides and Proteins: From Nature's Reservoir to the Laboratory and Beyond. *Front. Chem.* **2021**, *9*, 691532. [CrossRef]
- Lu, J.; Xu, H.; Xia, J.; Ma, J.; Xu, J.; Li, Y.; Feng, J. D- and Unnatural Amino Acid Substituted Antimicrobial Peptides With Improved Proteolytic Resistance and Their Proteolytic Degradation Characteristics. *Front. Microbiol.* 2020, *11*, 563030. [CrossRef] [PubMed]
- Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 1967, 27, 157–162. [CrossRef] [PubMed]
- Maluch, I.; Czarna, J.; Drag, M. Applications of Unnatural Amino Acids in Protease Probes. *Chem. Asian J.* 2019, 14, 4103–4113. [CrossRef] [PubMed]
- Poreba, M.; Salvesen, G.S.; Drag, M. Synthesis of a HyCoSuL peptide substrate library to dissect protease substrate specificity. *Nat. Protoc.* 2017, 12, 2189–2214. [CrossRef] [PubMed]
- 120. Vizovišek, M.; Vidmar, R.; Drag, M.; Fonović, M.; Salvesen, G.S.; Turk, B. Protease Specificity: Towards in vivo Imaging Applications and Biomarker Discovery. *Trends Biochem. Sci.* **2018**, *43*, 829–844. [CrossRef]
- 121. Liu, Y.; Patricelli, M.P.; Cravatt, B.F. Activity-based protein profiling: The serine hydrolases. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14694–14699. [CrossRef]
- 122. Smolewski, P.; Grabarek, J.; Halicka, H.D.; Darzynkiewicz, Z. Assay of caspase activation in situ combined with probing plasma membrane integrity to detect three distinct stages of apoptosis. *J. Immunol. Methods* **2002**, *265*, 111–121. [CrossRef]
- 123. Lai, Z.; Tan, P.; Zhu, Y.; Shao, C.; Shan, A.; Li, L. Highly Stabilized α-Helical Coiled Coils Kill Gram-Negative Bacteria by Multicomplementary Mechanisms under Acidic Condition. *ACS Appl. Mater. Interfaces* **2019**, *11*, 22113–22128. [CrossRef]
- 124. Sarwar Gilani, G.; Wu Xiao, C.; Cockell, K.A. Impact of Antinutritional Factors in Food Proteins on the Digestibility of Protein and the Bioavailability of Amino Acids and on Protein Quality. *Br. J. Nutr.* **2012**, *108*, S315–S332. [CrossRef]
- 125. Kiso, Y.; Matsumoto, H.; Mizumoto, S.; Kimura, T.; Fujiwara, Y.; Akaji, K. Small dipeptide-based HIV protease inhibitors containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic. *Pept. Sci.* **1999**, *51*, 59–68. [CrossRef]
- 126. Böhm, H.-J. Towards the automatic design of synthetically accessible protein ligands: Peptides, amides and peptidomimetics. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 265–272. [CrossRef] [PubMed]
- 127. Gentilucci, L.; De Marco, R.; Cerisoli, L. Chemical modifications designed to improve peptide stability: Incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Curr. Pharm. Des.* **2010**, *16*, 3185–3203. [CrossRef] [PubMed]
- 128. Quast, U.; Engel, J.; Steffen, E.; Mair, G.; Tschesche, H.; Jering, H. Kinetics of Binding of Bovine Trypsin-Kallikrein Inhibitor (Kunitz) in which the Reactive-Site Peptide Bond Lys-15-Ala-16 is Cleaved, to α-Chymotrypsin and β-Trypsin. *Eur. J. Biochem.* **1975**, 52, 505–510. [CrossRef] [PubMed]
- 129. Roberts, D.V.; Elmore, D.T. Kinetics and mechanism of catalysis by proteolytic enzymes. A comparison of the kinetics of hydrolysis of synthetic substrates by bovine α- and β-trypsin. *Biochem. J.* **1974**, *141*, 545–554. [CrossRef] [PubMed]
- 130. Yung-Chi, C.; Prusoff, W.H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, 22, 3099–3108. [CrossRef] [PubMed]
- Goettig, P.; Brandstetter, H.; Magdolen, V. Surface loops of trypsin-like serine proteases as determinants of function. *Biochimie* 2019, 166, 52–76. [CrossRef]
- 132. Lu, W.; Apostol, I.; Qasim, M.A.; Warne, N.; Wynn, R.; Zhang, W.L.; Anderson, S.; Chiang, Y.W.; Ogin, E.; Rothberg, I.; et al. Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **1997**, *266*, 441–461. [CrossRef]
- 133. Schellenberger, V.; Braune, K.; Hofmann, H.-J.; Jakubke, H.-D. The specificity of chymotrypsin. *Eur. J. Biochem.* **1991**, *199*, 623–636. [CrossRef]
- 134. Wysocka, M.; Lesner, A.; Legowska, A.; Jaśkiewicz, A.; Miecznikowska, H.; Rolka, K. Designing of substrates and inhibitors of bovine alpha-chymotrypsin with synthetic phenylalanine analogues in position P(1). *Protein Pept. Lett.* 2008, 15, 260–264. [CrossRef]
- 135. Huang, K.; James, M.N.G.; Lu, W.; Laskowski, M., Jr.; Anderson, S. Water molecules participate in proteinase-inhibitor interactions: Crystal structures of Leu¹⁸, Ala¹⁸, and Gly¹⁸ variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* 1995, 4, 1985–1997. [CrossRef]
- 136. Lee, T.-W.; Qasim, M.A.; Laskowski, M.; James, M.N.G. Structural Insights into the Non-additivity Effects in the Sequence-to-Reactivity Algorithm for Serine Peptidases and their Inhibitors. *J. Mol. Biol.* 2007, 367, 527–546. [CrossRef] [PubMed]
- 137. Ye, S.; Loll, B.; Berger, A.A.; Mülow, U.; Alings, C.; Wahl, M.C.; Koksch, B. Fluorine teams up with water to restore inhibitor activity to mutant BPTI. *Chem. Sci.* 2015, *6*, 5246–5254. [CrossRef] [PubMed]
- 138. Leppkes, J.; Dimos, N.; Loll, B.; Hohmann, T.; Dyrks, M.; Wieseke, A.; Keller, B.G.; Koksch, B. Fluorine-induced polarity increases inhibitory activity of BPTI towards chymotrypsin. *RSC Chem. Biol.* **2022**, *3*, 773–782. [CrossRef] [PubMed]
- 139. Bark, S.J.; Kent, S.B.H. Engineering an unnatural Nα-anchored disulfide into BPTI by total chemical synthesis: Structural and functional consequences. *FEBS Lett.* **1999**, *460*, 67–76. [CrossRef] [PubMed]
- Wang, B.; Brown, K.C.; Lodder, M.; Craik, C.S.; Hecht, S.M. Chemically mediated site-specific proteolysis. Alteration of protein-protein interaction. *Biochemistry* 2002, *41*, 2805–2813. [CrossRef] [PubMed]

- 141. Debowski, D.; Łukajtis, R.; Filipowicz, M.; Strzelecka, P.; Wysocka, M.; Łęgowska, A.; Lesner, A.; Rolka, K. Hybrid analogues of SFTI-1 modified in P1 position by β- and γ-amino acids and N-substituted β-alanines. *Pept. Sci.* 2013, 100, 154–159. [CrossRef]
- 142. Lukajtis, R.; Legowska, A.; Wysocka, M.; Debowski, D.; Lesner, A.; Rolka, K. Analogues of trypsin inhibitor SFTI-1 modified in the conserved P1' position by synthetic or non-proteinogenic amino acids retain their inhibitory activity. *J. Pept. Sci.* 2011, 17, 281–287. [CrossRef]
- 143. Fewer, D.P.; Jokela, J.; Rouhiainen, L.; Wahlsten, M.; Koskenniemi, K.; Stal, L.J.; Sivonen, K. The non-ribosomal assembly and frequent occurrence of the protease inhibitors spumigins in the bloom-forming cyanobacterium Nodularia spumigena. *Mol. Microbiol.* 2009, 73, 924–937. [CrossRef]
- 144. Ouyang, X.; D'Agostino, P.M.; Wahlsten, M.; Delbaje, E.; Jokela, J.; Permi, P.; Gaiani, G.; Poso, A.; Bartos, P.; Gulder, T.A.M.; et al. Direct pathway cloning and expression of the radiosumin biosynthetic gene cluster. Org. Biomol. Chem. 2023, 21, 4893–4908. [CrossRef]
- 145. Issac, M.; Aknin, M.; Gauvin-Bialecki, A.; De Voogd, N.; Ledoux, A.; Frederich, M.; Kashman, Y.; Carmeli, S. Cyclotheonellazoles A–C, Potent Protease Inhibitors from the Marine Sponge Theonella aff. swinhoei. J. Nat. Prod. 2017, 80, 1110–1116. [CrossRef]
- 146. Matthew, S.; Ross, C.; Rocca, J.R.; Paul, V.J.; Luesch, H. Lyngbyastatin 4, a Dolastatin 13 Analogue with Elastase and Chymotrypsin Inhibitory Activity from the Marine Cyanobacterium Lyngbya confervoides. J. Nat. Prod. 2007, 70, 124–127. [CrossRef] [PubMed]
- 147. Talukder, P.; Dedkova, L.M.; Ellington, A.D.; Yakovchuk, P.; Lim, J.; Anslyn, E.V.; Hecht, S.M. Synthesis of alanyl nucleobase amino acids and their incorporation into proteins. *Bioorg. Med. Chem.* **2016**, *24*, 4177–4187. [CrossRef] [PubMed]
- 148. Avrutina, O.; Schmoldt, H.-U.; Gabrijelcic-Geiger, D.; Le Nguyen, D.; Sommerhoff, C.P.; Diederichsen, U.; Kolmar, H. Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol. Chem.* **2005**, *386*, 1301–1306. [CrossRef]
- 149. Di Cera, E. Thrombin. Mol. Asp. Med. 2008, 29, 203–254. [CrossRef] [PubMed]
- 150. Figueiredo, A.C.; Clement, C.C.; Zakia, S.; Gingold, J.; Philipp, M.; Pereira, P.J.B. Rational Design and Characterization of D-Phe-Pro-D-Arg-Derived Direct Thrombin Inhibitors. *PLoS ONE* **2012**, *7*, e34354. [CrossRef] [PubMed]
- 151. De Marco, R.; Di Gioia, M.L.; Leggio, A.; Liguori, A.; Perri, F.; Siciliano, C.; Viscomi, M.C. A new non-natural arginine-like amino acid derivative with a sulfamoyl group in the side-chain. *Amino Acids* **2010**, *38*, 691–700. [CrossRef]
- 152. De Nanteuil, G.; Gloanec, P.; Lila, C.; Portevin, B.; Boudon, A.; Rupin, A.; Verbeuren, T.J. New tripeptidic thrombin inhibitors. Influence of P2 and P3 residues on activity and selectivity. *Bioorg. Med. Chem.* **1995**, *3*, 1019–1024. [CrossRef]
- 153. Portevin, B.; Lonchampt, M.; Canet, E.; De Nanteuil, G. Dual Inhibition of Human Leukocyte Elastase and Lipid Peroxidation: In vitro and in vivo Activities of Azabicyclo[2.2.2]octane and Perhydroindole Derivatives. J. Med. Chem. 1997, 40, 1906–1918. [CrossRef]
- 154. Deadman, J.; Claeson, G.; Scully, M.F. Structure/Function Aspects of Neutral P1 Residue Peptide Inhibitors of Thrombin. J. Enzym. Inhib. 1995, 9, 29–41. [CrossRef]
- 155. Tsuda, Y.; Szewczuk, Z.; Wang, J.; Yue, S.Y.; Purisima, E.; Konishi, Y. Interactions of Hirudin-Based Inhibitor with Thrombin: Critical Role of the IleH59 Side Chain of the Inhibitor. *Biochemistry* **1995**, *34*, 8708–8714. [CrossRef]
- 156. Lombardi, A.; Nastri, F.; Della Morte, R.; Rossi, A.; De Rosa, A.; Staiano, N.; Pedone, C.; Pavone, V. Rational Design of True Hirudin Mimetics: Synthesis and Characterization of Multisite-Directed α-Thrombin Inhibitors. *J. Med. Chem.* **1996**, *39*, 2008–2017. [CrossRef] [PubMed]
- 157. De Simone, G.; Lombardi, A.; Galdiero, S.; Nastri, F.; Morte, R.D.; Staiano, N.; Pedone, C.; Bolognesi, M.; Pavone, V. Hirunorms are true hirudin mimetics. The crystal structure of human α-thrombin-hirunorm V complex. *Protein Sci.* 1998, 7, 243–253. [CrossRef] [PubMed]
- 158. De Filippis, V.; Quarzago, D.; Vindigni, A.; Di Cera, E.; Fontana, A. Synthesis and Characterization of More Potent Analogues of Hirudin Fragment 1–47 Containing Non-Natural Amino Acids. *Biochemistry* **1998**, *37*, 13507–13515. [CrossRef] [PubMed]
- De Filippis, V.; Russo, I.; Vindigni, A.; Cera, E.D.; Salmaso, S.; Fontana, A. Incorporation of noncoded amino acids into the N-terminal domain 1-47 of hirudin yields a highly potent and selective thrombin inhibitor. *Protein Sci.* 1999, *8*, 2213–2217. [CrossRef]
- De Filippis, V.; De Boni, S.; De Dea, E.; Dalzoppo, D.; Grandi, C.; Fontana, A. Incorporation of the fluorescent amino acid 7-azatryptophan into the core domain 1-47 of hirudin as a probe of hirudin folding and thrombin recognition. *Protein Sci. A Publ. Protein Soc.* 2004, 13, 1489–1502. [CrossRef]
- 161. De Filippis, V.; Frasson, R.; Fontana, A. 3-Nitrotyrosine as a spectroscopic probe for investigating protein–protein interactions. *Protein Sci.* **2006**, *15*, 976–986. [CrossRef]
- 162. Steinmetzer, T.; Zhu, B.Y.; Konishi, Y. Potent Bivalent Thrombin Inhibitors: Replacement of the Scissile Peptide Bond at P1–P1' with Arginyl Ketomethylene Isosteres. *J. Med. Chem.* **1999**, *42*, 3109–3115. [CrossRef]
- 163. Brandstetter, H.; Turk, D.; Hoeffken, H.W.; Grosse, D.; Stürzebecher, J.; Martin, P.D.; Edwards, B.F.P.; Bode, W. Refined 2·3ÅX-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA: A starting point for improving antithrombotics. J. Mol. Biol. 1992, 226, 1085–1099. [CrossRef]
- 164. Bergner, A.; Bauer, M.; Brandstetter, H.; Stürzebecher, J.; Bode, W. The X-ray crystal structure of thrombin in complex with N alpha-2-naphthylsulfonyl-L-3-amidino-phenylalanyl-4-methylpiperidide: The beneficial effect of filling out an empty cavity. *J. Enzym. Inhib.* 1995, 9, 101–110. [CrossRef]
- Steinmetzer, T.; Schweinitz, A.; Künzel, S.; Wikström, P.; Hauptmann, J.; Stürzebecher, J. Structure-activity relationships of new NAPAP-analogs. J. Enzym. Inhib. 2001, 16, 241–249. [CrossRef]

- 166. Hauptmann, J.; Steinmetzer, T.; Vieweg, H.; Wikström, P.; Stürzebecher, J. Influence of Structural Variations in Peptidomimetic 4-Amidinophenylalanine-Derived Thrombin Inhibitors on Plasma Clearance and Biliary Excretion in Rats. *Pharm. Res.* 2002, 19, 1027–1033. [CrossRef] [PubMed]
- 167. Féthière, J.; Tsuda, Y.; Coulombe, R.; Konishi, Y.; Cygler, M. Crystal structure of two new bifunctional nonsubstrate type thrombin inhibitors complexed with human α-thrombin. *Protein Sci.* **1996**, *5*, 1174–1183. [CrossRef] [PubMed]
- 168. Tsuda, Y.; Cygler, M.; Gibbs, B.F.; Pedyczak, A.; Féthière, J.; Yue, S.Y.; Konishi, Y. Design of potent bivalent thrombin inhibitors based on hirudin sequence: Incorporation of nonsubstrate-type active site inhibitors. *Biochemistry* 1994, 33, 14443–14451. [CrossRef] [PubMed]
- 169. Grütter, M.G.; Priestle, J.P.; Rahuel, J.; Grossenbacher, H.; Bode, W.; Hofsteenge, J.; Stone, S.R. Crystal structure of the thrombinhirudin complex: A novel mode of serine protease inhibition. *EMBO J.* **1990**, *9*, 2361–2365. [CrossRef]
- Slon-Usakiewicz, J.J.; Purisima, E.; Tsuda, Y.; Sulea, T.; Pedyczak, A.; Féthière, J.; Cygler, M.; Konishi, Y. Nonpolar interactions of thrombin S' subsites with its bivalent inhibitor: Methyl scan of the inhibitor linker. *Biochemistry* 1997, 36, 13494–13502. [CrossRef]
- 171. Nieman, M.T.; Burke, F.; Warnock, M.; Zhou, Y.; Sweigart, J.; Chen, A.; Ricketts, D.; Lucchesi, B.R.; Chen, Z.; Di Cera, E.; et al. Thrombostatin FM compounds: Direct thrombin inhibitors—Mechanism of action in vitro and in vivo. *J. Thromb. Haemost. JTH* 2008, *6*, 837–845. [CrossRef]
- 172. Hanessian, S.; Ersmark, K.; Wang, X.; Del Valle, J.R.; Blomberg, N.; Xue, Y.; Fjellström, O. Structure-based organic synthesis of unnatural aeruginosin hybrids as potent inhibitors of thrombin. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3480–3485. [CrossRef]
- 173. Calisto, B.M.; Ripoll-Rozada, J.; Dowman, L.J.; Franck, C.; Agten, S.M.; Parker, B.L.; Veloso, R.C.; Vale, N.; Gomes, P.; de Sanctis, D.; et al. Sulfotyrosine-Mediated Recognition of Human Thrombin by a Tsetse Fly Anticoagulant Mimics Physiological Substrates. *Cell Chem. Biol.* 2021, *28*, 26–33.e28. [CrossRef]
- 174. Hosseini, M.; Jiang, L.; Sørensen, H.P.; Jensen, J.K.; Christensen, A.; Fogh, S.; Yuan, C.; Andersen, L.M.; Huang, M.; Andreasen, P.A.; et al. Elucidation of the Contribution of Active Site and Exosite Interactions to Affinity and Specificity of Peptidylic Serine Protease Inhibitors Using Non-Natural Arginine Analogs. *Mol. Pharmacol.* **2011**, *80*, 585–597. [CrossRef]
- 175. Zhao, B.; Xu, P.; Jiang, L.; Paaske, B.; Kromann-Hansen, T.; Jensen, J.K.; Sørensen, H.P.; Liu, Z.; Nielsen, J.T.; Christensen, A.; et al. A Cyclic Peptidic Serine Protease Inhibitor: Increasing Affinity by Increasing Peptide Flexibility. *PLoS ONE* 2014, 9, e115872. [CrossRef]
- 176. Modrzycka, S.; Kołt, S.; Polderdijk, S.G.I.; Adams, T.E.; Potoczek, S.; Huntington, J.A.; Kasperkiewicz, P.; Drąg, M. Parallel imaging of coagulation pathway proteases activated protein C, thrombin, and factor Xa in human plasma. *Chem. Sci.* **2022**, *13*, 6813–6829. [CrossRef]
- 177. Modrzycka, S.; Kołt, S.; Adams, T.E.; Potoczek, S.; Huntington, J.A.; Kasperkiewicz, P.; Drag, M. Fluorescent Activity-Based Probe To Image and Inhibit Factor XIa Activity in Human Plasma. *J. Med. Chem.* **2023**, *66*, 3785–3797. [CrossRef] [PubMed]
- Guillen Schlippe, Y.V.; Hartman, M.C.T.; Josephson, K.; Szostak, J.W. In vitro Selection of Highly Modified Cyclic Peptides That Act as Tight Binding Inhibitors. J. Am. Chem. Soc. 2012, 134, 10469–10477. [CrossRef] [PubMed]
- 179. Goettig, P.; Magdolen, V.; Brandstetter, H. Natural and synthetic inhibitors of kallikrein-related peptidases (KLKs). *Biochimie* **2010**, 92, 1546–1567. [CrossRef] [PubMed]
- Melo, R.L.; Barbosa Pozzo, R.C.; Pimenta, D.C.; Perissutti, E.; Caliendo, G.; Santagada, V.; Juliano, L.; Juliano, M.A. Human Tissue Kallikrein S1 Subsite Recognition of Non-Natural Basic Amino Acids. *Biochemistry* 2001, 40, 5226–5232. [CrossRef]
- Pimenta, D.C.; Melo, R.L.; Caliendo, G.; Santagada, V.; Fiorino, F.; Severino, B.; de Nucci, G.; Juliano, L.; Juliano, M.A. Design of inhibitors for human tissue kallikrein using non-natural aromatic and basic amino acids. *Biol. Chem.* 2002, 383, 853–857. [CrossRef]
- Guo, S.; Skala, W.; Magdolen, V.; Briza, P.; Biniossek, M.L.; Schilling, O.; Kellermann, J.; Brandstetter, H.; Goettig, P. A Single Glycan at the 99-Loop of Human Kallikrein-related Peptidase 2 Regulates Activation and Enzymatic Activity. J. Biol. Chem. 2016, 291, 593–604. [CrossRef]
- LeBeau, A.M.; Banerjee, S.R.; Pomper, M.G.; Mease, R.C.; Denmeade, S.R. Optimization of peptide-based inhibitors of prostatespecific antigen (PSA) as targeted imaging agents for prostate cancer. *Bioorg. Med. Chem.* 2009, 17, 4888–4893. [CrossRef]
- 184. Teufel, D.P.; Bennett, G.; Harrison, H.; van Rietschoten, K.; Pavan, S.; Stace, C.; Le Floch, F.; Van Bergen, T.; Vermassen, E.; Barbeaux, P.; et al. Stable and Long-Lasting, Novel Bicyclic Peptide Plasma Kallikrein Inhibitors for the Treatment of Diabetic Macular Edema. J. Med. Chem. 2018, 61, 2823–2836. [CrossRef]
- 185. Patel, S.; Homaei, A.; El-Seedi, H.R.; Akhtar, N. Cathepsins: Proteases that are vital for survival but can also be fatal. *Biomed. Pharmacother.* **2018**, *105*, 526–532. [CrossRef]
- Wysocka, M.; Łęgowska, A.; Bulak, E.; Jaśkiewicz, A.; Miecznikowska, H.; Lesner, A.; Rolka, K. New chromogenic substrates of human neutrophil cathepsin G containing non-natural aromatic amino acid residues in position P1 selected by combinatorial chemistry methods. *Mol. Divers.* 2007, 11, 93–99. [CrossRef]
- 187. de Veer, S.J.; White, A.M.; Craik, D.J. Sunflower Trypsin Inhibitor-1 (SFTI-1): Sowing Seeds in the Fields of Chemistry and Biology. *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 8050–8071. [CrossRef] [PubMed]
- 188. Łęgowska, A.; Dębowski, D.; Lesner, A.; Wysocka, M.; Rolka, K. Introduction of non-natural amino acid residues into the substrate-specific P1 position of trypsin inhibitor SFTI-1 yields potent chymotrypsin and cathepsin G inhibitors. *Bioorg. Med. Chem.* 2009, 17, 3302–3307. [CrossRef] [PubMed]

- St-Georges, C.; Désilets, A.; Béliveau, F.; Ghinet, M.; Dion, S.P.; Colombo, É.; Boudreault, P.-L.; Najmanovich, R.J.; Leduc, R.; Marsault, É. Modulating the selectivity of matriptase-2 inhibitors with unnatural amino acids. *Eur. J. Med. Chem.* 2017, 129, 110–123. [CrossRef]
- Groborz, K.; Kołt, S.; Kasperkiewicz, P.; Drag, M. Internally quenched fluorogenic substrates with unnatural amino acids for cathepsin G investigation. *Biochimie* 2019, 166, 103–111. [CrossRef]
- 191. Grzywa, R.; Burchacka, E.; Łęcka, M.; Winiarski, Ł.; Walczak, M.; Łupicka-Słowik, A.; Wysocka, M.; Burster, T.; Bobrek, K.; Csencsits-Smith, K.; et al. Synthesis of Novel Phosphonic-Type Activity-Based Probes for Neutrophil Serine Proteases and Their Application in Spleen Lysates of Different Organisms. *ChemBioChem* 2014, 15, 2605–2612. [CrossRef]
- 192. Asante, V.; Mortier, J.; Schlüter, H.; Koksch, B. Impact of fluorination on proteolytic stability of peptides in human blood plasma. *Bioorg. Med. Chem.* 2013, 21, 3542–3546. [CrossRef] [PubMed]
- 193. Donnelly, S.C.; MacGregor, I.; Zamani, A.; Gordon, M.W.; Robertson, C.E.; Steedman, D.J.; Little, K.; Haslett, C. Plasma elastase levels and the development of the adult respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* 1995, 151, 1428–1433. [CrossRef]
- 194. Asante, V.; Mortier, J.; Wolber, G.; Koksch, B. Impact of fluorination on proteolytic stability of peptides: A case study with α-chymotrypsin and pepsin. *Amino Acids* **2014**, *46*, 2733–2744. [CrossRef]
- 195. Kasperkiewicz, P.; Poreba, M.; Snipas, S.J.; Parker, H.; Winterbourn, C.C.; Salvesen, G.S.; Drag, M. Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. *Proc. Natl. Acad. Sci. USA* 2014, 111, 2518–2523. [CrossRef]
- 196. Lechtenberg, B.C.; Kasperkiewicz, P.; Robinson, H.; Drag, M.; Riedl, S.J. The Elastase-PK101 Structure: Mechanism of an Ultrasensitive Activity-based Probe Revealed. *ACS Chem. Biol.* **2015**, *10*, 945–951. [CrossRef] [PubMed]
- 197. Nakanishi, I.; Kinoshita, T.; Sato, A.; Tada, T. Structure of porcine pancreatic elastase complexed with FR901277, a novel macrocyclic inhibitor of elastases, at 1.6 Å resolution. *Biopolymers* **2000**, *53*, 434–445. [CrossRef]
- 198. Kasperkiewicz, P.; Poreba, M.; Snipas, S.J.; Lin, S.J.; Kirchhofer, D.; Salvesen, G.S.; Drag, M. Design of a Selective Substrate and Activity Based Probe for Human Neutrophil Serine Protease 4. *PLoS ONE* **2015**, *10*, e0132818. [CrossRef] [PubMed]
- 199. Kasperkiewicz, P.; Altman, Y.; D'Angelo, M.; Salvesen, G.S.; Drag, M. Toolbox of Fluorescent Probes for Parallel Imaging Reveals Uneven Location of Serine Proteases in Neutrophils. J. Am. Chem. Soc. 2017, 139, 10115–10125. [CrossRef]
- 200. Janiszewski, T.; Kołt, S.; Kaiserman, D.; Snipas, S.J.; Li, S.; Kulbacka, J.; Saczko, J.; Bovenschen, N.; Salvesen, G.; Drag, M.; et al. Noninvasive optical detection of granzyme B from natural killer cells with enzyme-activated fluorogenic probes. *J. Biol. Chem.* 2020, 295, 9567–9582. [CrossRef]
- 201. Vitte, J. Human mast cell tryptase in biology and medicine. Mol. Immunol. 2015, 63, 18-24. [CrossRef]
- Murakami, Y.; Takei, M.; Shindo, K.; Kitazume, C.; Tanaka, J.; Higa, T.; Fukamachi, H. Cyclotheonamide E4 and E5, New Potent Tryptase Inhibitors from an Ircinia Species of Sponge. J. Nat. Prod. 2002, 65, 259–261. [CrossRef]
- 203. Pierson, T.C.; Diamond, M.S. The continued threat of emerging flaviviruses. Nat. Microbiol. 2020, 5, 796–812. [CrossRef]
- Gouvea, I.E.; Izidoro, M.A.; Judice, W.A.S.; Cezari, M.H.S.; Caliendo, G.; Santagada, V.; dos Santos, C.N.D.; Queiroz, M.H.; Juliano, M.A.; Young, P.R.; et al. Substrate specificity of recombinant dengue 2 virus NS2B-NS3 protease: Influence of natural and unnatural basic amino acids on hydrolysis of synthetic fluorescent substrates. *Arch. Biochem. Biophys.* 2007, 457, 187–196. [CrossRef]
- Behnam, M.A.M.; Nitsche, C.; Vechi, S.M.; Klein, C.D. C-Terminal Residue Optimization and Fragment Merging: Discovery of a Potent Peptide-Hybrid Inhibitor of Dengue Protease. ACS Med. Chem. Lett. 2014, 5, 1037–1042. [CrossRef]
- Weigel, L.F.; Nitsche, C.; Graf, D.; Bartenschlager, R.; Klein, C.D. Phenylalanine and Phenylglycine Analogues as Arginine Mimetics in Dengue Protease Inhibitors. J. Med. Chem. 2015, 58, 7719–7733. [CrossRef] [PubMed]
- 207. Behnam, M.A.M.; Graf, D.; Bartenschlager, R.; Zlotos, D.P.; Klein, C.D. Discovery of Nanomolar Dengue and West Nile Virus Protease Inhibitors Containing a 4-Benzyloxyphenylglycine Residue. *J. Med. Chem.* **2015**, *58*, 9354–9370. [CrossRef] [PubMed]
- Rut, W.; Groborz, K.; Zhang, L.; Modrzycka, S.; Poreba, M.; Hilgenfeld, R.; Drag, M. Profiling of flaviviral NS2B-NS3 protease specificity provides a structural basis for the development of selective chemical tools that differentiate Dengue from Zika and West Nile viruses. *Antivir. Res.* 2020, 175, 104731. [CrossRef] [PubMed]
- 209. Rut, W.; Zhang, L.; Kasperkiewicz, P.; Poreba, M.; Hilgenfeld, R.; Drag, M. Extended substrate specificity and first potent irreversible inhibitor/activity-based probe design for Zika virus NS2B-NS3 protease. *Antivir. Res.* 2017, 139, 88–94. [CrossRef] [PubMed]
- Nitsche, C.; Onagi, H.; Quek, J.-P.; Otting, G.; Luo, D.; Huber, T. Biocompatible Macrocyclization between Cysteine and 2-Cyanopyridine Generates Stable Peptide Inhibitors. Org. Lett. 2019, 21, 4709–4712. [CrossRef]
- Wu, Z.; Yao, N.; Le, H.V.; Weber, P.C. Mechanism of autoproteolysis at the NS2–NS3 junction of the hepatitis C virus polyprotein. *Trends Biochem. Sci.* 1998, 23, 92–94. [CrossRef]
- 212. Urbani, A.; Biasiol, G.; Brunetti, M.; Volpari, C.; Di Marco, S.; Sollazzo, M.; Orrú, S.; Piaz, F.D.; Casbarra, A.; Pucci, P.; et al. Multiple Determinants Influence Complex Formation of the Hepatitis C Virus NS3 Protease Domain with Its NS4A Cofactor Peptide. *Biochemistry* 1999, 38, 5206–5215. [CrossRef]
- 213. Landro, J.A.; Raybuck, S.A.; Luong, Y.P.C.; O'Malley, E.T.; Harbeson, S.L.; Morgenstern, K.A.; Rao, G.; Livingston, D.J. Mechanistic Role of an NS4A Peptide Cofactor with the Truncated NS3 Protease of Hepatitis C Virus: Elucidation of the NS4A Stimulatory Effect via Kinetic Analysis and Inhibitor Mapping. *Biochemistry* 1997, *36*, 9340–9348. [CrossRef]

- 214. Frecer, V.r.; Kabeláč, M.; De Nardi, P.; Pricl, S.; Miertuš, S. Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus. *J. Mol. Graph. Model.* 2004, 22, 209–220. [CrossRef]
- 215. Arasappan, A.; Njoroge, F.G.; Chan, T.Y.; Bennett, F.; Bogen, S.L.; Chen, K.; Gu, H.; Hong, L.; Jao, E.; Liu, Y.T.; et al. Hepatitis C virus NS3-4A serine protease inhibitors: SAR of P2' moiety with improved potency. *Bioorg. Med. Chem. Lett.* 2005, 15, 4180–4184. [CrossRef]
- 216. Hagel, M.; Niu, D.; St Martin, T.; Sheets, M.P.; Qiao, L.; Bernard, H.; Karp, R.M.; Zhu, Z.; Labenski, M.T.; Chaturvedi, P.; et al. Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat. Chem. Biol.* 2011, 7, 22–24. [CrossRef] [PubMed]
- 217. Kataoka, H.; Ohe, T.; Takahashi, K.; Nakamura, S.; Mashino, T. Novel fullerene derivatives as dual inhibitors of Hepatitis C virus NS5B polymerase and NS3/4A protease. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 4565–4567. [CrossRef] [PubMed]
- 218. Taylor, J.G.; Zipfel, S.; Ramey, K.; Vivian, R.; Schrier, A.; Karki, K.K.; Katana, A.; Kato, D.; Kobayashi, T.; Martinez, R.; et al. Discovery of the pan-genotypic hepatitis C virus NS3/4A protease inhibitor voxilaprevir (GS-9857): A component of Vosevi[®]. *Bioorg. Med. Chem. Lett.* 2019, 29, 2428–2436. [CrossRef]
- 219. Basak, A.; Jean, F.; Seidah, N.G.; Lazure, C. Design and synthesis of novel inhibitors of prohormone convertases. *Int. J. Pept. Protein Res.* **1994**, *44*, 253–261. [CrossRef] [PubMed]
- Basak, A.; Schmidt, C.; Ismail, A.A.; Seidah, N.G.; Chrétien, M.; Lazure, C. Peptidyl substrates containing unnatural amino acid at the P'1 position are potent inhibitors of prohormone convertases. *Int. J. Pept. Protein Res.* 1995, 46, 228–237. [CrossRef] [PubMed]
- 221. Hardes, K.; Becker, G.L.; Lu, Y.; Dahms, S.O.; Köhler, S.; Beyer, W.; Sandvig, K.; Yamamoto, H.; Lindberg, I.; Walz, L.; et al. Novel Furin Inhibitors with Potent Anti-infectious Activity. *ChemMedChem* 2015, 10, 1218–1231. [CrossRef] [PubMed]
- Basak, S.; Mohottalage, D.; Basak, A. Multibranch and Pseudopeptide Approach for Design of Novel Inhibitors of Subtilisin Kexin Isozyme-1. Protein Pept. Lett. 2006, 13, 863–876. [CrossRef]
- Dwyer, M.A.; Lu, W.; Dwyer, J.J.; Kossiakoff, A.A. Biosynthetic phage display: A novel protein engineering tool combining chemical and genetic diversity. *Chem. Biol.* 2000, 7, 263–274. [CrossRef]
- Burchacka, E.; Sieńczyk, M.; Frick, I.-M.; Wysocka, M.; Lesner, A.; Oleksyszyn, J. Substrate profiling of Finegoldia magna SufA protease, inhibitor screening and application to prevent human fibrinogen degradation and bacteria growth in vitro. *Biochimie* 2014, 103, 137–143. [CrossRef]
- 225. Babkova, K.; Korabecny, J.; Soukup, O.; Nepovimova, E.; Jun, D.; Kuca, K. Prolyl oligopeptidase and its role in the organism: Attention to the most promising and clinically relevant inhibitors. *Future Med. Chem.* **2017**, *9*, 1015–1038. [CrossRef]
- 226. Portevin, B.; Benoist, A.; Rémond, G.; Hervé, Y.; Vincent, M.; Lepagnol, J.; De Nanteuil, G. New Prolyl Endopeptidase Inhibitors: In vitro and in vivo Activities of Azabicyclo[2.2.2]octane, Azabicyclo[2.2.1]heptane, and Perhydroindole Derivatives. *J. Med. Chem.* **1996**, *39*, 2379–2391. [CrossRef] [PubMed]
- 227. Racys, D.T.; Rea, D.; Fülöp, V.; Wills, M. Inhibition of prolyl oligopeptidase with a synthetic unnatural dipeptide. *Bioorg. Med. Chem.* 2010, *18*, 4775–4782. [CrossRef] [PubMed]
- 228. Miyamoto, T.; Katane, M.; Saitoh, Y.; Sekine, M.; Homma, H. Involvement of penicillin-binding proteins in the metabolism of a bacterial peptidoglycan containing a non-canonical d-amino acid. *Amino Acids* **2020**, *52*, 487–497. [CrossRef] [PubMed]
- 229. Lucana, M.C.; Arruga, Y.; Petrachi, E.; Roig, A.; Lucchi, R.; Oller-Salvia, B. Protease-Resistant Peptides for Targeting and Intracellular Delivery of Therapeutics. *Pharmaceutics* **2021**, *13*, 2065. [CrossRef]
- Massucci, M.T.; Giansanti, F.; Di Nino, G.; Turacchio, M.; Giardi, M.F.; Botti, D.; Ippoliti, R.; De Giulio, B.; Siciliano, R.A.; Donnarumma, G.; et al. Proteolytic activity of bovine lactoferrin. *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.* 2004, 17, 249–255. [CrossRef]
- 231. Kish-Trier, E.; Hill, C.P. Structural Biology of the Proteasome. Annu. Rev. Biophys. 2013, 42, 29-49. [CrossRef]
- 232. Gersch, M.; Stahl, M.; Poreba, M.; Dahmen, M.; Dziedzic, A.; Drag, M.; Sieber, S.A. Barrel-shaped ClpP Proteases Display Attenuated Cleavage Specificities. *ACS Chem. Biol.* 2016, *11*, 389–399. [CrossRef]
- Ju, Y.; He, L.; Zhou, Y.; Yang, T.; Sun, K.; Song, R.; Yang, Y.; Li, C.; Sang, Z.; Bao, R.; et al. Discovery of Novel Peptidomimetic Boronate ClpP Inhibitors with Noncanonical Enzyme Mechanism as Potent Virulence Blockers in vitro and in vivo. *J. Med. Chem.* 2020, 63, 3104–3119. [CrossRef]
- Kazmaier, U.; Junk, L. Recent Developments on the Synthesis and Bioactivity of Ilamycins/Rufomycins and Cyclomarins, Marine Cyclopeptides That Demonstrate Anti-Malaria and Anti-Tuberculosis Activity. *Mar. Drugs* 2021, 19, 446. [CrossRef]
- 235. Shin, M.; Watson, E.R.; Song, A.S.; Mindrebo, J.T.; Novick, S.J.; Griffin, P.R.; Wiseman, R.L.; Lander, G.C. Structures of the human LONP1 protease reveal regulatory steps involved in protease activation. *Nat. Commun.* **2021**, *12*, 3239. [CrossRef]
- 236. Babin, B.M.; Kasperkiewicz, P.; Janiszewski, T.; Yoo, E.; Drag, M.; Bogyo, M. Leveraging Peptide Substrate Libraries to Design Inhibitors of Bacterial Lon Protease. *ACS Chem. Biol.* **2019**, *14*, 2453–2462. [CrossRef] [PubMed]
- Thibaudeau, T.A.; Smith, D.M. A Practical Review of Proteasome Pharmacology. *Pharmacol Rev* 2019, 71, 170–197. [CrossRef]
 [PubMed]
- Gladysz, R.; Malek, N.; Rut, W.; Drag, M. Investigation of the P1' and P2' sites of IQF substrates and their selectivity toward 20S proteasome subunits. *Biol. Chem.* 2023, 404, 221–227. [CrossRef] [PubMed]
- 239. Zerfas, B.L.; Coleman, R.A.; Salazar-Chaparro, A.F.; Macatangay, N.J.; Trader, D.J. Fluorescent Probes with Unnatural Amino Acids to Monitor Proteasome Activity in Real-Time. *ACS Chem. Biol.* 2020, *15*, 2588–2596. [CrossRef] [PubMed]

- Zhang, X.; Adwal, A.; Turner, A.G.; Callen, D.F.; Abell, A.D. New Peptidomimetic Boronates for Selective Inhibition of the Chymotrypsin-like Activity of the 26S Proteasome. ACS Med. Chem. Lett. 2016, 7, 1039–1043. [CrossRef]
- 241. Geurink, P.P.; van der Linden, W.A.; Mirabella, A.C.; Gallastegui, N.; de Bruin, G.; Blom, A.E.M.; Voges, M.J.; Mock, E.D.; Florea, B.I.; van der Marel, G.A.; et al. Incorporation of non-natural amino acids improves cell permeability and potency of specific inhibitors of proteasome trypsin-like sites. *J. Med. Chem.* 2013, *56*, 1262–1275. [CrossRef]
- 242. Merz, T.; Heck, T.; Geueke, B.; Mittl, P.R.; Briand, C.; Seebach, D.; Kohler, H.-P.E.; Grütter, M.G. Autoproteolytic and Catalytic Mechanisms for the β-Aminopeptidase BapA—A Member of the Ntn Hydrolase Family. *Structure* 2012, 20, 1850–1860. [CrossRef]
- Heck, T.; Reimer, A.; Seebach, D.; Gardiner, J.; Deniau, G.; Lukaszuk, A.; Kohler, H.-P.E.; Geueke, B. β-Aminopeptidase-Catalyzed Biotransformations of β2-Dipeptides: Kinetic Resolution and Enzymatic Coupling. *ChemBioChem* 2010, 11, 1129–1136. [CrossRef]
- 244. Poreba, M.; Groborz, K.; Vizovisek, M.; Maruggi, M.; Turk, D.; Turk, B.; Powis, G.; Drag, M.; Salvesen, G.S. Fluorescent probes towards selective cathepsin B detection and visualization in cancer cells and patient samples. *Chem. Sci.* **2019**, *10*, 8461–8477. [CrossRef]
- 245. Tholen, M.; Yim, J.J.; Groborz, K.; Yoo, E.; Martin, B.A.; van den Berg, N.S.; Drag, M.; Bogyo, M. Design of Optical-Imaging Probes by Screening of Diverse Substrate Libraries Directly in Disease-Tissue Extracts. *Angew. Chem. Int. Ed. Engl.* 2020, 59, 19143–19152. [CrossRef]
- Poreba, M.; Mihelic, M.; Krai, P.; Rajkovic, J.; Krezel, A.; Pawelczak, M.; Klemba, M.; Turk, D.; Turk, B.; Latajka, R.; et al. Unnatural amino acids increase activity and specificity of synthetic substrates for human and malarial cathepsin C. *Amino Acids* 2014, 46, 931–943. [CrossRef] [PubMed]
- 247. Alves, L.C.; Melo, R.L.; Cezari, M.H.S.; Sanderson, S.J.; Mottram, J.C.; Coombs, G.H.; Juliano, L.; Juliano, M.A. Analysis of the S2 subsite specificities of the recombinant cysteine proteinases CPB of Leishmania mexicana, and cruzain of Trypanosoma cruzi, using fluorescent substrates containing non-natural basic amino acids. *Mol. Biochem. Parasitol.* 2001, 117, 137–143. [CrossRef] [PubMed]
- 248. Alves, L.C.; Melo, R.L.; Sanderson, S.J.; Mottram, J.C.; Coombs, G.H.; Caliendo, G.; Santagada, V.; Juliano, L.; Juliano, M.A. S1 subsite specificity of a recombinant cysteine proteinase, CPB, of Leishmania mexicana compared with cruzain, human cathepsin L and papain using substrates containing non-natural basic amino acids. *Eur. J. Biochem.* 2001, 268, 1206–1212. [CrossRef] [PubMed]
- Del Nery, E.; Alves, L.C.; Melo, R.L.; Cesari, M.H.; Juliano, L.; Juliano, M.A. Specificity of cathepsin B to fluorescent substrates containing benzyl side-chain-substituted amino acids at P1 subsite. *J. Protein Chem.* 2000, 19, 33–38. [CrossRef]
- 250. van der Linden, W.A.; Segal, E.; Child, M.A.; Byzia, A.; Drag, M.; Bogyo, M. Design and Synthesis of Activity-Based Probes and Inhibitors for Bleomycin Hydrolase. *Chem. Biol.* 2015, 22, 995–1001. [CrossRef]
- 251. Cuerrier, D.; Moldoveanu, T.; Campbell, R.L.; Kelly, J.; Yoruk, B.; Verhelst, S.H.L.; Greenbaum, D.; Bogyo, M.; Davies, P.L. Development of Calpain-specific Inactivators by Screening of Positional Scanning Epoxide Libraries. *J. Biol. Chem.* 2007, 282, 9600–9611. [CrossRef]
- 252. Renfrew, P.D.; Choi, E.J.; Bonneau, R.; Kuhlman, B. Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein-Peptide Interface Design. *PLoS ONE* 2012, *7*, e32637. [CrossRef]
- 253. Wang, Y.; Jia, S.; Tseng, B.; Drewe, J.; Cai, S.X. Dipeptidyl aspartyl fluoromethylketones as potent caspase inhibitors: Peptidomimetic replacement of the P2 amino acid by 2-aminoaryl acids and other non-natural amino acids. *Bioorg. Med. Chem. Lett.* 2007, 17, 6178–6182. [CrossRef]
- 254. Vickers, C.J.; González-Páez, G.E.; Wolan, D.W. Selective Detection of Caspase-3 versus Caspase-7 Using Activity-Based Probes with Key Unnatural Amino Acids. ACS Chem. Biol. 2013, 8, 1558–1566. [CrossRef]
- 255. Vickers, C.J.; González-Páez, G.E.; Litwin, K.M.; Umotoy, J.C.; Coutsias, E.A.; Wolan, D.W. Selective Inhibition of Initiator versus Executioner Caspases Using Small Peptides Containing Unnatural Amino Acids. ACS Chem. Biol. 2014, 9, 2194–2198. [CrossRef]
- 256. Poreba, M.; Kasperkiewicz, P.; Snipas, S.J.; Fasci, D.; Salvesen, G.S.; Drag, M. Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates. *Cell Death Differ.* **2014**, *21*, 1482–1492. [CrossRef] [PubMed]
- 257. Poreba, M.; Szalek, A.; Rut, W.; Kasperkiewicz, P.; Rutkowska-Wlodarczyk, I.; Snipas, S.J.; Itoh, Y.; Turk, D.; Turk, B.; Overall, C.M.; et al. Highly sensitive and adaptable fluorescence-quenched pair discloses the substrate specificity profiles in diverse protease families. *Sci. Rep.* 2017, 7, 43135. [CrossRef] [PubMed]
- Kasperkiewicz, P.; Kołt, S.; Janiszewski, T.; Groborz, K.; Poręba, M.; Snipas, S.J.; Salvesen, G.S.; Drag, M. Determination of extended substrate specificity of the MALT1 as a strategy for the design of potent substrates and activity-based probes. *Sci. Rep.* 2018, *8*, 15998. [CrossRef] [PubMed]
- 259. van de Plassche, M.A.T.; O'Neill, T.J.; Seeholzer, T.; Turk, B.; Krappmann, D.; Verhelst, S.H.L. Use of Non-Natural Amino Acids for the Design and Synthesis of a Selective, Cell-Permeable MALT1 Activity-Based Probe. J. Med. Chem. 2020, 63, 3996–4004. [CrossRef]
- Lee, J.; Bogyo, M. Synthesis and evaluation of aza-peptidyl inhibitors of the lysosomal asparaginyl endopeptidase, legumain. *Bioorg. Med. Chem. Lett.* 2012, 22, 1340–1343. [CrossRef]
- Rut, W.; Zmudzinski, M.; Snipas, S.J.; Bekes, M.; Huang, T.T.; Drag, M. Engineered unnatural ubiquitin for optimal detection of deubiquitinating enzymes. *Chem. Sci.* 2020, 11, 6058–6069. [CrossRef]
- Rajković, J.; Poreba, M.; Caglič, D.; Vidmar, R.; Wilk, A.; Borowik, A.; Salvesen, G.; Turk, V.; Drag, M.; Turk, B. Biochemical Characterization and Substrate Specificity of Autophagin-2 from the Parasite Trypanosoma cruzi. *J. Biol. Chem.* 2015, 290, 28231–28244. [CrossRef]

- 263. Breuning, A.; Degel, B.; Schulz, F.; Büchold, C.; Stempka, M.; Machon, U.; Heppner, S.; Gelhaus, C.; Leippe, M.; Leyh, M.; et al. Michael Acceptor Based Antiplasmodial and Antitrypanosomal Cysteine Protease Inhibitors with Unusual Amino Acids. *J. Med. Chem.* 2010, 53, 1951–1963. [CrossRef]
- 264. Pérez, B.C.; Teixeira, C.; Figueiras, M.; Gut, J.; Rosenthal, P.J.; Gomes, J.R.B.; Gomes, P. Novel cinnamic acid/4-aminoquinoline conjugates bearing non-proteinogenic amino acids: Towards the development of potential dual action antimalarials. *Eur. J. Med. Chem.* 2012, 54, 887–899. [CrossRef]
- 265. Joshi, R.P.; Schultz, K.J.; Wilson, J.W.; Kruel, A.; Varikoti, R.A.; Kombala, C.J.; Kneller, D.W.; Galanie, S.; Phillips, G.; Zhang, Q.; et al. AI-Accelerated Design of Targeted Covalent Inhibitors for SARS-CoV-2. J. Chem. Inf. Model. 2023, 63, 1438–1453. [CrossRef]
- Kronenberger, T.; Laufer, S.A.; Pillaiyar, T. COVID-19 therapeutics: Small-molecule drug development targeting SARS-CoV-2 main protease. *Drug Discov. Today* 2023, 28, 103579. [CrossRef] [PubMed]
- 267. Ma, Y.; Yang, K.S.; Geng, Z.Z.; Alugubelli, Y.R.; Shaabani, N.; Vatansever, E.C.; Ma, X.R.; Cho, C.-C.; Khatua, K.; Blankenship, L.; et al. The P3 O-Tert-Butyl-Threonine is Key to High Cellular and Antiviral Potency for Aldehyde-Based SARS-CoV-2 Main Protease Inhibitors. *bioRxiv* 2021, preprint. [CrossRef]
- Ma, Y.; Yang, K.S.; Geng, Z.Z.; Alugubelli, Y.R.; Shaabani, N.; Vatansever, E.C.; Ma, X.R.; Cho, C.-C.; Khatua, K.; Xiao, J.; et al. A multi-pronged evaluation of aldehyde-based tripeptidyl main protease inhibitors as SARS-CoV-2 antivirals. *Eur. J. Med. Chem.* 2022, 240, 114570. [CrossRef] [PubMed]
- Katagishi, D.; Yasuda, D.; Takahashi, K.; Nakamura, S.; Mashino, T.; Ohe, T. Fullerene derivatives as inhibitors of the SARS-CoV-2 main protease. *Bioorg. Med. Chem. Lett.* 2023, 80, 129121. [CrossRef] [PubMed]
- 270. van de Plassche, M.A.T.; Barniol-Xicota, M.; Verhelst, S.H.L. Peptidyl Acyloxymethyl Ketones as Activity-Based Probes for the Main Protease of SARS-CoV-2. *ChemBioChem Eur. J. Chem. Biol.* **2020**, *21*, 3383–3388. [CrossRef]
- Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Wang, X.; Jochmans, D.; Neyts, J.; Młynarski, W.; et al. SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* 2021, *17*, 222–228. [CrossRef]
 Color M. K. tala, T. Conse, M. Flacing, and for example imaging. *Nat. Chem. Biol.* 2021, *17*, 222–228. [CrossRef]
- 272. Goto, Y.; Katoh, T.; Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **2011**, *6*, 779–790. [CrossRef]
- 273. Miura, T.; Malla, T.R.; Owen, C.D.; Tumber, A.; Brewitz, L.; McDonough, M.A.; Salah, E.; Terasaka, N.; Katoh, T.; Lukacik, P.; et al. In vitro selection of macrocyclic peptide inhibitors containing cyclic γ2,4-amino acids targeting the SARS-CoV-2 main protease. *Nat. Chem.* 2023, *15*, 998–1005. [CrossRef]
- 274. Lorenz, I.C.; Marcotrigiano, J.; Dentzer, T.G.; Rice, C.M. Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature* **2006**, 442, 831–835. [CrossRef]
- Shaw, J.; Fishwick, C.W.G.; Harris, M. Epoxide based inhibitors of the hepatitis C virus non-structural 2 autoprotease. *Antivir. Res.* 2015, 117, 20–26. [CrossRef]
- 276. Fruton, J.S. A history of pepsin and related enzymes. Q. Rev. Biol. 2002, 77, 127–147. [CrossRef] [PubMed]
- Marciniszyn, J.; Hartsuck, J.A.; Tang, J. Mode of inhibition of acid proteases by pepstatin. J. Biol. Chem. 1976, 251, 7088–7094. [CrossRef] [PubMed]
- Fujinaga, M.; Chernaia, M.M.; Mosimann, S.C.; James, M.N.G.; Tarasova, N.I. Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* 1995, 4, 960–972. [CrossRef] [PubMed]
- 279. Kakizawa, T.; Sanjoh, A.; Kobayashi, A.; Hattori, Y.; Teruya, K.; Akaji, K. Evaluation of superior BACE1 cleavage sequences containing unnatural amino acids. *Bioorg. Med. Chem.* 2011, 19, 2785–2789. [CrossRef] [PubMed]
- Liao, Y.-F.; Wang, B.-J.; Hsu, W.-M.; Lee, H.; Liao, C.-Y.; Wu, S.-Y.; Cheng, H.-T.; Hu, M.-K. Unnatural Amino Acid-Substituted (Hydroxyethyl)urea Peptidomimetics Inhibit γ-Secretase and Promote the Neuronal Differentiation of Neuroblastoma Cells. *Mol. Pharmacol.* 2007, 71, 588–601. [CrossRef]
- 281. Iizuka, K.; Kamijo, T.; Kubota, T.; Akahane, K.; Umeyama, H.; Kiso, Y. New human renin inhibitors containing an unnatural amino acid, norstatine. *J. Med. Chem.* **1988**, *31*, 701–704. [CrossRef]
- 282. Paruszewski, R.; Jaworski, P.; Winiecka, I.; Tautt, J.; Dudkiewicz, J. New Renin Inhibitors with Pseudodipeptidic Units in P1–P1' and P2'–P3' Positions. *Chem. Pharm. Bull.* 2002, 50, 850–853. [CrossRef]
- Winiecka, I.; Jaworski, P.; Mazurek, A.P.; Marszałek, D.; Goldnik, A.; Sokulski, D. Novel renin inhibitors containing derivatives of N-alkylleucyl-β-hydroxy-γ-amino acids. J. Pept. Sci. Off. Publ. Eur. Pept. Soc. 2016, 22, 106–115. [CrossRef]
- Winiecka, I.; Marszałek, D.; Goldnik, A.; Jaworski, P.; Mazurek, A.P. Novel renin inhibitors containing a non-peptide aminoalkanoyl moiety at P1–P1' position. *Die Pharm.* 2014, 69, 263–270. [CrossRef]
- Kurt Yilmaz, N.; Swanstrom, R.; Schiffer, C.A. Improving Viral Protease Inhibitors to Counter Drug Resistance. *Trends Microbiol.* 2016, 24, 547–557. [CrossRef]
- 286. Thompson, W.J.; Ghosh, A.K.; Holloway, M.K.; Lee, H.Y.; Munson, P.M.; Schwering, J.E.; Wai, J.; Darke, P.L.; Zugay, J.; Emini, E.A.; et al. 3'-Tetrahydrofuranylglycine as a Novel, Unnatural Amino Acid Surrogate for Asparagine in the Design of Inhibitors of the HIV Protease. J. Am. Chem. Soc. 1993, 115, 801–803. [CrossRef] [PubMed]
- Ghosh, A.K.; Thompson, W.J.; Holloway, M.K.; McKee, S.P.; Duong, T.T.; Lee, H.Y.; Munson, P.M.; Smith, A.M.; Wai, J.M.; Darke, P.L.; et al. Potent HIV protease inhibitors: The development of tetrahydrofuranylglycines as novel P2-ligands and pyrazine amides as P3-ligands. J. Med. Chem. 1993, 36, 2300–2310. [CrossRef] [PubMed]

- Mimoto, T.; Kato, R.; Takaku, H.; Nojima, S.; Terashima, K.; Misawa, S.; Fukazawa, T.; Ueno, T.; Sato, H.; Shintani, M.; et al. Structure—Activity Relationship of Small-Sized HIV Protease Inhibitors Containing Allophenylnorstatine. *J. Med. Chem.* 1999, 42, 1789–1802. [CrossRef] [PubMed]
- Mimoto, T.; Hattori, N.; Takaku, H.; Kisanuki, S.; Fukazawa, T.; Terashima, K.; Kato, R.; Nojima, S.; Misawa, S.; Ueno, T.; et al. Structure-Activity Relationship of Orally Potent Tripeptide-Based HIV Protease Inhibitors Containing Hydroxymethylcarbonyl Isostere. *Chem. Pharm. Bull.* 2000, 48, 1310–1326. [CrossRef] [PubMed]
- 290. Nakatani, S.; Hidaka, K.; Ami, E.I.; Nakahara, K.; Sato, A.; Nguyen, J.-T.; Hamada, Y.; Hori, Y.; Ohnishi, N.; Nagai, A.; et al. Combination of Non-natural d-Amino Acid Derivatives and Allophenylnorstatine – Dimethylthioproline Scaffold in HIV Protease Inhibitors Have High Efficacy in Mutant HIV. J. Med. Chem. 2008, 51, 2992–3004. [CrossRef]
- Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Liu, B.; Topol, I.A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J.W. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* 1995, *3*, 581–590. [CrossRef]
- 292. Adachi, M.; Ohhara, T.; Kurihara, K.; Tamada, T.; Honjo, E.; Okazaki, N.; Arai, S.; Shoyama, Y.; Kimura, K.; Matsumura, H.; et al. Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. Proc. Natl. Acad. Sci. USA 2009, 106, 4641–4646. [CrossRef]
- Churchill, D.R.; Slade, P.M.; Youle, M.; Gazzard, B.G.; Weber, J.N. A phase I/II study of the safety and activity of a microsphere formulation of KNI-272 in patients with HIV-1 infection. *J. Antimicrob. Chemother.* 2001, 47, 353–355. [CrossRef]
- 294. Chen, M.-H.; Chang, S.-S.; Dong, B.; Yu, L.-Y.; Wu, Y.-X.; Wang, R.-Z.; Jiang, W.; Gao, Z.-P.; Si, S.-Y. Ahmpatinin iBu, a new HIV-1 protease inhibitor, from *Streptomyces* sp. CPCC 202950. *RSC Adv.* 2018, *8*, 5138–5144. [CrossRef]
- Schramm, H.J.; de Rosny, E.; Reboud-Ravaux, M.; Büttner, J.; Dick, A.; Schramm, W. Lipopeptides as dimerization inhibitors of HIV-1 protease. *Biol. Chem.* 1999, 380, 593–596. [CrossRef]
- 296. Mak, C.C.; Brik, A.; Lerner, D.L.; Elder, J.H.; Morris, G.M.; Olson, A.J.; Wong, C.-H. Design and synthesis of broad-Based monoand bi-cyclic inhibitors of FIV and HIV proteases. *Bioorg. Med. Chem.* 2003, *11*, 2025–2040. [CrossRef] [PubMed]
- 297. Strom, T.A.; Durdagi, S.; Ersoz, S.S.; Salmas, R.E.; Supuran, C.T.; Barron, A.R. Fullerene-based inhibitors of HIV-1 protease. *J. Pept. Sci.* 2015, *21*, 862–870. [CrossRef] [PubMed]
- Skwarecki, A.S.; Nowak, M.G.; Milewska, M.J. Amino Acid and Peptide-Based Antiviral Agents. *ChemMedChem* 2021, 16, 3106–3135. [CrossRef]
- 299. Kiso, Y. Design and synthesis of substrate-based peptidomimetic human immunodeficiency virus protease inhibitors containing the hydroxymethylcarbonyl isostere. *Pept. Sci.* **1996**, *40*, 235–244. [CrossRef]
- 300. Abdel-Rahman, H.M.; Kimura, T.; Hidaka, K.; Kiso, A.; Nezami, A.; Freire, E.; Hayashi, Y.; Kiso, Y. Design of inhibitors against HIV, HTLV-I, and Plasmodium falciparum aspartic proteases. *Biol. Chem.* 2004, 385, 1035–1039. [CrossRef]
- Nguyen, J.-T.; Zhang, M.; Kumada, H.-O.; Itami, A.; Nishiyama, K.; Kimura, T.; Cheng, M.; Hayashi, Y.; Kiso, Y. Truncation and non-natural amino acid substitution studies on HTLV-I protease hexapeptidic inhibitors. *Bioorg. Med. Chem. Lett.* 2008, 18, 366–370. [CrossRef]
- Hodder, A.N.; Sleebs, B.E.; Czabotar, P.E.; Gazdik, M.; Xu, Y.; O'Neill, M.T.; Lopaticki, S.; Nebl, T.; Triglia, T.; Smith, B.J.; et al. Structural basis for plasmepsin V inhibition that blocks export of malaria proteins to human erythrocytes. *Nat. Struct. Mol. Biol.* 2015, 22, 590–596. [CrossRef]
- 303. Worst, E.G.; Exner, M.P.; De Simone, A.; Schenkelberger, M.; Noireaux, V.; Budisa, N.; Ott, A. Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an *Escherichia coli* Cell-free Transcription-translation System. *J. Vis. Exp. JoVE* 2016, 114, e54273. [CrossRef]
- 304. Chang, M. Matrix metalloproteinase profiling and their roles in disease. RSC Adv. 2023, 13, 6304–6316. [CrossRef]
- 305. Leipoldt, F.; Santos-Aberturas, J.; Stegmann, D.P.; Wolf, F.; Kulik, A.; Lacret, R.; Popadić, D.; Keinhörster, D.; Kirchner, N.; Bekiesch, P.; et al. Warhead biosynthesis and the origin of structural diversity in hydroxamate metalloproteinase inhibitors. *Nat. Commun.* 2017, *8*, 1965. [CrossRef]
- Mustafa, S.; Koran, S.; Al Omair, L. Insights Into the Role of Matrix Metalloproteinases in Cancer and its Various Therapeutic Aspects: A Review. Front. Mol. Biosci. 2022, 9, 896099. [CrossRef] [PubMed]
- Song, J.; Tang, J.; Guo, F. Identification of Inhibitors of MMPS Enzymes via a Novel Computational Approach. Int. J. Biol. Sci. 2018, 14, 863–871. [CrossRef] [PubMed]
- 308. Schwartz, M.A.; Venkataraman, S.; Ghaffari, M.A.; Libby, A.; Mookhtiar, K.A.; Mallya, S.K.; Birkedal-Hansen, H.; Van Wart, H.E. Inhibition of human collagenases by sulfur-based substrate analogs. *Biochem. Biophys. Res. Commun.* 1991, 176, 173–179. [CrossRef] [PubMed]
- Ferry, G.; Boutin, J.A.; Atassi, G.; Fauchère, J.L.; Tucker, G.C. Selection of a histidine-containing inhibitor of gelatinases through deconvolution of combinatorial tetrapeptide libraries. *Mol. Divers.* 1997, 2, 135–146. [CrossRef]
- Levy, D.E.; Lapierre, F.; Liang, W.; Ye, W.; Lange, C.W.; Li, X.; Grobelny, D.; Casabonne, M.; Tyrrell, D.; Holme, K.; et al. Matrix Metalloproteinase Inhibitors: A Structure–Activity Study. J. Med. Chem. 1998, 41, 199–223. [CrossRef]
- 311. Rowsell, S.; Hawtin, P.; Minshull, C.A.; Jepson, H.; Brockbank, S.M.V.; Barratt, D.G.; Slater, A.M.; McPheat, W.L.; Waterson, D.; Henney, A.M.; et al. Crystal Structure of Human MMP9 in Complex with a Reverse Hydroxamate Inhibitor. *J. Mol. Biol.* 2002, 319, 173–181. [CrossRef]

- 312. Uttamchandani, M.; Wang, J.; Li, J.; Hu, M.; Sun, H.; Chen, K.Y.T.; Liu, K.; Yao, S.Q. Inhibitor Fingerprinting of Matrix Metalloproteases Using a Combinatorial Peptide Hydroxamate Library. *J. Am. Chem. Soc.* **2007**, *129*, 7848–7858. [CrossRef]
- Hayun, H.; Coban, M.; Bhagat, A.K.; Ozer, E.; Alfonta, L.; Caulfield, T.R.; Radisky, E.S.; Papo, N. Utilizing genetic code expansion to modify N-TIMP2 specificity towards MMP-2, MMP-9, and MMP-14. *Sci. Rep.* 2023, 13, 5186. [CrossRef]
- Tranchant, I.; Vera, L.; Czarny, B.; Amoura, M.; Cassar, E.; Beau, F.; Stura, E.A.; Dive, V. Halogen Bonding Controls Selectivity of FRET Substrate Probes for MMP-9. *Chem. Biol.* 2014, 21, 408–413. [CrossRef]
- 315. Takeuchi, T.; Hayashi, M.; Tamita, T.; Nomura, Y.; Kojima, N.; Mitani, A.; Takeda, T.; Hitaka, K.; Kato, Y.; Kamitani, M.; et al. Discovery of Aryloxyphenyl–Heptapeptide Hybrids as Potent and Selective Matrix Metalloproteinase-2 Inhibitors for the Treatment of Idiopathic Pulmonary Fibrosis. *J. Med. Chem.* 2022, 65, 8493–8510. [CrossRef]
- Mucha, A.; Cuniasse, P.; Kannan, R.; Beau, F.; Yiotakis, A.; Basset, P.; Dive, V. Membrane Type-1 Matrix Metalloprotease and Stromelysin-3 Cleave More Efficiently Synthetic Substrates Containing Unusual Amino Acids in Their P1' Positions. *J. Biol. Chem.* 1998, 273, 2763–2768. [CrossRef] [PubMed]
- 317. Holtz, B.; Cuniasse, P.; Boulay, A.; Kannan, R.; Mucha, A.; Beau, F.; Basset, P.; Dive, V. Role of the S1' Subsite Glutamine 215 in Activity and Specificity of Stromelysin-3 by Site-Directed Mutagenesis. *Biochemistry* 1999, 38, 12174–12179. [CrossRef] [PubMed]
- Meyer, B.S.; Rademann, J. Extra- and Intracellular Imaging of Human Matrix Metalloprotease 11 (hMMP-11) with a Cellpenetrating FRET Substrate. J. Biol. Chem. 2012, 287, 37857–37867. [CrossRef] [PubMed]
- Gall, F.M.; Hohl, D.; Frasson, D.; Wermelinger, T.; Mittl, P.R.E.; Sievers, M.; Riedl, R. Drug Design Inspired by Nature: Crystallographic Detection of an Auto-Tailored Protease Inhibitor Template. *Angew. Chem. Int. Ed. Engl.* 2019, 58, 4051–4055. [CrossRef] [PubMed]
- 320. Cowan, J.D.; Moss, L.M.; Andrews, C.R.; Musso, L.D.; McDougald, L.D.; Becherer, D.J.; Wiethe, W.R.; Rabinowitz, H.M.; Petrov, G.K.; Gaul, D.M.; et al. Substrate Specificity and Novel Selective Inhibitors of TNF-alpha Converting Enzyme (TACE) from Two-Dimensional Substrate Mapping. *Comb. Chem. High Throughput Screen.* 2005, *8*, 327–339. [CrossRef]
- Mallari, J.P.; Choy, C.J.; Hu, Y.; Martinez, A.R.; Hosaka, M.; Toriyabe, Y.; Maung, J.; Blecha, J.E.; Pavkovic, S.F.; Berkman, C.E. Stereoselective inhibition of glutamate carboxypeptidase by organophosphorus derivatives of glutamic acid. *Bioorg. Med. Chem.* 2004, 12, 6011–6020. [CrossRef]
- 322. Turner, A.J. Chapter 79—Aminopeptidase N. In *Handbook of Proteolytic Enzymes*, 3rd ed.; Rawlings, N.D., Salvesen, G., Eds.; Academic Press: Cambridge, MA, USA, 2013; pp. 397–403.
- Drag, M.; Bogyo, M.; Ellman, J.A.; Salvesen, G.S. Aminopeptidase Fingerprints, an Integrated Approach for Identification of Good Substrates and Optimal Inhibitors. J. Biol. Chem. 2010, 285, 3310–3318. [CrossRef]
- 324. Kiehstaller, S.; Ottmann, C.; Hennig, S. MMP activation–associated aminopeptidase N reveals a bivalent 14-3-3 binding motif. *J. Biol. Chem.* **2020**, 295, 18266–18275. [CrossRef]
- 325. Węglarz-Tomczak, E.; Poręba, M.; Byzia, A.; Berlicki, Ł.; Nocek, B.; Mulligan, R.; Joachimiak, A.; Drąg, M.; Mucha, A. An integrated approach to the ligand binding specificity of Neisseria meningitidis M1 alanine aminopeptidase by fluorogenic substrate profiling, inhibitory studies and molecular modeling. *Biochimie* 2013, 95, 419–428. [CrossRef]
- 326. Gajda, A.D.; Pawełczak, M.; Drag, M. Substrate specificity screening of oat (*Avena sativa*) seeds aminopeptidase demonstrate unusually broad tolerance in S1 pocket. *Plant Physiol. Biochem.* **2012**, *54*, 6–9. [CrossRef]
- 327. Velmourougane, G.; Harbut, M.B.; Dalal, S.; McGowan, S.; Oellig, C.A.; Meinhardt, N.; Whisstock, J.C.; Klemba, M.; Greenbaum, D.C. Synthesis of New (–)-Bestatin-Based Inhibitor Libraries Reveals a Novel Binding Mode in the S1 Pocket of the Essential Malaria M1 Metalloaminopeptidase. *J. Med. Chem.* 2011, 54, 1655–1666. [CrossRef] [PubMed]
- 328. Poreba, M.; McGowan, S.; Skinner-Adams, T.S.; Trenholme, K.R.; Gardiner, D.L.; Whisstock, J.C.; To, J.; Salvesen, G.S.; Dalton, J.P.; Drag, M. Fingerprinting the Substrate Specificity of M1 and M17 Aminopeptidases of Human Malaria, Plasmodium falciparum. PLoS ONE 2012, 7, e31938. [CrossRef] [PubMed]
- 329. Poreba, M.; Gajda, A.; Picha, J.; Jiracek, J.; Marschner, A.; Klein, C.D.; Salvesen, G.S.; Drag, M. S1 pocket fingerprints of human and bacterial methionine aminopeptidases determined using fluorogenic libraries of substrates and phosphorus based inhibitors. *Biochimie* **2012**, *94*, 704–710. [CrossRef] [PubMed]
- Byzia, A.; Szeffler, A.; Kalinowski, L.; Drag, M. Activity profiling of aminopeptidases in cell lysates using a fluorogenic substrate library. *Biochimie* 2016, 122, 31–37. [CrossRef] [PubMed]
- 331. Babaie, F.; Hosseinzadeh, R.; Ebrazeh, M.; Seyfizadeh, N.; Aslani, S.; Salimi, S.; Hemmatzadeh, M.; Azizi, G.; Jadidi-Niaragh, F.; Mohammadi, H. The roles of ERAP1 and ERAP2 in autoimmunity and cancer immunity: New insights and perspective. *Mol. Immunol.* 2020, 121, 7–19. [CrossRef]
- 332. Weglarz-Tomczak, E.; Vassiliou, S.; Mucha, A. Discovery of potent and selective inhibitors of human aminopeptidases ERAP1 and ERAP2 by screening libraries of phosphorus-containing amino acid and dipeptide analogues. *Bioorg. Med. Chem. Lett.* **2016**, 26, 4122–4126. [CrossRef]
- Fink, C.A.; Qiao, Y.; Berry, C.J.; Sakane, Y.; Ghai, R.D.; Trapani, A.J. New alpha-thiol dipeptide dual inhibitors of angiotensin-I converting enzyme and neutral endopeptidase EC 3.4.24.11. J. Med. Chem. 1995, 38, 5023–5030. [CrossRef]
- 334. Fink, C.A.; Carlson, J.E.; McTaggart, P.A.; Qiao, Y.; Webb, R.; Chatelain, R.; Jeng, A.Y.; Trapani, A.J. Mercaptoacyl Dipeptides as Orally Active Dual Inhibitors of Angiotensin-Converting Enzyme and Neutral Endopeptidase. J. Med. Chem. 1996, 39, 3158–3168. [CrossRef]

- 335. Vincent, M.; Marchand, B.; Rémond, G.; Jaguelin-Guinamant, S.; Damien, G.; Portevin, B.; Baumal, J.Y.; Volland, J.P.; Bouchet, J.P.; Lambert, P.H. Synthesis and ACE inhibitory activity of the stereoisomers of perindopril (S 9490) and perindoprilate (S 9780). Drug Des. Discov. 1992, 9, 11–28. [PubMed]
- 336. Wester, A.; Devocelle, M.; Tallant, E.A.; Chappell, M.C.; Gallagher, P.E.; Paradisi, F. Stabilization of Angiotensin-(1-7) by key substitution with a cyclic non-natural amino acid. *Amino Acids* **2017**, *49*, 1733–1742. [CrossRef]
- Dunn, F.W.; Dittmer, K. Action of carboxypeptidase toward peptides containing unnatural aromatic amino acids. *Science* 1950, 111, 173–175. [CrossRef] [PubMed]
- Dunn, F.W.; Smith, E.L. Action of Carboxypeptidase on Derivatives of Unnatural Amino Acids. J. Biol. Chem. 1950, 187, 385–392.
 [CrossRef] [PubMed]
- 339. Moriwaki, K.; Ogishima, T.; Ito, A. Analysis of Recognition Elements for Mitochondrial Processing Peptidase Using Artificial Amino Acids: Roles of the Intervening Portion and Proximal Arginine1. *J. Biochem.* **1999**, *126*, 874–878. [CrossRef]
- Martin, S.E.; Ganguly, T.; Munske, G.R.; Fulton, M.D.; Hopkins, M.R.; Berkman, C.E.; Black, M.E. Development of Inhibitor-Directed Enzyme Prodrug Therapy (IDEPT) for Prostate Cancer. *Bioconjug. Chem.* 2014, 25, 1752–1760. [CrossRef]
- 341. Nigam, P.K.; Nigam, A. Botulinum toxin. Indian J. Dermatol. 2010, 55, 8–14. [CrossRef] [PubMed]
- 342. Wang, D.; Krilich, J.; Baudys, J.; Barr, J.R.; Kalb, S.R. Enhanced detection of type C botulinum neurotoxin by the Endopep-MS assay through optimization of peptide substrates. *Bioorg. Med. Chem.* **2015**, *23*, 3667–3673. [CrossRef] [PubMed]
- Wang, D.; Baudys, J.; Hoyt, K.; Barr, J.R.; Kalb, S.R. Sensitive detection of type G botulinum neurotoxin through Endopep-MS peptide substrate optimization. *Anal. Bioanal. Chem.* 2019, 411, 5489–5497. [CrossRef] [PubMed]
- 344. Dadová, J.; Wu, K.-J.; Isenegger, P.G.; Errey, J.C.; Bernardes, G.J.L.; Chalker, J.M.; Raich, L.; Rovira, C.; Davis, B.G. Precise Probing of Residue Roles by Post-Translational β,γ-C,N Aza-Michael Mutagenesis in Enzyme Active Sites. ACS Cent. Sci. 2017, 3, 1168–1173. [CrossRef]
- 345. Osamura, T.; Okuda, M.; Yamaguchi, A.; Ohtake, K.; Sakamoto, K.; Takimura, Y. Variants of the industrially relevant protease KP-43 with suppressed activity under alkaline conditions developed using expanded genetic codes. *Biochem. Biophys. Rep.* 2019, 17, 93–96. [CrossRef]
- Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A. Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *ChemBioChem* 2017, 18, 1442–1447. [CrossRef]
- 347. Short, G.F.; Golovine, S.Y.; Hecht, S.M. Effects of Release Factor 1 on in vitro Protein Translation and the Elaboration of Proteins Containing Unnatural Amino Acids. *Biochemistry* **1999**, *38*, 8808–8819. [CrossRef]
- 348. Short, G.F.; Laikhter, A.L.; Lodder, M.; Shayo, Y.; Arslan, T.; Hecht, S.M. Probing the S1/S1' Substrate Binding Pocket Geometry of HIV-1 Protease with Modified Aspartic Acid Analogues. *Biochemistry* 2000, 39, 8768–8781. [CrossRef]
- 349. Ugwumba, I.N.; Ozawa, K.; de la Cruz, L.; Xu, Z.-Q.; Herlt, A.J.; Hadler, K.S.; Coppin, C.; Brown, S.E.; Schenk, G.; Oakeshott, J.G.; et al. Using a Genetically Encoded Fluorescent Amino Acid as a Site-Specific Probe to Detect Binding of Low-Molecular-Weight Compounds. ASSAY Drug Dev. Technol. 2010, 9, 50–57. [CrossRef]
- Walasek, P.; Honek, J.F. Nonnatural amino acid incorporation into the methionine 214 position of the metzincin Pseudomonas aeruginosa alkaline protease. BMC Biochem. 2005, 6, 21. [CrossRef] [PubMed]
- Ge, Y.; Fan, X.; Chen, P.R. A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. *Chem. Sci.* 2016, 7, 7055–7060. [CrossRef] [PubMed]
- 352. Brown, W.; Wesalo, J.; Tsang, M.; Deiters, A. Engineering Small Molecule Switches of Protein Function in Zebrafish Embryos. J. Am. Chem. Soc. 2023, 145, 2395–2403. [CrossRef] [PubMed]
- 353. Tang, S.; Beattie, A.T.; Kafkova, L.; Petris, G.; Huguenin-Dezot, N.; Fiedler, M.; Freeman, M.; Chin, J.W. Mechanism-based traps enable protease and hydrolase substrate discovery. *Nature* 2022, 602, 701–707. [CrossRef]
- 354. Fu, X.; Chang, Z. Biogenesis, quality control, and structural dynamics of proteins as explored in living cells via site-directed photocrosslinking. *Protein Sci.* 2019, *28*, 1194–1209. [CrossRef]
- 355. Sinz, A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: From single proteins to protein interaction networks. *Expert Rev. Proteom.* **2014**, *11*, 733–743. [CrossRef]
- 356. Aydin, Y.; Coin, I. Genetically encoded crosslinkers to address protein–protein interactions. Protein Sci. 2023, 32, e4637. [CrossRef]
- 357. Yi, D.; Xing, J.; Gao, Y.; Pan, X.; Xie, P.; Yang, J.; Wang, Q.; Gao, X. Enhancement of keratin-degradation ability of the keratinase KerBL from Bacillus licheniformis WHU by proximity-triggered chemical crosslinking. *Int. J. Biol. Macromol.* 2020, 163, 1458–1470. [CrossRef]
- 358. Cheng, Y.; Wu, J.; Han, Y.; Xu, J.; Da, Y.; Zhao, Q.; Guo, G.; Zhou, Y.; Chen, Y.; Liu, J.; et al. A CDR-based approach to generate covalent inhibitory antibody for human rhinovirus protease. *Bioorg. Med. Chem.* **2021**, *42*, 116219. [CrossRef] [PubMed]
- 359. Ge, X.; Wang, R.; Ma, J.; Liu, Y.; Ezemaduka, A.N.; Chen, P.R.; Fu, X.; Chang, Z. DegP primarily functions as a protease for the biogenesis of β-barrel outer membrane proteins in the Gram-negative bacterium *Escherichia coli*. *FEBS J.* 2014, 281, 1226–1240. [CrossRef] [PubMed]
- Fu, X.; Wang, Y.; Shao, H.; Ma, J.; Song, X.; Zhang, M.; Chang, Z. DegP functions as a critical protease for bacterial acid resistance. *FEBS J.* 2018, 285, 3525–3538. [CrossRef] [PubMed]
- Fu, X.; Wang, Y.; Song, X.; Shi, X.; Shao, H.; Liu, Y.; Zhang, M.; Chang, Z. Subunit interactions as mediated by "non-interface" residues in living cells for multiple homo-oligomeric proteins. *Biochem. Biophys. Res. Commun.* 2019, 512, 100–105. [CrossRef]

- 362. Liebsch, F.; Aurousseau, M.R.P.; Bethge, T.; McGuire, H.; Scolari, S.; Herrmann, A.; Blunck, R.; Bowie, D.; Multhaup, G. Full-length cellular β-secretase has a trimeric subunit stoichiometry, and its sulfur-rich transmembrane interaction site modulates cytosolic copper compartmentalization. *J. Biol. Chem.* 2017, 292, 13258–13270. [CrossRef]
- Nguyen, D.P.; Lusic, H.; Neumann, H.; Kapadnis, P.B.; Deiters, A.; Chin, J.W. Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-TRNA Synthetase/TRNACUA Pair and Click Chemistry. J. Am. Chem. Soc. 2009, 131, 8720. [CrossRef]
- 364. Kawamoto, S.A.; Coleska, A.; Ran, X.; Yi, H.; Yang, C.-Y.; Wang, S. Design of Triazole-Stapled BCL9 α-Helical Peptides to Target the β-Catenin/B-Cell CLL/lymphoma 9 (BCL9) Protein–Protein Interaction. J. Med. Chem. 2012, 55, 1137–1146. [CrossRef]
- 365. Schneider, T.; Schneider, D.; Rösner, D.; Malhotra, S.; Mortensen, F.; Mayer, T.U.; Scheffner, M.; Marx, A. Dissecting Ubiquitin Signaling with Linkage-Defined and Protease Resistant Ubiquitin Chains. *Angew. Chem. Int. Ed. Engl.* 2014, 53, 12925–12929. [CrossRef]
- 366. Abdelkader, E.H.; Feintuch, A.; Yao, X.; Adams, L.A.; Aurelio, L.; Graham, B.; Goldfarb, D.; Otting, G. Protein conformation by EPR spectroscopy using gadolinium tags clicked to genetically encoded p-azido-l-phenylalanine. *Chem. Commun.* 2015, 51, 15898–15901. [CrossRef]
- 367. Mahawaththa, M.C.; Lee, M.D.; Giannoulis, A.; Adams, L.A.; Feintuch, A.; Swarbrick, J.D.; Graham, B.; Nitsche, C.; Goldfarb, D.; Otting, G. Small neutral Gd(iii) tags for distance measurements in proteins by double electron–electron resonance experiments. *Phys. Chem. Chem. Phys.* 2018, 20, 23535–23545. [CrossRef] [PubMed]
- 368. Penna, E.; Cerciello, A.; Chambery, A.; Russo, R.; Cernilogar, F.M.; Pedone, E.M.; Perrone-Capano, C.; Cappello, S.; Di Giaimo, R.; Crispino, M. Cystatin B Involvement in Synapse Physiology of Rodent Brains and Human Cerebral Organoids. *Front. Mol. Neurosci.* 2019, 12, 195. [CrossRef] [PubMed]
- 369. Bard, J.A.M.; Martin, A. Recombinant Expression, Unnatural Amino Acid Incorporation, and Site-Specific Labeling of 26S Proteasomal Subcomplexes BT. In *The Ubiquitin Proteasome System: Methods and Protocols*; Mayor, T., Kleiger, G., Eds.; Springer: New York, NY, USA, 2018; pp. 219–236.
- 370. Bard, J.A.M.; Bashore, C.; Dong, K.C.; Martin, A. The 26S Proteasome Utilizes a Kinetic Gateway to Prioritize Substrate Degradation. *Cell* **2019**, 177, 286–298.e215. [CrossRef] [PubMed]
- Regev, O.; Linder, H.; Gur, E. Pup-Click—A New Chemoenzymatic Method for the Generation of Singly Pupylated Targets. Bioconjug. Chem. 2019, 30, 2909–2916. [CrossRef] [PubMed]
- 372. Exner, M.P.; Kuenzl, T.; To, T.M.T.; Ouyang, Z.; Schwagerus, S.; Hoesl, M.G.; Hackenberger, C.P.R.; Lensen, M.C.; Panke, S.; Budisa, N. Design of S-Allylcysteine in Situ Production and Incorporation Based on a Novel Pyrrolysyl-tRNA Synthetase Variant. *ChemBioChem* 2017, 18, 85–90. [CrossRef] [PubMed]
- 373. Koch, N.G.; Goettig, P.; Rappsilber, J.; Budisa, N. "Cold" Orthogonal Translation: Psychrophilic Pyrrolysyl-tRNA Synthetase as Efficient Tool for Expanding the Genetic Code. *bioRxiv* 2023, preprint. [CrossRef]
- 374. Turk, N.; Raza, A.; Wuytens, P.; Demol, H.; Daele, M.V.; Detavernier, C.; Skirtach, A.; Gevaert, K.; Baets, R. Waveguide-based surface-enhanced Raman spectroscopy detection of protease activity using non-natural aromatic amino acids. *Biomed. Opt. Express* 2020, 11, 4800–4816. [CrossRef]
- 375. Hostetler, Z.M.; Cory, M.B.; Jones, C.M.; Petersson, E.J.; Kohli, R.M. The Kinetic and Molecular Basis for the Interaction of LexA and Activated RecA Revealed by a Fluorescent Amino Acid Probe. ACS Chem. Biol. 2020, 15, 1127–1133. [CrossRef]
- 376. Poreba, M. Protease-activated prodrugs: Strategies, challenges, and future directions. FEBS J. 2020, 287, 1936–1969. [CrossRef]
- 377. Garner, P.; Sherry, B.; Moilanen, S.; Huang, Y. In vitro stability of α-Helical peptide nucleic acids (αPNAs). *Bioorg. Med. Chem. Lett.* 2001, *11*, 2315–2317. [CrossRef]
- 378. Radchenko, T.; Brink, A.; Siegrist, Y.; Kochansky, C.; Bateman, A.; Fontaine, F.; Morettoni, L.; Zamora, I. Software-aided approach to investigate peptide structure and metabolic susceptibility of amide bonds in peptide drugs based on high resolution mass spectrometry. *PLoS ONE* 2017, *12*, e0186461. [CrossRef] [PubMed]
- 379. Radchenko, T.; Fontaine, F.; Morettoni, L.; Zamora, I. WebMetabase: Cleavage sites analysis tool for natural and unnatural substrates from diverse data source. *Bioinformatics* **2019**, *35*, 650–655. [CrossRef] [PubMed]
- Radchenko, T.; Fontaine, F.; Morettoni, L.; Zamora, I. Software-aided workflow for predicting protease-specific cleavage sites using physicochemical properties of the natural and unnatural amino acids in peptide-based drug discovery. *PLoS ONE* 2019, 14, e0199270. [CrossRef] [PubMed]
- 381. Courtney, T.; Deiters, A. Recent advances in the optical control of protein function through genetic code expansion. *Curr. Opin. Chem. Biol.* **2018**, *46*, 99–107. [CrossRef] [PubMed]
- 382. Fok, J.A.; Mayer, C. Genetic-Code-Expansion Strategies for Vaccine Development. ChemBioChem 2020, 21, 3291–3300. [CrossRef] [PubMed]
- Rezhdo, A.; Islam, M.; Huang, M.; Van Deventer, J.A. Future prospects for noncanonical amino acids in biological therapeutics. *Curr. Opin. Biotechnol.* 2019, 60, 168–178. [CrossRef]
- Zhu, C.; Xu, L.; Chen, L.; Zhang, Z.; Zhang, Y.; Wu, W.; Li, C.; Liu, S.; Xiang, S.; Dai, S.; et al. Epitope-Directed Antibody Elicitation by Genetically Encoded Chemical Cross-Linking Reactivity in the Antigen. ACS Cent. Sci. 2023, 9, 1229–1240. [CrossRef]
- 385. Benner, S.A. Expanding the genetic lexicon: Incorporating non-standard amino acids into proteins by ribosome-based synthesis. *Trends Biotechnol.* **1994**, *12*, 158–163. [CrossRef]

- 386. Chen, Y.; He, X.; Ma, B.; Liu, K.; Gao, T.; Niu, W.; Guo, J. Noncanonical amino acid mutagenesis in response to recoding signal-enhanced quadruplet codons. *Nucleic Acids Res.* 2022, 50, e94. [CrossRef]
- 387. Anderson, J.C.; Magliery, T.J.; Schultz, P.G. Exploring the limits of codon and anticodon size. *Chem. Biol.* 2002, *9*, 237–244. [CrossRef]
- Hartman, M.C.T.; Josephson, K.; Lin, C.-W.; Szostak, J.W. An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides. PLoS ONE 2007, 2, e972. [CrossRef] [PubMed]
- Kaiser, F.; Krautwurst, S.; Salentin, S.; Haupt, V.J.; Leberecht, C.; Bittrich, S.; Labudde, D.; Schroeder, M. The structural basis of the genetic code: Amino acid recognition by aminoacyl-tRNA synthetases. *Sci. Rep.* 2020, 10, 12647. [CrossRef]
- Hartman, M.C.T.; Josephson, K.; Szostak, J.W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. USA* 2006, 103, 4356–4361. [CrossRef] [PubMed]
- Krishnakumar, R.; Prat, L.; Aerni, H.R.; Ling, J.; Merryman, C.; Glass, J.I.; Rinehart, J.; Söll, D. Transfer RNA misidentification scrambles sense codon recoding. *Chembiochem* 2013, 14, 1967–1972. [CrossRef]
- 392. Eremeeva, E.; Herdewijn, P. Non canonical genetic material. Curr Opin Biotechnol 2019, 57, 25–33. [CrossRef] [PubMed]
- Völler, J.-S.; Budisa, N. Coupling genetic code expansion and metabolic engineering for synthetic cells. *Curr. Opin. Biotechnol.* 2017, 48, 1–7. [CrossRef]
- 394. Agostini, F.; Sinn, L.; Petras, D.; Schipp, C.J.; Kubyshkin, V.; Berger, A.A.; Dorrestein, P.C.; Rappsilber, J.; Budisa, N.; Koksch, B. Multiomics Analysis Provides Insight into the Laboratory Evolution of *Escherichia coli* toward the Metabolic Usage of Fluorinated Indoles. ACS Cent Sci 2021, 7, 81–92. [CrossRef]
- 395. Dragosits, M.; Mattanovich, D. Adaptive laboratory evolution—Principles and applications for biotechnology. *Microb. Cell Factories* **2013**, *12*, 64. [CrossRef] [PubMed]
- 396. Hoesl, M.G.; Oehm, S.; Durkin, P.; Darmon, E.; Peil, L.; Aerni, H.-R.; Rappsilber, J.; Rinehart, J.; Leach, D.; Söll, D.; et al. Chemical Evolution of a Bacterial Proteome. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 10030–10034. [CrossRef]
- 397. Tolle, I.; Oehm, S.; Hoesl, M.G.; Treiber-Kleinke, C.; Peil, L.; Bozukova, M.; Albers, S.; Bukari, A.-R.A.; Semmler, T.; Rappsilber, J.; et al. Evolving a mitigation of the stress response pathway to change the basic chemistry of life. *Front. Synth. Biol.* **2021**, *1*, 1248065.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.