

Article

The Repository of Amino Acids and Their Modifications—A Tool for Enlarging the Space of Food-Derived and Other Peptides in the BIOPEP-UWM Database

Anna Iwaniak, Piotr Minkiewicz * and Małgorzata Darewicz

Department of Food Biochemistry, University of Warmia and Mazury in Olsztyn, Plac Cieszyński 1, 10-726 Olsztyn, Poland; ami@uwm.edu.pl (A.I.); darewicz@uwm.edu.pl (M.D.)

* Correspondence: minkiew@uwm.edu.pl

Abstract

Peptides are the most extensively studied bioactive compounds derived from food. They are analyzed using, e.g., in silico strategy. The BIOPEP-UWM database has become a standard tool in computer-aided peptide research. The aim of this study was to equip this database with a tool enabling the annotation and processing of peptide or protein sequences containing modified amino acid residues as well as other residues. The most recent section of the BIOPEP-UWM, i.e., the repository of amino acids and modifications, apart from 20 proteinogenic amino acids, annotates non-amino acid moieties or residues subjected to enzymatic or chemical modifications (phosphorylation, oxidation, hydroxylation, acylation, etc.). This part of BIOPEP-UWM provides the following information: Compound ID, name, symbol in a special code, InChIKey identifier, SMILES representation, number in the PubChem database (CID), formula, and IDs in other databases. The search options include ID in the repository, name, symbol in a biological code, InChIKey, PubChem Compound identifier (CID), and chemical formula. Peptide or protein sequences annotated using symbols from the repository are utilized by all applications available in the BIOPEP-UWM database. The database meets contemporary trends involving modifications of amino acid residues in the bioinformatic analysis of peptides and proteins.

Keywords: amino acids; bioactive peptides; chemical modifications; CHUCKLES format; database; food peptides; SMILES code

Academic Editor: Burkhard
Poeggeler

Received: 17 June 2026
Revised: 16 July 2026
Accepted: 23 July 2026
Published: 27 July 2026

Copyright: © 2026 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/).

1. Introduction

Contemporary peptide science is interested not only in peptides containing canonical (proteinogenic) amino acids, but also in compounds containing non-canonical or modified amino acids as well as other residues belonging to many classes of organic compounds [1–4]. The inclusion of monomers other than canonical amino acids expands the accessible peptide sequence space [5]. To date, peptides with non-canonical or modified amino acids have received less attention in food science than in pharmacological or medical research. Most known food-derived peptides consist exclusively of proteinogenic amino acids. Nevertheless, food proteins and peptides undergo thermally induced chemical modifications, examples of which were described by Brown et al. [6]. Food products contain peptides with chemical composition not restricted to canonical

amino acids. As an example, casein hydrolysates contain, among others, phosphopeptides, e.g., mineral-binding peptides [7,8]. In turn, taste-affecting peptides may contain monomers other than canonical amino acids, such as those described by Ueda et al. [9] or Hashizume et al. [10]. Another example includes the antioxidative peptides described by Vitale et al. [11].

Peptide science employs various *in silico* methods to support experimental research [3,4,12–17]. Most software currently used for computer-aided research on peptides relies primarily on amino acid sequences. The usefulness of cheminformatic tools within this approach, particularly for studies with peptides containing non-canonical or modified amino acids, has been pointed out by Minkiewicz et al. [13]. Table 1 provides examples of bioinformatics and cheminformatics tools suitable for research with peptides whose composition is not restricted to canonical amino acids only.

Table 1. List of databases, programs, and websites cited in this article (accessed 6 July 2026).

Name	Website	Ref.
ADMETlab 3.0	https://admetlab3.scbdd.com/	[18]
BIOPEP-UWM	https://biochemia.uwm.edu.pl/en/biopep-uwm-2/	[19]
CAS	https://commonchemistry.cas.org/	[20]
CFM-ID	https://cfmid.wishartlab.com/	[21]
ChEMBL	https://www.ebi.ac.uk/chembl/db/	[22]
ChemSpider	https://www.chemspider.com/	[23]
CycPeptMPDB	http://cycpeptmpdb.com/	[24]
DFBP	http://www.cqudfbp.net/	[25]
HMDB	https://www.hmdb.ca/	[26]
IIIT, Raghava's group	https://webs.iiitd.edu.in/raghava/webservices.html	
iNclusive	https://non-canonical-aas.biologie.uni-freiburg.de/	[27]
LIPID MAPS	https://www.lipidmaps.org/	[28]
NCBI Protein	https://www.ncbi.nlm.nih.gov/protein/	[29]
Norine	https://bioinfo.cristal.univ-lille.fr/norine/	[30]
Open Babel	https://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html	[31]
pepADMET	https://pepadmet.ddai.tech	[32]
PepFoundry	https://github.com/BilodeauGroup/PepFoundry	[33]
PepINVENT	https://www-cohsoftware.ch.cam.ac.uk/; https://github.com/MolecularAI/PepINVENT/	[34]
PepSMI	https://www.novoprolabs.com/tools/convert-peptide-to-smiles-string	
PepTherDia	https://peptherdia.herokuapp.com/	[35]
Peptipedia	https://app.peptipedia.cl/	[36]
Protein Line Notation (PLN) Guide	http://www.biochemfusion.com/doc/PLN_Guide/PLN_Guide.html	
PubChem	https://pubchem.ncbi.nlm.nih.gov/	[37]
SATPdb	http://crdd.osdd.net/raghava/satpdb/; https://webs.iiitd.edu.in/raghava/vaxinpad/index1.html	[38]
SwissDock	https://www.swissdock.ch/	[39]
SwissLipids	https://www.swisslipids.org/#/	[40]
SwissSideChain	https://www.swissidechain.ch/	[41]
SwissTargetPrediction	https://www.swisstargetprediction.ch/	[42]
UniProt	https://www.uniprot.org/	[43]

Several databases contain peptides composed not only of canonical amino acids. Examples include SATPdb [38], Norine [30], PepTherDia [35], CycPeptMPDB [24], and

Peptipedia [36] databases. The DFBP database contains γ -glutamyl peptides [25]. Canonical and non-canonical amino acids are cataloged in specialized databases, such as SwissSideChain [41] or iNclusive [27]. They are also annotated in general chemical compound databases focused on low-molecular-mass compounds, including PubChem [37], Chemical Abstract Service (CAS) [20], ChemSpider [23], or ChEMBL [22].

Publicly available datasets of monomers (amino acids and other components of peptides), associated with peptide databases are valuable resources for users. As peptide databases continue to expand, increasing demands are placed on their search functionalities. Finding short peptides (especially dipeptides) within databases containing more than thousands of compounds requires the possibility to search for exact sequence matches, i.e., peptides of identical length and sequence. Otherwise, the query results in all peptides containing dipeptide as a fragment (including from dozens to even thousands of irrelevant sequences). The application of the above option to peptides containing monomers beyond canonical (proteinogenic) amino acids requires a univocal code for such amino acids. Their symbols should be readily available to users before initiating the search. Otherwise, users must first identify a peptide containing monomers other than a canonical amino acid to determine the symbol required to query the database. Databases such as Norine [30] and CycPeptMPDB [24], or the PepFoundry program [33], which provide separate datasets of monomers, represent models of best practice and served as reference databases in the present study.

The annotation of peptides containing non-canonical amino acids requires using symbols that go beyond amino acid sequences. Codes such as HELM [44–46], BILN [47], MAP [48], Protein Line Notation (PLN), and our own proposal [49] represent attempts to introduce annotation schemes capable of describing peptides using symbols beyond those corresponding to canonical amino acids. Chemical codes, based on explicit atom-level representations, such as SMILES [50], serve as a universal language for annotating the full diversity of chemical compounds, including peptides. The above-mentioned notation is currently the most extensively used chemical language [51,52]. Among the tools summarized in Table 1, the SMILES code is utilized by databases such as the Chemical Abstracts Service (CAS) [20], Norine [30], PubChem [37], ChemSpider [23], and ChEMBL [22] as well as by programs such as ADMETLab 3.0 [18], CFM-ID [21], SwissTargetPrediction [42], SwissDock [39], and PepADMET [32]. SMILES representations are also recommended as an input for machine-learning models involving peptides that contain non-canonical and modified amino acids [53] and may be converted into many other chemical formats using programs such as OpenBabel [31].

The BIOPEP-UWM database is one of the standard tools in research on food-derived peptides [19]. User expectations regarding its further development were summarized in a review article published by Mamoudou and Mune Mune [16], who highlighted, among other priorities, the expansion of the range of annotated bioactivities and the inclusion of peptides containing non-canonical amino acids. Although the most extensively studied bioactivities, such as the inhibition of angiotensin-converting enzymes (EC 3.4.15.1), dipeptidyl peptidase IV (EC 3.4.14.5) and antioxidative activity, continue to dominate in the BIOPEP-UWM database, the number of annotated activities has been steadily increasing. In 2019, peptides annotated in the BIOPEP-UWM database of bioactive peptides represented 42 activity categories [54], whereas more recently (1 October 2025) this number has risen to 96. In contrast, some recently proposed classification systems include fewer activity categories than those currently presented in the BIOPEP-UWM [55]. Several less extensively studied peptide bioactivities are represented both by short peptides, containing 2 or 3 canonical amino acid residues (likely to be found in food protein sequences) and compounds containing other monomers. Examples include inhibitors of glutamate carboxypeptidase II (EC 3.4.17.21) [56] or tripeptidyl peptidase II

(EC 3.4.14.10) [57]. Another expectation identified by Mamoudou and Mune Mune [16] was the expansion of the BIOPEP-UWM database through the inclusion of peptides containing non-canonical and modified amino acids. However, extending the set of monomers in BIOPEP-UWM to include non-canonical or modified amino acids and other residues requires the implementation of a biological code compatible with workflows involving peptide and protein sequences [19,54]. In addition, the structural features and chirality of each monomer should be described as comprehensively as possible, in accordance with recommendations of López-López and coworkers [3]. Another essential element is the extension of the existing application for translating peptide sequences into the SMILES code [54] by enabling the conversion of a set of monomers beyond canonical amino acids and their enantiomers. The fourth requirement is making monomer data directly accessible to users, as described by Flissi et al. [30], Li et al. [24], and Garzon-Otero et al. [33]. Finally, the repository should facilitate straightforward incorporation of newly described monomers as they become available. The BIOPEP-UWM repository of amino acids and modifications was launched in 2023 as an attempt to address the above requirements and has been announced in our earlier publications [19,58], but is not described in detail to date.

The aim of this article is to provide a detailed description of the BIOPEP-UWM repository of amino acids and modifications and its potential for research in the area of peptide science.

2. Construction of the Repository

Data concerning the structure of monomers were retrieved from the literature or from existing databases, including PubChem in particular. The following information was manually mined from the above resources: compound name and structural representations. The biological code used for the annotation of compounds is an extension of our previous proposal [49]. Rules of the code, illustrated by examples of monomers, are available on the BIOPEP-UWM repository of amino acids and modifications page, via the Tab entitled “Annotation of the amino acids and their modifications” (see Figures 1 and 2). This tab provides access to a downloadable pdf file. This option allowing the publication of the above rules enables their update and extension if necessary. Briefly, canonical amino acids and their D-enantiomers are annotated using a one-letter code, while other amino acids are annotated using symbols containing more characters. Acyl and alkyl residues are described using an extension of the code designed for the description of fatty acids in lipid databases [28,40]. Abbreviations of names are used to denote monosaccharide residues.

The application for translating amino acid sequences into the SMILES code is based on the CHUCKLES format, described by Siani et al. [59] and recommended in other publications [33,34,54,60,61]. Initially, this format has been designed to enable the construction of SMILES representations of peptides. The workflow for the arrangement of amino acid representations in the above format is well established and has been described by Siani et al. [59] and later by Minkiewicz et al. [61]. SMILES strings of amino acids bound via the α -peptide bonds were arranged using a typical order: amine group, asymmetric carbon atom, side chain, and carboxyl group. A general order of symbols in the CHUCKLES format is as follows: representation of a functional group 1, representation of monomer core structure, and representation of a functional group 2 ended by oxygen symbol. Monomer core in the case of α -amino acids includes an asymmetric carbon atom and a side chain. Functional group 1, in compounds annotated to date in the BIOPEP-UWM repository of amino acids and modifications, comprises $-\text{NH}_2$, $-\text{NH}-$, $-\text{OH}$, or $-\text{SH}$. A monomer core is characteristic for individual compounds. Functional group 2 comprises $-\text{COOH}$, $-\text{OH}$ or $-\text{CSOH}$ to date. The notation including the terms functional

group 1 and functional group 2 is an extension of the proposal of Garzon Otero et al. [33]. Functional group 1 or 2 in residues that are N- or C-terminal modifications, respectively, can be omitted. Examples of the SMILES representations of monomers are presented in the Supplementary Materials (Part S1). SMILES representations of canonical amino acids (20 compounds) and their D-enantiomers (19 compounds) were publicly available in the (unfortunately recently withdrawn) source code of the CycloPs program [60] and in the SwissSideChain database [37]. All other SMILES representations of amino acids and other monomers required rearrangement. The order of representation of groups order is changed if amino acid residues form other kinds of bonds (see Supplementary Materials Parts S1 and S2). SMILES representations of N-terminal acyl groups possess carboxyl groups on the right. Representations of alkyl or aryl groups at the N-terminus are finished by the symbol of the hydroxyl group, whereas at the C-terminus they start from a hydroxyl or amine group. In monosaccharide residues at the N-terminus, the symbol of the hydroxyl group bound to an anomeric carbon atom is on the right, whereas in residues that are parts of modified amino acid residues this is on the left. We used the Kekule version of SMILES for annotation of aromatic compounds, as recommended by Hähnke et al. [62] and applied it to the PubChem database. SMILES representations of amino acids and other monomers were constructed or adapted to the CHUCKLES version manually and verified as recommended in our earlier works [49,61]. The molecule editor MarvinSketch 17.28 (ChemAxon, Budapest, Hungary) served to convert SMILES representations into the InChI code [63] and InChIKey. Two correct SMILES representations of the same molecule are converted to the same InChIKey [63], indicating that the adaptation to the CHUCKLES format is appropriate. Another criterion of the correctness of SMILES representations was for the identity of InChIKeys to match with those available in the databases (PubChem and others). The molecule editor also served to check the correctness of the structure and configuration of chirality centers in the modified amino acid residues and other compounds. There are two possibilities: the order of substituents according to the priority rules of Cahn et al. [64] may be changed due to the modification or may remain the same. A change in substituents' order leads to a change in the absolute configuration (from Rectus to Sinister or from Sinister to Rectus). Finally, if peptides containing residues beyond canonical amino acids possess correct SMILES representations and hence InChIKeys, it is possible to find them in a databases (e.g., PubChem). Examples of the manual validation of SMILES representations of monomers are presented in the Supplementary Materials (Part S2). To date, we have only annotated the BIOPEP-UWM amino acids without any known errors identified during curation. We have not found errors in SMILES peptide representations. This does not imply that the BIOPEP-UWM database is generally free of errors, but manual curation allows them to be corrected as quickly as possible after detection. This is achievable for a relatively small dataset. Manual curation is not enough in the case of datasets containing at least thousands of compounds, such as those used for design of the PepINVENT program [34].

The algorithm for calculating molecular masses of particular monomers is currently outlined as follows. Molecular mass increments of residues built into peptide sequences are calculated based on their formulas, being part of the InChI representations [62]. Atomic masses were introduced according to Danikiewicz [65]. Monoisotopic masses involving the most abundant isotopes and chemical (average) masses accounting for isotope distribution were used. The molecular mass increment of a residue was calculated using Equation (1), whereas the molecular mass of a peptide was calculated using Equation (2).

$$MI = \sum nM_e - M_{H_2O} \quad (1)$$

where

M_e —atomic mass of a given element;

n —number of atoms of a given element in a molecule;

M_{H_2O} —molecular mass of water.

$$M_{pept} = \sum MI + M_{H_2O} \quad (2)$$

where

MI—mass increment of a particular residue.

Proteins [current number: 773]	Please cite the following paper if you are using BIOPEP-UWM database : Minkiewicz P, Iwaniak A, Darewicz M. BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. International Journal of Molecular Sciences. 2019; 20(23):5978. doi: 10.3390/ijms20235978
Bioactive peptides [current number: 5787]	
Allergenic proteins with their epitopes [current number: 136]	Please cite the following paper if you are using BIOPEP-UWM database of allergenic proteins and their epitopes: Dziuba M., Minkiewicz P., Dąbek M., Peptides, specific proteolysis products as molecular markers of allergenic proteins – in silico studies. Acta Scientiarum Polonorum Technologia Alimentaria, 12, 2013, 101-112
Sensory peptides and amino acids [current number: 614]	Please cite the following paper if you are using BIOPEP-UWM database of sensory peptides and amino acids: Iwaniak A., Minkiewicz P., Darewicz M., Sieniawski K., Starowicz P., 2016. BIOPEP database of sensory peptides and amino acids. Food Research International, 85, 155-161, doi: 10.1016/j.foodres.2016.04.031
BIOPEP-UWM Virtual [current number: 704]	Our new database. Please cite the following paper if you are using BIOPEP-UWM Virtual database : Minkiewicz P, Iwaniak A, Darewicz M., 2022. BIOPEP-UWM Virtual—A Novel Database of Food-Derived Peptides with In Silico-Predicted Biological Activity. Applied Sciences, 12(14):7204, doi: 10.3390/app12147204.
1 BIOPEP-UWM repository of amino acids and modifications [current number: 230]	BIOPEP-UWM repository of amino acids and modifications.
Submit new peptide sequence	Here you can send us the peptide sequence that is currently not present in our database[s]. Thank you :)

Number of visits from 2024.01.01 - 105915 [Open BIOPEP-UWM in new window](#) Last update - 2026-07-06

Figure 1. Screenshot of a homepage of the BIOPEP-UWM database (accessed 6 July 2026). (1) Tab of the repository of amino acids and modifications providing information concerning amino acids and other monomers that are building blocks of peptides.

The term “mass increment” is understood as the mass of a monomer residue being part of a peptide or a protein chain. The difference between mass increment and molecular mass corresponds to the removal of a water molecule. This mass defect corresponds to the formation of a peptide bond. This method of calculation implies the use of compounds containing hydroxyl groups (alcohols or phenols) as precursors for N-terminal amine group modifications although it does not correspond to the actual chemical reactions. The mass defect corresponding to the removal of a water molecule is not always enough to annotate cyclic or branched peptides. An example of the calculation of peptide molecular mass using Equations (1) and (2) is provided in the Supplementary File, Part S3.

The BIOPEP-UWM database has been constructed using the WordPress 5.7 content management system, whereas accompanying applications were written in the PHP language, as stated previously [66].

BIOPEP-UWM: Repository of amino acids and modifications

1

Annotation of amino acids and their modifications

Number of amino acids in database: 230

	ID	Symbol	Name	Chem. mass	Monois. mass
Amino acid / modification data	239	{ph!}	Phenol as a precursor of N-terminal phenyl group	76.0958	76.0312
Amino acid / modification data	240	{C4:0!}	Butan-1-ol as a precursor of N-terminal butyl group	56.1060	56.0624
Amino acid / modification data	241	{A[3!cC6:0]}	Cyclohexylalanine	153.2209	153.1150
Amino acid / modification data	242	{Y[10!C1;11!ph]}	O-benzyltyrosine	253.2952	253.1099
Amino acid / modification data	243	{~!C2:0}	Aminoethane as a part of C-terminal amide group	27.0682	27.0472
Amino acid / modification data	244	{~!C3:0[3ph]}	3-Phenylpropan-1-amine as a part of C-terminal amide group	117.1905	117.0940
Amino acid / modification data	245	{A[3!βnph]}	3-(2-Naphthyl)-L-alanine	197.2321	197.0838
Amino acid / modification data	246	{Aad}	L-2-Aminoadipic acid	143.1402	143.0580
Amino acid / modification data	247	{C2:0[2SH]}	Thioglycolic acid – N-terminal modification	74.1027	73.9826
Amino acid / modification data	248	{H3PO4}	Phosphoric acid	79.9799	79.9663
Amino acid / modification data	249	{F[7NO2]}	p-Nitrophenylalanine	192.1711	192.0533
Amino acid / modification data	250	{A[3!(4)tz]}	3-(4-thiazolyl)-L-alanine	154.1905	154.0200
Amino acid / modification data	251	{A[3!(4)itz]}	3-(4-isothiazolyl)-L-alanine	154.1905	154.0200
Amino acid / modification data	252	{(S)βS}	(S)β-Serine	87.0771	87.0319
Amino acid / modification data	253	{(R)βS}	(R)β-Serine	87.0771	87.0319

First Previous Page: OK Next Last
Page 13 / 16

2

Search: by: ID GO ☐ exact
 ID (e.g. 2572)
 Symbol
 Name (e.g. lactorphin)
 InChIKey
 PubChem CID
 Compound formula

Figure 2. Screenshot of a fragment of the list of compounds in the BIOPEP-UWM database of amino acids and modifications. (1) The link to a .pdf file entitled “Annotation of amino acids and modifications”, containing a detailed description of the rules of the biological code used in the database and a translation of this code into SMILES. (2) Search engine for amino acids and other monomers. A more detailed description of search options is presented in Table 2, and examples of search results are provided in Part S4 of the Supplementary Materials (accessed 6 July 2026).

Table 2. Search options in the BIOPEP-UWM repository of amino acids and modifications (accessed 6 July 2026). Examples of search results are presented in the Supplementary Materials, Part S4.

Search Option	Comment
ID	ID of the compound may be found on the page showing the peptide or protein containing residues other than unmodified, canonical amino acids (tab: additional information). Layout of peptide information is described in the Minkiewicz et al. article [54].
Symbol	Compound symbol may be found on the page showing the peptide or protein containing residues other than unmodified, canonical amino acids (tab: additional information).
Name	Name of the compound may be found on the page showing the peptide or protein containing residues other than unmodified, canonical amino acids (tab: additional information).

InChIKey	Univocal representation of chemical compounds that are suitable as a query in database searching [62].
PubChem CID	Compound ID in the PubChem database [37]; the biggest database of chemical compounds, commonly used as a reference database for other cheminformatics and bioinformatics tools.
Compound formula	Compound formula written as follows: carbon, hydrogen and other elements in order of increasing atomic numbers (e.g., C ₃ H ₈ NO ₅ PS for compound with ID 264). Formula written using the above rule, e.g., part of InChI representation of compound [62] and is used in major chemical databases, such as PubChem [37] or ChemSpider [23].

3. Overview of the Repository

The location of the BIOPEP-UWM repository of amino acids and modifications on the website is presented in Figure 1. A fragment of the list of compounds annotated in the repository is presented in Figure 2. The layout of this list is consistent with the standard format used throughout the BIOPEP-UWM database [54,66]. The list includes links to the pages of individual compounds, the ID in the repository, symbols in the biological code, as well as the compound names and molecular masses.

The BIOPEP-UWM repository of amino acids and modifications contains 20 canonical amino acids and their 19 D-enantiomers, bound via α -peptide bond. It comprises not only amino acid residues but also building blocks representing N- or C-terminal modifications, e.g., phenol as a precursor of the N-terminal phenyl group (ID 239) or butanol as a precursor of the N-terminal butyl group (ID 240) (see Figure 2).

Monomers belonging to classes other than canonical amino acids and their enantiomers are summarized in Table 3. The first class comprises amino acids bound via bonds other than α -peptide. They are represented in the repository by separate entries reflecting their different modes of attachment. For example, glutamic acid is annotated as an amino acid residue bound via the α -carboxyl group (ID 68) or via the γ -carboxyl group (ID 77). Although these two annotations differ in their symbols and SMILES representations, which are adapted to the CHUCKLES format [59], all other information associated with IDs 68 and 77 is identical. The framework enabling the introduction of peptides containing peptide bonds other than those formed between α -carboxyl and α -amine groups also addressed expectations expressed by BIOPEP-UWM users [67]. The remaining monomers are classified according to their chemical structure rather than their origin (natural or synthetic). Modified canonical amino acids and other (non-canonical) amino acids are the largest classes of compounds in the repository. Compounds beyond amino acids (N- and C-terminal modifications) have previously been included in the list of monomers of the Norine database [30], but their SMILES representations are not arranged in the CHUCKLES format in this database. Examples of rearranged SMILES strings of N-terminal acyl groups were included in the compound dataset described by Orsi and Reymond [5]. The inclusion of monomers, such as alkyl, aryl or monosaccharide residues, with SMILES representations, arranged in the CHUCKLES format represents a distinctive strength of the BIOPEP-UWM repository.

Table 3. Number of monomers belonging to particular classes beyond canonical amino acids and their enantiomers (version from 6 July 2026).

Class of Monomers Beyond Canonical Amino Acids and Their Enantiomers, Bound via α -	Number of Monomers
Peptide Bonds	in the Class
Amino acids bound via bonds other than α -peptide	6
Chemical or enzymatic modifications of canonical amino acids	101
Other (non-canonical) amino acids	48
Acyl residues (N-terminal)	8

Alkyl residues (N-terminal or C-terminal)	10
Aryl residues (N-terminal or C-terminal)	5
Monosaccharide residues (N-terminal)	7
Inorganic acids	2
Other residues	4

The tab entitled “Annotation of amino acids and modifications” provides a link to a file, containing the rules of the BIOPEP-UWM code, mentioned in Section 2. One important change has been introduced compared with the initial version [49], used in our previous publications [58,68]: the characters indicating the start and the end of multi-character symbols are now “{” and “}” instead of “<” and “>”, respectively. This change is justified by the fact that the symbols “<” and “>” have other important functions in informatics and are thus not generally used for encoding biopolymers (except for Protein Line Notation). In contrast, symbols “{” and “}” are typically used for biopolymer annotation; their application is relatively easy and thus are desired from the perspective of cooperation between biochemists and informaticians [69]. The BIOPEP-UWM code may serve not only to encode peptides, but also to annotate other compounds, even without amino acid residues.

A summary of external IDs of compounds is presented in Table 4. Most compounds (205 from 230) are annotated in other databases, especially in PubChem. Only one compound is present in ChemSpider and not present in PubChem. This fact is the rationale for the choice of PubChem CID (Compound ID) as a query for search in the repository.

Table 4. Number of external IDs in the BIOPEP-UWM repository of amino acids and modifications (data from 6 July 2026).

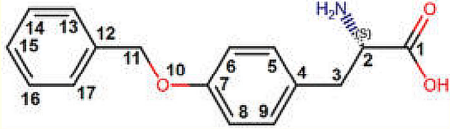
Database	Number of IDs
CAS	168
ChEMBL	86
ChemSpider	186
HMDB	76
Norine	55
PubChem	204
SwissSideChain	55
No IDs in the above databases	26

An example of information about an individual compound annotated in the BIOPEP-UWM repository of amino acids and modifications is presented in Figure 3 and Table 5, based on O-benzyltyrosine (ID 242), which is an amino acid residue occurring in a synthetic peptide that inhibits tripeptidyl peptidase II—EC 3.4.14.10 [51] (peptide ID in the BIOPEP-UWM database of bioactive peptides—10804).

Information available on the website includes compound ID, name, symbol in the biological code, InChIKey, SMILES representation, arranged according to CHUCKLES format rules [59], as well as Compound ID (CID) in the PubChem database and molecular formula. If a given compound is not annotated in PubChem, “0” is displayed instead of the CID. Any of the above information, except SMILES, can be used as a query in the repository search engine (see Table 2). Compound information includes residue mass increments, calculated according to Equation (1) (see Section 2), and the compound’s structure with continuous atom numbering, also used in symbols. The continuous numbering system is explained in our previous publication [49] and was designed to overcome the limitations of traditional atom numbering, which allows for the assigning

of the same number to a few atoms in the same molecule. Continuous atom numbering enables the unambiguous localization of any modification in the amino acid residue and facilitates machine-readability of amino acid representations compared with traditional conventions. It is not the first proposal of atom numbering other than the traditional approach, i.e., that recommended by the International Union of Pure and Applied Chemistry. Unambiguous atom numbering has been proposed in InChI [62] and ReactionCode [70] representations. Our approach seems to be more intuitive than the alternatives, which may enhance human-readability of the code, although it may be less universally applicable than the numbering systems used in InChI and ReactionCode.

BIOPEP-UWM: Amino acids Data

ID	242	Name	O-benzyltyrosine
Symbol	{Y[10!C1;11!ph]}		
InChIKey	KAFHLONDOVSENM-HNNXBMFYSA-N		
SMILES	N[C@@H](CC1=CC=C(C=C1)OCC1=CC=C(C=C1))C(=O)O		
PubChem CID	725701		
Compound formula	C16H17NO3		
Chemical Mass	253.2952	Monoisotopic Mass	253.1099
Structure			

[reference](#)

[additional information](#)

[database references](#)

[screen and print BIOPEP-UWM dataset of amino acids and modifications](#)

[Browse entire list of compounds](#)

Figure 3. Screenshot of data of O-benzyltyrosine (ID 242; “Structure” tab) as an example of a non-canonical amino acid (accessed 6 July 2026).

Table 5. Information about O-benzyltyrosine (ID 242) as an example of a non-canonical amino acid (see Figure 3) (data from 6 July 2026).

ID	242
Name	O-Benzyltyrosine
Symbol	{Y[10!C1;11!ph]}
InChIKey	KAFHLONDOVSENM-HNNXBMFYSA-N
SMILES	N[C@@H](CC1=CC=C(C=C1)OCC1=CC=C(C=C1))C(=O)O
PubChem CID	725701
Compound formula	C16H17NO3
Residue chemical mass increment (Da)	253.2952
Residue monoisotopic mass increment (Da)	253.1099
Additional information	
SMILES:	
N[C@@H](CC1=CC=C(C=C1)OCC1=CC=C(C=C1))C(=O)O	
InChI=1S/C16H17NO3/c17-15(16(18)19)10-12-6-8-14(9-7-12)/20-11-13-4-2-1-3-5-13/h1-9,15H,10-11,17H2,(H,18,19)/t15-	

/m0/s1

InChIKey=KAFHLONDOVSENM-HNNXBMFYSA-N

Database reference:

CAS: Registry No 16652-64-5

ChEMBL: ID ChEMBL147608

ChemSpider: ID 633643

PubChem: CID 725701

Table 5 provides complete information about a compound, available via the tab entitled “Screen and print BIOPEP-UWM dataset of amino acids and modifications”. Apart from the information mentioned in Figure 3, it includes the content of the tab “Additional information” with compound representations in chemical codes and other information if necessary. The “Database reference” tab contains IDs of a given compound in other databases (if available). Examples of peptides annotated using multi-character symbols from the BIOPEP-UWM repository of amino acids and modifications are presented in Table 6, whereas a summary of applications of the symbols from the repository of amino acids and modifications in other parts of the BIOPEP-UWM database can be found in Table 7. Information about non-canonical amino acids or other residues present in a given peptide, such as their names and IDs in the repository of amino acids and modifications (right column in Table 6), is available via the “Additional information” tab.

Table 6. Examples of peptides from the BIOPEP-UWM database of bioactive peptides containing residues other than canonical amino acids, bound via α -peptide bonds (accessed 6 July 2026).

ID of Peptide ¹	Sequence	Residues Other Than Canonical Amino Acids Bound via α -Peptide Bonds (Symbols, Names, IDs in the Repository of Amino Acids and Modifications and Monomer Classes According to Table 3)
10390	{C3:1(2E)[3ph;6OH;7OH]}AH~residue)	{C3:1(2E)[3ph;6OH;7OH]}—caffeic acid; ID 107 (monomer class: acyl ~—C-terminal amide group, ID 75 (monomer class: other residues)
10432	{C1}MLF	{C1}—Formic acid—N-terminal modification; ID 157 (monomer class: acyl residue)
10582	{GABA}{Thea}P	{GABA}— γ -Aminobutyric acid; ID 196 (monomer class: non-canonical amino acid) {Thea}—Theanine; ID 208 (monomer class: non-canonical amino acid)
10765	{C4:0!}{H2CO3}V{Nle}	{C4:0!}—N-terminal butyl group; ID 240 (monomer class: alkyl residue) {H2CO3}—Carbonic acid; ID 228 (monomer class: inorganic acid) {Nle}—Norleucine; ID 231 (monomer class: non-canonical amino acid)
10862	{E;5}E	{E;5}—Glutamic acid bound via γ -carboxyl group; ID 77 (monomer class: amino acids bound via bonds other than α -peptide)

¹ ID in the BIOPEP-UWM database of bioactive peptides.

Table 7. Number of items annotated in the BIOPEP-UWM using symbols beyond canonical amino acids and their enantiomers (accessed 6 July 2026).

Class of Compounds	Total Number of Compounds Within the Class	Number of Compounds Containing Monomers Other Than Canonical Amino Acids and Their D-Enantiomers
Proteins	773	34
Bioactive peptides	5787	142
Allergenic proteins	136	0

Sensory peptides and amino acids	614	18
Virtually bioactive peptides ¹	704	0

¹ Peptides with bioactivity predicted in silico (e.g., using molecular docking), but not verified experimentally. Such peptides are annotated in the BIOPEP-UWM Virtual database [66].

The role of the repository of amino acids and modifications in the BIOPEP-UWM database is shown in Figure 4. Symbols from the repository are available for use across all databases within the BIOPEP-UWM. Symbols of monomers, listed in the repository, may serve to construct query sequences for sequence-based searches. Search options, listed in our previous publications [19,54], utilize sequences annotated using symbols of non-canonical amino acids or non-amino acid residues. Search results that include symbols of monomers other than canonical amino acids can be exported to Microsoft Excel files. In the case of profiles showing the potential biological activity of a peptide or protein fragments, the results can be sorted by ID, activity or sequence prior to export. Incorporating these symbols into the application system associated with the BIOPEP-UWM is the major strength of the repository of amino acids and modifications. However, matching modified peptides with protein sequences, annotated using exclusively canonical amino acids, such as those retrieved from the UniProt database, remains a challenge. As pointed out previously [49], this would require the conversion of modified peptides into their corresponding unmodified sequences. Nevertheless, the use of symbols other than canonical amino acids is essential for an exact search option (see Figure 2).

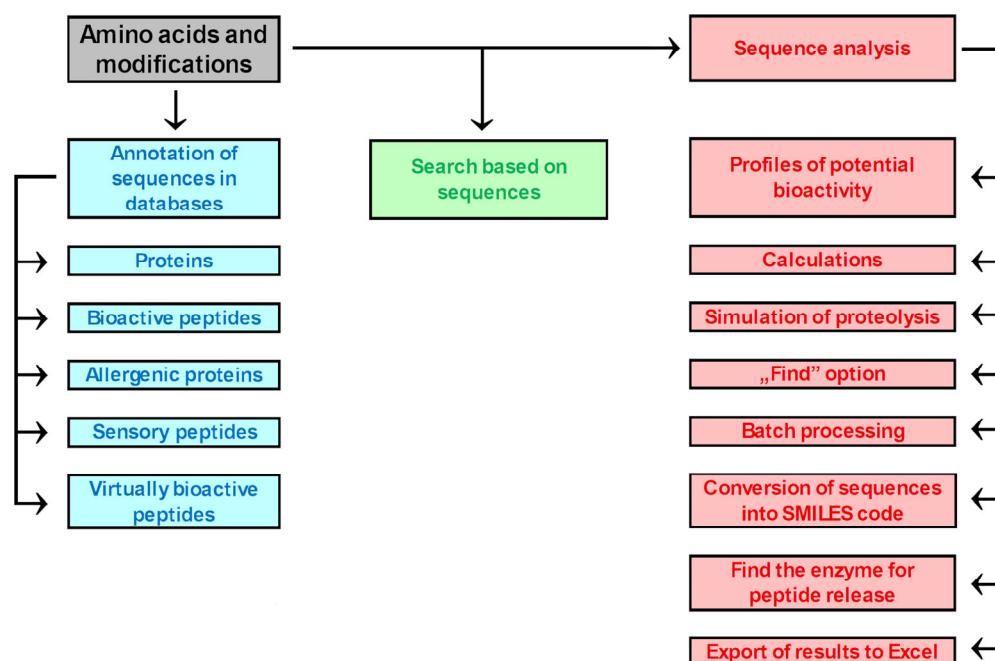


Figure 4. Application of symbols from the section “Repository of amino acids and modifications” in the BIOPEP-UWM database and associated programs (version from 6 July 2026). Examples of use of particular applications are presented in the Supplementary Materials Part S5.

Comparison of BILN, HELM, MAP, and PLN codes and versions of SMILES (CHUCKLES format) used in existing datasets is presented in Table 8.

Table 8. Comparison of encoding monomers using a biological code and a CHUCKLES format of the SMILES code used in the BIOPEP-UWM and other biological codes and versions of the CHUCKLES format (version from 6 July 2026).

Code or Version of CHUCKLES Format	Comparison with the Code or CHUCKLES Used in the BIOPEP-UWM
BILN	BILN code offers the possibility to annotate linear, cyclic and branched peptides containing diverse set of monomers, i.e., offers more options than the BIOPEP-UWM code. BILN allows sequence alignments.
HELM	HELM code is to date more universal than the BIOPEP-UWM code, but not compatible with one-letter notation of peptide and protein sequences written using symbols of canonical amino acids. It is more complicated than the BIOPEP-UWM code in the case of sequence matching and proteolysis simulation.
MAP	Format is extension of conventional sequences written in a one-letter code. In contrast to the BIOPEP-UWM code, it offers the possibility to annotate cyclic peptides with a diverse set of monomers. MAP enables the annotation of additional information together with a sequence.
PLN	In contrast to the BIOPEP-UWM code, PLN enables the annotation of cyclic peptides containing a broad array of monomers (not only amino acids).
CHUCKLES format in the PepFoundry dataset	CHUCKLES format version from PepFoundry is more easily readable by humans although more complicated for informaticians than that used in the BIOPEP-UWM. It fosters potential to annotate monomers beyond amino acids.
CHUCKLES format in the dataset of Orsi and Reymond [5]	Format proposed by Orsi and Reymond is simple compared with the classical CHUCKLES format used in the BIOPEP-UWM, but less universal. It is unable to annotate bonds with chirality centers, such as glycosidic bonds (see Supplementary Materials Part S6).

The above-mentioned codes offer an opportunity to annotate branched or cyclic peptides, not available to date in the BIOPEP-UWM code. In turn, the latter is applied for sequence matching, conversion into SMILES and proteolysis simulation, i.e., for work in applications available in the BIOPEP-UWM database. The above options are more important for the BIOPEP-UWM users (predominantly food scientists) than the possibility of describing branched and cyclic amino acids.

All options of the BIOPEP-UWM were initially designed for work with sequences containing canonical amino acids only [54]. Taking the above into account, we need a code as similar as possible to the conventional single letter code used in the UniProt and peptide databases such as DFBP, Peptipedia, SATPdb and many others. Using sequences containing amino acid symbols separated by dashes or dots (in BILN and HELM, respectively) in options presented in Figure 4 and Part S5 of the Supplementary Materials would require conversion to traditional amino acid sequences and hence cause additional complications in all programs associated with the BIOPEP-UWM database.

Another limitation of the BIOPEP-UWM code is the fact that its rules are restricted to the molecules present in the repository of amino acids and modifications. Inclusion of new classes of compounds may require adding new rules.

BILN and HELM codes are applied in programs that serve to predict physicochemical properties of peptides and features important from the drug discovery standpoint [4,71–74]. The HELM code is also used in databases, such as CycPepDB [24] and PubChem [37]. A comparison of the MAP format with HELM and BILN has been discussed by Shendre and coworkers [48]. We can expect broader applications of these codes in the future.

4. Examples of the Compounds Annotated in the Repository

The biological activity of peptides containing modified amino acid residues can be simulated or predicted *in silico*, even if there is no protocol for the synthesis or

identification of particular compounds. The structures from the BIOPEP-UWM database may be used for computational studies, such as molecular docking [12], to account for all details of the structure and the chirality of amino acids and other residues that are fragments of linear peptides. In the case of the examples mentioned below, this includes all isomers or sugar moieties or chirality of sulfur atoms in sulfoxides. Results from such computational studies, involving these and other compounds as peptide building blocks, can be annotated in the BIOPEP-UWM Virtual database [66].

Two examples illustrating the potential of the BIOPEP-UWM repository of amino acids and modifications in the context of recommendations for annotating complete information about compound structure and chirality [3], are presented in Figures 5 and 6.

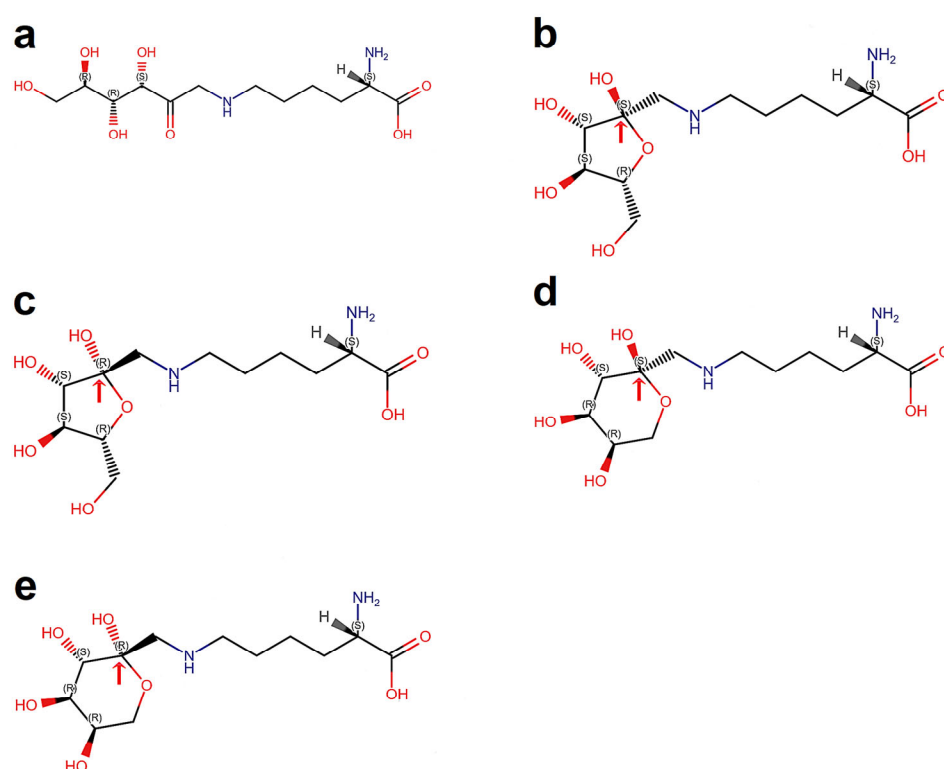


Figure 5. Structures of products of lysine glycation, annotated in the BIOPEP-UWM repository of amino acids and modifications: (a) N(6)-D-fructosyl-L-lysine (ID 145); (b) N(6)-α-D-fructofuranosyl-L-lysine (ID 146); (c) N(6)-β-D-fructofuranosyl-L-lysine (ID 147); (d) N(6)-α-D-fructopyranosyl-L-lysine (ID 270); and (e) N(6)-β-D-fructopyranosyl-L-lysine (ID 271). Symbols (R) and (S) indicate configuration around asymmetric atoms “Rectus” and “Sinister”, respectively. Arrows indicate anomeric carbon atoms.

Figure 5 shows the structures of five isomers of D-fructosyl-L-lysine, a product of non-enzymatic glycation. According to current knowledge [75,76], reaction products of hexoses with compounds containing amine groups may exist as α- and β-anomers of pyranose or furanose rings, as well as in linear form. Pyranoses are the most stable and hence the most abundant isomers of glycation products; however, other isomers can also occur in significant amounts. Bearing this in mind, five isomers of D-fructosyl-L-lysine have been included in the repository section. Other glycated amino acids of interest, such as those described by Kijewska et al. [77], can also be annotated. To the best of our knowledge, the influence of the type of sugar ring and the configuration around the anomeric carbon atom on the biological activity of peptides containing glycated lysine has not yet been investigated.

Figure 6 illustrates examples of cysteine derivatives, including sulfoxides. The phenomenon of methionine oxidation is well documented [78,79]. Cysteine derivatives with modified sulfhydryl groups are substrates for the same type of reaction. Sulfoxides, formed upon the oxidation of amino acids and their derivatives, possess asymmetric sulfur atoms. The chirality of the sulfur atom in methionine sulfoxide was first described in 1947 [80]. Any sulfoxide with two different substituents around the sulfur atom forms two enantiomers or diastereoisomers. The chirality of sulfur atoms is in the focus of interest to chemists [81,82] in the context of designing new potential drugs. However, to the best of our knowledge, no studies so far have addressed the chirality of sulfur atoms in food peptides or its influence on their bioactivity. Although the biological activity of peptides containing oxidized methionine has been reported [83,84], the configuration of substituents around asymmetric sulfur atoms has not been specified. Furthermore, while methods of separation and identification of methionine sulfoxide diastereoisomers have been published [85,86], studies involving the identification of chirality of sulfur atoms in residues that are parts of food peptides are currently lacking. On the other hand, configuration of substituents around asymmetric atoms in compound representations may be omitted if the existing information about chirality is incomplete. The above-presented option enables the annotation of peptides containing methionine sulfoxide without defined configuration around the sulfur atom [83,84].

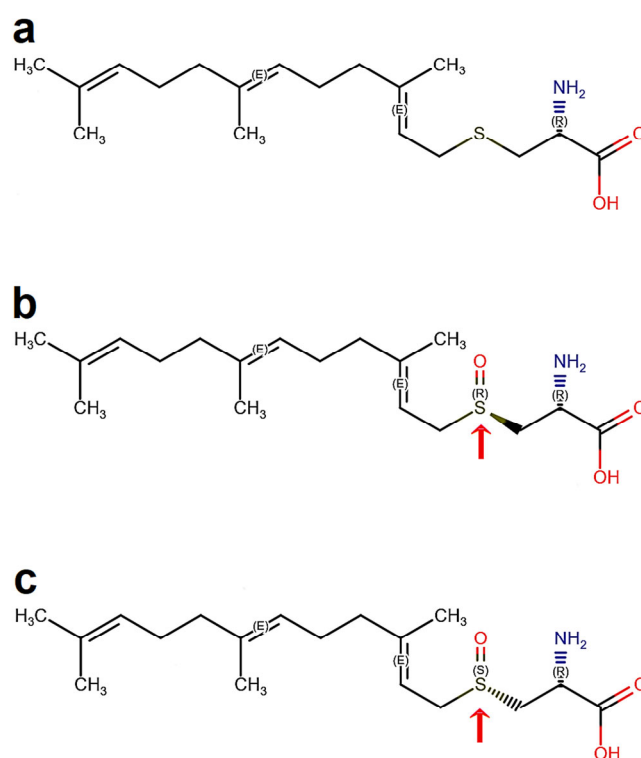


Figure 6. Structures of cysteine derivatives, annotated in the BIOPEP-UWM repository of amino acids and modifications: (a) S-farnesylcysteine (ID 158); (b) S-farnesylcysteine (R)-sulfoxide (ID 198); and (c) S-farnesylcysteine (S)-sulfoxide (ID 199). Symbols (R), (S), and (E) indicate configuration “Rectus” or “Sinister” around asymmetric atoms and configuration “Entgegen” around the double bond, respectively. Arrows indicate asymmetric sulfur atoms.

5. Present and Possible Applications of the BIOPEP-UWM Repository of Amino Acids and Modifications

The first applications of the repository of amino acids and modifications have been reported by Minkiewicz et al. [58], Vuong et al. [68], and Duliński et al. [87]. All these publications described *in silico* studies with protein sequences, modified by the insertion of post-translational modifications, such as cysteine residues involved in the formation of disulfide bonds, phosphorylated serine and threonine residues or N-terminal pyroglutamic acid. Research with protein sequences included calculations of quantitative parameters describing proteins as precursors of bioactive peptides and simulations of proteolysis. Incorporating amino acid sequences with modifications allows for the avoidance of false-positive results regarding the occurrence of bioactive peptides in amino acid sequences of proteins. Minkiewicz et al. [58] additionally presented predictions of the bioactivity of peptides containing phosphorylated amino acids and pyroglutamic acid residues, made using the SwissTargetPrediction program.

The modified protein sequences, described by Minkiewicz et al. [58] and Vuong et al. [68], are available in the BIOPEP-UWM database of proteins (ID: 1834-1846, 1849-1850 and 1859-1867). Symbols of modifications have been changed compared with those included in the articles cited above (Symbols: “{” and “}” instead of “<” and “>” respectively). Peptides annotated using the older version of the BIOPEP-UWM code are also available in the Peptipedia database [36].

Direct incorporation of modifications of amino acid residues into amino acid sequences represents a potential advancement in protein sequence annotation compared with the standard system used in the NCBI Protein database [29] or UniProt [43], and identifies their availability for bioinformatics tools, such as searches for bioactive fragments or proteolysis simulations. The current state-of-the-art in sequence analysis [88] involves traditionally annotated proteins. In turn, the applicability of notations used in the BIOPEP-UWM is restricted to linear peptides and does not encode all information displayed by PLN, BILN, or MAP codes. It concerns information about the location of disulfide bonds or peptide cyclization. Among these notations, MAP is being progressively introduced in databases maintained by the G.P.S. Ragahava's group at Indraprastha Institute for Information Technology (IIIT) in Delhi, India, whereas PLN representations of peptides are provided in the PubChem database [37]. López-López et al. [3] have pointed out problems with human-readability as a weak point of codes when they are extensions of a single-letter code of protein and peptide sequences. In our approach, we have tried to mitigate this problem by, for instance, adapting compound encoding rules applied in lipid databases, such as SwissLipids [40] or LIPID MAPS [28], for the annotation of acyl and alkyl residues. A formal comparison of human-readability between our proposed code and HELM, BILN, PLN or MAP should be performed by independent users, with experience in various systems of annotation of non-canonical and modified amino acids.

Cooperation of BIOPEP-UWM with other databases, datasets or programs utilizing non-canonical amino acids may require adaptation of the biological code or SMILES representations. Recently (6 July 2026), making a search in the Peptipedia required using the older version of the BIOPEP-UWM code (using characters “<” and “>” to indicate the start and the end of the monomer symbol, respectively). In turn, possible data exchange between the BIOPE-UWM and PepFoundry would require adaptation of SMILES strings. SMILES representations of amino acids in the PepFoundry dataset are written in the CHUCKLES format, but use symbols [NH2:1], [NH1:1] and [C:2](=O)O for the primary or secondary amine group forming a peptide bond with a preceding amino acid residue and a carboxyl group forming the peptide bond with the following residue in the peptide

sequence, respectively [33]. Such symbols are not used in the BIOPEP-UWM database. SMILES representations of amino acids and N-terminal acyl groups in the CHUCKLES format in the set described by Orsi and Reymond [5] contain carboxyl groups denoted as C(=O) instead of C(=O)O (such notation corresponds to the aldehyde group). Data exchange between these sets would require either the removal or addition of a symbol of the terminal oxygen atom. SMILES representations of amino acids in the iNclusive [27] and the dataset for the translation of the BILN code [47] are not written using the CHUCKLES format.

Although BIOPEP-UWM remains a peptide database, it may also serve to construct SMILES strings for other compounds (e.g., peptidomimetics). Examples of non-peptide chemical compounds, encoded using the BIOPEP-UWM code and SMILES representations obtained by our application, are provided in Part S6 of the Supplementary Materials.

6. Final Remarks

Until recently, the BIOPEP-UWM database processed only sequences composed of proteinogenic amino acids or their D-enantiomers. The set of monomers that can serve as peptide components has been significantly enlarged and may be continuously enriched. The database aligns with contemporary trends involving modifications of amino acid residues in bioinformatic analysis of peptides and proteins. The enriched set of building blocks enables annotating linear peptides containing canonical and non-canonical amino acids as well as other residues. Possible future development of the code describing peptides and their monomers entails providing an opportunity for encoding branched and cyclic peptides and for applications for translation of their sequences into the SMILES code and the simulation of proteolysis as well as developing profiles of the potential biological activity of such peptides. Another avenue of repository expansion involves the enrichment of the diversity of monomers, which includes implementation of the possibility for encoding peptides and their monomers. Users of the BIOPEP-UWM database are encouraged to submit new sequences of bioactive peptides and their building blocks to further expand the database.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/a19080628/s1>.

Author Contributions: Conceptualization, A.I., P.M. and M.D.; methodology, A.I., P.M. and M.D.; software A.I., P.M. and M.D.; validation, P.M.; formal analysis, A.I., P.M. and M.D.; investigation, A.I., P.M. and M.D.; resources, A.I., P.M. and M.D.; data curation, A.I. and P.M.; writing—original draft preparation, A.I., P.M. and M.D.; writing—review and editing, A.I., P.M. and M.D.; visualization, A.I., P.M. and M.D.; supervision, P.M.; project administration, M.D.; funding acquisition, A.I. and M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by designated subsidy of the Minister of Science and Higher Education Republic of Poland, task entitled: The Research Network of Life Sciences Universities for the Development of the Polish Dairy Industry—Research Project and by Minister of Science under “the Regional Initiative of Excellence Program”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data is available in the article, Supplementary Materials and the BIOPEP-UWM database.

Acknowledgments: We thank Mariusz Falkowski and Krzysztof Sieniawski (Enter Krystyna Sieniawska, Olsztyn, Poland) for IT support and also Monika Hryniewicz, Monika Pliszka, and Marta Godlewska for their help in inserting new data to the database.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACS	American Chemical Society
ADMET	Absorption, distribution, metabolism, excretion, toxicity
BILN	Boehringer Ingelheim Line Notation
CAS	Chemical Abstract Service
CFM	Competitive Fragmentation Modeling
CID	Compound ID (in the PubChem database)
CycPeptMPDB	Cyclic Peptide Membrane Permeability Database
DFBP	Database of Food Bioactive Peptides
EMBL	European Molecular Biology Laboratory
HELM	Hierarchical Editing Language for Macromolecules
HMDB	Human Metabolome Database
IIIT	Indraprastha Institute for Information Technology Delhi
InChI	International Chemical Identifier
MAP	Modification and Annotation in Proteins
NCBI	National Center for Biotechnology Information
PepTherDia	Peptide Therapeutics and Diagnostics
PHP	Personal Homepage
PLN	Protein Line Notation
PLoS	Public Library of Science
SATPdb	Structurally Annotated Therapeutic Peptide Database
SMILES	Simplified Molecular Input Line Entry System
UWM	University of Warmia and Mazury in Olsztyn

References

1. Ding, Y.; Ting, J.P.; Liu, J.; Al-Azzam, S.; Pandya, P.; Afshar, S. Impact of non-proteinogenic amino acids in the discovery and development of peptide therapeutics. *Amino Acids* **2020**, *52*, 1207–1226. <https://doi.org/10.1007/s00726-020-02890-9>.
2. van Walraven, N.; FitzGerald, R.J.; Danneel, H.-J.; Amigo-Benavent, M. Bioactive peptides in cosmetic formulations: Review of current in vitro and ex vivo evidence. *Peptides* **2025**, *193*, 171440. <https://doi.org/10.1016/j.peptides.2025.171440>.
3. López-López, E.; Sánchez, J.P.; Martínez-Cortés, M.S.; de la Fuente-Nunez, C.; Medina-Franco, J.L. Exploring and expanding the chemical multiverse of peptides. *Chem. Sci.* **2026**, *17*, 1461–1479. <https://doi.org/10.1039/D5SC04465K>.
4. Erckes, V.; Abderrahmane, M.; Jusot, M.; Steuer, C.; Ochoa, R. Peptide cheminformatics tools: Making computational tasks accessible in peptide drug discovery. *Drug Discov. Today* **2026**, *31*, 104612. <https://doi.org/10.1016/j.drudis.2026.104612>.
5. Orsi, M.; Reymond, J.-L. Navigating a 1E+60 chemical space of peptide/peptoid oligomers. *Mol. Inform.* **2025**, *44*, e202400186. <https://doi.org/10.1002/minf.202400186>.
6. Brown, C.N.; Dupre, R.A.; Ebmeier, C.C.; Patil, S.; Smith, B.; Mattison, C.P. Heating differentiates pecan allergen stability: Car i 4 is more heat labile than Car i 1 and Car i 2. *Food Sci. Nutr.* **2025**, *13*, e4747. <https://doi.org/10.1002/fsn3.4747>.
7. FitzGerald, R.J. Potential uses of caseinophosphopeptides. *Int. Dairy J.* **1998**, *8*, 451–457. [https://doi.org/10.1016/S0958-6946\(98\)00068-5](https://doi.org/10.1016/S0958-6946(98)00068-5).
8. Tenenbaum, M.; Deracinois, B.; Dugardin, C.; Auger, J.; Baniel, A.; Boulier, A.; Flahaut, C.; Ravallec, R.; Cudennec, B. Digested casein phosphopeptides impact intestinal calcium transport in vitro. *Food Funct.* **2024**, *15*, 8104–8115. <https://doi.org/10.1039/D4FO01637H>.
9. Ueda, Y.; Sakaguchi, M.; Hirayama, K.; Miyajima, R.; Kimizuka, A. Characteristic flavor constituents in water extract of garlic. *Agric. Biol. Chem.* **1990**, *54*, 163–169. <https://doi.org/10.1080/00021369.1990.10869909>.
10. Hashizume, K.; Ito, T.; Nagae, Y.; Tokiwano, T. Quantitation and sensory properties of three newly identified pyroglutamyl oligopeptides in sake. *Biosci. Biotechnol. Biochem.* **2019**, *83*, 357–364. <https://doi.org/10.1080/09168451.2018.1530095>.

11. Vitale, G.A.; Scarpato, S.; Mangoni, A.; D'Auria, M.V.; Della Sala, G.; de Pascale, D. Enhanced molecular networking shows *Microbacterium* sp. V1 as a factory of antioxidant proline-rich peptides. *Mar. Drugs* **2023**, *21*, 256. <https://doi.org/10.3390/md21040256>.
12. Vidal-Limon, A.; Aguilar-Toalá, J.E.; Liceaga, A.M. Integration of molecular docking analysis and molecular dynamics simulations for studying food proteins and bioactive peptides. *J. Agric. Food Chem.* **2022**, *70*, 934–943. <https://doi.org/10.1021/acs.jafc.1c06110>.
13. Minkiewicz, P.; Iwaniak, A.; Darewicz, M. Contemporary bioinformatics and cheminformatics support for food peptidomics. *Curr. Opin. Food Sci.* **2024**, *56*, 101125. <https://doi.org/10.1016/j.cofs.2024.101125>.
14. Cheng, S.; Chen, Y.; Cao, Y.; Huang, X.; Liu, B.; Liu, M.; Jiang, S.; Jiang, L. Deep learning in food Science: Innovative approaches for predicting and simulating food-derived protein–peptides. *Trends Food Sci. Technol.* **2025**, *164*, 105216. <https://doi.org/10.1016/j.tifs.2025.105216>.
15. Singh, B.P.; Paul, S.; Ferreira-Santos, P.; Bhushan, B.; Hati, S.; Goel, G.; Udenigwe, C. In silico and molecular docking approaches in food-derived bioactive peptide discovery: Trends, challenges, and prospects. *Food Res. Int.* **2025**, *222*, 117637. <https://doi.org/10.1016/j.foodres.2025.117637>.
16. Mamoudou, H.; Mune Mune, M.A. AI-driven bioactive peptide discovery of next-generation metabolic biotherapeutics. *Appl. Food Res.* **2025**, *5*, 101291. <https://doi.org/10.1016/j.afres.2025.101291>.
17. Lima de Oliveira, E.C.; Auzier, J.; Pereira Coelho, G.; do Nascimento, L.D.; Lima e Lima, A.H.; Flexa Rodrigues, C.M.; De Spiegeleer, A.; Wynendaele, E.; Sales, C.; De Spiegeleer, B.; et al. Navigating in the chemical space of peptides: Computational strategies and molecular features to unveil their functional and drug-like properties. *Phys. Chem. Chem. Phys.* **2026**, *28*, 11519–11545. <https://doi.org/10.1039/D5CP04611D>.
18. Fu, L.; Shi, S.; Yi, J.; Wang, N.; He, Y.; Wu, Z.; Peng, J.; Deng, Y.; Wang, W.; Wu, C.; et al. ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Res.* **2024**, *52*, W422–W431. <https://doi.org/10.1093/nar/gkae236>.
19. Iwaniak, A.; Minkiewicz, P.; Darewicz, M. BIOPEP-UWM database—Present and future. *Curr. Opin. Food Sci.* **2024**, *55*, 101108. <https://doi.org/10.1016/j.cofs.2023.101108>.
20. Jacobs, A.; Williams, D.; Hickey, K.; Patrick, N.; Williams, A.J.; Chalk, S.; McEwen, L.; Willighagen, E.; Walker, M.; Bolton, E.; et al. CAS Common Chemistry in 2021: Expanding access to trusted chemical information for the scientific community. *J. Chem. Inf. Model.* **2022**, *62*, 2737–2743. <https://doi.org/10.1021/acs.jcim.2c00268>.
21. Wang, F.; Allen, D.; Tian, S.; Oler, E.; Gautam, V.; Greiner, R.; Metz, T.O.; Wishart, D.S. CFM-ID 4.0—A web server for accurate MS-based metabolite identification. *Nucleic Acids Res.* **2022**, *50*, W165–W174. <https://doi.org/10.1093/nar/gkac383>.
22. Zdrasil, B.; Felix, E.; Hunter, F.; Mannes, E.J.; Blackshaw, J.; Corbett, S.; de Veij, M.; Ioannidis, H.; Mendez Lopez, D.; Mosquera, J.F.; et al. The ChEMBL Database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **2024**, *52*, D1180–D1192. <https://doi.org/10.1093/nar/gkad1004>.
23. Williams, A.; Tkachenko, V. The Royal Society of Chemistry and the delivery of chemistry data repositories for the community. *J. Comput. Aided Mol. Des.* **2014**, *28*, 1023–1030. <https://doi.org/10.1007/s10822-014-9784-5>.
24. Li, J.; Yanagisawa, K.; Sugita, M.; Fujie, T.; Ohue, M.; Akiyama, Y. CycPeptMPDB: A comprehensive database of membrane permeability of cyclic peptides. *J. Chem. Inf. Model.* **2023**, *63*, 2240–2250. <https://doi.org/10.1021/acs.jcim.2c01573>.
25. Qin, D.; Bo, W.; Zheng, X.; Hao, Y.; Li, B.; Zheng, J.; Liang, G. DFBP: A comprehensive database of food-derived bioactive peptides for peptidomics research. *Bioinformatics* **2022**, *38*, 3275–3280. <https://doi.org/10.1093/bioinformatics/btac323>.
26. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. <https://doi.org/10.1093/nar/gkab1062>.
27. Icking, L.-S.; Riedlberger, A.M.; Krause, F.; Widder, J.; Frederiksen, A.S.; Stockert, F.; Spädt, M.; Edel, N.; Armbruster, D.; Forlani, F.; et al. iNclusive: A database collecting useful information on non-canonical amino acids and their incorporation into proteins for easier genetic code expansion implementation. *Nucleic Acids Res.* **2024**, *52*, D476–D482. <https://doi.org/10.1093/nar/gkad1090>.
28. Conroy, M.J.; Andrews, R.M.; Andrews, S.; Cockayne, L.; Dennis, E.E.; Fahy, E.; Gaud, C.; Griffiths, W.J.; Jukes, G.; Kolchin, M.; et al. LIPID MAPS: Update to databases and tools for the lipidomics community. *Nucleic Acids Res.* **2024**, *52*, D1677–D1682. <https://doi.org/10.1093/nar/gkad896>.
29. Sayers, E.W.; Bolton, E.E.; Fine, A.M.; Kelly, C.; Kim, S.; Landrum, M.; Lathrop, S.; Malheiro, A.; Murphy, T.D.; Phan, L.; et al. Database resources of the National Center for Biotechnology Information in 2026. *Nucleic Acids Res.* **2026**, *54*, D20–D27. <https://doi.org/10.1093/nar/gkaf1060>.

30. Flissi, A.; Ricart, E.; Campart, C.; Chevalier, M.; Dufresne, Y.; Michalik, J.; Jacques, P.; Flahaut, C.; Lisacek, F.; Leclère, V.; et al. Norine: Update of the nonribosomal peptide resource. *Nucleic Acids Res.* **2020**, *48*, D465–D469. <https://doi.org/10.1093/nar/gkz1000>.
31. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. <https://doi.org/10.1186/1758-2946-3-33>.
32. Tan, X.; Liu, Q.; Zhou, M.; Fang, Y.; Ouyang, D.; Zeng, W.; Dong, J. pepADMET: A novel computational platform for systematic ADMET evaluation of peptides. *J. Chem. Inf. Model.* **2026**, *66*, 936–946. <https://doi.org/10.1021/acs.jcim.5c02518>.
33. Garzon Otero, D.; Akbari, O.; Mandapati, A.; Bilodeau, C. PepFoundry: A pipeline for building machine-learning ready representations of nonstandard peptides containing cycles, nonnatural residues, polymer units, and more. *J. Chem. Inf. Model.* **2026**, *66*, 1264–1273. <https://doi.org/10.1021/acs.jcim.5c02629>.
34. Geylan, G.; Janet, J.P.; Tibo, A.; He, J.; Patronov, A.; Kabeshov, M.; Czechtizky, W.; David, F.; Engkvist, O.; De Maria, L. PepINVENT: Generative peptide design beyond natural amino acids. *Chem. Sci.* **2025**, *16*, 8682–8696. <https://doi.org/10.1039/D4SC07642G>.
35. D'Aloisio, V.; Dognini, P.; Hutcheon, G.A.; Coxon, C.R. PepTherDia: Database and structural composition analysis of approved peptide therapeutics and diagnostics. *Drug Discov. Today* **2021**, *26*, 1409–1419. <https://doi.org/10.1016/j.drudis.2021.02.019>.
36. Cabas-Mora, G.; Daza, A.; Soto-García, N.; Garrido, V.; Alvarez, D.; Navarrete, M.; Sarmiento-Varón, L.; Sepúlveda Yañez, J.H.; Davari, M.D.; Cadet, F.; et al. Peptipedia v2.0: A peptide sequence database and user-friendly web platform. A major update. *Database* **2024**, *2024*, baae113. <https://doi.org/10.1093/database/baae113>.
37. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2025 update. *Nucleic Acids Res.* **2025**, *53*, D1516–D1525. <https://doi.org/10.1093/nar/gkae1059>.
38. Singh, S.; Chaudhary, K.; Dhanda, S.K.; Bhalla, S.; Usmani, S.S.; Gautam, A.; Tuknait, A.; Agrawal, P.; Mathur, D.; Raghava, G.P.S. SATPdb: A database of structurally annotated therapeutic peptides. *Nucleic Acids Res.* **2016**, *44*, D1119–D1126. <https://doi.org/10.1093/nar/gkv1114>.
39. Bugnon, M.; Röhrig, U.F.; Goullieux, M.; Perez, M.A.S.; Daina, A.; Michielin, Q.; Zoete, V. SwissDock 2024: Major enhancements for small-molecule docking with attracting cavities and AutoDock Vina. *Nucleic Acids Res.* **2024**, *52*, W324–W332. <https://doi.org/10.1093/nar/gkae300>.
40. Aimo, L.; Liechti, R.; Hyka-Nouspikel, N.; Niknejad, A.; Gleizes, A.; Götz, L.; Kuznetsov, D.; David, F.P.A.; Gisou van der Goot, F.; Riezman, H.; et al. The SwissLipids knowledgebase for lipid biology. *Bioinformatics* **2015**, *31*, 2860–2866. <https://doi.org/10.1093/bioinformatics/btv285>.
41. Gfeller, D.; Michielin, Q.; Zoete, V. SwissSidechain: A molecular and structural database of non-natural sidechains. *Nucleic Acids Res.* **2013**, *41*, D327–D332. <https://doi.org/10.1093/nar/gks991>.
42. Daina, A.; Michielin, Q.; Zoete, V. SwissTargetPrediction: Updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* **2019**, *47*, W357–W364. <https://doi.org/10.1093/nar/gkz382>.
43. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **2025**, *53*, D609–D617. <https://doi.org/10.1093/nar/gkae1010>.
44. Zhang, T.; Li, H.; Xi, H.; Stanton, R.V.; Rotstein, S.H. HELM: A hierarchical notation language for complex biomolecule structure representation. *J. Chem. Inf. Model.* **2012**, *52*, 2796–2806. <https://doi.org/10.1021/ci3001925>.
45. Milton, J.; Zhang, T.; Bellamy, C.; Swayze, E.; Hart, C.; Weisser, M.; Hecht, S.; Rotstein, S. HELM software for biopolymers. *J. Chem. Inf. Model.* **2017**, *57*, 1233–1239. <https://doi.org/10.1021/acs.jcim.6b00442>.
46. Noisier, A.F.M.; Ghiandoni, G.M.; Pinto-Gil, K.; Ingar, A.; Broddefalk, J.; Bolt, H.; Brive, L. Non-natural amino acid nomenclature for HELM monomer registration. *ChemRxiv* **2026**. <https://doi.org/10.26434/chemrxiv.10001730/v4>.
47. Fox, T.; Bieler, M.; Haebel, P.; Ochoa, R.; Peters, S.; Weber, A. BILN: A human-readable line notation for complex peptides. *J. Chem. Inf. Model.* **2022**, *62*, 3942–3947. <https://doi.org/10.1021/acs.jcim.2c00703>.
48. Shendure, A.; Mehta, N.K.; Rathore, A.S.; Kumar, N.; Patiyl, S.; Raghava, G.P.S. MAP Format for representing chemical modifications, annotations, and mutations in protein sequences: An extension of the FASTA format. *arXiv*. **2025**. <https://doi.org/10.48550/arXiv.2505.03403>.
49. Minkiewicz, P.; Darewicz, M.; Iwaniak, A.; Turło, M. Proposal of the annotation of phosphorylated amino acids and peptides using biological and chemical codes. *Molecules* **2021**, *26*, 712. <https://doi.org/10.3390/molecules26030712>.
50. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. <https://doi.org/10.1021/ci00057a005>.

51. David, L.; Thakkar, A.; Mercado, R.; Engkvist, O. Molecular representations in AI-driven drug discovery: A review and practical guide. *J. Cheminform.* **2020**, *12*, 56. <https://doi.org/10.1186/s13321-020-00460-5>.
52. Wigh, D.S.; Goodman, J.M.; Lapkin, A.A. A review of molecular representation in the age of machine learning. *WIREs Comput. Mol. Sci.* **2022**, *12*, e1603. <https://doi.org/10.1002/wcms.1603>.
53. Fernández-Díaz, R.; Ochoa, R.; Hoang, T.; Lopez, V.; Shields, D. How to build machine learning models able to extrapolate from standard to modified peptides. *J. Cheminform.* **2025**, *17*, 185. <https://doi.org/10.1186/s13321-025-01115-z>.
54. Minkiewicz, P.; Iwaniak, A.; Darewicz, M. BIOPEP-UWM database of bioactive peptides: Current opportunities. *Int. J. Mol. Sci.* **2019**, *20*, 5978. <https://doi.org/10.3390/ijms20235978>.
55. Asim, M.N.; Asif, T.; Mehmood, F.; Dengel, A. Peptide classification landscape: An in-depth systematic literature review on peptide types, databases, datasets, predictors architectures and performance. *Comput. Biol. Med.* **2025**, *188*, 109821. <https://doi.org/10.1016/j.combiomed.2025.109821>.
56. Robinson, M.B.; Blakely, R.D.; Couto, R.; Coyle, J.T. Hydrolysis of the brain dipeptide N-acetyl-L-aspartyl-L-glutamate. *J. Biol. Chem.* **1987**, *262*, 14498–14506. [https://doi.org/10.1016/S0021-9258\(18\)47823-4](https://doi.org/10.1016/S0021-9258(18)47823-4).
57. Ganellin, C.R.; Bishop, P.B.; Bambal, R.B.; Chan, S.M.; Law, J.K.; Marabout, B.; Luthra, P.M.; Moore, A.N.; Peschard, O.; Bourgeat, P.; et al. Inhibitors of tripeptidyl peptidase II. 2. Generation of the first novel lead inhibitor of cholecystokinin-8-inactivating peptidase: A strategy for the design of peptidase inhibitors. *J. Med. Chem.* **2000**, *43*, 664–674. <https://doi.org/10.1021/jm990226g>.
58. Minkiewicz, P.; Darewicz, M.; Iwaniak, A. In silico analysis of individual fractions of bovine casein as precursors of bioactive peptides—Influence of post-translational modifications. *Appl. Sci.* **2023**, *13*, 8091. <https://doi.org/10.3390/app13148091>.
59. Siani, M.A.; Weininger, D.; Blaney, J.M. CHUCKLES: A method for representing and searching peptide and peptoid sequences on both monomer and atomic levels. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 588–593. <https://doi.org/10.1021/ci00019a017>.
60. Duffy, F.J.; Verniere, M.; Devocelle, M.; Bernard, E.; Shields, D.C.; Chubb, A.J. CycloPs: Generating virtual libraries of cyclized and constrained peptides including nonnatural amino acids. *J. Chem. Inf. Model.* **2011**, *51*, 829–836. <https://doi.org/10.1021/ci100431r>.
61. Minkiewicz, P.; Iwaniak, A.; Darewicz, M. Annotation of peptide structures using SMILES and other chemical codes—practical solutions. *Molecules* **2017**, *22*, 2075. <https://doi.org/10.3390/molecules22122075>.
62. Hähnke, V.D.; Kim, S.; Bolton, E.E. PubChem chemical structure standardization. *J. Cheminform.* **2018**, *10*, 36. <https://doi.org/10.1186/s13321-018-0293-8>.
63. Goodman, J.M.; Pletnev, I.; Thiessen, P.; Bolton, E.; Heller, S.R. InChI version 1.06: Now more than 99.99% reliable. *J. Cheminform.* **2021**, *13*, 40. <https://doi.org/10.1186/s13321-021-00517-z>.
64. Cahn, R.S.; Ingold, C.; Prelog, V. Specification of molecular chirality. *Angew. Chem. Int. Ed.* **1966**, *5*, 385–415. <https://doi.org/10.1002/anie.196603851>.
65. Danikiewicz, W. *Spektrometria Mas*; PWN: Warsaw, Poland, 2020. (In Polish)
66. Minkiewicz, P.; Iwaniak, A.; Darewicz, M. BIOPEP-UWM Virtual—A novel database of food-derived peptides with in silico-predicted biological activity. *Appl. Sci.* **2022**, *12*, 7204. <https://doi.org/10.3390/app12147204>.
67. Huang, Y.-P.; Masarweh, C.; Paviani, B.; Mills, D.A.; Barile, D. Exploring bioactive compounds in chickpea and bean aquafaba: Insights from glycomics and peptidomics analyses. *Food Chem.* **2024**, *460*, 140635. <https://doi.org/10.1016/j.foodchem.2024.140635>.
68. Vuong, T.T.; Dupre, R.A.; He, Z.; Minkiewicz, P.; Darewicz, M.; Mattison, C.P. Cottonseed vicilin proteins: Potential allergen risk and source of bioactive peptides. *ACS Food Sci. Technol.* **2025**, *5*, 502–511. <https://doi.org/10.1021/acsfoodscitech.4c00577>.
69. Koehler Leman, J.; Weitzner, B.D.; Renfrew, P.D.; Lewis, S.M.; Moretti, R.; Watkins, A.M.; Mulligan, V.K.; Lyskov, S.; Adolf-Bryfogle, J.; LaBonte, J.W.; et al. Better together: Elements of successful scientific software development in a distributed collaborative community. *PLoS Comput. Biol.* **2020**, *16*, e1007507. <https://doi.org/10.1371/journal.pcbi.1007507>.
70. Delannée, V.; Nicklaus, M.C. ReactionCode: Format for reaction searching, analysis, classification, transform, and encoding/decoding. *J. Cheminform.* **2020**, *12*, 72. <https://doi.org/10.1186/s13321-020-00476-x>.
71. Ochoa, R. Pepfun 2.0: Improved protocols for the analysis of natural and modified peptides. *Future Drug Discov.* **2023**, *5*, FDD82. <https://doi.org/10.4155/fdd-2023-0004>.
72. Ochoa, R.; Brown, J.B.; Fox, T. pyPept: A python library to generate atomistic 2D and 3D representations of peptides. *J. Cheminform.* **2023**, *15*, 79. <https://doi.org/10.1186/s13321-023-00748-2>.
73. Ochoa, R.; Deibler, K. PepFuNN: Novo Nordisk open-source toolkit to enable peptide in silico analysis. *J. Peptide Sci.* **2025**, *31*, e3666. <https://doi.org/10.1002/psc.3666>.

74. Frolov, A.I.; Morley, J.; Gomes dos Santos, A.; Genapathy, S.; Ghiandoni, G.M. Peptide-Tools–Web server for calculating physicochemical properties of peptides. *J. Chem. Inf. Model.* **2026**, *66*, 16–27. <https://doi.org/10.1021/acs.jcim.5c02296>.
75. Mossine, V.V.; Mawhinney, T.P. 1-Amino-1-deoxy-d-fructose (“Fructosamine”) and its derivatives. *Adv. Carbohydr. Chem. Biochem.* **2010**, *64*, 291–402. [https://doi.org/10.1016/S0065-2318\(10\)64006-1](https://doi.org/10.1016/S0065-2318(10)64006-1).
76. Mossine, V.V.; Barnes, C.L.; Mawhinney, T.P. N-(1-Deoxy- α -D-tagatopyranos-1-yl)-N-methylaniline (“D-Tagatose-N-methylaniline”). *Molbank* **2018**, *2018*, M994. <https://doi.org/10.3390/M994>.
77. Kijewska, M.; Nuti, F.; Wierzbicka, M.; Waliczek, M.; Ledwoń, P.; Staśkiewicz, A.; Real-Fernandez, F.; Sabatino, G.; Rovero, P.; Stefanowicz, P.; et al. An optimised di-boronate-ChemMatrix affinity chromatography to trap deoxyfructosylated peptides as biomarkers of glycation. *Molecules* **2020**, *25*, 755. <https://doi.org/10.3390/molecules25030755>.
78. Brot, N.; Weissbach, H. Biochemistry of methionine sulfoxide residues in proteins. *BioFactors* **1991**, *3*, 91–96. [https://doi.org/10.1016/0968-0004\(82\)90204-3](https://doi.org/10.1016/0968-0004(82)90204-3).
79. Yan, C.; He, B.; Wang, C.; Li, W.; Tao, S.; Chen, J.; Wang, Y.; Yang, L.; Wu, Y.; Wu, Z.; et al. Methionine in embryonic development: Metabolism, redox homeostasis, epigenetic modification and signaling pathway. *Crit. Rev. Food Sci. Nutr.* **2025**, *65*, 8051–8074. <https://doi.org/10.1080/10408398.2025.2491638>.
80. Lavine, T.F. The formation, resolution, and optical properties of the diastereoisomeric sulfoxides derived from L-methionine. *J. Biol. Chem.* **1947**, *169*, 477–491. [https://doi.org/10.1016/S0021-9258\(17\)30864-5](https://doi.org/10.1016/S0021-9258(17)30864-5).
81. Han, J.; Soloshonok, V.A.; Klika, K.D.; Drabowicz, J.; Wzorek, A. Chiral sulfoxides: Advances in asymmetric synthesis and problems with the accurate determination of the stereochemical outcome. *Chem. Soc. Rev.* **2018**, *47*, 1307–1350. <https://doi.org/10.1039/C6CS00703A>.
82. Tsuzuki, S.; Kano, T. Asymmetric synthesis of chiral sulfimides. *Asian J. Org. Chem.* **2025**, *14*, e202400765. <https://doi.org/10.1002/ajoc.202400765>.
83. Jin, Y.; Yan, J.; Yu, Y.; Qi, Y. Screening and identification of DPP-IV inhibitory peptides from deer skin hydrolysates by an integrated approach of LC–MS/MS and in silico analysis. *J. Funct. Foods.* **2015**, *18*, 344–357. <https://doi.org/10.1016/j.jff.2015.07.015>.
84. Lima, B.; Ricci, M.; Garro, A.; Juhász, T.; Szigyártó, I.C.; Papp, Z.I.; Feresin, G.; de la Torre, J.G.; Cascales, J.L.; Fülöp, L.; et al. New short cationic antibacterial peptides. Synthesis, biological activity and mechanism of action. *Biochim. Biophys. Acta Biomembranes.* **2021**, *1863*, 183665. <https://doi.org/10.1016/j.bbmem.2021.183665>.
85. Tarrago, L.; Péterfi, Z.; Lee, B.C.; Michel, T.; Gladyshev, V.N. Monitoring methionine sulfoxide with stereospecific mechanism-based fluorescent sensors. *Nat. Chem. Biol.* **2015**, *11*, 332–338. <https://doi.org/10.1038/nchembio.1787>.
86. Raskatov, J.A.; Virgil, S.; Lee, H.-W.; Henling, L.M.; Chan, K.; Kuhn, A.J.; Foley, A.R. A facile method for the separation of methionine sulfoxide diastereomers, structural assignment, and DFT analysis. *Chem. Eur. J.* **2020**, *26*, 4467–4470. <https://doi.org/10.1002/chem.201904848>.
87. Duliński, R.; Zdaniewicz, M.; Byczyński, Ł.; Żuk-Gołaszewska, K.; Bukowska, B. Application of HPSEC technique and in silico analysis in the evaluation of bioactive peptides and polysaccharide profile in wort supplemented with malted and unmalted hemp seeds. *Molecules* **2025**, *30*, 3676. <https://doi.org/10.3390/molecules30183676>.
88. Asim, M.N.; Asif, T.; Hassan, F.; Dengel, A. Protein sequence analysis landscape: A systematic review of task types, databases, datasets, word embeddings methods, and language models. *Database* **2025**, *2025*, baaf027. <https://doi.org/10.1093/database/baaf027>.

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.