

Research

Effective screening for hypolipidemic bile acid-binding peptides by hierarchical clustering-based catalog array

Ryuji Kato<sup>1,2</sup>, Kei Kanie<sup>1,2</sup>, Chiaki Kaga<sup>1</sup>, Mina Okochi<sup>1,3</sup>, Satoshi Nagaoka<sup>4</sup>, Hiroyuki Honda<sup>1,5\*</sup>

<sup>1</sup> Dept. of Biotechnology, Graduate School of Engineering, Nagoya University, Nagoya, Japan

<sup>2</sup> Dept. of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, Nagoya, Japan

<sup>3</sup> Dept. of Chemical Science and Engineering, School of Materials and Chemical Technology, Tokyo Institute of Technology, Japan

<sup>4</sup> Dept. of Applied Life Science, Gifu University, Gifu, Japan

<sup>5</sup> MEXT Innovative Research Center for Preventive Medical Engineering, Nagoya University, Nagoya, Japan

**Abstract**

Peptide arrays, which are high-throughput combinatorial screening platforms, have a limited capacity for exploring vast amino acid sequence combinations. To mitigate such limitations, we propose the concept of “catalog array,” comprised of bioinformatically enriched minimal peptides representing maximal physicochemical variety. For this method, hierarchical clustering was introduced to group and select peptides with similar physicochemical properties to be synthesized on the array. Indeed, clustering determined that all 3-mer peptides could be narrowed to 47 groups using the 13 amino acid indices. We subsequently evaluated the validity of our array by screening for bile acid-binding proteins. As a target bio-functional peptide, bile acid-binding peptides have recently gained attention due to their hypolipidemic properties. Notably, the catalog array provided up to four-fold higher screening efficacy (23.2% efficacy) over simple random screening (6.9%), enabling the identification of peptides with greater bile acid-binding capacities than the previously reported 6-mer peptide VAWWMY.

Keywords: Hyperlipidemia; Bile acid binding peptide; Peptide array; Hierarchical clustering; Screening

**INTRODUCTION**

Hypercholesterolemia is a primary risk factor for cardiovascular disease [1-3]. Recently, bile acid was identified as a plausible target for lowering cholesterol, as bile acid-binding agents alter the way cholesterol is absorbed [4-9]. Specifically, such agents, including cholestyramine, are designed to bind bile acids with high affinity within the intestine [10], thereby disrupting bile acid micelles, which are the key complexes that capture and target dietary cholesterol to specific transporters for absorption. Thus, disruption of micelle formation leads to decreased cholesterol absorption. Moreover, agents with high bile acid binding capacities prevent the re-absorption of bile acids (95–98% of secreted bile acids) in the lower intestine, leading to bile acid excretion.

Lastly, in addition to inhibiting bile acid re-absorption, bile acid-binding peptides can also trigger the activation of cholesterol 7 $\alpha$ -hydroxylase in the liver, promoting the synthesis of new bile acids by consuming stored cholesterol.

**Corresponding Author:** Ryuji Kato, Dept. of Biotechnology, Graduate School of Engineering, Nagoya University, Nagoya, Japan, Dept. of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, Nagoya, Japan E-mail: kato-r@ps.nagoya-u.ac.jp

**Received date:** July 25, 2018; **Accepted date:** August 06, 2018; **Published date:** August 07, 2018.

**DOI:** <https://doi.org/10.31546/BBAB.1001>

Previously, Choi et al. reported that a hydrophobic peptide derived from the soy protein glycinin binds strongly to bile acid in vitro [11, 12]. Nagaoka S. et al. found for the first time that VAWWMY (soystatin) derived from soybean glycinin acted as an inhibitor of cholesterol absorption in vivo [13]. There are so many in vitro active soypeptides, soystatin is an only in vivo active hypocholesterolemic peptides derived from soybean [14]. Given the growing requirements for daily preventive therapies for lifestyle related diseases, hypolipidemic peptides could yield new avenues for medical dietary interventions. Notably, the use of such naturally occurring peptides should be more safe and effective than protein hydrolysates [14-18], which contain multiple unknown components in addition to the active molecules.

Combinatorial peptide library technologies (e.g., SPOT peptide array) [19-22] offer advantages for high-throughput screening and analysis of bile acid-binding peptides. The ability to make positional amino acid substitutions and to examine peptides of different lengths has helped to advance our understanding of the structural rules of peptide interactions with various substrates. However, the greatest disadvantage of array-type peptide libraries is the limited size of the screenable library compared to the total number of peptide varieties. Essentially, the number of possible amino acid sequences ( $20^n$  for a peptide that is  $n$  residues long) is too vast to cover exhaustively, both in terms of physical array area and preparation costs. In this study, to compensate for these size limitations, we propose a hierarchical clustering-based array design, which we refer to as the catalog array.

Hierarchical clustering, also known as the unsupervised data segmentation method, groups objects into subsets, called "clusters," using multiple information variants. In bioinformatics, this method is most commonly applied to interpret similarities in gene expression profiles among the samples [23, 24]. With hierarchical clustering, all objects in a data set can be described in a dendrogram showing each similarity. Therefore,

data can be objectively divided into several groups. Since our previous studies showed that positional information for amino acid indices (i.e., physicochemical property information) were predictive of short peptide affinity (<16-mer)<sup>24-28</sup>, we hypothesized that clustering based on such amino acid indices would identify peptides of similar affinity. If peptides with similar affinity could subsequently be grouped successfully by in silico clustering (with small deviations in each cluster), the peptide library could be sized-down by selecting a few representative peptides to reduce inherent redundancy within the screening.

In this report, we verify the effectiveness of our clustering-enriched catalog array design technique by screening for bile acid-binding peptides.

### Materials and Methods

**Selection of amino acid indices:** Variables for conversion of amino acids in peptide sequences were obtained from the AAindex1 reported in Genome Net Japan, organized by Kyoto University (<http://www.genome.ad.jp/dbget/aaindex.html>) [30]. A total of 544 amino acid indices (Table SI) registered in the database (version 9.1, as of January 2008) were found to comprise 21 clusters, with high correlation (Figure S1). To represent the remaining 13 independent clusters covering the majority of indices, 13 indices were selected (schematic illustration in Figure 1A; detailed list in Table SII): [1] Isoelectric point [31], [2] Normalized van der Waals volume [32], [3] Alpha-helix indices for beta-proteins [33], [4] Beta-strand indices for beta-proteins [33], [5] Side-chain contribution to protein stability [34], [6] The stability scale from the knowledge-based atom-atom potential [35], [7] Hydrophathy index [36], [8] Normalized frequency of turn [37], [9] Free energy in beta-strand region [38], [10] Free energy in alpha-helical region [38], [11] Polarity [31], [12] Side chain interaction parameter [39], and [13] Amino acid distribution [40]. Each amino acid of each sequenced peptide sequenced was numbered from the N to C-terminus using each of the 13 indices. Therefore, each amino acid index profile for each sequence included<sup>39</sup> parameters (3 positions  $\times$  13 indices).

**Hierarchical clustering:** Average-linkage hierarchical clustering was calculated using Cluster 3.0 software, distributed by Michiel de Hoon et al. at the University of Tokyo, Human Genome Center (<http://bonsai.ims.utokyo.ac.jp/~mdehoon/software/cluster/software.htm>). Results from hierarchical clustering analysis were visualized with the open source tool Maple Tree, version 0.2.3.2 BETA (<http://mapletree.sourceforge.net>) (See Supplementary Dataset S1 for .cdt and .gtr files). Detailed sequence data analysis using clustering results was performed with an original source program written in C.

**Peptide array synthesis:** The peptide array was designed and synthesized by Fmoc synthesis, according to the manufacturer's protocol but with some modifications, as reported in previous works (24–26). The first catalog array consisted of 846 trimer peptides (triplicate spots for each of six representative peptides for each of 47 clusters; see Table SV for peptide list), which represented the physicochemical properties of 8,000 peptides (203 combinations). The second catalog array consisted of 750 peptides (3 spots  $\times$  25 peptides  $\times$  10 clusters; see Table SV for peptide list). For comparison to the catalog array, three different types of random arrays (designated random 1–3), consisting of trimer peptides randomly selected from the 8,000 peptides *in silico*, were generated with no redundancy (see Table SV for peptide list). For reproducibility, three copies of each array were prepared. All spots were randomly distributed during array synthesis. In each array, positive controls (VAWWMY) and negative controls (linker only) were included with eight internal control sequences (see supplementary information).

**Bile acid binding assay with peptide array:**

The detailed binding assay protocol and the methods used for data analysis are described in the supplementary material. Briefly, peptide arrays were hybridized with taurocholate (Sigma-Aldrich, St. Louis, MO, USA), anti-cholic acid (Cosmo Bio, Tokyo, Japan), and Alexa 488-conjugated anti-

rabbit IgG, respectively. The peptide spots were scanned with a FLA-7000 device (Fujifilm, Tokyo, Japan) and analyzed with ArrayGauge Ver.2.1 software (Fujifilm).

**Results**

***Designing the catalog array by two-step hierarchical clustering:***

To maximize library diversity and minimize redundant examination of biased types of peptides, we utilized hierarchical clustering to group peptides by similarities in bile acid binding affinity. A schematic flow diagram illustrating the design of the catalog array is shown Figure 1. An initial hierarchical clustering step (clustering of AA indices) was performed to reduce the number of informative variables and to identify the representative indices of those that are clearly independent (Figure S1). For effective clustering, such independent numerical variables are required to provide the optimal variability in the physicochemical aspects of a peptide. Thirteen of the most independent amino acid indices were selected from the first hierarchical clustering to represent 544 indices in the AA index database (Tables SI and SII).

After the initial clustering, each peptide sequence was converted into numerical values at the sequence conversion step (Figure 1B). The sequences covering trimer variety (8,000 = 203) were then converted into 39 parameters (3 positions  $\times$  13 indices). The peptides, represented by the resulting variables, were then subjected to a second hierarchical clustering (clustering of peptides) step. Representative peptides from each cluster were randomly selected for construction of a minimal library to be incorporated into the catalog array. The selected peptides provided maximal coverage of the variability in peptide binding properties using the fewest number of peptides (Figure 1C).

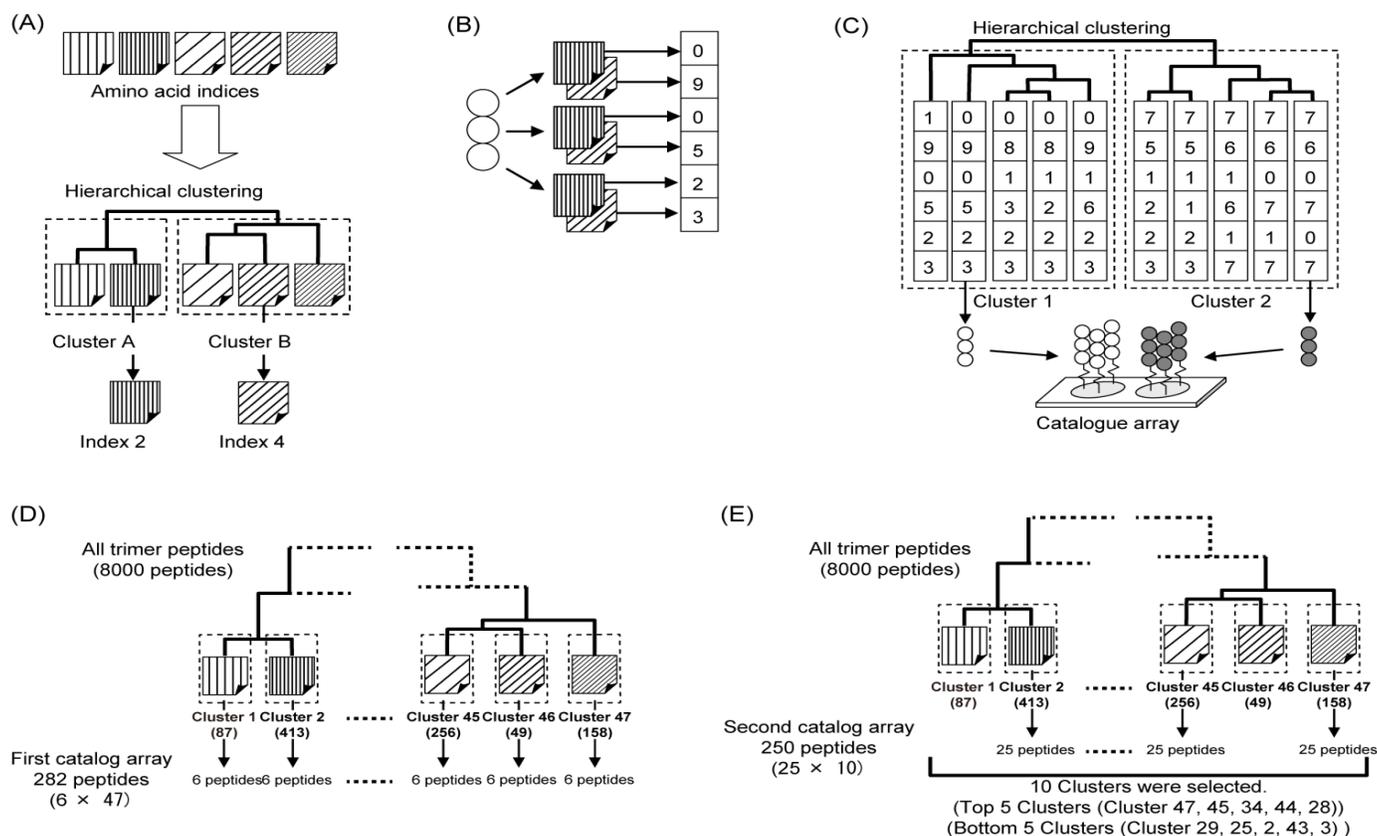
A third hierarchical clustering step, clustering of all trimer peptides, was subsequently applied to the numerically converted peptides for physicochemical grouping. Forty-seven clusters were selected to represent the 47 distinct physicochemical property

groups (Table SIII). Six representative peptides from each selected clusters were randomly selected for synthesis for the first catalog array (Figure 1D). After narrowing the first catalog array to 10 clusters, 25 representative peptides from each cluster were again randomly selected for the second catalog array (Figure 1D). The variety in the number of members per cluster indicated that the physicochemical properties of trimer peptides are greatly biased; there was a 22-fold difference in the

number of cluster members between the maximum cluster and the minimum cluster (detailed list in Tables SIII and SIV). Three clusters (no. 8, no. 9, and no. 31) had fewer than 20 members, while the largest clusters (no. 2, no. 20 and no. 38) had greater than 300 members. Notably, while such huge biases in physicochemical properties suggest that peptides with rare physicochemical properties will rarely appear in random screening, the catalog array can account for these biases by evenly weighting each variety, thereby enabling balanced testing

Figure 1. Graphic depiction of the flow of the catalog array-based screening method.

Notes: (A) First hierarchical clustering step for the selection of amino acid indices. Based on the clustering results, a representative index from each cluster is chosen. (B) Peptide sequence conversion by amino acid indices. In this case, one residue is converted into two numerical variables. (C) Second hierarchical clustering step for the selection of representative peptides for the catalog array. In this case, two clusters were obtained from the clustering of 10 peptides. From each cluster, a representative peptide is randomly selected for synthesis on catalog array. (D) Schematic images of the first and second catalog array designs. Forty-seven clusters were selected to represent the 47 distinct physicochemical property groups. Six representative peptides were randomly selected from each cluster for synthesis on the first catalog array (six peptides  $\times$  47 clusters = 282 peptides). By limiting the first catalog array to 10 clusters (top five and bottom five clusters), 25 representative peptides from each cluster were again randomly selected for synthesis on the second catalog array (25 peptides  $\times$  10 clusters = 250 peptides).



### ***Physicochemical preferences of bile acid-binding peptides interpreted from catalog array***

The catalog array screening has the advantage of interpreting properties that commonly appear in peptides with high affinities for bile acids. To evaluate the bile acid binding properties of cataloged trimer peptides, we hybridized bile acids (taurocholic acids) to the first catalog array (six peptides from each of the 47 clusters). Table SIV lists the average fluorescence value for each of the 47 clusters. The low coefficient of variation (CV) value in each of the 28 clusters indicates that six randomly selected representative peptides from the same cluster share similar affinities for bile acids (Table I). These findings support our hypothesis that the affinity of short peptides may be clustered by their physicochemical properties, also suggesting that peptides in the same cluster could be represented by several candidates to enable screening of diverse physicochemical properties.

From the clusters in the top or bottom rankings, we identified characteristic amino acid index profiles underlying the interaction between peptides and bile acids, which could be interpreted as principles of physicochemical preferences (Figure 2). The five most stable clusters (cluster no. 48, 46, 35, 45, and 29) shared several common structural characteristics related to high affinity bile acids; each of these clusters encoded an N-terminal Pro, with Tyr/Phe/Ile/Met/Leu/Val/Trp at any position. These amino acids share a high value for three indices: side-chain contribution to protein stability, stability scale based on knowledge-based atom-atom potential, and hydrophobicity. Interestingly, aromatic residues (Trp, Tyr, and Phe) had a strong positive effect on bile acid binding capacity. Likewise, the five least stable clusters (cluster no. 30, 26, 2, 44, and 3) exhibited common structural characteristics related to low affinity bile acids: possession of a Cys or Pro at the C-terminal end and Asp/Glu/Ala/Gly/Ser/Gln/Thr in the N-terminal position. These properties have low values for three indices: normalized van der Waals volume, beta-strand indices from beta-proteins, and stability scale based on knowledge-based atom-atom

potential. This is the first time that such amino acid preferences contributing to bile acid interactions with trimer peptides have been characterized and analyzed. As such, these results provide principles for the future design of bile acid-binding molecules, including potential drugs for the treatment of high cholesterol.

### ***Focused screening of bile acid-binding peptides based on the catalog array results.***

To further examine the physicochemical properties that are important for bile acid binding capacity, we expanded the screening size within the specific clusters identified in the initial catalog array. We selected the five most stable (cluster no. 48, 46, 35, 45, and 29) and five least stable (cluster no. 30, 26, 2, 44, and 3) clusters, and randomly selected 25 new representative peptides to synthesize on the second catalog array.

According to our hypothesis, if the hierarchical clustering of 8,000 peptides were successful, the peptides selected from the once examined clusters should show similar affinities to bile acids. As shown in Figure 3A and 3B, the newly selected representatives from the most stable clusters exhibited approximately 2-fold higher (average 1,633) affinities for bile acid than those of the least stable clusters (average 894), and this difference was statistically significant ( $P = 10^{-30}$ ) (Table SV), suggesting that the peptides in the stable clusters bind to bile acids with significantly greater strength. This concordance with the results of first catalog array strongly supports the validity of our approach of hierarchical clustering of short peptides by physicochemical properties.

From this focused screening, we obtained six new trimer peptides (WIY, YYW, WRW, PWR, MFW, and WFM) that exhibited extremely high bile acid binding capacities (Table SV). Interestingly, these trimers had higher bile acid affinities than the longer bile acid-binding peptide VAWWMY, which was previously shown to strongly suppress cholesterol absorption in liver, intestine, and serum by simple oral administration [13]. Trimers YYW, WFW, WIY, WRW, and WYM were grouped in the top 2 clusters (no. 47 and 45), while trimer PWR was grouped in

the fifth cluster from the top (no. 28) from the first catalog screening (Table SV). These results again indicate that the amino acid profile of the top clusters (Figure 2) was an important characteristic for identifying peptides with high affinities for bile acid.

### **Screening efficacy of the peptide catalog array**

The concept behind the catalog array is that numerous peptide variations can be covered using a limited number of representative peptides through successful clustering. By clustering, it was expected that a minimal number of representative peptides could provide maximal levels of independence and variation to enhance screening efficacy. To confirm the screening effectiveness of the catalog array, we synthesized three random arrays (random 1–3) for comparison. Each random array consisted of 125 peptides randomly selected from the 8,000 possible trimer variations, without redundancy. Figure 3C shows the distributions of the bile acid binding affinities of the peptides in these random arrays; each of these peptides was distributed within the low affinity range (average fluorescent values, random 1 = 1,347; random 2 = 1,257; and random 3 = 1,032). Notably, the bile acid binding affinities of the peptides in the top five clusters in the catalog array exhibited significantly higher affinities ( $P = 3.1 \times 10^{-13}$ ; Figure 3A, 3C) for bile acid than those in the random arrays. Meanwhile, the peptides in the bottom five clusters exhibited significantly lower affinities ( $P = 4.29 \times 10^{-27}$ ) than those in the random arrays (Figure 3B, 3C). Moreover, of the 375 randomly selected peptides on the random arrays, no peptide had a higher affinity for bile acid than the strong bile acid-binding peptide VAWWMI, and the yield efficacy to obtain high affinity peptides (fluorescence intensity > 2,000) was only 6.9%. In contrast, catalog array screening, using one-third the number of peptides (125), succeeded in identifying six peptides with stronger affinities for bile acid than VAWWMI (yield efficacy = 23.2%). These results further support our hypothesis that the affinities of short peptides are predominantly controlled by amino acid physiochemical properties, and that clustering by those properties can effectively minimize the number

of candidate peptides synthesized for the array while maximizing the screening efficacy.

### **Construction of a bile acid-binding peptide using trimer peptides as building blocks**

To apply the principles elucidated by the catalog array to the design of novel bile acid-binding peptides, we attempted to build longer peptides by connecting successful trimer peptides as “building blocks”. For example, using four high-affinity trimers (WYM, YMW, MWM, and WMI), we synthesized WYMW (4-mer), WYMWMI (5-mer), and WYMWMI (6-mer) by merging the overlapping amino acids of distinct trimers. In the same manner, we enumerated the potential combinations of top cluster peptides *in silico*. A total of 164 possible tetramers, 157 pentamers, and 186 hexamers were subsequently identified and synthesized onto a peptide array for screening. The five novel peptides with the highest bile acid binding properties are shown in Table II. By joining bile acid-binding trimers from bile acid-binding clusters, an elongated peptide molecule with high bile acid binding affinity (exceeding that of VAWWMI) could be obtained at an efficiency of 36% (57/159 peptides). This result suggests that a protein consisting of bile acid-binding peptides from bile acid-binding clusters is likely to exhibit a high bile acid binding affinity, both as an intact peptide and as resultant trimers after digestion.

### **Discussion**

In this study, we introduced *in silico* hierarchical clustering to design an enriched array that compensates for the volume limitations of array-based screens. To our knowledge, this is the first study to combine the clustering strategy to design an efficient peptide library for peptide array screening.

In the present study, our catalog array successfully identified both effective and ineffective “clusters” of bile acid-binding peptides (Figure 2). Analysis of these clusters also highlighted the optimal physiochemical characteristics of peptides exhibiting high binding affinities for bile acid.

Typically, peptides that bind to bile acid are highly hydrophobic 4), such as the previously reported

peptide VAWWMY, which was isolated from the most hydrophobic region of the soybean glycinin subunit 11). Through our screening, however, we found that this single simple property was insufficient to predict the strength of interactions between peptides and bile acids. For example, several candidate peptides with extreme hydrophobicity (IVV, VFI, VLF, ILM, and MLI) exhibited half the bile acid binding capacity as those in the top five clusters selected using the catalog array (Table SVI). Furthermore, peptides from the strongest bile acid binding clusters had relatively hydrophilic properties compared to those made up of only hydrophobic residues. Our clustering-based screening was effective for visualizing and interpreting such complex guiding principles for desirable and undesirable molecular properties by leveraging the positive and negative informational advantages provided by array-based screening (Figure 1).

Notably, we demonstrated that peptides exhibiting high affinities for bile acid could serve as “building blocks” for designing longer bile acid-binding peptides (Table SVII). Such an approach should be helpful in designing new peptides or molecules for a variety of applications, with minimum screening cost and time. We also expect that the concept of a “building block” could be further applied in reverse to score unknown proteins exhibiting hypolipidemic effects. Moreover, our observations support the concept of characterizing elongated peptides with digestion product trimers that also have significant bile acid binding properties. Given that bile acid-binding peptides act to lower cholesterol in the intestine, where the digestion of proteins produces many trimer peptides, evaluating the prevalence of peptide blocks belonging to bile acid-binding clusters within larger proteins could serve as a novel index for characterizing hypolipidemic foods. Indeed, the idea of “screening natural resources by peptide clusters” could comprise a bottom-up alternative to the

classical HPLC-based strategy for finding novel hypolipidemic natural resources.

While there are many advantages to the peptide catalog array method proposed here, this strategy does have several restrictions. First, peptide catalog arrays can only be designed using short peptides, as clustering of longer peptides would require an extremely high computational capacity. As such, other clustering methods should be chosen for investigating longer peptides to reduce these computational demands. In addition, determination of final cluster numbers is a critical problem when evaluating peptides with high sequence variation. Since the final cluster numbers in this study were determined intentionally according to experimental size, criteria to optimize the final cluster number should be introduced as a goal in future work.

Second, the catalog array strategy, which utilizes amino acid indices for property description, can only be applied to investigating interactions using short peptides. Since the behavior of long peptides is strongly influenced by their secondary sequences and three-dimensional conformations, more informative indices (e.g., describing secondary sequences) should be introduced. Lastly, catalog array screening should be limited to studying interactions in which affinity is predominantly controlled by physicochemical properties and which are amenable to multiple mutations, as the aim of clustering is to evaluate pools of similar peptides, as opposed to one specific peptide.

In spite of these limitations, our catalog array-based screening method provides a powerful new strategy for understanding the interactions between small peptides and molecules. We expect that the structural rules elucidated in this study will contribute to the design of novel peptide drugs and functional dietary supplements for preventing lifestyle-related hyperlipidemia.

Table 1 Ranking of peptide clusters by average bile acid-binding capacity

Ranking	Cluster No.	Average bile acid binding ratio [-] <sup>‡</sup>	S.D. <sup>§</sup>	Number of cluster members	Ranking	Cluster No.	Average bile acid binding ratio [-] <sup>‡</sup>	S.D. <sup>§</sup>	Number of cluster members
1	47*	1.23	0.39	158	15	37	0.54	0.12	329
2	45*	0.89	0.22	256	16	5	0.53	0.09	108
3	34*	0.84	0.37	275	17	4	0.52	0.29	204
4	44*	0.80	0.37	82	18	33	0.52	0.19	72
5	28*	0.78	0.21	152	19	24	0.50	0.14	292
6	19	0.74	0.12	373	20	20	0.50	0.24	228
7	38	0.71	0.33	198	21	1	0.49	0.15	87
8	14	0.70	0.19	204	22	23	0.46	0.12	88
9	30	0.68	0.16	152	23	18	0.46	0.08	124
10	9	0.64	0.06	20	24	29 <sup>†</sup>	0.46	0.24	209
11	42	0.63	0.11	43	25	25 <sup>†</sup>	0.43	0.07	124
12	27	0.61	0.17	209	26	2 <sup>†</sup>	0.41	0.16	413
13	36	0.57	0.03	74	27	43 <sup>†</sup>	0.41	0.14	74
14	40	0.54	0.18	213	28	3 <sup>†</sup>	0.41	0.08	82

\*Top five clusters. †Bottom five clusters. ‡Average and §standard deviation of bile acid binding ratio are the average of bile acid ratios of six peptides selected from each cluster and assayed in the first catalog array. For each peptide, a bile acid ratio was obtained from the median of nine fluorescence intensity values (triplicate spots from three different arrays) normalized against the average fluorescence intensity of the linker only spot. Linker only, 1.00; VAWWMY, 2.95. Of the 47 clusters, those with coefficient of variation (CV) values > 20 were eliminated from the ranking.

Table 2 Top five peptides designed by combining trimer bile acid-binding peptides as building blocks

Ranking	Sequence	Length	Bile acid binding ratio [-] <sup>‡</sup>
Top 1	PWRW	4	5.36
Top 2	PWRI	4	4.16
Top 3	WYMW	4	3.95
Top 4	WWIY	4	3.68
Top 5	YYWM	4	3.62
Top 1	WYMWM	5	4.70
Top 2	YWYY	5	4.62
Top 3	YYWY	5	4.61
Top 4	YYWMM	5	4.54
Top 5	YYWMQ	5	4.47
Top 1	PWYMW	6	5.67
Top 2	PWYYW	6	5.41
Top 3	YYWYY	6	5.25
Top 4	AWYMW	6	5.24
Top 5	WYMWMQ	6	5.21

\*For each peptide, a bile acid ratio was obtained from the median of nine fluorescence intensity values (triplicate spots from three different arrays) normalized against the average fluorescence intensity of the linker only spot (1.00). Linker only, 1.00; VAWWMY, 3.38. See Table SVI for other peptides. Figure 1. Type your caption here. Obtain permission and include the acknowledgement required by the copyright holder if a figure is being reproduced from another source.

Figure 2. Profile of amino acid indices.

Notes: (A) Heat map profile of the amino acid indices for the five clusters with the highest bile acid binding capacities. (B) Heat map profile of the amino acid indices of the five clusters with the lowest bile acid binding capacities. Color gradations to green and red indicate lower negative (<-1.5) and higher positive values (>+1.5), respectively. The black color indicates the middle value (0). Rows, 39 variables representing 13 amino acid indices × 3 positions; columns, peptide sequences (see Table SIII). See also Figure S3 for detailed clustering results.

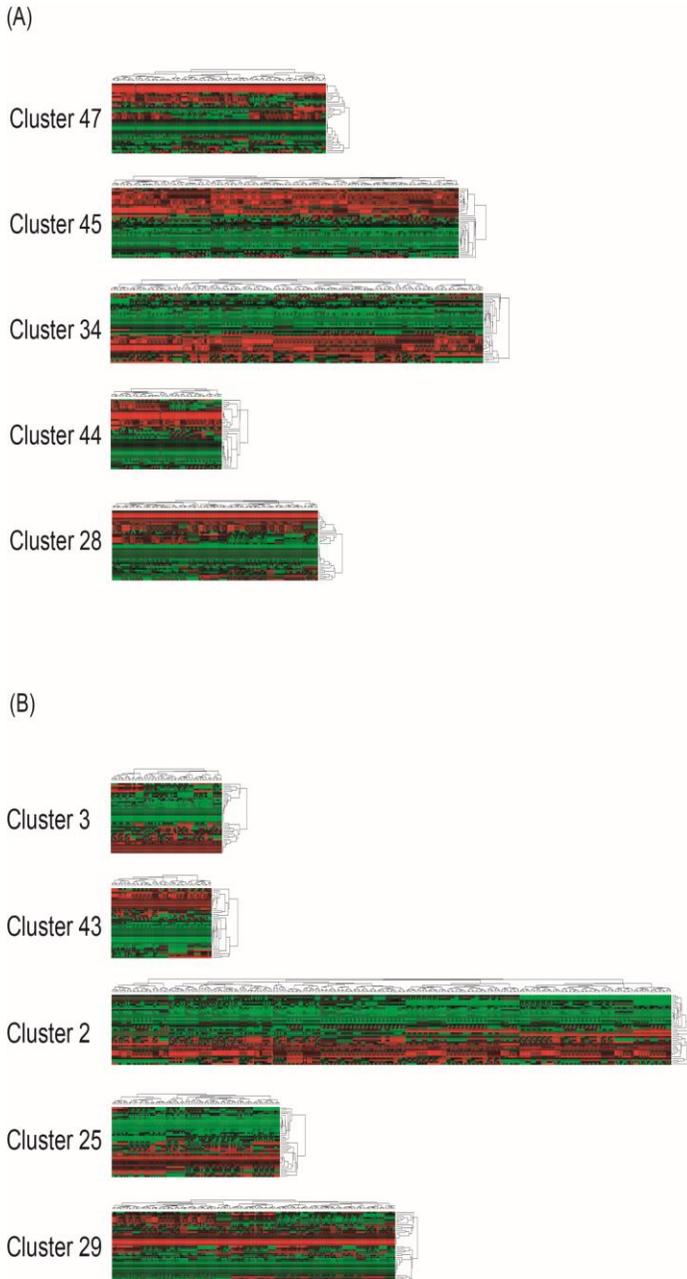
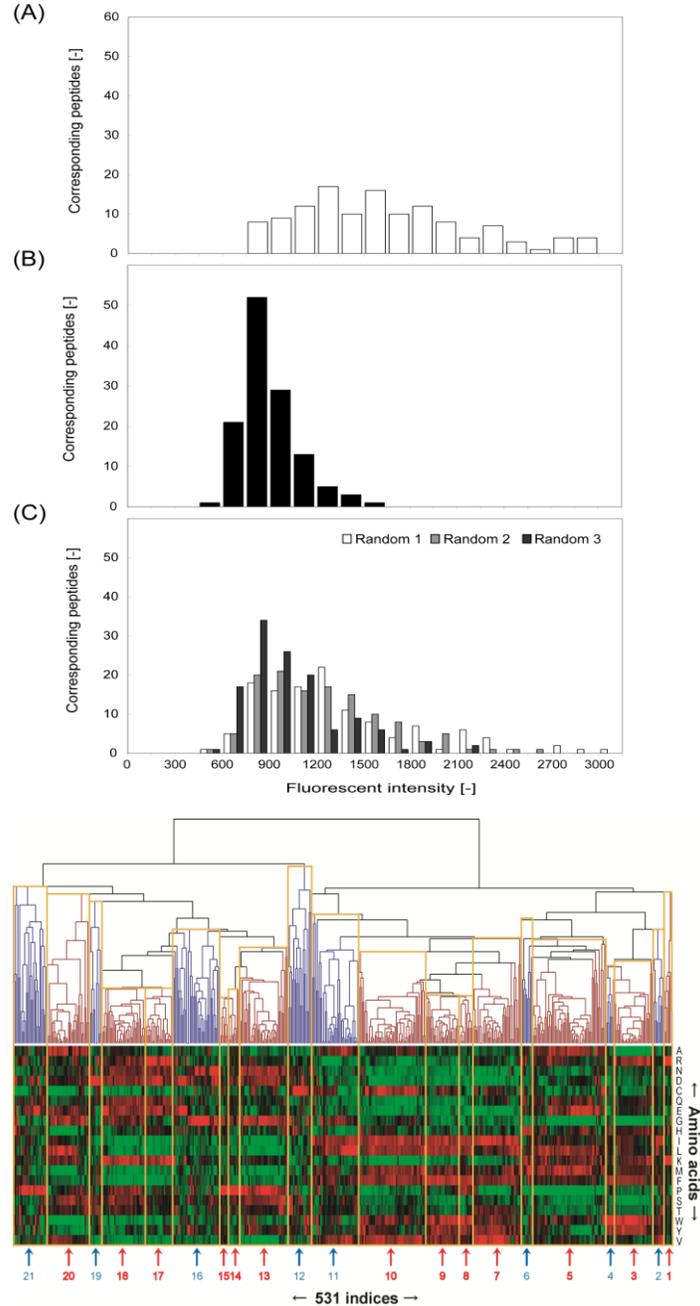


Figure 3. Distribution of the bile acid binding capacities of the peptides on array.

Notes: Bile acid binding capacities were calculated as the median FL3 value of nine peptide spots (see Materials and Methods). (A and B) Distribution of the bile acid binding capacities of 125 peptides randomly selected from the five clusters associated with the (A) highest (no. 47, 45, 34, 44, and 28) and (B) lowest bile acid binding capacities (no. 3, 43, 2, 25, and 29). (C) Three distributions of the 125 peptides randomly selected from the pool of 8,000 peptides.



**Acknowledgments:**

We greatly appreciate Dr. Mitoshi Kunimatsu for his assistance in establishing the peptide assay platform.

**Author contribution:**

R.K., C.K., M.O., S.N. and H.H. designed research, K.K. performed the experiments, R.K., and K.K. analyzed data, and R.K. wrote the manuscript. All authors have read and approved the final manuscript.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding:**

This study was partially supported by the [Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science and Ministry of Education, Science, Sports and Culture] under Grant [No. 20650074]; [Grant-in-Aid for Exploratory Research] under Grant [No.20380075-01].

**References:**

1. Beynen AC, Katan MB, Van Zutphen LF. Hypo- and hyperresponders: individual differences in the response of serum cholesterol concentration to changes in diet. *Adv Lipid Res.* 1987;22:115-171.
2. Burnett JR, Huff MW. Cholesterol absorption inhibitors as a therapeutic option for hypercholesterolaemia. *Expert Opin Investig Drugs.* 2006;15:1337-1351.
3. Assmann G, Kannenberg F, Ramey DR, et al. Effects of ezetimibe, simvastatin, atorvastatin, and ezetimibe-statin therapies on non-cholesterol sterols in patients with primary hypercholesterolemia. *Curr Med Res Opin.* 2008;24:249-259.
4. Iwami K, Sakakibara, K., and Ibuki, F. Involvement of post-digestion 'Hydrophobic' peptides in plasma cholesterol-lowering effect of dietary plant proteins. *Agricultural and Biological Chemistry.* 1986;50:1217-1222.
5. Iwami K, Kitagawa M, Ibuki F. Effect of dietary proteins and/or their digestive products on intestinal taurocholate absorption. *J Nutr Sci Vitaminol (Tokyo).* 1990;36 Suppl 2:S141-146.
6. Kramer W, Wess G, Schubert G, et al. Liver specific drug targeting by coupling to bile-acids. *J Biol Chem.* 1992;267:18598-18604.
7. Wilson MD, Rudel LL. Review of cholesterol absorption with emphasis on dietary and biliary cholesterol. *J Lipid Res.* 1994;35:943-955.
8. Kramer W, Glombik H. Bile acid reabsorption inhibitors (BARI): novel hypolipidemic drugs. *Curr Med Chem.* 2006;13:997-1016.
9. Thomas C, Pellicciari R, Pruzanski M, et al. Targeting bile-acid signalling for metabolic diseases. *Nat Rev Drug Discovery.* 2008;7: 678-693.
10. Kobayashi M, Ikegami H, Fujisawa T, et al. Prevention and treatment of obesity, insulin resistance, and diabetes by bile acid-binding resin. *Diabetes.* 2007;56:239-247.
11. Choi S, Adachi M, Utsumi S. Identification of the bile acid-binding region in the soy glycinin A1aB1b subunit. *Biosci Biotechnol Biochem.* 2002;66:2395-2401.
12. Choi S, Adachi M, 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.* 2004;68:1980-1983.

13. Nagaoka S., Nakamura A., Shibata H., Kanamaru Y. Soystatin (VAWWMY), a novel bile acid-binding peptide, decreased micellar solubility and inhibited cholesterol absorption in rats. *Biosci. Biotechnol. Biochem.* 2010; 1738-1741.
14. Nagaoka S. Structure-function properties of hypolipidemic peptides. *J. Food Biochem.* 2018, e12539.
15. Sugano M, Goto S, Yamada Y, et al. Cholesterol-lowering activity of various undigested fractions of soybean protein in rats. *J Nutr.* 1990; 120:977-985.
16. Lovati MR, Manzoni C, Gianazza E, et al. Soy protein peptides regulate cholesterol homeostasis in Hep G2 cells. *J Nutr.* 2000;130:2543-2549.
17. Nagaoka S, Futamura Y, Miwa K, et al. Identification of novel hypocholesterolemic peptides derived from bovine milk beta-lactoglobulin. *Biochem Biophys Res Commun.* 2001;281:11-17.
18. Tovar AR, Murguia F, Cruz C, et al. A soy protein diet alters hepatic lipid metabolism gene expression and reduces serum lipids and renal fibrogenic cytokines in rats with chronic nephrotic syndrome. *J Nutr.* 2002;132:2562-2569.
19. Frank R. Spot-Synthesis - an Easy Technique for the Positionally Addressable, Parallel Chemical Synthesis on a Membrane Support. *Tetrahedron.* 1992;48:9217-9232.
20. Frank R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports-principles and applications. *J Immunol Methods.* 2002; 267:13-26.
21. Kramer A, Schneider-Mergener J. Synthesis and screening of peptide libraries on continuous cellulose membrane supports. *Methods Mol Biol.* 1998;87:25-39.
22. Reineke U, Volkmer-Engert R, Schneider-Mergener J. Applications of peptide arrays prepared by the SPOT-technology. *Curr Opin Biotechnol.* 2001; 12:59-64.
23. Hastie T, Tibshirani R, Eisen MB, et al. 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol.* 2000;1:RESEARCH0003.
24. Chiang DY, Brown PO, Eisen MB. Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles. *Bioinformatics.* 2001;17 Suppl 1:49-55.
25. Kato R, Kaga C, Kunimatsu M, et al. Peptide array-based interaction assay of solid-bound peptides and anchorage-dependant cells and its effectiveness in cell-adhesive peptide design. *J Biosci Bioeng.* 2006;101:485-495.
26. Tomita Y, Kato R, Okochi M, et al. A Motif Detection and Classification Method for Peptide Sequences Using Genetic Programming. *J Biosci Bioeng.* 2008;106:154-161.
27. Kaga C, Okochi M, Nakanishi M, et al. Screening of a novel octamer peptide, CNSCWSKD, that induces caspase-dependent cell death. *Biochemical and Biophysical Research Communications.* 2007;362:1063-1068.
28. Noguchi H, Kato R, Hanai T, et al. Hidden Markov model-based prediction of antigenic peptides that interact with MHC class II molecules. *J Biosci Bioeng.* 2002;94:264-270.

29. Kaga C, Okochi M, Tomita Y, et al. Computationally assisted screening and design of cell-interactive peptides by a cell-based assay using peptide arrays and a fuzzy neural network algorithm. *Biotechniques*. 2008;44:393-402.
30. Kawashima S, Pokarowski P, Pokarowska M, et al. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res*. 2008;36:202-205.
31. Zimmerman JM, Eliezer N, Simha R. The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol*. 1968;21:170-201.
32. Fauchere JL, Charton M, Kier LB, et al. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int J Pept Protein Res*. 1988;32:269-278.
33. Geisow MJR, R.D.B. Amino acid preferences for secondary structure vary with protein class. *Int J Biol Macromol*. 1980;2:387-389.
34. Takano K, Yutani K. A new scale for side-chain contribution to protein stability based on the empirical stability analysis of mutant proteins. *Protein Eng*. 2001;14:525-528.
35. Zhou H, Zhou Y. Quantifying the effect of burial of amino acid residues on protein stability. *Proteins*. 2004;54:315-322.
36. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157:105-132.
37. Crawford JL, Lipscomb WN, Schellman CG. The reverse turn as a polypeptide conformation in globular proteins. *Proc Natl Acad Sci U S A*. 1973;70:538-542.
38. Munoz V, Serrano L. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. *Proteins*. 1994;20:301-311.
39. Krigbaum WR, Komoriya A. Local interactions as a structure determinant for protein molecules: II. *Biochim Biophys Acta*. 1979;576:204-248.
40. Jukes TH, Holmquist R, Moise H. Amino acid composition of proteins: Selection against the genetic code. *Science*. 1975;189:50-51.