Food Chemistry 234 (2017) 431–438

Contents lists available at ScienceDirect

Food Chemistry

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

# A study to evaluate the potential of an *in silico* approach for predicting dipeptidyl peptidase-IV inhibitory activity *in vitro* of protein hydrolysates

Tzu-Yuan Wang [a], Cheng-Hong Hsieh [b], Chuan-Chuan Hung [c,d], Chia-Ling Jao [e], Pei-Yi Lin [d], You-Liang Hsieh [b], Kuo-Chiang Hsu [b,c,d,*]

[a] Division of Endocrine and Metabolism, China Medical University Hospital, 2 YuDe Road, Taichung 40447, Taiwan
[b] Department of Health and Nutrition Biotechnology, Asia University, 500, Lioufeng Road, Taichung 41354, Taiwan
[c] Food Safety and Inspection Center, Asia University, 500, Lioufeng Road, Taichung 41354, Taiwan
[d] Department of Nutrition, China Medical University, 91, Hsueh-Shih Road, Taichung 40402, Taiwan
[e] Department of Food and Beverage Management, Tung Fang Design University, 110, Tung-Fang Road, Kaohsiung 82941, Taiwan

ARTICLE INFO

*Article history:*
Received 21 June 2016
Received in revised form 2 May 2017
Accepted 6 May 2017
Available online 8 May 2017

*Keywords:*
*In silico* analysis
Dipeptidyl peptidase IV inhibitor
Dietary protein
Correlation analysis

ABSTRACT

A total of 294 edible protein sequences and 5 commercial proteases listed in the BIOPEP database were analyzed *in silico*. The frequency (*A*), a parameter *in silico* described previously, was examined further to calculating the ratio of truncated peptides with Xaa-proline and/or Xaa-alanine to all peptide fragments in a protein hydrolyzed with a protease, using the BIOPEP database. Then the *in vitro* DPP-IV inhibitory activity was determined using the same 15 protein and protease combinations to evaluate their relationship. The result shows that *A* values considering the number of Xaa-proline + Xaa-alanine exhibited a strong correlation with *in vitro* DPP-IV inhibition rates by Pearson's correlation analysis ($r = 0.6993$; $P < 0.05$). Therefore, the *in silico* approach is effective to predict DPP-IV inhibitory activities *in vitro* of protein hydrolysates.



## 1. Introduction

Dipeptidyl peptidase IV (DPP-IV; EC 3.4.14.5) is a serine protease that preferentially cleaves post-proline or alanine dipeptides from the N-terminus of polypeptides and proteins (Cunningham & O'Connor, 1997). This multifunctional enzyme, expressed in body fluids and a variety of tissues, including liver, kidney and small intestine (Darmoul et al., 1994; Holst & Deacon, 1998), had been shown to degrade chemokines, neuropeptides and many hormones, such as glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) (Lambeir et al., 2001). GIP and GLP-1, the incretin hormones, are released in response to food consumption to stimulate β-cell to secret insulin (Hansotia & Drucker, 2005). Therefore, the use of DPP-IV inhibitors is a novel therapeutic approach for management of type 2 diabetes (Deacon & Holst, 2006).

In addition to synthetic compounds (Gooben & Gräber, 2012), bioactive peptides derived from various food proteins, such as milk proteins (Lacroix & Li-Chan, 2012a; Nongonierma & FitzGerald, 2014), gelatin (Huang, Hung, Jao, Tung, & Hsu, 2014; Li-Chan, Huang, Jao, Ho, & Hsu, 2012), meat proteins (Lafarga, O'Connor, & Hayes, 2014), have been widely studied to show DPP-IV inhibitory activity (Power, Nongonierma, Jakeman, & FitzGerald, 2014).

*In silico* analysis, a bioinformatics-driven approach, has been demonstrated to be useful in predicting the potential of proteins as precursors of DPP-IV inhibitory peptides by using several parameters (Jin, Yan, Yu, & Qi, 2015; Lacroix & Li-Chan, 2012b; Lafarga et al., 2014; Nongonierma & FitzGerald, 2014; Nongonierma, Mooney, Shields, & FitzGerald, 2014). A parameter, frequency of occurrence, was used to quantify the potential of selected proteins comprising the amount of fragments matching peptides with DPP-IV inhibitory activity reported in the literature. According to the frequency of occurrence, caseins from cow's milk, collagens from bovine meat and salmon (Lacroix & Li-Chan, 2012b), and the small subunit of sorghum ribulose bisphosphate carboxylase (RuBisCO) (Udenigwe, Gong, & Wu, 2013) appeared as potential sources of DPP-IV inhibitors. An advanced parameter, potency index (PI), was developed to take into account both the frequency of occurrence and potency ($IC_{50}$ value) of the DPP-IV inhibitory peptides (Nongonierma & FitzGerald, 2014), and

* Corresponding author at: Department of Nutrition, China Medical University, No. 91, Hsueh-Shih Road, Taichung 40402, Taiwan.
*E-mail address:* kchsu@mail.cmu.edu.tw (K.-C. Hsu).

http://dx.doi.org/10.1016/j.foodchem.2017.05.035
0308-8146/

therefore, canola, chicken egg, oat and wheat were studied as potential sources of DPP-IV inhibitory peptides. Molecular docking simulations have been utilized to elucidate the peptide sequences may actually be able to interact with the proteins showing specific biological activity (Sánchez-Rivera, Martínez-Maqueda, Cruz-Huerta, Miralles, & Recio, 2014). However, no direct correlation was observed between *in vitro* DPP-IV inhibitory activity of some tri-peptides and the Vina scores (predicted affinity) obtained by molecular docking of those tri-peptides to the active site of DPP-IV and (Nongonierma et al., 2014). There have been some other bioinformatics driven methods developed, such as the quantitative structure-activity relationship (QSAR) (Zhou, Yang, Ren, Whang, & Tian, 2013), Food-Wiki database (FoodWikiDB) (Holton, Vijayakumar, & Khaldi, 2013), and peptide array technology (Wang et al., 2013), but the prospects and limitations of these methods as successful screening approaches still needed more studies to verify.

To date, the potential of the hydrolysates obtained from the combinations of protein sources and given proteases as DPP-IV inhibitors predicted by *in silico* approaches was not successfully established, and the correlation between the parameter used in *in silico* analysis and the *in vitro* DPP-IV inhibitory activity has not been determined. As it has been well-known that DPP-IV has a specificity with a preference for substrates by removing proline, alanine, serine and hydroxyproline in sequence as the second N-terminal residue (Davy et al., 2000), we hypothesized that protein hydrolysates with high contents of peptides having Xaa-Pro and Xaa-Ala may have the potential to be DPP-IV inhibitors (Nongonierma & FitzGerald, 2014). In our previous study, we adopted an *in silico* approach by using a novel parameter the frequency (*A*), calculating the ratio of the number of truncated peptides with Xaa-proline and Xaa-alanine to all peptide fragments from a protein hydrolyzed with a specific protease, to predict the potential of the hydrolysates from several food protein sources as DPP-IV inhibitors. The results showed that there was a positive correlation between *A* values of the hydrolysates from several protein sources and their *in vitro* DPP-IV inhibitory activities (Hsieh et al., 2016). Therefore, the aim of this study was to use the *in silico* parameter described previously by considering numbers of peptides with Xaa-Pro and Xaa-Ala in a protein hydrolyzed with a proteases, and the correlation between the *in silico* frequency and *in vitro* DPP-IV inhibitory activity was also determined.

## 2. Materials and methods

### 2.1. Materials and reagents

The protein sources, β-lactoglobulin (L3908, from bovine milk), α-casein (C6780, from bovine milk), β-casein (C6905, from bovine milk), κ-casein (C0406, from bovine milk), myoglobin (M0630, from equine skeletal muscle), and lactoferrin (L1294, from human), were purchased from Sigma-Aldrich (St. Louis, MO, USA). The commercial proteases, bromelain (B4882, from pineapple stem, ≥3 units/mg protein), papain (P3375, from papaya latex, 1.5–10 units/mg solid), thermolysin (P1512, from *Geobacillus stearothermophilus*, 30–175 units/mg), pepsin (P7000, from porcine gastric mucosa, ≥250 units/mg solid) and trypsin (T8802, from bovine pancreas, ≥10,000 BAEE units/mg protein) were also obtained from Sigma-Aldrich.

### 2.2. In silico analysis

#### 2.2.1. Proteins and proteolysis simulation

The amino acid sequences of 294 proteins from edible sources were taken selected from a total number of 707 proteins in BIOPEP database, available at http://www.uwm.edu.pl/biochemia/ (accessed between March and June 2014). The edible proteins include 80 animal sources and 214 plant sources, and their amino acid sequences were listed in Table S1. Before *in vitro* analysis of DPP-IV inhibitory activity, the sequences of substitute proteins were searched in both BIOPEP database and UniProt Knowledgebase of ExPASy Proteomics Server available at http://expasy.org/ in case the protein sources were not commercially available.

The proteolysis simulation process of all protein sources by bromelain (EC 3.4.22.4), papain (EC 3.4.22.2), thermolysin (EC 3.4.24.27), pepsin (EC 3.4.23.1), or trypsin (EC 3.4.21.4) was performed using BIOPEP Enzyme(s) Action program (Minkiewicz, Dziuba, Iwaniak, Dziuba, & Darewicz, 2008).

#### 2.2.2. Analysis

We used the following parameter to characterize the potential of the proteins as the precursors of DPP-IV inhibitory peptides. The parameter was introduced to the ratio of the number of the peptides with Pro (Xaa-Pro), Ala (Xaa-Ala) or Pro + Ala (Xaa-Pro + Xaa-Ala) as the penultimate N-terminal residues to the total number of total peptide fragments released by proteases. The value (*A*) of the frequency was used to quantify the potential of the proteins to serve as the DPP-IV inhibitory peptide precursors and calculated based on the following equation:

$$A(\%) = (d/F) \times 100$$

where *d* is the number of the truncated peptides with Pro, Ala or Pro + Ala as the second N-terminal residues; *F* is the number of total peptide fragments counted by considering free amino acids and peptides released by given proteases.

After *in silico* analysis, all the combinations of protein sources and proteases were listed and ranked by *A* values (Table S2), and they were divided into three groups: high frequency ($A > 10\%$), medium frequency (*A* values between 5 and 10%) and low frequency ($A \leq 5\%$). Fifteen protein and protease combinations were selected from the three groups for *in vitro* analysis. In case the protein sources were not commercially available, the substitute proteins were used in consideration of the availability for purchase and their amino acid sequences have been listed in database.

### 2.3. In vitro analysis

#### 2.3.1. Sample preparation

All proteins were dissolved in 25-fold volume (w/v) of $ddH_2O$. Each protein solution was adjusted by 2 N NaOH or 2 N HCl to the optimal pH of a given protease and then incubated for 20 min at the optimal temperature of the protease. The hydrolysis reaction was initiated by the addition of the enzyme solution (enzymes in powder form dissolved in $ddH_2O$, 1:1 w/v) at enzyme/substrate ratio of 3%: bromelain (pH 6.7, 45 °C), papain (pH 6.2, 25 °C) and thermolysin (pH 8.0, 70 °C). After hydrolysis for up to 4 h, the obtained hydrolysates were heated in boiling water for 15 min to inactivate enzymes, cooled in ice water for 10 min, adjusted to pH 7.0 with 2 N NaOH, and then centrifuged (Centrifuge 05P-21, Hitachi Ltd., Katsuda, Japan) at 13,000*g* and 4 °C for 15 min. The proteins hydrolyzed with pepsin and trypsin were started by the addition of pepsin (pH 2.0, 37 °C) for up to 4 h. After heating, cooling and centrifuging at the same condition described above, the hydrolysate solutions were adjusted pH to 7.5 with 2 N NaOH and then added with trypsin (pH 7.5, 37 °C) and the hydrolysis process lasted for up to 4 h. Then, the hydrolysates were heated, cooled, pH adjusted to 7.0 and centrifuged. The supernatant of the hydrolysates was lyophilized and stored at −20 °C.