



Structural characteristics of food protein-originating di- and tripeptides using principal component analysis

Anna Iwaniak¹ · Monika Hrynkiewicz¹ · Justyna Bucholska¹ · Małgorzata Darewicz¹ · Piotr Minkiewicz¹

Received: 2 March 2018 / Revised: 25 April 2018 / Accepted: 27 April 2018 / Published online: 10 May 2018
© The Author(s) 2018

Abstract

This paper presents the use of principal component analysis (PCA) to study the effect of specific physicochemical attributes on bitterness of di- and tripeptides originating from food proteins. Peptide sequences were derived from the BIOPEP-UWM database of sensory peptides and amino acids. Descriptors defining the physicochemical properties of amino acids forming the analyzed peptides were study variables. They were derived from ProtScale program and Biological Magnetic Resonance Data Bank. Finally, PCA was carried out for 51 dipeptides/12 variables, and 51 tripeptides/18 variables using STATISTICA[®] 13.1 software. PCA allowed reducing the input datasets to 4 principal components (PCs) for dipeptides and to 5 PCs for tripeptides. The impact of the following properties on the bitterness of peptides was observed: relatively high molecular weight, bulkiness, increasing number of carbon and hydrogen atoms of amino acids forming the sequences. These properties characterized the N- (negative correlations) and C-terminal residue (positive correlations) of both di- and tripeptides. An additional property affecting peptide bitterness was amino acids' hydrophobicity. Our results were consistent with scientific reports on structure–bitterness of peptides. Thus, we find PCA a chemometric approach helpful in broadening the knowledge about the function of peptides resulting from their chemical nature.

Keywords Bitter peptides · Food proteins · PCA · BIOPEP-UWM database

Introduction

Food provides the organism with energy as well as essential components necessary to maintain the life functions [1]. Apart from having the nutritional value, such food components may affect their sensory properties. Among them taste is considered as one of the most important factors determining attractiveness of food products and consumer preferences [2]. Chemical compounds that occur in food and affect its taste include, e.g., carbohydrates, lipids, organic acids, salts, amino acids, and nucleotides [3]. According to scientific reports, peptides—which are products of the hydrolysis of proteins—show all five taste sensations, i.e., bitterness,

saltiness, sourness, sweetness, and umami [4–6]. So far, peptides have not been found as the major tastant molecules [4]; however, it needs to be emphasized that they serve various biological functions, e.g., reduce blood pressure, cholesterol and glucose levels, as well as act as antioxidative, antibacterial, antithrombotic agents [5, 7]. Therefore, the influence of such peptides on protein-originating food taste may be considered important for scientists and food technologists who strive to produce foods beneficial for health and attractive to consumers. It especially concerns food products the undesired taste of which may result from the presence of bitter peptides [8].

A continuously increasing load of information about biological phenomena (e.g., derived from experimental data) has contributed to the development of bioinformatic (in silico) tools which, e.g., support the research by helping plan the experiments and/or analyze data and interpret the results. Examples of such tools include databases of biological and chemical information of compounds, programs for the prediction of physicochemical properties of molecules, for the simulation of processes taking place in the living systems, and for the analysis of the function of a molecule taking

Anna Iwaniak and Monika Hrynkiewicz contributed equally to this work.

✉ Anna Iwaniak
ami@uwm.edu.pl

¹ Faculty of Food Science, Chair of Food Biochemistry, University of Warmia and Mazury in Olsztyn, Pl. Cieszyński 1, Kortowo, 10-726 Olsztyn, Poland

into attention its structure [9, 10]. The latter one concerns also the food protein-originating peptides. Thus, scientists often use chemometric and cheminformatic tools to analyze the structure–function relationships between peptides [11]. The above-mentioned disciplines are useful in explaining processes and/or phenomena [9] by transformation of the multi-dimensional datasets to find general regularities between the variables and/or to define some attributes (variables) that may potentially affect properties of a molecule [10, 12]. Such an approach may then be used to analyze the bitter taste of peptides considering their structure defined by numerical physicochemical attributes.

Chemometric methods applied to study peptides that serve various functions include multivariate analyses, e.g., principal component analysis (PCA) [10, 13], partial least square regression (PLSR) [14], and artificial neural networks (ANN) [15]. The first one (PCA) is considered as probably the oldest and best-known chemometric methodology employed in data analysis in many scientific disciplines [16] such as biological, chemical, geographical, psychological, social, and also agricultural sciences including food research [17]. According to Jolliffe [17], the potential PCA application in other disciplines has not been fully exploited yet and it is to be expected that the list of areas will increase.

Briefly, PCA concerns the reduction of datasets and/or data dimensionality. The input variables are transformed to the reduced linear combinations called principal components (PCs) [18]. PCs explaining the largest portion of variation facilitate detection of the relationships in the multivariate dataset, which were not initially recognized in the input data matrix. Moreover, the first few PCs are enough to explain majority of variation and thus provide a lot of information in the interpretation of the classified objects (e.g., peptides) [12].

Bitterness is the most frequently occurring taste sensation among tastant peptides [4], as evidenced by the literature addressing sensory peptides as well as databases of peptide sequences like BIOPEP-UWM [19]. The multiplicity of bitter peptides derived from food protein sources is the likely reason for the use of bio- and cheminformatic methods in their analysis. Such methodologies enable finding some structural regularities between the sequences and their bitterness. The knowledge about the structural nature of peptides that tend to be bitter may be useful to, e.g., food technologists who aim to eliminate (if needed) the undesired, protein-originating taste of food [5].

The accessibility to databases and *in silico* programs providing the information about the bitter peptide sequences as well as variables describing the physicochemical properties of single amino acids forming the peptides encouraged us to undertake the study aimed at applying the PCA to define the properties represented by PCs and deciding about the bitterness of food protein-originating di- and tripeptides.

Materials and methods

Datasets and variables for elaboration of data matrices

Bitter di- and tripeptides were derived from the BIOPEP-UWM database of sensory peptides and amino acids available at: <http://www.uwm.edu.pl/biochemia> [19] (accessed: September 2017) to create two separate datasets (consisting of 51 sequences each). Dipeptides were represented by the following sequences: RR, PP, RG, FF, RP, PR, RF, KP, PK, GR, GP, FG, GF, YG, GY, YY, FV, FL, FI, FP, AF, VF, LF, IF, YF, PF, KF, YP, VY, GV, VL, VI, VE, IV, LE, VD, GE, AD, LD, LL, LG, GL, II, IG, GI, LV, LI, IL, EI, EY, and EF. Tripeptides' dataset comprised of: PPP, FFF, FGG, GFG, GGF, FFG, FGF, GFF, RGP, PGR, RPG, GRP, RPF, YGG, GYG, GGY, YYG, YGY, GYY, YYY, FIV, FPP, FPK, PFP, VIF, KPF, PPF, KPK, PPG, PGG, PGP, GPG, GGP, PFP, VYP, YPF, PGI, VVV, DLL, GGV, GVV, EGG, ELL, PIP, LLL, LGG, GLG, LLG, LGL, GLL, and GGL sequences. Literature data about all peptides taken for the analysis are provided in the BIOPEP-UWM database of sensory peptides and amino acids [19]. Variables (i.e., attributes and descriptors) reflected the following physicochemical properties of each amino acid forming di- and/or tripeptide sequence: molecular weight (MW), bulkiness (B) [20], polarity (P) [21], hydrophobicity (H_{KD}) [22] as well as the numbers of carbon (nCat) and hydrogen atoms (nHat). Their abbreviations are given in parentheses. Moreover, the suffixes (“-1”, “-2”, and/or “-3”) were assigned to each attribute to distinguish the specific property of N-, C-, and/or middle amino acid. Thus, in the case of dipeptides, suffixes “-1” and “-2” described the property of N- and C-terminal amino acid, respectively, whereas for tripeptides, the property of N-, middle and C-terminal amino acid was distinguished by suffix “-1”, “-2”, and “-3”, respectively. Descriptors such as MW, B, P and H_{KD} were obtained from the ProtScale program available at: <http://web.expasy.org/protscale> [23], whereas nCat and nHat were taken from the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu/ref_info/aadata.dat) [24]. Both programs were accessed in September 2017.

Finally, the dipeptide data matrix was composed of 51 samples/objects (i.e., peptides) and 12 variables, while the tripeptide matrix contained 51 objects and 18 variables. It was the exact set of peptide sequences and variables that had been used in our work concerning the application of multiple linear regression (MLR) to analyze structure–bitterness relationship of peptides [25].

Principal component analysis (PCA)

The PCA was performed for bitter di- and tripeptides using STATISTICA®13.1. software by opening “Statistics” menu and then running the protocol “Principal Components & Classification” being a part of the module called “Mult/Exploratory”. The maximal number of components was set at 11 and 11, respectively, and minimal eigenvalue was 1.0 both for di- and tripeptide datasets. The standardization of variables was made by default in STATISTICA®13.1. To check if the PCA was an appropriate method to analyze our peptide datasets, the following calculations were performed: KMO (Kaiser–Meyer–Olkin) measure of sampling adequacy and Bartlett’s test of sphericity [26]. These tests were conducted using STATISTICA®13.1 software. According to statistical handbooks, the normal distribution was not required when performing PCA [17].

Results and discussion

Bitter di- and tripeptide separate datasets were studied using PCA for few reasons. First, they were the most numerous group of bitter peptides listed in BIOPEP-UWM database comparing to longer chain sequences [19] and such sets of di- and tripeptides enabled to obtain well-conditioned matrices [25]. According to scientific reports, chemometric models developed for each subset analyzed generate less predictability errors [27]. Second, many bitter di- and tripeptides tend to exhibit additional biological function, mainly enzyme inhibiting, antioxidative, and/or antithrombotic. The knowledge about structural characteristics of such peptides is in the focus of scientists aiming to develop technologies of production of health-beneficial and taste-attractive food components [5]. As mentioned above, the data matrices were composed of 51 bitter dipeptides and 12 variables, and 51 tripeptides and 18 variables. According to the literature, the minimal number of samples (i.e., cases, observations) should be 50 to get clear and recognizable PCA results [28]. The results concerning the correctness of using PCA to our datasets are given in Table 1.

Table 1 Results of KMO and Bartlett’s tests calculated for bitter di- and tripeptides

Test	Dipeptides	Tripeptides
KMO index	0.54	0.45
Bartlett’s test of sphericity		
Approximate χ^2	720.99	2986.09
Degrees of freedom	66	153
Significance	0.00	0.00

According to the literature, the KMO index ranges from 0 to 1. The higher the KMO index value, the better the foundations to use PCA. This chemometric method is recommended for dataset analysis when KMO index is higher than 0.5 [26]. Bartlett’s test is a test of hypothesis saying that the data matrix is an identity matrix, which means that the values of all correlation coefficients are 0.0 and there are no significant correlations between the variables. Following this rule, the null hypothesis (H_0) meant that the correlation matrix is an identity matrix, whereas an alternative hypothesis (H_A) is in contradiction to H_0 . Thus, rejection of H_0 would confirm the correctness of PCA application in our study [26]. When looking at our results (see Table 1), we found that KMO indices for bitter di- and tripeptides were 0.54 and 0.45, respectively. Results of Bartlett’s test revealed that χ^2 was 720.99 for bitter dipeptides and 2986.09 for tripeptides. The p values calculated for both peptide populations were 0.00 and were lower than 0.05. It meant that there were no grounds for accepting H_0 , which confirmed the appropriateness of PCA application for studying the main factors affecting the bitterness of peptides analyzed. Although the KMO index for bitter tripeptides was lower than 0.5, we decided to continue PCA. The following issues were found sufficient to recommend PCA in our study: the advantageous number of samples, the adequate results of Bartlett’s test for both peptide datasets, and value of the KMO index obtained for dipeptides.

The important aspect of PCA is to specify the PCs obtained, which may not be easy and thus some approaches were proposed aiming to help their interpretation [17]. The results of PCA presenting the eigenvalues, percentages of variance, and cumulative variance calculated for bitter di- and tripeptides are shown in Table 2. The number of PCs in di- and tripeptide models was determined based on the percentages of variance explained [29]. According to the

Table 2 Values of parameters determining the number of PCs selected for bitter di- and tripeptides data interpretation

Principal component (PC)	Eigenvalue	Variance (%)	Cumulative variance (%)
Dipeptides			
PC-I	3.84	31.99	31.99
PC-II	2.99	24.92	56.91
PC-III	2.09	17.43	74.34
PC-IV	1.54	12.82	87.16
Tripeptides			
PC-I	4.03	22.39	22.39
PC-II	3.74	20.75	43.14
PC-III	3.52	19.56	62.70
PC-IV	1.96	10.92	73.62
PC-V	1.60	8.88	82.50

literature, the number of PCs may be described by cumulative variance, exceeding a threshold value depending on the specificity of data analyzed. Usually, the sufficient threshold value of cumulative variance is at least 70% and in some cases 80% [30]. Thus, to solve the problem of the number of PCs selection, we decided to assume 75% of the cumulative variance as the threshold value.

In our study, over 75% of the cumulative variance was assigned to four (dipeptides) and five (tripeptides) PCs. They explained 87.16 and 82.50% of the cumulative variance, respectively. The number of PCs obtained for di- and tripeptides was also confirmed both by “scree” test proposed by Cattell and by Kaiser’s criterion (data not shown) [29]. The first one is the graphical interpretation of eigenvalues. The rocky slope assigned to the given eigenvalue is the point indicating the number of PCs selected for data interpretation [31]. According to the criterion proposed by Kaiser, those PCs which eigenvalues are greater than 1.0 are taken for the interpretation of results [32].

The composition of individual PCs with the variables representing the physicochemical properties of amino acids present in bitter peptide sequences is shown in Table 3. The following functions of the statistical software were used to define the specific variables in PCs: “Factor & coordinates of variables” and “Contributions of variables”. The first represents correlation coefficients between the variables and PC, whereas the latter shows determination coefficients. These correlation coefficients which absolute value was ≥ 0.60 and were highlighted in red by software were found as the most important in a given PC.

As mentioned above, four PCs explained the majority of variation in the case of bitter dipeptides. Positive correlations (in brackets) between the first component (PC-I) and the properties of C-terminal amino acid concerned MW-2 (0.67), B-2 (0.62), nCat-2 (0.62), and nHat-2 (0.67). Negative correlations between PC-I and variables were related to the properties of N-terminal amino acid in a dipeptide sequence. They were MW-1 (−0.72), B-1 (−0.69), nCat-1 (−0.76), and nHat-1 (−0.73). Determination coefficients ranged between 0.10 and 0.15 which meant that the impact of the above-mentioned variables in forming PC-I was between 10 and 15%. PC-II reflected other property of the C-terminal residue in bitter dipeptide like H_{KD-2} (0.70; $R^2 = 0.16$). PC-III included the attributes concerning the N-terminal residue such as P-1 (0.89) and H_{KD-1} (−0.91). These attributes had a relatively high impact on PC-III formation (38 and 40%, respectively). The last PC (PC-IV) contained the variable describing polarity of C-terminal amino acid ($R = -0.73$, $R^2 = 0.34$). The structure of dipeptide datasets taken for PCA revealed “at a glance” more factors affecting the bitterness of sequences analyzed than in the case of MLR. For example, MLR showed statistically significant impact of B-1 and nCat-2 on the bitter taste of

Table 3 Variable composition of individual PCs obtained for bitter di- and tripeptide datasets

Principal component (PC)	Variable	Correlation coefficient (R)	Determination coefficient (R^2)
Dipeptides			
PC-I	MW-1	−0.72	0.14
	B-1	−0.69	0.12
	nCat-1	−0.76	0.15
	nHat-1	−0.73	0.14
	MW-2	0.67	0.12
	B-2	0.62	0.10
	nCat-2	0.62	0.10
PC-II	nHat-2	0.67	0.12
	H_{KD-2}	0.70	0.16
PC-III	P-1	0.89	0.38
	H_{KD-1}	−0.91	0.40
PC-IV	P-2	−0.73	0.34
Tripeptides			
PC-I	MW-2	0.79	0.15
	B-2	0.73	0.13
	nCat-2	0.77	0.15
	nHat-2	0.81	0.16
PC-II	MW-3	0.89	0.21
	B-3	0.92	0.23
	nCat-3	0.91	0.22
PC-III	nHat-3	0.91	0.22
	MW-1	−0.76	0.16
	B-1	−0.77	0.17
PC-IV	nCat-1	−0.77	0.17
	nHat-1	−0.75	0.16
PC-IV	H_{KD-1}	0.66	0.22
PC-V	P-1	0.73	0.34

dipeptides. Next step, i.e., stepwise regression, revealed that additional variables such as nCat-1, nCat-2, and nHat could also be important in the explanation of structure–bitterness of dipeptides [25].

In the case of bitter tripeptides, the first three PCs (PC-I, PC-II, and PC-III) were related to attributes such as MW, B, nCat, and nHat. These variables were assigned to the middle, C-, and N-terminal position described by PC-I, PC-II, and PC-III, respectively. These attributes were also found as variables affecting the tripeptides’ bitterness using stepwise regression [25]. All correlation coefficients between PC and individual variables were positive in the case of PC-I and PC-II (see Table 3). Determination coefficients (R^2) ranged between 0.13 and 0.16 (PC-I) and 0.21–0.23 (PC-II). The correlation coefficients between PC-III and the above-mentioned variables were negative (−0.76, −0.77, −0.77, and −0.75, respectively). The impact of MW-1, B-1, nCat-1, and nHat-1 on PC-III formation was between 16 and 17%

($R^2 = 0.16–0.17$). PC-IV included hydrophobicity (H_{KD}) ($R = 0.66$, $R^2 = 0.22$), whereas PC-V polarity (P) ($R = 0.73$, $R^2 = 0.34$). These both attributes concerned the amino acid located at the N-terminal position of a bitter tripeptide.

The correlation between input variables and PCs obtained for bitter di- and tripeptides is shown in Fig. 1a and b, respectively. It is a plane projection of data presenting the distribution of variables which form the first two PCs. In the case of dipeptides, the first two PCs explained 56.91% of variance whereas in the case of tripeptides they explained 43.14% of the variance. Every single variable was represented by a vector. The direction and length of each vector

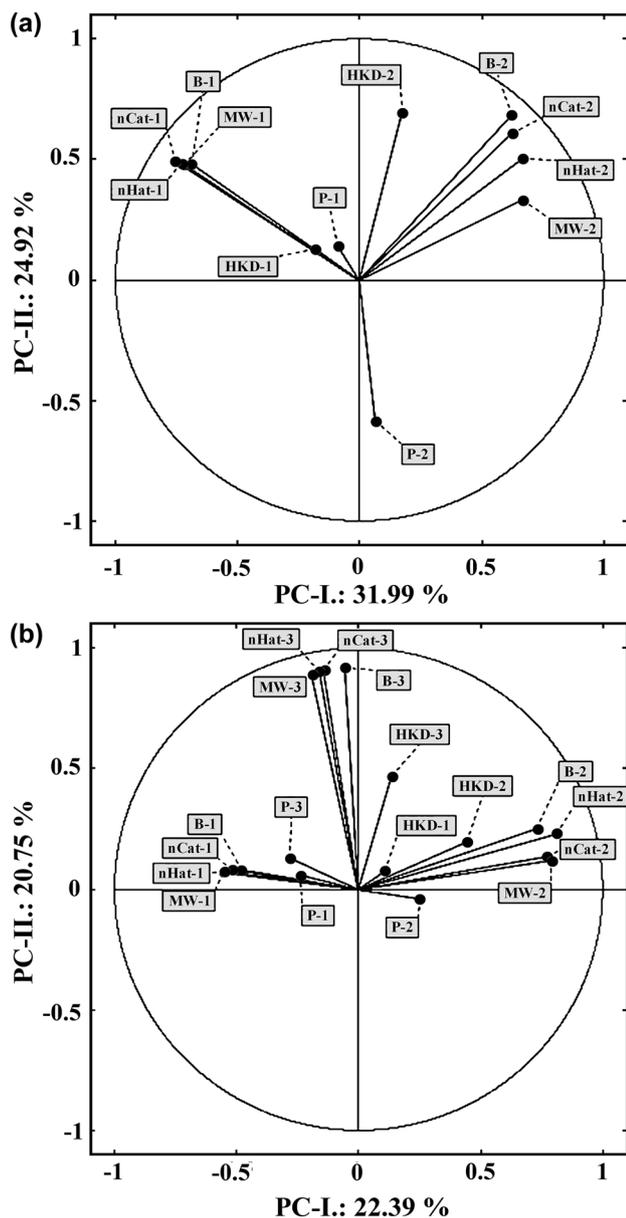


Fig. 1 Circle plot of vectors obtained for bitter dipeptides (a) and tripeptides (b)

inform about the impact of attributes on the individual PC [33]. As seen in Fig. 1a, b, the majority of variables were located near the circle which meant that the information “coded” in the variable was explained by PCs. The vicinity of the variables indicated positive correlations between them, e.g., MW-1, B-1, nCat-1, and nHat-1 (Fig. 1a). Vectors located vertically indicated no correlations between the attributes (see MW-1 and MW-2 in Fig. 1a). Negative correlations between the attributes were represented by vectors located opposite to each other [33], like MW-2 and MW-1 in the tripeptide plot (Fig. 1b).

Our chemometric analysis of di- and tripeptides revealed the molecular weight (MW) of amino acids in their sequences to be one of the crucial factors affecting peptide bitterness. Similar results were also observed by Iwaniak et al. [25] when analyzing bitter di- and tripeptides using MLR. This attribute (MW) appeared in PC-I (dipeptide dataset) and in the first three PCs of the tripeptide dataset. The positive correlations were assigned to the C-terminal residue of a peptide sequence (both datasets) and also to the middle position (tripeptides). Relatively high molecular weight of a given amino acid resulted also from the increasing number of carbon and hydrogen atoms, which was also revealed in the obtained PCs. Kim and Li-Chan [34] reported that high molecular weight of amino acids had the impact on peptides’ bitterness, which also influenced their hydrophobicity. Peptides composed of F and/or Y (with a relatively high molecular weight) were reported to be bitter [34]. Our two datasets contained peptides possessing an amino acid with ring (F, Y and/or P). Moreover, according to scientific reports, the presence of hydrophobic residue is indicative of the bitter taste of a peptide. It was found that although bitter peptides have no specific primary structures, they contain amino acids with hydrophobic side chains [35]. Chemometric studies carried out on the sequences composed of 2–14 amino acids showed that their bitterness was associated with the presence of hydrophobic/bulky and basic/hydrophilic residue located at C and N end of a peptide chain, respectively. Further PLSR analysis of the above-mentioned peptides revealed that their bitter taste is related to the hydrophobicity, molecular mass, and number of amino acids of the whole sequence rather than to specific property of a single amino acid [36]. Our results confirmed that bitterness of di-/tripeptide depended on the particular properties resulting from the presence of specific residues in a peptide sequence. The majority of di- and tripeptides of our datasets were composed of hydrophobic C-terminal amino acid with a bulky chain, e.g., V, L, and/or I.

According to Soltani et al. [27], hydrophobicity and an increasing number of carbon atoms in the side chains of amino acids forming peptides affect the bitterness of peptides. These attributes are related to the bulkiness and molecular weight of amino acids. It was indicated that to

be bitter, a peptide should ideally possess basic and bulky amino acid at N terminus, and hydrophobic amino acid at C terminus of its sequence [37]. When looking at, e.g., the dipeptide sequences analyzed, they were composed of N-terminal R or K—basic amino acids, with rather hydrophilic character. It is consistent with the PCA results showing the negative correlation between PC-II and hydrophobicity of the N-terminal amino acid.

Some of the di- and tripeptides analyzed contained G residue. According to the scientific reports [37], the location of GG fragment in a tripeptide chain may decide about its bitterness. It was reported that GG motif located after N-terminal position of longer peptide sequence decreased the bitterness of the whole sequence, whereas position 1–2 of GG had a positive impact on its taste [37]. Our tripeptide dataset contained 5 peptides with the *Aa*-GG motif and 5 sequences possessed the GG-*Aa* fragment (where “*Aa*” means any amino acids) (see “Materials and methods”). All of them were reported in the literature as bitter. When looking at their bitterness measure—Rcaf. value defined as the ratio of bitterness to 1 mM caffeine solution; Rcaf = 1.0 [34]—the peptides with GG-*Aa* sequences (“more bitter”) had Rcaf. between 0.03 (GGV) and 0.67 (GGY). The Rcaf. of the *Aa*-GG peptides (“less bitter”) ranged between 0.01 (LGG) and EGG (0.67) (data on Rcaf. were taken from the BIOPEP-UWM database of sensory peptides and amino acids). Based on Rcaf. of peptides, it can be observed that the sequences with both *Aa*-GG and GG-*Aa* can be bitter at comparable level of bitterness. It may suggest that the presence of an additional amino acid (“*Aa*”) in a tripeptide sequence could intensify its bitterness, which can be explained by the use of chemometric methods like PCA. For example, peptides GGF and GGY were found as bitter like GG-*Aa* type peptides due to the presence of bulky and high molecular weight residues [34], which was also observed in our PCA results (see Table 3).

Our literature search for the examples of PCA application to study the impact of physicochemical properties of specific amino acids on peptide bitterness showed no publication concerning this subject. The results of PCA concerned mostly peptides in general, including those derived from foods. For example, Nongonierma and FitzGerald [36] summarized the examples of descriptors obtained by means of PCA reflecting the physicochemical character of proteinogenic amino acids. These descriptors were applied for peptide quantitative structure–activity relationship (QSAR) modeling [36]. Another studies involving PCA quantitative descriptors of protein-originating amino acids were applied for PLSR of ACE-inhibiting, bitter-tasting and thromboplastin-inhibiting peptides [13]. PCA was applied by Iwaniak [38] to group ACE inhibitory peptides for families classified by their physicochemical properties. This analysis allowed finding some model ACE

inhibitors that were searched in food-originating protein sequences [38]. Using PCA, Alizadeh-Pasdar et al. [39] studied Raman spectra to analyze the impact of, e.g., κ -carrageenan, pH, temperature on the structure of whey proteins.

To summarize, PCA known as the oldest and best-known multivariate methodology [15] may be also used to study the chemical nature of food-originating peptides. Moreover, PCA allows to show so-called the structure of the dataset and recognize the importance of specific variables forming PCs. It can be especially useful when other chemometric methods (e.g., MLR, see discussion above) may not “immediately” indicate which attributes affect the property of a molecule. Thus, PCA can be applied as the initial study helping to broaden the knowledge about the molecules. Using PCA as the initial step in continuing chemometric studies is consistent with the Wold’s postulate “use the resulting principal components to guide your continued investigations or chemical experimentation, not as end in itself” [40]. Regardless of the potential and practical application of PCA, our results show that this method may be used to find the relationships between variables to discover the structure of the dataset analyzed.

Final remarks

PCs obtained for di- and tripeptides (4 and 5, respectively) led to observe the impact of the following properties on the bitter taste of peptides analyzed: relatively high molecular weight, bulkiness, and increasing number of carbon and hydrogen atoms of amino acid residues forming the sequences of peptides. Such properties were typical of the C-terminal residue of the di- and tripeptide chain (positive correlations). The above-mentioned properties were also confirmed for the N-terminal residue of di- and tripeptide but the correlations were negative. Our results revealed the importance of hydrophobicity in the bitterness of peptides analyzed. It confirmed observations of other authors who analyzed the impact of the chemical nature of peptides on their bitterness using other bio-/cheminformatic tools. Thus, PCA was found a useful method enabling to distinguish the major properties which determine the bitterness of di- and tripeptides derived from food proteins.

Funding This study was funded by the Warmia and Mazury University in Olsztyn (Project number: 17.610.014-300).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Compliance with ethics requirements This article does not contain any studies with human or animal subject.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Hall KD, Heymsfield SB, Kemnitz JW, Klein S, Schoeller DA, Speakman JR (2012) Energy balance and its components: implications for body weight regulation. *Am J Clin Nutr* 95(4):989–994
- Vabø M, Hansen H (2014) The relationships between food preferences and food choice: a theoretical discussion. *Int J Bus Soc Sci* 5(7):145–157
- Chaudhari N, Roper SD (2010) The cell biology of taste. *J Cell Biol* 190(3):285–296
- Temussi PA (2012) The good taste of peptides. *J Pept Sci* 18(2):73–82
- Iwaniak A, Minkiewicz P, Darewicz M, Hryniewicz M (2016) Food protein-originating peptides as tastants—physiological, technological, sensory, and bioinformatic approaches. *Food Res Int* 89:27–38
- Zhao CJ, Schieber A, Gänzle MG (2016) Formation of taste-active amino acids, amino acid derivatives and peptides in food fermentations—a review. *Food Res Int* 89:39–47
- Li-Chan ECY (2015) Bioactive peptides and protein hydrolysates: research trends and challenges for application as nutraceuticals and functional food ingredients. *Curr Opin Food Sci* 1:28–37
- Kęska P, Stadnik J (2017) Taste-active peptides and amino acids of pork meat as components of dry-cured pork meat: an in-silico study. *J Sens Stud* 32:e12301
- Iwaniak A, Minkiewicz P, Darewicz M, Protasiewicz M, Mogut D (2015) Chemometrics and cheminformatics in the analysis of biologically active peptides from food sources. *J Funct Foods* 16:334–351
- Scotti L, Júnior FJBM., Ishiki HM, Ribeiro FF, Duarte MC, Santana GS, Oliveira TB, Formiga Melo Diniz MF, Quintans-Júnior LJ, Scotti MT (2018) In: Grumezescu AM, Holban AM (eds) Computer-aided drug design studies in food chemistry. Natural and artificial flavoring agents and food dyes. Elsevier BV, Amsterdam, pp 261–297
- Udenigwe CC, Aluko RE (2011) Chemometric analysis of the amino acid requirements of antioxidant food protein hydrolysates. *Int J Mol Sci* 12(5):3148–3161
- Abdi H, Williams LJ (2010) Principal component analysis. *WIREs Comput Stat* 2(4):433–459. <https://doi.org/10.1002/wics.101>
- Shu M, Mei H, Yang S, Liao L, Li Z (2009) Structural parameter characterization and bioactivity simulation based on peptide sequence. *QSAR Comb Sci* 28(1):27–35
- Wu J, Aluko RE, Nakai S (2006) Structural requirements of angiotensin-I converting enzyme inhibitory peptides: quantitative structure-activity relationship study on di- and tripeptides. *J Agric Food Chem* 54:732–738
- Lata S, Sharma BK, Raghava GPS (2007) Analysis and prediction of antibacterial peptides. *BMC Bioinform* 8:236
- Demšar U, Reades J, Manley E, Batty M (2017) Revisiting the past: replicating fifty-year-old flow analysis using contemporary taxi flow data. *Ann Am Assoc Geogr* <https://doi.org/10.1080/24694452.2017.1374164>
- Jolliffe IT (2002) Mathematical and statistical properties of population principal components. In: Jolliffe IT (ed) *Principal Component Analysis*. Springer Series in Statistics, 2nd edn Springer, New York, pp 1–518
- Kim H-J (2008) Common factor analysis versus principal component analysis: choice for symptom cluster research. *Asian Nurs Res* 2(1):17–24
- Iwaniak A, Minkiewicz P, Darewicz M, Sieniawski K, Starowicz P (2016) BIOPEP database of sensory peptides and amino acids. *Food Res Int* 85:155–161
- Zimmerman JM, Eliezer N, Simha RJ (1968) The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol* 21(2):170–201
- Grantham R (1974) Amino acid difference formula to explain protein evolution. *Science* 185(4154):862–864
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157(1):105–132
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) *The proteomics protocols handbook*. Humana Press, New York, pp. 571 – 60
- Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Markley JL (2008) BioMagResBank. *Nucl Acids Res* 36:D402–D408
- Iwaniak A, Hryniewicz M, Bucholska J, Minkiewicz P, Darewicz M (2018) Understanding the nature of bitter-taste di- and tripeptides derived from food proteins based on chemometric analysis. *J Food Biochem* 42:e12500
- Bartlett MS (1954) A note on multiplying factors for various chi-squared approximations. *J R Stat Soc Series B* 16:296–298
- Soltani S, Haghaei H, Shayanfar A, Vallipour J, Zeynali KA, Jouyban A (2013) QSBP study of bitter peptides: application of GA-PLS in combination with MLR, SVM, and ANN approaches. *BioMed Res Int* 2013:501310. <https://doi.org/10.1155/2013/501310>
- Arrindell WA, Van Den Ende J (1985) An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Appl Psychol Meas* 9(2):165–178
- Stanisz A (2007) Statistical course with an application of STATISTICA PL on medicine examples. In: Stanisz A (ed) *Principal component analysis*. StatSoft Cracow, Poland, pp 165–181 (in Polish)
- O'Rourke N, Hatcher L, Stepanski EJ (2005) Principal component analysis. In: O'Rourke N, Hatcher L, Stepanski EJ (eds) *A step-by-step approach to using SAS for univariate and multivariate statistics*. 2nd edn, SAS Institute, Cary, pp 453–455
- Bryant FB, Yarnold PR (1995) Principal-Components Analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR (eds) *Reading and understanding multivariate statistics*. American Psychological Association, Washington, DC, pp 99–136
- Jackson DA (1993) Stopping rules in principal components analysis: a comparison of heuristic and statistical approaches. *Ecology* 74(8):2204–2214
- Milovanović I, Mišan A, Simeunović J, Kovač D, Jambrec D, Mandić A (2015) Determination of volatile organic compounds in selected strains of Cyanobacteria. *J Chem* 2015:969542. <https://doi.org/10.1155/2015/969542>
- Kim H-J, Li-Chan ECY (2006) Application of Fourier transform Raman spectroscopy for prediction of bitterness of peptides. *Appl Spectrosc* 60(11):1297–1306
- Meahashi K, Huang L (2009) Bitter peptides and bitter taste receptors. *Cell Mol Life Sci* 66(10):1661–1671
- Nongonierma AB, FitzGerald RJ (2016) Learnings from quantitative structure-activity relationship (QSAR) studies with respect to food protein-derived bioactive peptides: A review. *RSC Adv* 6:75400–75413
- Pripp AH, Ardö S (2007) Modelling relationship between angiotensin-(I)-converting enzyme inhibition and the bitter taste of peptides. *Food Chem* 102(3):880–888

38. Iwaniak A (2011) Analysis of relationship between the structure and ACE inhibitory activity food derived peptides. The evaluation of suitability of *in silico* methods in the research on protein precursors of bioactive peptides. UWM Olsztyn, Poland, pp 1–152 (in Polish)
39. Alizadeh-Pasdar N, Nakai S, Li-Chan ECY (2002) Principal component analysis of Raman spectra to study the effects of pH, heating and κ -carrageenan on whey protein structure. *J Agric Food Chem* 50(21):6042–6052
40. Wold S (1987) Principal component analysis. *Chemometr Intell Lab Syst* 2(1–3):37–52