Journal of Molecular Structure 998 (2011) 53–61

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstruc

Quantitative structure–activity relationship study of antioxidative peptide by using different sets of amino acids descriptors

Yao-Wang Li [a,b], Bo Li [a,b,*], Jiguo He [a], Ping Qian [c]

[a] College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China
[b] Key Laboratory of Functional Dairy, Ministry of Education, Beijing 100083, China
[c] The Quartermaster Equipment Institute, G.L.D., Beijing 100010, China

ARTICLE INFO

Article history:
Received 23 February 2011
Received in revised form 16 April 2011
Accepted 6 May 2011
Available online 11 May 2011

Keywords:
Quantitative structure–activity relationship
PLS
Antioxidative peptide
Descriptors
Structural characteristics

ABSTRACT

A database consisting of 214 tripeptides which contain either His or Tyr residue was applied to study quantitative structure–activity relationships (QSAR) of antioxidative tripeptides. Partial Least-Squares Regression analysis (PLSR) was conducted using parameters individually of each amino acid descriptor, including Divided Physico-chemical Property Scores (DPPS), Hydrophobic, Electronic, Steric, and Hydrogen (HESH), Vectors of Hydrophobic, Steric, and Electronic properties (VHSE), Molecular Surface-Weighted Holistic Invariant Molecular (MS-WHIM), isotropic surface area–electronic charge index (ISA–ECI) and *Z*-scale, to describe antioxidative tripeptides as ***X***-variables and antioxidant activities measured with ferric thiocyanate methods were as ***Y***-variable. After elimination of outliers by Hotelling's T^2 method and residual analysis, six significant models were obtained describing the entire data set. According to cumulative squared multiple correlation coefficients (R^2), cumulative cross-validation coefficients (Q^2) and relative standard deviation for calibration set (RSD_c), the qualities of models using DPPS, HESH, ISA–ECI, and VHSE descriptors are better ($R^2 > 0.6$, $Q^2 > 0.5$, $RSD_c < 0.39$) than that of models using MS-WHIM and *Z*-scale descriptors ($R^2 < 0.6$, $Q^2 < 0.5$, $RSD_c > 0.44$). Furthermore, the predictive ability of models using DPPS descriptor is best among the six descriptors systems (cumulative multiple correlation coefficient for predict set (R^2_{ext}) > 0.7). It was concluded that the DPPS is better to describe the amino acid of antioxidative tripeptides. The results of DPPS descriptor reveal that the importance of the center amino acid and the N-terminal amino acid are far more than the importance of the C-terminal amino acid for antioxidative tripeptides. The hydrophobic (positively to activity) and electronic (negatively to activity) properties of the N-terminal amino acid are suggested to play the most important significance to activity, followed by the hydrogen bond (positively to activity) of the center amino acid. The N-terminal amino acid should be a high hydrophobic and low electronic amino acid (such as Ala, Gly, Val, and Leu); the center amino acid would be an amino acid that possesses high hydrogen bond property (such as base amino acid Arg, Lys, and His). The structural characteristics of antioxidative peptide be found in this paper may contribute to the further research of antioxidative mechanism.

1. Introduction

The hydrolysates from various proteins, such as soybean, casein, bullfrog, royal jelly, venison, r-lactalbumin, myofibrillar, rice endosperm, have been shown to have antioxidant activities against the peroxidation of lipids or radical scavenging activities [1–9]. Thus, a number of antioxidative peptides, usually composed of 3–16 amino acid residues, have been isolated and identified from these hydrolysates, and their antioxidant activities have been investigated to gain insight into the antioxidative mechanism of peptides. Several amino acid residues, such as His, Met, Tyr, Cys, and Trp, are generally accepted as antioxidants in spite of their pro-oxidative effects in some cases [10]. Among these peptides, some contained hydrophobic amino acids (Val or Leu) at the N-terminus, but it was reported that the deletion of the N-terminal Leu had no effect on the activity [1,11]. Some peptides that include basic amino acid residues such as His and Lys possess high antioxidant activities [11], however, the peptides containing mainly acidic amino acid residues (Glu, Asp) also possess high antioxidant activities [4]. The results obtained so far indicate that the antioxidant activities of peptides are attributed to the cooperative effects of the whole amino acid sequence of peptide, but there is still little information concerning the structural characteristics of antioxidative peptides.

Antioxidative peptides are potentially novel antioxidants and radical scavengers and understanding and predicting peptide

* Corresponding author at: College of Food Science and Nutritional Engineering, China Agricultural University, No. 17, Qinghua East Road, Haidian District, Beijing 100083, China. Tel.: +86 10 62736243.
E-mail address: libo@cau.edu.cn (B. Li).

0022-2860/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.molstruc.2011.05.011

structures responsible for different antioxidant activities is therefore of interest. QSAR modeling provides methodology to find mathematical expressions for such relationships which may be useful for estimating activities of any compound and for predicting structures of high activity. In medical sciences and toxicology, QSAR modeling has been used extensively for such purposes. The idea behind this methodology is that biological activity is a function of chemical structures that can be described by molecular or physicochemical variables, e.g. electronic attributes, hydrophobicity and steric properties [12].

Popular molecular structural characterization approaches are mainly typed for two typical kinds; one is based on two-dimensional (2D) structures and the other three-dimensional (3D) spatial conformations [13]. For the 2D descriptor, traditional 2D molecular structural descriptors mainly include topological index and physicochemical parameters. The pioneer Sneath obtained some semi-quantitative descriptors from physicochemical properties of 20 natural amino acids and applied to oxytocin vasopressine analogues [14]. Kidera et al. extracted 10 orthogonal factors based on 188 kinds of properties of 20 natural acids in 1985 [15]. *Z*-scale descriptor is popular in recent researches. Hellberg et al. achieved *Z*-scale that contains three scales from 29 physicochemical variables of 20 coded amino acids by principal components analysis (PCA) in 1987 and *Z*-scale descriptor was firstly validated by the use in QSAR of Oxytocin Analogues [16], and later was successfully applied to QSAR models of bitter tasting di-peptides (BTD) and angiotensin converting enzyme (ACE) inhibitors successfully by followers later [17,18]. In 1998, Sandberg et al. extended the *Z*-scale and developed a new *Z*-scale that contains five scales from 29 physicochemical variables of 87 amino acids and was applied to Quantitative Sequence-Activity Modeling (QSAM) of 89 elastase substrate analogues and QSAM of 29 neurotensin analogues [19]. Mei et al. derived VHSE (Vectors of Hydrophobic, Steric, and Electronic properties) from the 18 hydrophobic properties, 17 steric properties, and 15 electronic properties [20]. In 2009, Tian et al. developed DPPS scale based on 23 kinds of electronic properties, 37 kinds of steric properties, 54 kinds of hydrophobic properties and 5 kinds of hydrogen bond properties of amino acids [21] and Shu et al. derived HESH descriptor from 171 physicochemical properties of 20 coded amino acids for structural parameter characterization and bioactivity simulation based on peptide sequence [22]. Tian et al. also established *T*-scale as a novel vector of topological descriptor for amino acids based on 67 kinds of structural topological variables of 135 amino acids [23]. Liu et al. derived a novel molecular holographic distance vector (MHDV) to characterize the structures of the peptide molecules and employed to relate to biological activities of the peptides by means of principal component regression (PCR) method [24]. Based on the distance between atoms and the electro-negativity of each atom, a new set of descriptors call the molecular electro-negativity edge vector (VMEE) be developed by Li et al. and be applied to describe the molecular structure of ACE inhibitors and bitter tasting di-peptides [25]. There were reports of the use of other descriptors such as VSTV [26], SZOTT [27], FASGAI [28], and VSW [29]. For the 3D descriptor, in 1995, Collantes et al. developed new 3D descriptors for QSAR studies of peptide analogues, named ISA (isotropic surface area)–ECI (electronic charge index) scales with parameterizations that base on three-dimensional structure of amino acid side chains [30]. Bravi et al. also developed 3D descriptors called MS-WHIM which focuses on the Molecular Electrostatic Potential (MEP) calculation on the molecular surface [31].

Model-based methods can be successfully used for design, optimization molecular structure with high bioactivity. Ideal models would be those that could quantitatively relate structure to activity or property. However, building sound models requires suitable parameters related with chemical structures. Therefore, the key issue is to choose the most appropriate amino acid descriptors. Peptide QSAR modeling has been especially useful for antimicrobial, ACE-inhibitory and bitter tasting peptides, but could easily be expanded to antioxidative peptides within food research. In this manuscript, two kinds of descriptors, including six amino acid descriptors, have been applied to antioxidative peptides, and the qualities of models using six descriptors were investigated. The physicochemical properties involved in antioxidant activity, and potential applications of QSAR modeling within antioxidative peptides are discussed.

2. Materials and methods

2.1. Dataset prepared

The tripeptide database composed of 214 peptides was obtained from the published literature [32], and antioxidant activities were measured by the ferric thiocyanate methods [33], and the results are shown as relative activities by adjusting the control to 1.0 (Table 1). The activities of tripeptides were extracted from the graph by xyExtract software to make up of the ***Y*** value matrix of 214 tripeptides. Model validation is an absolutely necessary step in QSAR modeling. In this study, the database was randomly divided into calibration set and prediction set with 2:1, that means calibration set is 143 samples and prediction set is 71 samples. The calibration dataset was used to establish QSAR models and perform internal validation. On the base of internal validation, external validation was also performed using the prediction dataset. Several evaluation functions were used to evaluate predictive power of the resulting QSAR models.

2.2. Descriptors of amino acids

The first step is to map structural characteristics of target sequences into numerical variables, and during this process, useful information is supposed to maximally keep in the model when filtering noise. Six descriptors of amino acid have been chosen from references (Table 2). They can be categorized in two groups. One group is physiochemical descriptors of amino acids, including DPPS, HESH, VHSE, and *Z*-scale; the other is 3D descriptors, including ISA–ECI and MS-WHIM.

The approach to QSAR modeling in this study is to quantitatively characterize properties of the individual amino acids in the peptide sequence. Each variable express the amino acid position and the amino acid property. The number of amino acids in tripeptides is 3, so the total number variables of the peptide is 3 multiply the number of descriptor' variable. Such as *Z*-scale has three variables, and then the variable number of peptides was $3 \times 3 = 9$, and the variables of peptides formed the ***X*** matrix, and the sequence of variables was from N-terminus to C-terminus. And the antioxidant activity makes up the ***Y*** matrix. The proper descriptors have been selected base on physicochemical means of descriptor and the quality and predictive ability of PLS model.

2.3. PLS modeling

The details about PLS could be found in references [34,35]. All PLSR analysis between amino acid descriptors matrix ***X*** and antioxidant activity matrix ***Y*** were processed by PLS program which was written by us on Matlab software (7.6.0 version, The Mathworks, Inc.). Meanwhile, the internal validation and the external validation were designed to evaluate the quality and predictive ability of model in the PLS program, all variables were centered and so had an equal participation in the model, and the sample database was divided into the calibration set and prediction set automatically. For the cal-