Journal of Bioscience and Bioengineering
VOL. 127 No. 3, 366–371, 2019

www.elsevier.com/locate/jbiosc

# Searching for high-binding peptides to bile acid for inhibition of intestinal cholesterol absorption using principal component analysis

Masako Ito,[1,2] Kazunori Shimizu,[1] and Hiroyuki Honda[1,*]

*Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, Nagoya 464-8603, Japan*[1] *and Industrial Research Center, Aichi Center for Industry and Science Technology, 1-157-1 Onda-cho, Kariya, Aichi 448-0013, Japan*[2]

Received 8 May 2018; accepted 13 August 2018
Available online 18 September 2018

**We previously proposed a new method for exploring functional peptides using both spot-synthesis peptide libraries and principal component analysis (PCA). Here, we applied these methods to determine if high-binding 6-mer peptides can be used on bile acid for the inhibition of intestinal cholesterol absorption. We used a binding assay of 512 basal 4-mer peptides to bile acid, and from these selected high-binding and low-binding peptides. PCA was performed on data from both these binding groups and many physicochemical variables of the 512 peptides tested, and then through this, the variables were reduced to two principal components (PCs). The peptides were plotted on the PCA chart, and we identified distinct clusters of high- and low-binding regions. These PCA regions were applied to 6-mer random peptides, and we identified 6-mer peptides with high and low binding capacity to bile acid. We confirmed that the average fluorescence intensity of high-binding peptides was 3.0-fold higher than that of low-binding peptides. We succeeded in identifying 6-mer peptides with high and low-binding affinity based on the PCA analysis of 4-mer peptides. These results were compared and discussed with regard to those acquired by our previous computational analysis based on neural networks.**



[**Key words:** Peptide; Bile acid; Peptide array; Principal component analysis; Cholesterol]

Peptide arrays have received much attention as tools for identifying linear peptides with specific activities, such as antibody binding, enzyme inhibition, cell surface binding, and so on. Our group has already identified several high-binding peptides, with high affinity to various targets including for example angiotensin (1), amylase (2), IgG (3,4), IL-2 (4), bile acid (5), apoptosis-inducing peptides (6,7), and cell adhesive peptides (8,9).

Cholesterol absorption in the upper intestine is facilitated by bile acid micelles, which can capture cholesterol. Inhibition of bile acid micelle formation can suppress intestinal cholesterol absorption (10). Thus, bile acid binding peptides could be expected to inhibit cholesterol absorption. It was reported that VAWWMY (Val-Ala-Trp-Trp-Met-Tyr), the hydrophobic 6-mer peptide in the glycinin A1a and B1b subunits of soybean protein, has inhibitory effects on cholesterol absorption and lowers serum cholesterol (11–13). Therefore, our group screened 6-mer peptides, replacing each position of VAWWMY with another amino acid (unpublished data). Our group also performed computationally assisted screening from a random 6-mer library (2212 sequences), and found a positional rule for identifying high-binding peptides. According to that rule, about 33% of the peptides from the spot-synthesized peptide array evaluation were indicated to have high-binding tendencies (5).

Given that many known peptides might bind to bile acid, it is worth screening 6-mer peptides and longer peptides for such activity. The spot-synthesized peptide array is useful to evaluate not only short peptides but also long ones. However, screening from long peptide libraries is difficult because such libraries are very large. For example, there are 25.6 billion peptides in the 8-mer peptide library.

Given the above, our group created a new methodology for functional peptide screening. Principal component analysis (PCA) was conducted as a means of reducing sequence variables of 512 4-mer peptides classified based on their amino acids' physicochemical characteristics into fewer and simpler ordinate variables (14). From the previous spot-synthesized peptide array evaluation of 512 peptides, high- and low-binding rules for peptide targets were determined. These rules were adapted to the 8-mer peptide library, and we found high- and low-binding peptides, for which, information was extracted and evaluated. With this methodology, our group previously succeeded in identifying 8-mer high-binding peptides to IL2 and IgG from a 4-mer small peptide library (512 sequences). In this study, we applied this new methodology to explore the peptides binding to bile acid, i.e., we screened 6-mer high-binding peptides to bile acid from a 4-mer small peptide library. In addition, the rule so as to identify high binding peptides was compared to results from our previous computational analysis based on neural networks.

## MATERIALS AND METHODS

**Preparation of peptide array** The peptide array was prepared as described in our previous study (14). A cellulose membrane (grade 542, Whatman, Maidstone, UK) was activated using β-alanine as the basal spacer. Fluorenylmethoxycarbonyl

* Corresponding author. Tel.: +81 52 789 3215; fax: +81 789 3214.
*E-mail address:* honda@chembio.nagoya-u.ac.jp (H. Honda).

1389-1723/$ – see front matter 
https://doi.org/10.1016/j.jbiosc.2018.08.006

amino acids (Fmoc-aa), adjusted to 1.5 M, were spotted using a peptide auto-spotter (ASP222, Intavis AG, Köln, Germany). The membrane was deprotected with 20% piperidine and washed with *N,N′*-dimethyl formamide and methanol. We repeated these steps four or six times to make 4-mer or 6-mer peptides. After the last Fmoc-aa were spotted, the side-chain protecting groups were finally removed by washing with diethyl ether and methanol.

**Preparation of basal 4-mer peptide library** The 4-mer peptide library consisted of 512 peptides, for which four kinds of physicochemical properties were extracted (14), as shown in Table S1.

**Binding assay of peptide arrays to bile acid** Binding assay of the peptide arrays to bile acid was performed as described in our previous study (5). After removing side-chain protecting groups by overnight incubation, the peptide array membranes were washed several times with phosphate-buffered saline (PBS, pH 7.0) to remove the last traces of reagents. They were then soaked in 1% bovine serum albumin solution (BSA, Wako Pure Chemical Industries, Osaka, Japan) in PBS for 24 h to block unspecific adsorption. After washing with PBS once for 5 min, and then again twice for 2 min at room temperature, the arrays were incubated with 0.01 M bile acid solution in PBS for 1 h at 37°C, with rotation at 50 rpm using a shaker (TAL RS310 LM, Thomas Kagaku Co., Ltd., Tokyo, Japan). Taurocholic acid (T-4009, Sigma Chemical, St. Louis, MO, USA) was herein used as a typical bile acid. Unbound taurocholic acid was removed by washing the array with PBS. The peptide arrays were then incubated with primary antibodies against cholic acid from rabbit (FKA502, Cosmo Bio Co., Ltd., Tokyo, Japan) for 1 h at 37°C, with rotation at 50 rpm. Unbound primary antibody was removed by washing the array with 0.05% Tween 20-PBS. Next, the peptide arrays were incubated with secondary antibody from goat conjugated with Alexa 488 (ab150117, Abcam, Cambridge, UK) in PBS for 1 h at 37°C, with rotation at 50 rpm using a shaker. Unbound secondary antibody was removed by washing the array once for 10 min at room temperature and twice for 30 min at 37°C with 0.05% Tween20-PBS.

The fluorescence intensity of each peptide spot was calibrated using a Typhoon FLA-2500 biomolecular imager (GE Healthcare Life Sciences, Buckinghamshire, UK), with ImageQuantTL software (GE Healthcare Life Sciences).

The amount of binding of the bile acid to a peptide spot was calculated from the fluorescence intensities of the peptide spots as follows:

$$\begin{aligned}&(\text{Fluorescence intensity of binding the bile acid to a peptide spot})\\&= (\text{Fluorescence intensity of peptide spot}) - (\text{Fluorescence intensity of peptide}\\&\quad \text{spot without primary antibodies against the bile acid}) \qquad (1)\end{aligned}$$

**Principal component analysis** PCA was performed with R (ver. 3.4.3) using the variables shown in Table S2. The variables for each peptide included polarity (15), side-chain contribution to protein stability (16), isoelectric point (15), hydropathy index (17), scaled side chain hydrophobicity value (18), molecular weight (19), and number of groups in each peptide because 20 amino acids were divided into 4 groups in previous study (14). These indices without molecular weight and number of groups were referred from amino acid index (AAindex, https://www.genome.jp/aaindex/). AAindex is a well-known and convenient database of numerical indices representing various physicochemical and biochemical properties of amino acids and pairs of amino acids.

To determine the best combination of variables for determining bile acid binding affinity from peptide libraries, we performed seven PCAs using different combinations of the variables shown in Table S2. Then, we calculated Pearson's correlation coefficients of the relationship between fluorescence intensities and each of the principal components, PC1 or PC2. Based on this, we determined that run no. 2 (containing 16 variables) provided the best-fitting set of variables (Table S2). The proportions of variance and cumulative proportions were calculated from the PCA with R.

**Analysis of a high- and low-binding peptide rules for bile acid** We determined the rules for identifying high- and low-binding affinity to bile acid from the PC scores from a basal 4-mer basal peptide library. First, we calculated the average and standard deviation (SD) of the fluorescence intensities. Because the fluorescence intensity of peptides were widely distributed, we defined that high-binding peptides have higher intensity than the average +1 SD, and the low-binding peptides have lower intensity than the average −1 SD. Peptides with higher or lower binding activity would be selected when 2 SD was used. However, we used 1SD because number of high binding peptides was too few and there was nothing of low binding peptides in the present study. Second, we generated score plots of each of PC1 and PC2, and plotted all of peptides including high-binding peptides and the low-binding peptides. Since high-binding peptides were relatively localized in restricted area, the rectangle restricted area of PC1-PC2 was identified in which is surrounding both the scale of PC1 from the average +1 SD to the average −1 SD and the scale of PC2 from the average +1 SD to the average −1 SD. The area corresponds to the high-binding rule. In the same manner, the rectangle restricted area for low-binding peptides was identified and the area also corresponds to low-binding rule. Third, we calculated PC scores of the 4-mer and 6-mer random peptides using the best combination of variables (run no. 2, with 16 variables, Table S2) and adopted these rules. The region outside of that encompassed by these rules was identified as the "other rule" zone.

In this manner, we selected the high-binding, low-binding, and other peptides from the 4-mer (Table S3) and 6-mer random peptide libraries (Table S4) using basal information from the 4-mer peptide libraries.

## RESULTS

**Selection of variables for peptide sequence analysis** The PCA using the basal 512 4-mer peptides was performed using seven physicochemical properties of the amino acids of which peptides consisted. These included average, maximum, and minimum values of each peptide's polarity, side-chain contribution to protein stability, isoelectric point, hydropathy index, partition coefficient (Log P), molecular weight, the number of different groups in the peptide.

We first performed seven PCAs (run no. 1 to 7) using 9 to 18 variables (Table S2), and the cumulative proportions of variance explained by the first two PCs were calculated from the PC scores (Table 1). The cumulative proportions for first two PCs varied from 0.56 to 0.67. Those of runs no. 2, 5, 6, and 7 were higher than those of the others.

Next, we calculated correlation coefficients between fluorescence intensities of peptides and their values for PC1 or PC2 (Table 1), according to methods used in our previous study (14). Herein, we used some indices defined as physicochemical properties of peptides, and never used fluorescent intensity of peptides for PCA. Neither of principal components was determined so as to correlate with fluorescent intensity against bile acid. Therefore, there was no correlation between the fluorescence intensities and PC1 because all values were under 0.1. In the case of the correlation between fluorescence intensities and PC2, the coefficients' values exceeded 0.3 in runs no. 2 to 4. Because of the higher cumulative proportions of variance explained by the first two PCs in run no. 2 (Fig. 1 and Table 1), we selected run no. 2 for further PCA.

**Determination of high- and low-binding rules for bile acid from the PCA using the basal 512 4-mer peptides** To construct high- and low-binding rules, high and low binding peptides should be identified. In the screening of functional peptides, however, it is difficult to define high binding peptide and low-binding peptide. In previous study, we defined top 30

**TABLE 1.** The cumulative proportions of variance explained and the Pearson's correlation coefficients between the fluorescence intensities of peptides and PCA scores using 16 variables.

| Run no. | Number of variables | Cumulative proportion | Coefficients of correlation between fluorescence intensities and PCA | |
|---|---|---|---|---|
| | | | PC1 | PC2 |
| 1 | 18 | 0.56 | 0.000 | 0.199 |
| 2 | 16 | 0.64 | 0.007 | 0.323 |
| 3 | 16 | 0.56 | 0.053 | 0.302 |
| 4 | 16 | 0.57 | 0.006 | 0.330 |
| 5 | 13 | 0.64 | 0.005 | 0.293 |
| 6 | 13 | 0.64 | 0.004 | 0.008 |
| 7 | 9 | 0.67 | 0.040 | 0.273 |

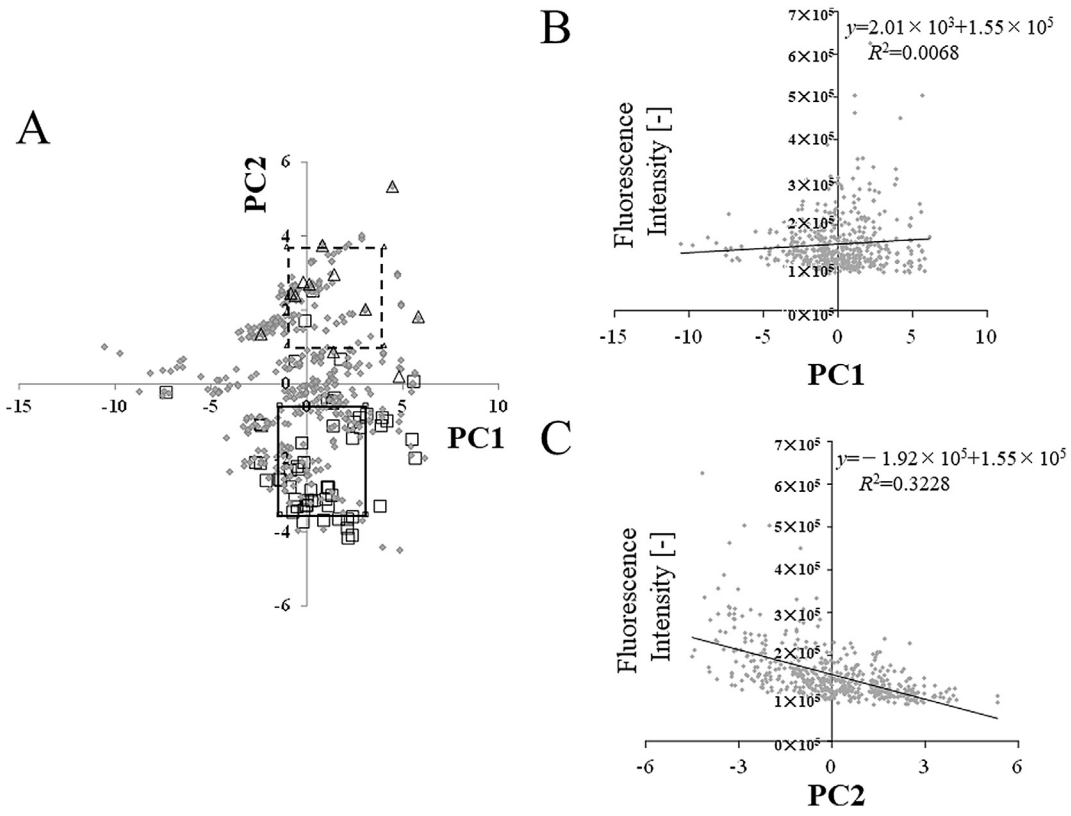


FIG. 1. Principal component analysis (PCA) of the basal 512 4-mer peptides using 16 variables. PC scores (PC1 and PC2) of the 512 4-mer peptides are plotted as gray diamonds. Open squares and open triangles are high- and low-binding peptides, respectively, for bile acid as determined based on binding assay, and these are overlaid with gray diamonds (A). The large square drawn with a solid line is the high-binding rule area, and the one drawn with a dashed line is the low-binding rule area. The relationship between the fluorescence intensities of the basal 512 4-mer peptides to bile acid and the PC1 (B) and PC2 (C) are plotted. The solid lines and formulas in panels B and C show the approximate best-fit line and its formula for each scatterplot.

and bottom 30 binding peptides as high and low binding peptides, respectively. Those were not any statistical reason. In this study, since random peptide library was used for peptide screening, statistical consideration is possible to select high and low binding peptide. We calculated the average and SD of the fluorescence intensities. We decided that high-binding peptides were those with intensities over the average +1 SD, and low-binding peptides were those with intensities under the average −1 SD. Table S5 shows the 59 high-binding peptides and 12 low-binding peptides. We then plotted them over the PC graph using run no. 2 as the basis of the plot (Fig. 1). We determined the high-binding rule region from the portion of the PCA plot overlapped by the scattered points representing high-binding peptides. Specifically, the high-binding rule encompassed the region within PC1 values of −1.43 to 3.06 and with PC2 values of −0.58 to −3.53 (Fig. 1). In the same manner, the low-binding rule encompassed the region within PC1 values of −1.05 to 4.03 and PC2 values of 1.04−3.72 (Fig. 1).

**Analysis of binding affinities from the high- and low-binding rules using 500 random 4-mer peptides and 500 random 6-mer peptides** Next, we prepared 500 random 4-mer peptides and 500 random 6-mer peptides, and calculated PC scores on them using the same variables as previously (run no. 2). We then adopted the high- and low-binding rule on the PC graphs of 500 random 4-mer peptides and 500 random 6-mer peptides (Fig. 2). From these graphs, we could estimate the high- and low-binding 4-mer and 6-mer peptides from the random libraries. We estimated that there were 89 high-binding peptides (17.8% of 500 peptides) and 34 low-binding peptides (6.8% of 500 peptides) from the 4-mer random library (Table S3), and there were 94 high-binding peptides (18.8% of 500 peptides) and 19 low-binding peptides (3.8% of 500 peptides) from the 6-mer random library (Table S4).

**Analysis of binding affinities from the high- and low-binding rules using new 4-mer and 6-mer random peptides** We chose 90 peptides from a random library of 500 random 4-mer peptides. Specifically, there were 24 peptides selected of the 89 high-binding peptides identified, 8 of the 34 low-binding peptides, and 58 of another 377 peptides were selected (Table S6). In the same manner, we chose 90 peptides from the random library of 6-mer peptides (Table S6). We actually synthesized each of the 90 peptides identified by these methods and used them in a 4-mer or 6-mer peptide array, and then performed a binding assay to evaluate their binding affinity to bile acid.

We then compared the fluorescence intensities between high- and low-binding peptides. Among the 4-mer peptides, the average fluorescence intensity of 24 high-binding peptides (37,860) was the highest of the three groups, and was 8.3-fold higher than that of 8 low-binding peptides (4542, Table 2). Among the 6-mer peptides, the average fluorescence intensity of 24 high-binding peptides (8202) was the highest of the three groups, and was 3.0-fold higher than that of 6 low-binding peptides (2752, Table 2). We have not yet been checked adsorption inhibiting effect of high binding peptides in vivo in animal model. However, we have demonstrated the disintegration of cholesterol-micelle with high-binding three peptides to bile acid selected from high-binding rules in 6-mer random library. Top three 6-mer peptides identified here was selected and the assay was carried out. Then those peptides showed the similar disintegration activity as VAWWMY (data not shown).

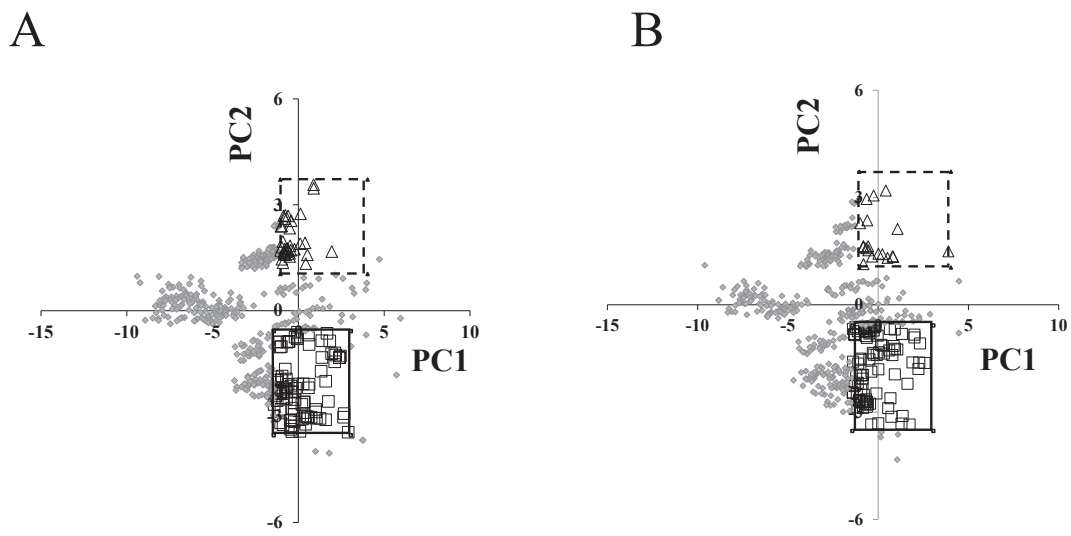


FIG. 2. PC score plot of the 500 peptides selected randomly using 16 variables. The left panel is a PC score plot of the 500 4-mer peptides (A) and the right panel is a PC score plot of the 500 6-mer peptides (B) as gray diamonds. Open squares and open triangles are estimated high- and low-binding peptides, respectively, and these are overlaid on gray diamonds. The large square drawn with a solid line is the high-binding rule area, and the one drawn with a dashed line is the low-binding rule area.

Therefore, we concluded that we could filter the high- and low-binding 6-mer peptides for bile acid using binding rules determined from the basal 512 4-mer peptide library.

## DISCUSSION

Our group created the new methodology used in this study to find high-binding peptides with affinity for binding to a specific target (14). We first categorized 20 amino acids into four groups based on their physicochemical characteristics. Second, we selected 512 4-mer peptides which had amino acids corresponding to those of each group. PCA was conducted to reduce the dimensionality of the variables of these 512 4-mer peptides. From the spot-synthesized peptide array evaluation of 512 peptides, high- and low-binding rules for a specific target were determined. These rules were adopted to different peptide libraries and were still able to identify high-binding peptides to the target. In this study, we applied this new methodology to find high-binding peptides to bile acid.

We first performed seven PCAs (run no. 1 to 7) using 9 to 18 variables (Table S2), and we selected run no. 2 from both of the cumulative proportions of variance and correlation coefficients of PC1 or PC2 against fluorescence intensities of peptides. There was no correlation between the fluorescence intensities and PC1 because all values were under 0.1, and the correlation with PC2 was relatively high. In order to understand the results of PCA, we investigated coefficient matrix of principal components in PCA accomplished here. As a results, we found that PC1 was attributed to hydrophobic properties and PC2 was electrostatic properties as shown in Table S7. The binding ability of peptides to bile acid has no correlation with PC1, but has a slight correlation with PC2. This suggests that electrostatic property of peptides correlate linearly with binding activity. PC2 negatively correlated with isoelectric point and PC2 also negatively correlated with binding activity as shown in Fig. 1. Consequently binding activity to bile acid will be considered to positively correlate with isoelectric point of peptide.

**TABLE 2.** The averages and SDs of the fluorescence intensities and numbers of peptides in different binding groups.

| | | Ave | SD | Number of peptides |
|---|---|---|---|---|
| 4-mer peptides | High-binder | 37860 | 27665 | 24 |
| | Low-binder | 4542 | 2828 | 8 |
| | Others | 10290 | 15661 | 58 |
| 6-mer peptides | High-binder | 8202 | 5893 | 24 |
| | Low-binder | 2752 | 805 | 6 |
| | Others | 6313 | 6754 | 60 |

Ave, average; SD, standard deviation.

From the results of our PCAs and the spot-synthesized peptide array evaluation of 512 peptides, we determined high- and low-binding rules for peptides to bile acid. Those rules were adopted to different peptide libraries of 500 4-mer and 6-mer peptides each, selected randomly. From these libraries, we could estimate that 89 and 94 peptides were high-binding (17.8% and 18.8% of 500 peptides) and 34 and 19 were low-binding peptides (6.8% and 3.8% of 500 peptides) from the 4-mer (Table S3) and 6-mer (Table S4) random libraries, respectively. Table 2 shows the average fluorescence intensity of each of the high-binding peptides, low-binding peptides, and others actually synthetized and assayed by the spot-synthesized peptide array evaluation. Among the 4-mer peptides, the average fluorescence intensity of 24 peptides estimated to be high-binding (37,860) was the highest of the three groups, and was 8.3-fold higher than that of 8 peptides estimated to be low-binding (4542). Among the 6-mer peptides, the average fluorescence intensity of the 24 peptides estimated to be high-binding (8202) was the highest of the three groups, and was 3.0-fold higher than that of the 6 peptides estimated to be low-binding (2752). As the peptides estimated to be high-binding by this methodology were also later shown to actually have high-binding affinity to bile acid, this methodology is useful to extract not only high-binding 4-mer peptides but also longer peptides for binding to bile acid. In this study, we could identify 24 peptides as probably high-binding peptides from out of 500 randomly selected 4-mer peptides, i.e., a 4.8% discovery rate. Therefore, it could be said that 9600 peptides among the 160,000 known 4-mer peptides are possibly high-binding.

It was reported previously that VAWWMY shows high binding activity to bile acid and has cholesterol-lowing capacity (11–13). VAWWMY is a functional peptide, the hydrophobic 6-mer peptides of conglycinin of soybean protein. This peptide sequence is not yet optimized. In our previous paper (5), we easily obtained alternative peptides with higher binding activity from amino acid substituting

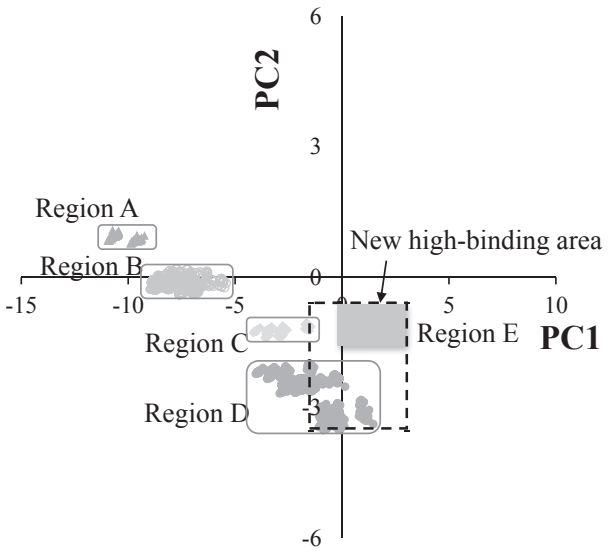


FIG. 3. PC score plot of the 2352 high-binding 4-mer peptides using run no. 2 using 16 variables. The large square drawn with a dashed line is the high-binding rule area. Peptides are clearly separated into four clusters or regions, regions A to D. Closed triangles are region A, open circles are region B, closed diamonds are region C, and closed circles are region D. The gray square is a new high-binding area (region E).

peptides. Therefore, we attempted screening of novel peptide with high binding activity from random peptide library and compered them with that of VAWWMY. We synthesized VAWWMY and did binding assays for bile acid. Among the 24 peptides estimated to be high-binding from the 500 random 6-mer library (Table S6), the number with higher fluorescence intensities than VAWWMY was 8, or 33.3% of the 24 peptides. Interestingly, our group previously had the same result. We constructed a positional rule for 6-mer peptides that have high-binding affinities to bile acid from 2212 peptides selected randomly with neural network analysis (5). According to this work, 32.5% of peptides has a higher binding activity to bile acid than VAWWMY.

According to the positional rule, the first 4 amino acids in a peptide would be important to binding of bile acid. There were 2352 peptides fitting the rule as described in a previous study (5), although all 2352 peptides do not always show high-binding activity. Here, we calculated PC scores of 2352 high-binding 4-mer peptides using variables in our run no. 2. There were four clusters (i.e., regions) clearly separated on the PCA plot of 2352 high-binding peptides (Fig. 3), and each region's score range of PC1 and PC2 are shown in Table S8. Of these, 96 peptides (4.1%) were in region A, 1227 (52.2%) were in region B, 147 (6.3%) were in region C, and 882 (37.5%) were in region D. Number of high binding peptides in region D was higher than that of region A. The reason why this tendency occurred might be owing that high binding peptides were frequently existed also in the peripheral region near by our identified region. As described in Fig. 3, part of region D overlapped with our high-binding rule. Therefore, we think this area is the most important one to identify high-binding peptides. In addition, a new high-binding area, region E, was found by this methodology. It was the region within PC1 values of about −0.26 to 3.06 and PC2 values of about −1.6 to −0.58, within the high-binding rule.

On the other hand, the greatest number of peptides, 1227 peptides (52.2% of 2352 peptides) fell within region B. Nagaoka et al. identified 4 high-binding peptides to bile acid, PWWWMY, IPWYFY, VIWWFK, IYWYMY (unpublished data). With PCA, we found that 3 peptides of these 4 were in region B. Herein, we identified the high-binding region to be within PC1 values of −1.43 to 3.06 and PC2 values of −0.58 to −3.53 (Fig. 1), and this did not overlap with region B. Moreover, we investigated high-binding peptide among 2352 peptide, of which binding activity was higher than VAWWMY as an original peptide. Those were 95 peptides and surprisingly 49 of them were included in region D, whereas 40 and 6 of them were in region B and C, respectively (Tables S8 and S9). The ratio of peptides including in region D was reached to be 51.6% of high binding peptides identified by rule 3. The rule 3 was an amino acid rule of high binding peptides finally acquired by neural network in our previous study (5). The detail rule was the follows; first amino acid was selected from W, F, I, Y, L, V, P, H, M,K, A, R (Trp, Phe, Ile, Tyr, Leu, Val, Pro, His, Met, Lys, Ala and Arg), the second amino acid was selected from W, F, I, Y, L, V, P, the third amino acid was selected from W, Y, R, F, the forth amino acid was selected from W, F, I, Y, L, V, P, and the fifth and sixth amino acid was selected from all amino acids. We reported that rule 3 as identified in previous study was an important rule to search a 6-mer high binding peptides. Here we strongly suggested that region D, not other regions, could point out one of the most important part of rule 3 for identifying higher binding peptides to bile acid.

Our proposed method is an inductive approach for peptide screening. The binding activities of peptides were gathered to be used for peptide sequence analysis and inductive consideration on characteristics of peptide with high binding activity was attempted. Because it is so difficult that molecular recognition between taurocholic acid and fitting peptide are deduced by general phenomenon such as electrostatic force or hydrophobic interaction. Although computational approach based on structural chemistry is one of the candidate deductive approaches, neither of taurocholic acid and peptides have a low degree of freedom because both of them are small molecules. Therefore, suitable fitting structure of those binding complex could be never determined. Our method is a challenging approach for inductive finding of suitable functional peptide. In this manuscript, however, we demonstrated high binding region was successfully pointed out in PC plot by using 512 tetramer peptide library. We well understood that our approach is inductive and plural proof of concept are necessary for acceptance of our approach.

In a previous study (14), our group created this methodology and certified its ability to be used to find high-binding peptides to IL-2 and IgG. The above results showed that this methodology is also useful to find high-binding peptides to bile acid. Since oligopeptides work as functional peptides and it is very meaningful to identify such peptides and their functions, this methodology might be useful to find variously sized high-binding peptides to different targets.

To summarize, by conducting the spot-synthesized peptide array and PCA of 512 4-mer peptides, as well as finding a high-binding rule for peptides to bile acid, we could extract various sizes of high-binding peptides to a specific target effectively. This methodology has the potential to allow exploration of peptide binding to other targets of biological and medical relevance.

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jbiosc.2018.08.006.

## ACKNOWLEDGMENTS

This work was partially supported by JSPS KAKENHI (Grant Numbers: JP16H04575). We would like to thank Editage for English language editing.

## References

1. **Kato, R., Kunimatsu, M., Fujimoto, S., Kobayashi, T., and Honda, H.:** Angiotensin II inhibitory peptide found in the receptor sequence using peptide array, Biochem. Biophys. Res. Commun., **315**, 22–29 (2004).

2. **Ochiai, T., Sugita, T., Kato, R., Okochi, M., and Honda, H.:** Screening of α-amylase inhibitor peptide by photolinke-peptide array, Biosci. Biotechnol. Biochem., **76**, 819–824 (2012).
3. **Sugita, T., Katayama, M., Okochi, M., Ichihara, T., Kato, R., and Honda, H.:** Screening of peptide ligands that bind to the Fc region of IgG using peptide array and its application to affinity purification of antibody, Biochem. Eng. J., **79**, 33–40 (2013).
4. **Kume, A., Okochi, M., Shimizu, K., Yoshida, Y., and Honda, H.:** Development of a tactical screening method to investigate the characteristics of functional peptides, Biotechnol. Bioprocess Eng., **21**, 119–127 (2016).
5. **Takeshita, T., Okochi, M., Kato, R., Kaga, C., Tomita, Y., Nagaoka, S., and Honda, H.:** Screening of peptides with a high affinity to bile acids using peptide arrays and a computational analysis, J. Biosci. Bioeng., **112**, 92–97 (2011).
6. **Okochi, M., Nakanishi, M., Kato, R., Kobayashi, T., and Honda, H.:** High-throughput screening of cell death inducible short peptides from TNF-related apoptosis-inducing ligand sequence, FEBS Lett., **580**, 885–889 (2006).
7. **Kaga, C., Okochi, M., Nakanishi, M., Hayashi, H., Kato, R., and Honda, H.:** Screening of a novel octamer peptide, CNSCWSKD, that induces caspase-dependent cell death, Biochem. Biophys. Res. Commun., **362**, 1063–1068 (2007).
8. **Okochi, M., Nomura, S., Kaga, C., and Honda, H.:** Peptide array-based screening of human mesenchymal stem cell-adhesive peptides derived from fibronectin type III domain, Biochem. Biophys. Res. Commun., **371**, 85–89 (2008).
9. **Kanie, K., Narita, Y., Owaki, J., Zhao, Y., Kuwabara, F., Satake, M., Honda, S., Kaneko, H., Yoshioka, T., Okochi, M., Honda, H., and Kato, R.:** Collagen type IV-specific tripeptides for selective adhesion of endothelial and smooth muscle cells, Biotechnol. Bioeng., **109**, 1808–1816 (2012).
10. **Altmann, S. W., Davis, H. R., Zhu, L. J., Yao, X., Hoos, L. M., Tetzloff, G., Iyer, S. P. N., Maguire, M., Golovko, A., Zeng, M., and other 3 authors:** Niemann–Pick C1 like 1 protein is critical for intestinal cholesterol absorption, Science, **303**, 1201–1204 (2004).
11. **Nagaoka, S., Nakamura, A., Shibata, H., and Kanamaru, Y.:** Soystatin (VAWWMY),a novel bile acid-binding peptide, decreased micellar solubility and inhibitedcholesterol absorption in rats, Biosci. Biotechnol. Biochem., **74**, 1738–1741 (2010).
12. **Choi, S.-K., Adachi, M., and Utsumi, S.:** Identification of the bile acid-binding region in the soy glycinin A1aB1b subunit, Biosci. Biotechnol. Biochem., **66**, 2395–2401 (2002).
13. **Choi, S.-K., Adachi, M., and Utsumi, S.:** Improved bile acid-binding ability of soybean glycinin Ala polypeptide by the introduction of a bile acid-binding peptide (VAWWMY), Biosci. Biotechnol. Biochem., **68**, 1980–1983 (2004).
14. **Kume, A., Kawai, S., Kato, R., Iwata, S., Shimizu, K., and Honda, H.:** Exploring of high binding octamer peptides based on principal component analysis of tetramer peptides, J. Biosci. Bioeng., **113**, 230–238 (2017).
15. **Zimmerman, J. M., Eliezer, N., and Simha, R.:** The characterization of amino acid sequences in protein by statistical methods, J. Theor. Biol., **21**, 170–201 (1968).
16. **Takano, K. and Yutani, K.:** A new scale for side-chain contribution to protein stability based on the empirical stability analysis of mutant proteins, Protein Eng., **14**, 525–528 (2001).
17. **Kyte, J. and Doolittle, R. F.:** A simple method for displaying the hydropathic character of a protein, J. Mol. Biol., **157**, 105–132 (1982).
18. **Black, S. D. and Mould, D. R.:** Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications, Anal. Biochem., **193**, 72–82 (1991).
19. **Zamyatnin, A. A.:** Protein volume in solution, Prog. Biophys. Mol. Biol., **24**, 107–123 (1972).