Research Article

Journal of PeptideScience

Received: 21 September 2010 Revised: 30 October 2010 Accepted: 22 November 2010 Published online in Wiley Online Library: 14 April 2011

(wileyonlinelibrary.com) DOI 10.1002/psc.1345

Structure–activity relationship study of antioxidative peptides by QSAR modeling: the amino acid next to *C*-terminus affects the activity

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

Screening, isolation and *in vitro* assays have been used for characterization of antioxidative peptides derived from food proteins, and incompatible deductions of structural characteristics derived from the isolated peptides have been brought forward. However, there is still little information concerning the structure-activity relationship of antioxidative peptides. QSAR modeling was performed, respectively, on synthetic tripeptides and tetrapeptides related to LLPHH. According to cumulative squared multiple correlation coefficients (R^2), cumulative cross-validation coefficients (Q^2) and relative standard deviation for calibration set (RSD_c), two credible models for tripeptide and tetrapeptide databases, respectively, have been built with partial least squares (PLS) regression (R^2 for models of tripeptide and tetrapeptide are 0.744 and 0.943, Q^2 are 0.631 and 0.414, and RSD_c are 0.323 and 0.111, respectively). Meanwhile, according to the cumulative multiple correlation coefficient for the predictive set (R^2_{ext}) and the relative standard deviation for the predictive set (RSD_p), the predictive ability of the model for tripeptides also is excellent (R^2_{ext} and RSD_p are 0.719 and 0.450, respectively). Hydrogen bond property and hydrophilicity of the amino acid residue next to the *C*-terminus, and the hydrophobicity as well as electronic propertyof the *N*-terminus are more significant; meanwhile, the electronic property of the *C*-terminus is beneficial for antioxidant activity. The structural characteristics we found are very useful in understanding and predicting the peptide structures responsible for activity and development of functional foods with peptides as active compounds, or antioxidative peptides as alternatives to other antioxidants. Copyright © 2011 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: : antioxidant activity; peptides; structure-activity relationship; PLS; QSAR

Introduction

The hydrolyzates from various proteins, such as soybean, casein, Alaska Pollack, alfalfa leaf, bullfrog, royal jelly, venison, pork collagen, r-lactalbumin, myofibrillar, rice endosperm, and Yellow Stripe Trevally [1–15], have been shown to have antioxidant activities against the peroxidation of lipids or radical scavenging activities. Thus, a number of antioxidative peptides, usually composed of 3–16 amino acid residues, have been isolated and identified from these hydrolyzates, and their antioxidant activities 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 [16,17]. Among these peptides, some contained hydrophobic amino acids (Val or Leu) on the *N*-terminus, but it was reported that the deletion of the *N*-terminus Leu had no effect on the activity [1,18]. Some peptides that include basic amino acid residues such as His and Lys possess high antioxidant activities [18,19]; however, some contained mainly acidic amino acid residues (Glu and Asp) [5]. 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.

At present, the main strategy has been to identify and characterize the antioxidative peptides from the hydrolyzate of protein, but methods describing the relationship between peptide structure and antioxidative activity are needed to predict the antioxidant potential of food protein hydrolyzates. The Muramoto researching group tried to investigate the residue–activity relationship of antioxidative peptides by combinatorial chemistry [1,2,20]. Incipiently, they isolated six antioxidative peptides from soybean protein. On the basis of the smallest peptide, LLPHH, 28 synthetic peptides were constructed and their antioxidant activities compared. The results indicated that *N*-terminal Leu had no effect, and His and Pro played important roles in the antioxidant activity. Further study with 22 synthetic peptides containing His residues demonstrated that His-containing peptides can act as a metal-ion chelator, an active-oxygen quencher, and a hydroxyl radical scavenger [2]. Subsequently, the Muramoto group constructed two series of tripeptide libraries (totally 222 peptides) prepared by the combinatorial chemistry to explore the antioxidative properties of peptides. Among the tested

* *Correspondence to: Bo Li, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, 100083, China. E-mail: libo@cau.edu.cn*

a *College of Food Science and Nutritional Engineering, China Agricultural University, Qinghua East Road 17, Haidian District, Beijing 100083, People's Republic of China*

b *The Quartermaster Equipment Institute, G.L.D, Beijing 100010, People's Republic of China*

peptides, Tyr-(His, Lys, Arg)-Tyr was found to have the highest antioxidant activity. However, taking into account the large amount of theoretical possible peptides, i.e. 400 dipeptides, 8000 tripeptides, 160 000 tetrapeptides, etc., the examination of all possible peptides to find highly efficient antioxidants would be a daunting task.

One approach is to develop statistical models that predict the relationship between structure (amino acid sequence) and activity. The approach is referred to as QSAR modeling. Such models could then be used for prediction and synthesis of bioactive peptides as well as to contribute to biochemical understanding of which peptides show activity. In medical sciences and toxicology, QSAR modeling has been used extensively for such purposes. A sub-field, peptide QSAR, has been successfully established on ACE-inhibitory peptides [21,22], circular peptides [23], bitter-tasting dipeptides, and bradykinin-potentiating pentapeptides [24]. The idea behind this methodology is that biological activity is a function of chemical structures which could be described by molecular or physicochemical variables, e.g. electronic attributes, hydrophobicity, and steric properties [25,26], resulting in multivariate data. A key issue is to choose the most appropriate amino acid descriptors for QSAR modeling. So far, large number sets of descriptors for amino acids are available from research literature and databases, e.g. Z-scale, t-scales, ISA-ECI (isotropic surface area-electronic charge index), Molecular Surfaces-Weighted Holistic Invariant Molecular (MS-WHIM) scores, Vectors of Hydrophobic, Steric, and Electronic properties (VHSE), Vector of Structural and Topological Variables (VSTV), two-dimensional or three-dimensional descriptors, etc. [27–29]. In order to understand the structural characteristics of peptides with definite physiochemical meanings, as well as contributing to a better insight into biochemical mechanisms, we selected in this article a set of physicochemical descriptors named Divided Physico-chemical Property Scores (DPPS) descriptors, which relate to hydrophobic properties, steric properties, electronic properties, and hydrogen bond.

As more specific information is needed to understand relationships between peptide structures and their antioxidant activities, the aim of this study was to develop the structural characteristics of antioxidative peptide sequence, to predict the amino acid sequence of antioxidative peptides, as well as provide a better understanding of the physicochemical mechanisms involved.

Materials and Methods

The tripeptide database was obtained from the report of Saito *et al*. [20], and antioxidant activity was measured by the ferric thiocyanate method. The antioxidant activities of tripeptides in the aqueous autoxidation system of linoleic acid were extracted from the graph using xyExtract software (Campina Grande, Paraíba, Brazil) to make up the ***Y*** value matrix of 214 tripeptides. The tetrapeptide database, in which the activity was also measured by the ferric thiocyanate method, was obtained from the report of Chen *et al*. [1].

The DPPS descriptors [30] (Table 1) were applied to describe the database of tripeptides and tetrapeptides. The DPPS descriptors include hydrophobic property, steric property, electronic property and hydrogen bond property describing the structure of peptides. It was developed by Tian *et al*. 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. *V1–V4* represent the electronic property of amino acid, *V5–V6* steric property, *V7–V8* hydrophobicity, and *V9–V10* hydrogen bond, respectively. Each amino acid was described by 10 scores, and then the tripeptide had $3 \times 10 = 30$ descriptors, and was expressed as *X1–X30*. *X1–X10* described the *N*-terminal amino acid residue, *X11–X20* the second amino acid residue from the *N*-terminus to the *C*-terminus, and *X21–X30* described the *C*-terminal amino acid residue. The 30 variables constituted the ***X*** matrix. So it included 40 variables being expressed as *X1–X40* for the tetrapeptide from the *N*-terminus to the *C*-terminus. Partial least squares (PLS) multiple linear regression analysis with leave-one-out validation was performed using the Matlab7.0 software (The Mathworks, Inc., Natick, Massachusetts, USA). Model validation is an absolutely necessary step in QSAR modeling. In this study, the tripeptide database was randomly divided into calibration set and predictive set with 2 : 1. It means that the calibration set has 143 samples and the predictive set has 71 samples. The calibration dataset was used to establish QSAR models and perform internal validation. On the basis of internal validation, external validation was also performed using the predictive dataset. The tetrapeptide database was only used to build the model and for internal validation.

R^2 is the coefficients of determination for the regression of observed *versus* calculated activities of the calibration set, which can amount to ***Y***, 'explained' in terms of the sum of squares of ***Y*** and provide an estimate of the multiple fit. F-test and relative standard deviation for calibration set (RSD_c) were used with R^2 as they could evaluate and compare the quality of fitting between the observed values and calculated values before and after outliers were excluded in the calibration dataset. RSD_c and relative standard deviation for predictive set (RSD_p) are convenient to compare the quality of different models as it is a relative value (dimensionless). The higher the R^2 and F value with lower RSD_c value, the better fitting ability of the model. Q^2 (the cross-validation R^2) is a parameter to determine the actual predictive power and stability of a QSAR model, which is based on predictive residual sum of squares (PRESS) and $Q^2 = (1 - \text{PRESS/RSS})$. PRESS is the squared difference between observed and predicted values for the data kept out of the model fitting, and RSS is the residual sum of squares. It has been pointed out in recent literature that Q^2 is insufficient to determine the actual predictive power of a QSAR model and that external validation is required [31,32]. In this study, R^2_{ext} and RSD_p of the predictive set were also calculated to evaluate the quality of fitting between observed values and predicted values as external validation.

Outliers have a great negative effect on the quality of a model. For obtaining a good model and in an attempt to improve the predictive ability of the model, outliers are excluded before modeling for calibration set. In this manuscript, the Hotelling's T^2 [30] and residue analysis have been applied to eliminate the outliers [33].

Results and Discussion

Relationship of Antioxidant Activities of Tripeptides and Tetrapeptide with their Electronic Property, Steric Property, Hydrophobicity, and Hydrogen Bond

The antioxidative tripeptide libraries were obtained from the report of Saito and others [20]. The database, which was used in this study, was made up of two libraries. One was a library of 108 peptides containing either His or Tyr residues. In this library, each tripeptide contains 2 His or Tyr, the other position was 1 of the

Table 1. The DPPS descriptors of amino acids [31]

Amino acids	Abbreviation	Electronic property				Steric property		Hydrophobicity		Hydrogen bond	
		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Ala	A	−1.02	−2.88	−0.56	0.36	−6.15	−1.68	0.04	−2.51	−1.94	−0.01
Arg	R	1.99	4.13	−4.41	−1.02	4.78	3.04	−9.06	6.71	4.41	0.07
Asn	N	−2.19	1.86	0.38	−0.13	−2.30	1.41	−5.71	−1.11	1.73	−0.19
Asp	D	−6.60	3.32	1.61	0.36	−3.25	1.95	−7.36	0.14	1.24	−0.15
Cys	C	0.21	1.12	3.42	−0.68	−2.27	−1.22	3.11	−2.98	−1.70	1.57
Gln	Q	−0.47	1.16	−0.57	0.69	0.39	1.93	−5.46	−0.84	1.93	0.85
Glu	E	−5.39	0.65	−0.98	1.39	−0.23	2.51	−6.84	−0.68	1.41	1.28
Gly	G	−2.86	−5.00	−2.97	0.53	−11.45	1.89	−2.11	−3.99	−2.16	−0.76
His	H	0.73	2.68	−0.66	−1.89	1.60	1.13	−1.94	−0.11	0.44	0.15
Ile	I	1.91	−3.13	0.01	1.14	2.70	−4.55	8.93	0.18	−1.10	−0.76
Leu	L	1.64	−2.57	0.00	1.35	2.62	−2.65	7.72	0.05	−1.03	−1.81
Lys	K	2.47	1.54	−4.28	−0.86	2.77	2.06	−6.18	2.05	2.19	−1.65
Met	M	1.93	−0.01	1.21	0.99	2.79	−0.56	5.33	−0.87	−0.99	−1.09
Phe	F	2.68	0.84	2.22	0.71	5.02	−0.30	8.60	1.13	−1.40	−0.28
Pro	P	0.45	−2.89	1.77	−5.81	−3.79	−0.61	0.70	1.21	−1.67	1.79
Ser	S	−1.76	−0.19	1.06	−0.69	−5.72	0.14	−4.14	−2.42	−0.13	0.69
Thr	T	−0.55	−0.66	0.13	−0.31	−2.76	−1.56	−2.46	−2.12	0.17	0.08
Trp	W	3.88	1.78	1.68	2.00	9.31	0.89	7.53	4.27	−0.23	−1.42
Tyr	Y	2.10	1.26	1.15	0.91	5.90	0.74	3.71	3.32	0.25	1.33
Val	V	0.83	−3.02	−0.22	0.97	0.05	−4.55	5.61	−1.41	−1.44	0.30

Table 2. Statistical parameters of PLS model in QSAR of antioxidative tripeptides and tetrapeptides

Statistic	n[a]	R^{2b}	Q^{2c}	R^{2}_{ext} [d]	M[e]	F[f]	RSDc[g]	RSDp[h]
Tripeptide	102	0.744	0.631	0.719	7	73.615	0.323	0.450
Tetrapeptide	12	0.943	0.414	–	3	44.336	0.111	–

[a] The number of samples in the study.
[b] The cumulative multiple correlation coefficient for calibration set.
[c] The cumulative cross-validation coefficients for calibration set.
[d] The cumulative multiple correlation coefficient for predictive set.
[e] The number of principal components.
[f] The F value in F-test.
[g] The relative standard deviation of calibration set.
[h] The relative standard deviation of predictive set.

other 19 natural amino acids in tripeptide sequence for observing the influence of the other 19 amino acids at different positions on the antioxidant activity. Another was a library of 114 peptides related to Pro–His–His, which had been identified as an active core of the antioxidative peptide. In this study, the approach to QSAR modeling is to quantitatively characterize properties of the individual amino acid in the peptide sequence. Each variable in the dataset expressed the sequence position and an amino acid property. The DPPS descriptors, which include electronic property, steric property, hydrophobicity, and hydrogen bond, were chosen as the amino acid descriptors (Table 1). Then three amino acid positions in the sequence and ten descriptors give a dataset with 30 variables (*X1–X30*), expressing the chemical structure of the tripeptides. The variables together with the response (antioxidant activities) were modeled by PLS. The statistical parameters of PLS are shown in Table 2. From Table 2, the value of R^2, RSD_c, and Q^2 are 0.744, 0.323, and 0.631 respectively, which means that the extracted principal components have positive effects on the predictive ability of the QSAR model. Figure 1(A) is a plot of the calculated activities against observed activities for the tripeptide calibration set, indicating a linear relation, and most samples are closely dispersed along an origin-passed line forming an angle of 45°, with only a few outliers especially caused by samples. It reviews that the activities for the 102 tripeptides are modeled well, although some systematic differences exist between them. The quality of the model for the tetrapeptide is better than that for the tripeptide (R^2 is 0.943 and 0.744, Q^2 is 0.414 and 0.631 for tetrapeptide and tripeptide, respectively).

A predictive set of 71 peptides was marked off randomly. The predictive set was applied to validate the predictive power of the QSAR model. It yielded an R^2 of 0.719 and an RSD_p of 0.450 respectively; the result illustrated that the model was creditable. The relationship between observed and predicted activity for the predictive set is illustrated in Figure 1(B). The results show that the model proposed in this study has potential for estimating activities of peptides and for predictive structures of high activity.

From the loading plot of the first two components given in Figure 2(A), the important contributions to the model are *X20, X8, X2, X4, X1, X7, X11, X28, X24, X5, X15, X17,* and *X19*. Among these variables, *X20, X4, X1, X7, X11, X28, X5,* and *X19* contribute

Journal of PeptideScience

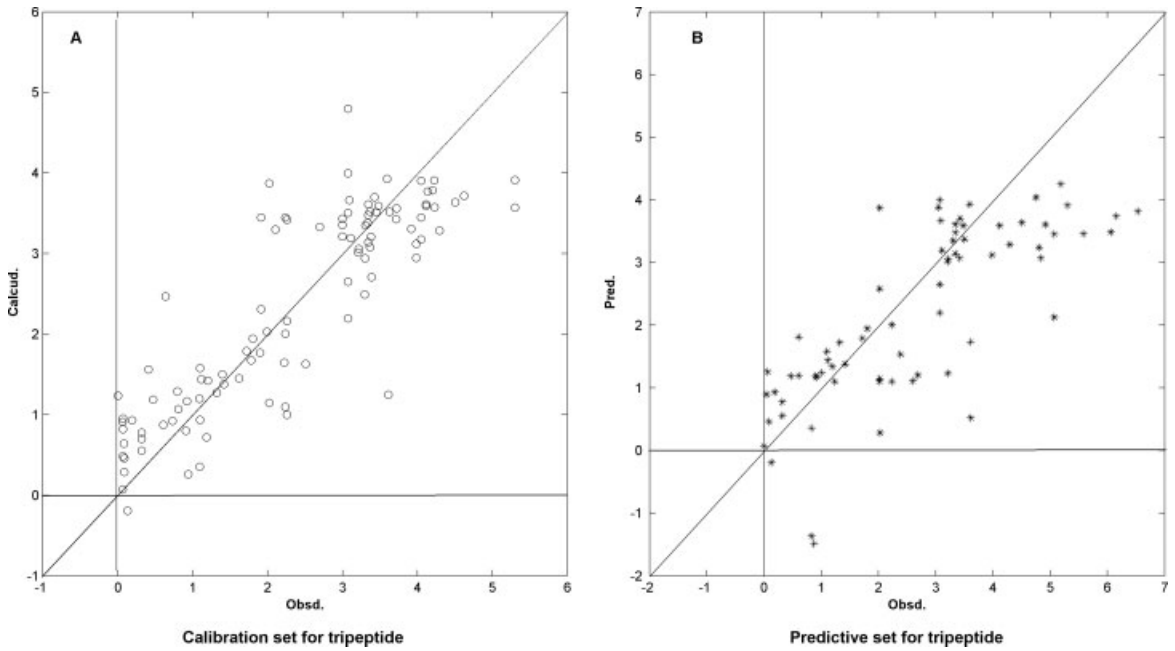

Figure 1. The calculated value *versus* observed value plot of calibration set (A) and the predicted value versus observed value plot of predictive set in tripeptides (B).

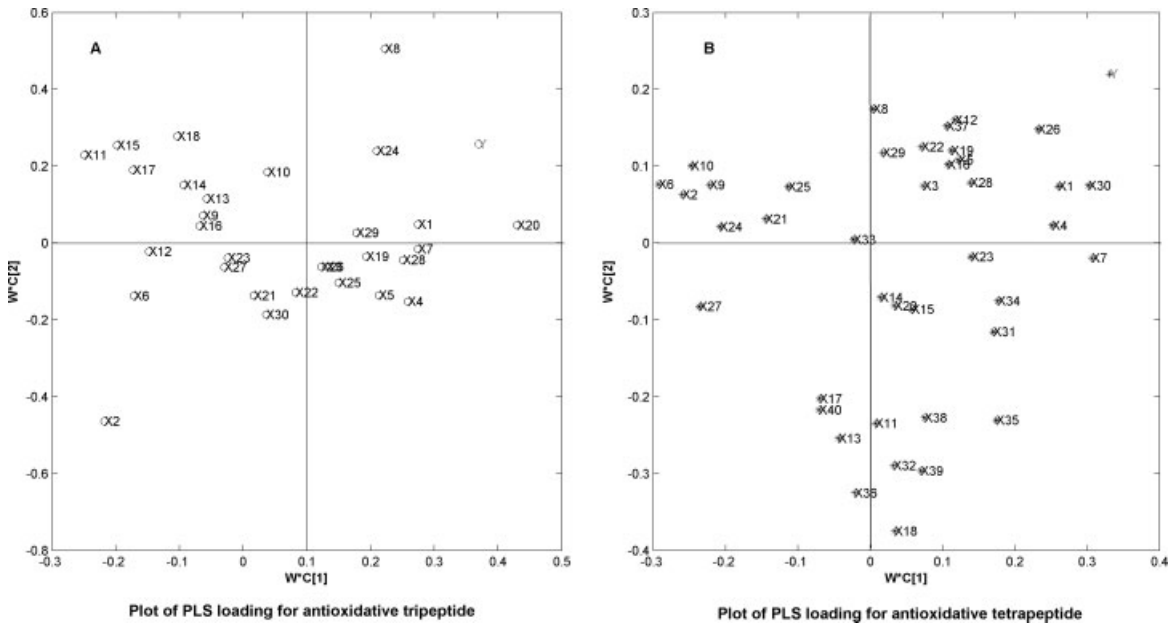

Figure 2. The PLS loading plot for antioxidative tripeptides (A) and tetrapeptides (B).

to the first component, *X20, X4, X1, X7, X28, X5,* and *X19* have the positive correlation with *Y* and *X11* has the negative correlation with *Y*. Variables *X8, X2, X24,* and *X15* contribute to the second component, *X8, X24* and *X15* have a positive correlation with *Y* and *X2* has a negative correlation with *Y*. The size of the contribution for variables to activity is revealed in the variable importance in project (VIP) plot (Figure 3(A)). The order of the important variables is *X20*>*X8*>*X2*>*X4*>*X1*>*X7*>*X11*>*X28*>*X24*>*X5*>*X15*>*X17*>*X19* (VIP value>1). *X20 and X19* illustrate the significance of the hydrogen bond of the second amino acid, the *X8, X7*, and *X2, X4, X1* show the importance of hydrophobicity and electronic property in the *N*-terminal amino acid, respectively; for the *C*-terminal amino acid, mainly focus is on the hydrophobicity and electronic property. Meanwhile, the *X20, X8, X4, X7, X28, X24,* and *X19* are the positive effect on antioxidant activity; the *X2, X1, X11, X5, X15,* and *X17* are the negative effect on antioxidant activity (Figure 4(A)). That is to say, the antioxidant activity of peptides is mainly related to the hydrogen bond and electronic property of the second amino acid residue, hydrophobicity and electronic property in the *N*-terminal amino acid, electronic property and hydrophobicity (positive contribution to antioxidant activity) in the *C*-terminal amino acid residue as well.

Looking at the loading (Figure 2(B)) and VIP plots (Figure 3(B)) of the tetrapeptide, it is found that the variables *X7, X30, X6, X2, X1, X10,*

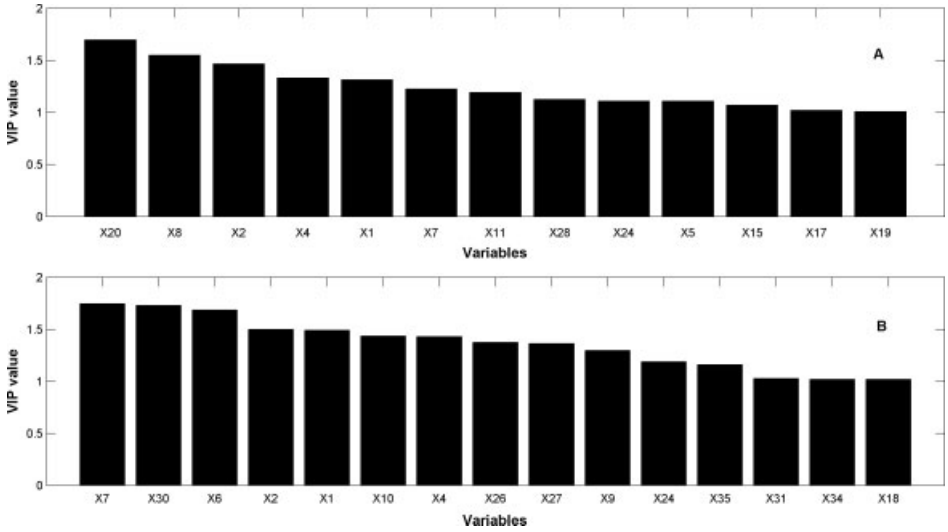

Figure 3. The PLS VIP plot for tripeptide and tetrapeptide The VIP plot of tripeptide is expressed in (A) and tetrapeptide in (B).

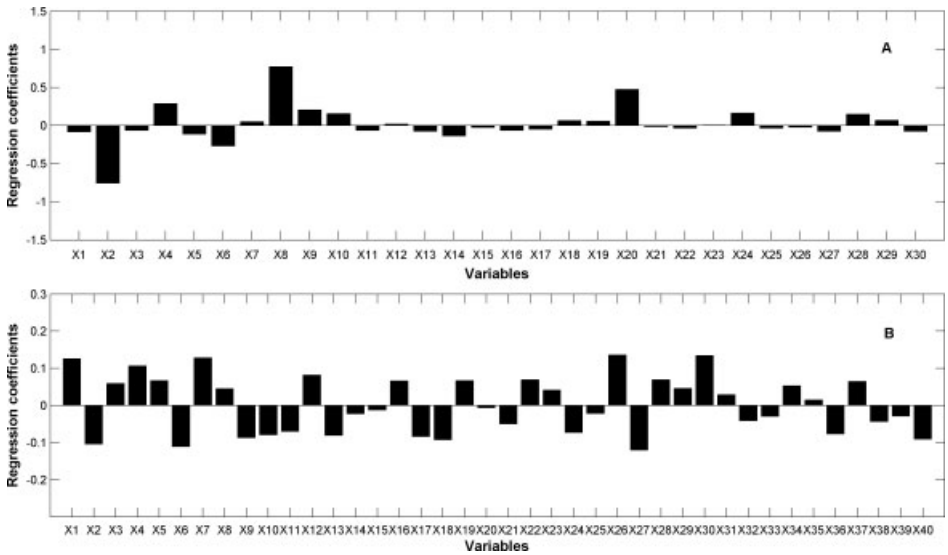

Figure 4. The plot of standard regression coefficients of variables for tripeptides (A) and tetrapeptides (B).

X4, X26, X27, X9, X24, X35, X31, X34, and *X18* have great correlation with *Y*. In the other view, *X7, X30, X1, X4, X26, X35, X31, and X34* have a positive effect on antioxidant activity; meanwhile, the *X6, X2, X10, X27, X9, X24,* and *X18* have a negative effect on antioxidant activity (Figure 4(B)). According to the position in amino acid sequence, the hydrophobicity (*X7*), steric property (*X6*), electronic property (*X2, X1*, and *X4*), and hydrogen bond property (*X10* and *X9*) of the *N*-terminal amino acid residue are very important. Meanwhile, in the third amino acid residue, hydrogen bond (*X30*), steric property (*X26*), hydrophobicity (*X27*), and electronic property (*X24*) play an important role in the antioxidant activity. It follows that the steric property (*X35*) and electronic property (*X31* and *X34*) in the *C*-terminus also contribute to antioxidant activity.

Structural Analysis of Tripeptides and Tetrapeptides using QSAR Modeling

According to the QSAR modeling resulting from the tripeptide database, the amino acid in the second position of the amino acid sequence plays a key role in antioxidant activity. However, in the amino acid sequence of the tripeptide, the amino acid in the second position is next to the *N*-terminus, as well as to the *C*-terminus. It is difficult to assess which of them is more important for the antioxidative peptide composed of more than three amino acid residues. A set of 14 tetrapeptides was then analyzed by QSAR modeling with the same descriptors.

Four amino acid positions in the sequence and ten descriptors give a dataset with 40 variables (*X1*–*X40*, serial number begins from *N*-terminal amino acid residue) expressing the chemical structure of the tetrapeptides. The statistical parameter of the PLS model is shown in Table 2. According to the R^2 ($R^2 = 0.943$), the QSAR model of tetrapeptides is said to have a high predictive power, as well as contributing to biochemical understanding of which peptides show activity. The calculated and observed antioxidant activities of 12 tetrapeptides (two samples were excluded as outliers) were shown in Table 3.

From the QSAR results of tripeptides and tetrapeptides, it is found that the structural profile from the QSAR modeling

Journal of PeptideScience

Table 3. The observed and calculated antioxidant activities of 12 tetrapeptides

Samples	Observed value	Calculated value	Samples	Observed value	Calculated value
HPLH	0.6	0.450	LPHH	2.7	2.522
HHLP	0.9	0.948	HHPL	2.9	2.525
HLPH	1.2	1.296	LHPH	2.9	3.056
PLHH	1.6	1.588	YLYP	2	2.005
HPHL	1.6	1.991	LYPY	2.4	2.424
LLHH	2.4	2.425	LPYY	2.5	2.472

of tetrapeptides is similar to that obtained from tripeptides. All of them reveal the importance of the *N*-terminus and the *C*-terminus. To the *N*-terminus, hydrophobicity and electronic property have a positive effect on antioxidant activity for the tripeptides and tetrapeptides. To the *C*-terminus, the results of tripeptides and tetrapeptides indicate the contribution of the electronic property (positive effect on antioxidant activity). Besides that, for the tetrapeptides, the third amino acid residue which is near the *C*-terminus is very important for antioxidant activity. The hydrogen bond and steric property of the third amino acid residue have a positive effect on antioxidant activity, and the hydrophobic property and electronic property have a negative effect on antioxidant activity. Meanwhile, the hydrogen bond in the second amino acid for the tripeptide has a great positive significance to antioxidant activity. The hydrogen bond properties extracted from five properties about the hydrogen bond are as follows: number of hydrogen bond donors, number of hydrogen bond acceptors, hydrogen bond donor factors, hydrogen bond acceptor factors, and information measure for those extended without the H-bond [34]. From the results of the tripeptide and tetrapeptide, it was found that the variable *V10* (*X20* in tripeptide and *X30* in tetrapeptide) has a great contribution to the antioxidant activity in the DPPS descriptor, which means mainly the information measure for those extended without the H-bond according to the analysis of the principal component *V10*.

According to the result from the QSAR modeling of tripeptides and tetrapeptides, the amino acid residues next to the *C*-terminus and the *N*-terminal amino acid residue are important for antioxidant activity followed by the *C*-terminus amino acid residue. The common structural characteristics of the antioxidative peptide are as follows: the amino acid residue next to the *C*-terminus not only has hydrogen bond property and electronic property but also has hydrophilic property; the *N*-terminal amino acid residue is hydrophobic and has electronic property; moreover, the electronic property of the *C*-terminus contributes to antioxidant activity.

Illustration of the Structural Characteristics of Antioxidative Peptides

According to the results from QSAR modeling, the amino acid residue next to the *C*-terminus and the *N*-terminus amino acid residue are more important than the *C*-terminus in contributing to antioxidant activity. The amino acid residue next to the *C*-terminus with hydrogen bond property and lower hydrophobicity will have higher antioxidant activity, the basic amino acid which has the lowest hydrophobicity and the highest hydrogen bond value will be most suitable for high antioxidant activity; to the *N*-terminus, the hydrophobicity is very prominent. Meanwhile, the *C*-terminus possesses electronic property which is favorable. It should be the first time that such structural characteristics of antioxidative peptides are being brought forward, and these characteristics corroborate the experimental results of earlier studies (Table 4).

The amino acid residue next to the *C*-terminus possesses hydrogen bond property which has an impressive effect on antioxidant activity. According to the results of experiments by Saito and others from which the tripeptide database was obtained, when the second position *X* is R, K, or H in pattern Y*X*Y, the relative activity is the highest. The activity of LH*X*, PH*X*, or RH*X* is remarkably higher than that of LW*X*, PW*X*, or RW*X*, respectively (*X*, any of 20 amino acids). Saito and others concluded that the tested peptides Tyr-(His, Lys, Arg)-Tyr showed the highest antioxidant activity [20]. These are also consistent with the typical structural characteristics of antioxidative peptides found in this study.

For the peptides composed of four or more amino acids, all of the four properties of the amino acid residue next to the *C*-terminus contribute to the activity. And the bigger the value of hydrogen bond and steric properties and lower the hydrophobicity it has, the higher the activity. Therefore, the basic and acidic amino acids (R, K, H, E, and D) and the other hydrophilic amino acids (T, S, N, and Q) have either a high value of hydrogen bond property or a high steric property, with low hydrophobicity. In this position, they would have higher activity than the other amino acids. Table 4 summarizes parts of peptide sequences with their antioxidant activities from the literature. Most peptides are accorded with these structural characteristics: the peptides from β-lactoglobulin hydrolyzate contain E, R, and D in the position next to the *C*-terminus, respectively [6]; and each peptide from prawn contains K. In the same position of peptides from casein, the amino acids are E [3], P, and Q [35] respectively.

The hydrophobicity of the *N*-terminal amino acid residue is also very important for antioxidant activity. The bigger the hydrophobicity, the higher is the activity. This is consistent with the earlier studies. Many studies speculated that the amino acid with hydrophobic property on the *N*-terminus (A, V, L, etc), played an important role in antioxidant activity [18,19]. After analysing the peptide sequences from different protein sources in Table 4, we can conclude that most antioxidative peptides have a hydrophobic amino acid on the *N*-terminus.

The electronic property of amino acid on the *C*-terminus has some effect on antioxidant activity. The bigger the value of electronic property, the higher is the activity. In other words, the *C*-terminus is a polar position, which is thus affected by electrostatic potential to some extent. Then amino acids W, E, L, I, M, V, Y etc. are suitable in the *C*-terminus. In the former studies, some researchers speculated that an amino acid on the *C*-terminus would play an important role in activity. Suetsuna and others [3] separated and identified a radical scavenging peptide, YFYPEL, from casein hydrolyzate, and it was confirmed that EL on the *C*-terminus mainly contributed to the antioxidant activity. Kim and others [13] speculated that the hydrophobic property of an amino acid on the *C*-terminus, e.g. Val, Leu, had a distinct effect on the activity based on the antioxidative peptides from venison hydrolyzate. However, as the result of QSAR modeling, the electronic property of the amino acid on the *C*-terminus contributed to antioxidant activity. We can find many peptides with W, E, L, I, Y, Q, and G in the *C*-terminus obtained from royal jelly and soybean protein (Table 4).

Chen and others [1] investigated 28 synthetic peptides based on an antioxidative peptide (LLPHH); they found that the deletion of the *C*-terminus His decreased the activity, whereas the deletion of the *N*-terminus Leu had no effect. It seemed that the *C*-terminus

Table 4. The peptide comes from various sources

No.	Sequence	Activity	Method[a]	Source	Literature
1	YVEEL	0.799	ORAC	β-lactoglobulin	[6]
2	MHIRL	0.306			
3	WYSLAMAASDI	2.621			
4	YFYPEL	79.2	SOSA		[3]
5	FYPEL	127.5			
6	YPEL	189.3			
7	PEL	306		Caseins	
8	VKEAMAPK	0.95	DPPH		[35]
9	AVPYPQR	1			
10	KVLPVPQK	0.99			
11	VLPVPQK	1.05			
12	VHDY		DPPH	Tuna cooking juice	[36]
13	FGHPY		Hydroxyl radical	Blue mussel	[37]
14	IKK		FTC	Prawn	[38]
15	FKK				
16	FIKK				
17	LDR	85.1	DETBA	Royal jelly	[12]
18	KNYP	80.6			
19	RYN	65.2			
20	GVPSS	65.3			
21	LPHVP	37.9			
22	ALPHVP	64.1			
23	YEG	35.5			
24	EIPHD	72.4			
25	IEIPHD	34.4			
26	SDQ	30.5			
27	VDTEQ	40.2			
28	IDGES	31.1			
29	IDGE	42.7			
30	EIPHD	35.0			
31	WNEH	38.1			
32	YEEN	35.2			
33	FDD	49.7			
34	YDY	36.6			
35	HEW	57.8			
36	VNPHDHQN			Soybean peptides	[18]
37	LVNPHDHQN				
38	LLPHH				
39	LLPHHADADY				
40	VIPAGYP				

[a] ORAC: Oxygen radical antioxidant capacity; SOSA: superoxide anion scavenging activity; DPPH: 1, 1-diphenyl-2-picrylhydrazyl; FTC: ferric thiocyanate; DETBA: 1, 3-diethyl-2-thiobarbituric acid.

related to the activity, while the *N*-terminus had no effect. However, this is a feint and it can be clarified easily by using the structural characteristic. After deletion of the *C*-terminus His, LLPH was obtained. Compared to LLPHH, the second amino acid residue from the *C*-terminus of LLPH was changed from H to P. As His has better parameters of V4, V6, and V7 than Pro (except V10), LLPH would have lower activity. In the same way, after deletion of the *N*-terminal Leu, the structural characteristic did not change from the new peptide LPHH to the original peptide LLPHH.

Saito and others [20] pointed out that two Tyr-containing tripeptides (such as Y***X***Y) showed higher activities than two His-containing tripeptides (H*X*H). According to the structural characteristic, the hydrophobicity of an amino acid on the *N*-terminus is most important for activity based on the same amino acid in the central position, and Y on the *N*-terminus has stronger hydrophobicity than H. It can be predicted that Y***X***Y had higher activity than H*X*H.

To sum up, using descriptors that have definite physiochemical meanings has a better insight in biochemical mechanisms. Although the importance of the *N*-terminus has been presented [39,40], the significance of the amino acid residue next to the *C*-terminus should not be overstated. The structural characteristics of peptides with high antioxidant activity were as follows: hydrogen bond and hydrophilic amino acid residue in position next to the *C*-terminus, hydrophobic amino acid residue in the *N*-terminus and electronic amino acid residue in the *C*-terminus. So, some statements, such as the antioxidative peptides scavenging lipid-

derived radicals due to unique high content of hydrophobic amino acid residues in their sequences [41], could be incorrect.

Cacciuttolo and others [42] reported that tyrosine, tryptophan, and phenylalanine, which have aromatic residues, can make reactive oxygen species stable through electron transfer. Da'valos and others [39] reported that among the amino acids, Trp, Tyr, and Met showed the highest antioxidant activity, followed by Cys, His, and Phe. The rest of the amino acids (Arg, Asn, Gln, Asp, Pro, Ala, Val, Lys, Ile, Tre, Leu, Glu, and Gly) did not exhibit antioxidant activity. Therefore, several amino acids such as Tyr, Met, His, Cys, and Trp were generally accepted as antioxidants contributing to the activities of the identified peptides [6,18,19]. And the antioxidative peptides containing His had been attributed to His due to the proton-donation ability of the His imidazole group. However, according to the physicochemical properties, His, Lys, Tyr, Cys, and Pro have a higher value of hydrogen bond (*V10*) and some other parameters (Table 1); they surely play a very important role, but only in the position next to the *C*-terminus. We speculated that it would form a hydrogen bond with receptor in the position; for example, Tyr is a hydrogen bond donor and interacts with carboxyl oxygen [43].

Why does the antioxidative peptide have such chemical structural characteristics? The antioxidative mechanism involved is still not well understood. Murase and others [44] reported that the accessibility of carnosine and His toward the peroxyl radical were increased through increased hydrophobicity. As the antioxidant activities of peptides were examined in an aqueous autoxidation system of linoleic acid, it was presumed that the hydrophobicity of amino acid residues on the *N*-terminus position formed a hydrophobic environment and it was easy for the peptide to enter into the lipid system. The electronic amino acid on the *C*-terminus and the amino acid next to the *C*-terminus with hydrophilicity made it easy for the peptides to provide the hydrogen atom for quenching the free radical. In view of the physicochemical properties, the structural characteristics of hydrophobic amino acid–hydrophilic amino acid alternation meet the requirements of the emulsifying agent. And the emulsifying effect and stability have a great influence to exert the antioxidative effect. There are some studies on protein hydrolyzates that have the emulsifying ability, such as alcalase-hydrolyzed potato proteins [45]. At a given pH value, the solubility and emulsifying ability index of these polypeptides were closely related to their relative contents of acidic (and basic) amino acids [46]. Meanwhile, the best stabilities of the multiple emulsions were obtained when there was a similarity between the hydrophobic part of the emulsifier and the oil phase [47]. Owing to acidic and basic amino acid residue (electronic amino acid) in peptide sequence, it was also a dipole moment compound, which means easy integration with polar compound. Such structures bring antioxidative peptides easily to attach the object, so as to react with the free radical and improve the probability of collision.

Conclusions

The structural characteristics will provide guides for evaluation of food proteins as potential precursors of antioxidative peptides, and predict the possible release of bioactive peptides from various proteins using proteolytic enzymes with different specificity. In this article, QSAR modeling is shown to provide an efficient methodology for elucidating SAR, which makes it possible to develop and predict peptide and protein structures with antioxidant activities. The importance of the second amino acid residue from the *C*-terminus is suggested. Ideal antioxidative properties of oligopeptides were derived from a delicate distribution of polar and hydrophobic residue.

However, the structure of the peptides in the database was comparatively simple. Further work is to establish the enlarged peptides database, including di-, tri-, tetra-, penta-, hexa-, hepta-, octa-, and decapeptides, and different combinations of the peptides with different peptide lengths. The source of peptides should cover meat, fish, collagen, soybean protein, casein, whey protein, etc.

Acknowledgement

This work is financially supported by the National Natural science Foundation Committee of People's Republic of China (No. 20976205).

References

1 Chen H, Muramoto K, Yamauchi F, Nokihara K. Antioxidant activity of designed peptides based on the antioxidative peptide isolated from digests of a soybean protein. *J. Agric. Food Chem.* 1996; **44**: 2619–2623.
2 Chen H, Muramoto K, Yamauchi F, Fujimoto K, Nokihara K. Antioxidative properties of histidine-containing peptides designed from peptide fragments found in the digests of a soybean protein. *J. Agric. Food Chem.* 1998; **46**: 49–53.
3 Suetsuna K, Ukeda H, Hirotomo O. Isolation and characterization of free radical scavenging activities peptides derived from casein. *J. Nutr. Biochem.* 2000; **11**: 128–131.
4 Kim S, Kim Y, Byun H, Nam K, Joo D, Shahidi F. Isolation and characterization of antioxidative peptides from gelatin hydrolysate of Alaska pollack skin. *J. Agric. Food Chem.* 2001; **49**: 1984–1989.
5 Saiga A, Tanabe S, Nishimura T. Antioxidant activity of peptides obtained from porcine myofibrillar proteins by protease treatment. *J. Agric. Food Chem.* 2003; **51**: 3661–3667.
6 Hernández-Ledesma B, Dávalos A, Bartolomé B, Amigo L. Preparation of antioxidant enzymatic hydrolysates from a-lactalbumin and b-lactoglobulin. Identification of active peptides by HPLC-MSMS. *J. Agric. Food Chem.* 2005; **53**: 588–593.
7 Je J, Park P, Kim S. Antioxidant activity of a peptide isolated from Alaska pollack (*Theragra chalcogramma*) frame protein hydrolysate. *Food Res. Intern.* 2005; **38**: 45–50.
8 Je J, Qian Z, Byun H, Kim S. Purification and characterization of an antioxidant peptide obtained from tuna backbone protein by enzymatic hydrolysis. *Process Biochem.* 2007; **42**: 840–846.
9 Je J, Qian Z, Kim S. Antioxidant peptide isolated from muscle protein of Bullfrog, Rana catesbeiana Shaw. *J. Med. Food.* 2007; **10**: 401–407.
10 Qian Z, Jung W, Kim S. Free radical scavenging activity of a novel antioxidative peptide purified from hydrolysate of bullfrog skin, Rana catesbeiana Shaw. *Bioresour. Technol.* 2008; **99**: 1690–1698.
11 Xie Z, Huang J, Xu X, Jin Z. Antioxidant activity of peptides isolated from alfalfa leaf protein hydrolysate. *Food Chem.* 2008; **111**: 370–376.
12 Guo H, Kouzuma Y, Yonekura M. Structures and properties of antioxidative peptides derived from royal jelly protein. *Food Chem.* 2009; **113**: 238–245.
13 Kim E, Lee S, Jeon B, Moon S, Kim B, Park T, Han J, Park P. Purification and characterisation of antioxidative peptides from enzymatic hydrolysates of venison protein. *Food Chem.* 2009; **114**: 1365–1370.
14 Klompong V, Benjakul S, Yachai M, Visessanguan W, Shahidi F, Hayes KD. Amino acid composition and antioxidative peptides from protein hydrolysates of yellow stripe trevally (*Selaroides leptolepis*). *J. Food Sci.* 2009; **74**: C126–C133.
15 Zhang J, Zhang H, Wang L, Guo X, Wang X, Yao H. Isolation and identification of antioxidative peptides from rice endosperm protein enzymatic hydrolysate by consecutive chromatography and MALDI-TOF/TOF MS/MS. *Food Chem.* 2010; **119**: 226–234.
16 Marcuse R. Antioxidative effect of amino-acids. *Nature* 1960; **186**: 886–887.
17 Yamaguchi N. studies on antioxidative activities of amino compounds on fats AND oils. PART I. oxidation of methionine

during course of autoxidation of linoleic acid. *Nippon Shokuhin Kogyo Gakkai.* 1971; **18**: 313–318.
18 Chen H, Muramoto K, Yamauchi F. Structural analysis of antioxidative peptides from soybean beta.-conglycinin. *J. Agric. Food Chem.* 1995; **43**: 574–578.
19 Tsuge N, Eikawa Y, Nomura Y, Yamamoto M, Sugisawa K. Antioxidative activity of peptides prepared by enzymatic hydrolysis of egg-white albumin. *Nippon Nogeikagaku Kaishi* 1991; **65**: 1635–1641.
20 Saito K, Jin D, Ogawa T, Muramoto K, Hatakeyama E, Yasuhara T, Nokihara K. Antioxidative properties of tripeptide libraries prepared by the combinatorial chemistry. *J. Agric. Food Chem.* 2003; **51**: 3668–3674.
21 Pripp AH, Isaksson T, Stepaniak L, Sørhaug T. Quantitative structure-activity relationship modelling of ACE-inhibitory peptides derived from milk proteins. *Eur. Food Res. Technol.* 2004; **219**: 579–583.
22 Pripp AH, Ardo Y. Modelling relationship between angiotensin-(I)-converting enzyme inhibition and the bitter taste of peptides. *Food Chem.* 2007; **102**: 880–888.
23 Nystrom A, Andersson PM, Lundstedt T. Multivariate data analysis of topographically modified α-melanotropin analogues using auto and cross auto covariances (ACC). *Quant. Struct-Act. Rel.* 2000; **19**: 264–269.
24 Mei H, Liao ZH, Zhou Y, Li SZ. A new set of amino acid descriptors and its application in peptide QSARs. *Pept. Sci.* 2005; **80**: 775–786.
25 Hansch C, Leo A. *Exploring QSAR: Fundamentals and Application in Chemistry and Biology*. American Chemical Society: Washington, DC, 1995.
26 King R, Hirst J, Sternberg M. New approaches to QSAR: neural networks and machine learning. *Perspect. Drug. Discov. Des.* 1993; **1**: 279–290.
27 Tomii K, Kanehisa M. Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. *Protein Eng.* 1996; **9**: 27–36.
28 Venkatarajan MS, Braun W. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical–chemical properties. *J. Mol. Model.* 2001; **7**: 445–453.
29 Kawashima S, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res.* 2000; **28**: 374.
30 Tracy ND, Young JC, Mason RL. Multivariate control charts for individual observations. *J. Qual. Technol.* 1992; **24**: 88–95.
31 Golbraikh A, Tropsha A. Beware of q2! *J. Mol. Graph. Model.* 2002; **20**: 269–276.
32 Tropsha A, Gramatica P, Gombar V. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* 2003; **22**: 69–77.
33 Topp R, Gómez G. Residual analysis in linear regression models with an interval-censored covariate. *Stat. Med.* 2004; **23**: 3377–3391.
34 Tian F, Yang L, Lv F, Yang Q, Zhou P. Zhou P, In silico quantitative prediction of peptides binding affinity to human MHC molecule: an intuitive quantitative structure–activity relationship approach. *Amino Acids* 2009; **36**: 535–554.
35 Rival SG, Boeriu CG, Wichers HJ. Caseins and casein hydrolysates. 2. Antioxidative properties and relevance to lipoxygenase inhibition. *J. Agric. Food Chem.* 2001; **49**: 295–302.
36 Hsu K, Lu G, Jao C. Antioxidative properties of peptides prepared from tuna cooking juice hydrolysates with orientase. *Food Res. Intern.* 2009; **42**: 647–652.
37 Jung W, Rajapakse N, Kim S. Antioxidative activity of a low molecular weight peptide derived from the sauce of fermented blue mussel, *Mytilus edulis*. *Eur. Food Res. Technol.* 2005; **220**: 535–539.
38 Suetsuna K. Antioxidant peptides from the protease digest of prawn (*Penaeus japonicus*) muscle. *J. Mar. Biotechnol.* 2000; **2**: 5–10.
39 Dávalos A, Miguel M, Bartolomé B, López-Fandiño R. Antioxidant activity of peptides derived from egg white proteins by enzymatic hydrolysis. *J. Food Prot.* 2004; **67**: 1939–1944.
40 Kitts DD, Weiler K. Bioactive proteins and peptides from food sources. Applications of bioprocesses used in isolation and recovery. *Curr. Pharm. Design.* 2003; **9**: 1309.
41 Rajapaksea N, Mendisa E, Byunb H, Kim S. Purification and in vitro antioxidative effects of giant squid muscle peptides on free radical-mediated oxidative systems. *J. Nutr. Biochem.* 2005; **16**: 562–569.
42 Cacciuttolo MA, Trinh L, Lumpkin JA, Rao G. Hyperoxia induces DNA damage in mammalian cells. *Free Radic. Biol. Med.* 1993; **14**: 267–276.
43 Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol.* 1995; **13**: 587–622.
44 Murase H, Nagao A, Terao J. Antioxidant and emulsifying activity of N-(long-chain-acyl)histidine and N-(long-chain-acyl)carnosine. *J. Agric. Food Chem.* 1993; **41**: 1601–1604.
45 Nieto G, Castillo M, Xiong YL, Varez D, Payne FA, Garrido MD. Antioxidant and emulsifying properties of alcalase-hydrolyzed potato proteins in meat emulsions with different fat concentrations. *Meat Sci.* 2009; **83**: 24–30.
46 Yuan DB, Yang XQ, Tang CH, Zheng ZX, Wei-Min Ahmad I, Yin SW. Physicochemical and functional properties of acidic and basic polypeptides of soy glycinin. *Food Res. Intern.* 2009; **42**: 700–706.
47 Magdassi S, Frenkel M, Garti N. Correlation between nature of emulsifier and multiple emulsion stability. *Drug Dev. Ind. Pharm.* 1985; **11**: 791–798.

