Combinatorial Bead-Based Peptide Libraries Improved for Rapid and Robust Screenings

Joo-Eun Jee, Yi Li Ang, Junhoe Cha, Mei Wei Ang, Jingjing Ling, Jaehong Lim* and Su Seong Lee*

Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, The Nanos 138669, Singapore

Abstract: In pursuit of utilizing combinatorial peptide libraries on beads, rapid and robust screening is one of the key steps for the success of high-throughput process. We have introduced improved structural features that greatly facilitate a MALDI-MS/MS-based sequencing, associated with easy and fast synthesis and analysis of such libraries. Whilst commonly used MS-based analysis involves in sophisticated procedures such as ladder synthesis, encoding tags are not required in our MS/MS-based sequencing platform. Fragment peaks in an acquired MS/MS should be outstanding in line with correct identification of parent mass in the preceding MS. To meet these requirements a one-bead-one-compound (OBOC) peptide library was designed by placing a positively charged arginine at C-terminus. As well as enhancing the overall ionization efficiency, arginine appended in all y-ion fragments generates a series of doublet peaks under MS/MS environments, which can speed up the sequencing process in conjunction with high accuracy. It is another strong benefit that the designed library significantly suppresses the adverse formation of sodium ion adducts, which seriously jeopardizes the sequencing, especially of peptides containing negatively charged amino acids. A peptide library constructed with D-amino acids was applied to screening against a clinically significant biomarker, C-reactive protein (CRP). Through the screening of focused libraries narrowed down from a comprehensive library, several hexamer peptide ligands were successfully identified and their binding affinity and specificity towards CRP were validated by surface plasmon resonance (SPR) and dot blot experiments.

Keywords: Bead-based, combinatorial, C-reactive protein, MALDI-MS/MS, peptide library, screening, sequencing.

INTRODUCTION

With a given target biomarker such as protein, the eventual aim of screening a combinatorial peptide library is to fish out specific peptide sequences that bind strongly and selectively. It is, however, challenging to identify short peptide ligands of which affinity and specificity are high enough through one round of screening campaign using a comprehensive pentamer/hexamer peptide library. Once some initial peptide ligands are identified and validated, their binding property should be further harnessed through several methods such as *in situ* click screen [1-3]. Therefore, it is of utmost importance to identify good initial short peptide ligands in a high-throughput fashion for the development of peptide-based ligands equipped with excellent binding affinity and specificity. Robust peptide sequencing lies in the most essential process of the entire screening campaign so that the elaborated capture agents may be used in practical applications such as *in vitro* diagnostics.

Although there are several known methods for peptide sequencing that include ladder synthesis [4] and partial Edman degradation/MS sequencing [5], *de novo* peptide sequencing based on the analysis of MS/MS obtained by MALDI-TOF/TOF experiments provides a simple and efficient solution. However, this high-throughput-oriented process still requires modification in terms of raising accuracy, as well as consuming less time and labor, in order to be more practical for a wide range of applications. In MS/MS-based *de novo* sequencing, a typical strategy employed is to obtain a mass spectrum first to identify the parent mass, and then the corresponding MS/MS that deploys the peaks fragmented from the selected parent mass [6]. Sequence is subsequently elucidated by the differences in mass values of neighboring fragment peaks in the MS/MS. It is therefore obvious that efficiency and accuracy of the sequencing will depend significantly on the quality of both MS and MS/MS. However, the signal intensities of the spectra in MALDI-TOF/TOF experiments are often varying and unpredictable, which is known to come from several intrinsic properties of peptides. These properties include length [6], charge (*e.g.*, guanidinium in arginine [7, 8]), and hydrophobicity [7, 9, 10] of the residual functionality along with secondary structure of particular peptides [11].

One of the central problems associated with sequencing the peptides from an OBOC library *via* MALDI-TOF/TOF [12, 13] is that the peptide fragments containing acidic residues tend to have low signal intensities in positive ion acquisition mode of MALDI-MS experiments. The negatively-charged groups often reduce ionization efficiency resulting in low signal-to-noise ratio and increase adversely the tendencies of sodium ion adduct formation [14]. Another important issue to address is that a wide generation of a series of multiple ions disperse the ion current hence to reduce the overall signal intensity. There are a few existing methods to improve the peak intensities of MALDI-MS. Kim *et al.* reported that the peptide signals were enhanced by 20 to 35-fold upon picolinamidination [15, 16]. Pashkova *et al.* reported that the tagging of coumarin at N-terminus of peptide chain enhanced the MALDI-MS signal intensity by

*Address correspondence to these authors at the Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, The Nanos 138669, Singapore; Tel: +65-6824-7179; Fax: +65-6478-9080;
E-mails: (Jaehong Lim) jhlim@ibn.a-star.edu.sg,
(Su Seong Lee) sslee@ibn.a-star.edu.sg

1875-5402/14 $58.00+.00

up to 40-fold [17]. Introduction of coumarin at N-terminus containing a sulfonate group and guanidination of terminal amine group could also enhance MS/MS signal intensities [18]. These enhancement methods involve the modification of the N-terminus of the peptide chains. However, these methods are not desirable for ligand- or capture agent-related applications [19] using OBOC libraries because structural modifications apparently interfere with the ligand-substrate binding modes. Peptide linkers with a positively charged arginine have been utilized as an alternative method to enhance the ionization efficiency of peptides as well as to avoid the peaks from matrix clusters [20-23]. However, such types of peptide linkers were designed for MS-based sequencing of libraries on beads *via* ladder synthesis method. A new strategy is required to facilitate MS/MS-based sequencing *via* facile identification of fragmented peptide peaks in regards to enhancing the ionization efficiency by the presence of a positively charged moiety.

Herein, we report an improved strategy that facilitates the screening of bead-based libraries in conjunction with efficient sequencing of peptides based on MALDI-MS/MS experiments. Arginine is utilized as an ionization promoter at C-terminus and excluded from the diversity elements that constitute variable domain. This modification resulted in significant enhancement of signal-to-noise ratio as well as easy identification of fragmented peptide peaks due to the formation of characteristic doublet y ions generated by the loss of ammonia from the arginine residue at C-terminus [24]. Sodium ion adduct formation is also greatly suppressed by the presence of arginine. This peptide library, improved for rapid and robust screenings, was successfully applied to efficiently identifying hexamer ligands that possess high binding affinity and specificity towards a target biomarker CRP through the facilitated MS/MS-based peptide sequencing in combination with our previously reported method to differentiate isobaric amino acids.

MATERIAS AND METHODS

General

N-methylpyrrolidone (NMP), diethylether, dichloromethane (DCM) and Fmoc-γ-Abu-OH (GABA) were purchased from Merck. Fmoc-protected D-amino acids (Fmoc-D-AAs) were purchased from GL Biochem. TentaGel S amino resin was purchased from Rapp Polymere. α-Cyano-4-hydroxycinnamic acid (CHCA) was purchased from Bruker. EZ-Link NHS-Biotin reagent was purchased from Thermo Scientific. Unless otherwise specified, chemicals were purchased from Aldrich. CRP was purchased from Calbiochem. H-albumin was purchased from Fitzgerald Industries. IgG and IgA (FC fragment) as well as streptavidin-HRP were purchased from Abcam. α-fetoprotein (AFP) was purchased from YO Proteins AB. Prostate-specific antigen (PSA) was purchased from Raybiotech. Bovine serum albumin (BSA) and bovine carbonic anhydrase II (bCA II) were purchase from Sigma-Aldrich. MALDI-MS and MS/MS experiments were performed using ultrafleXtreme™ TOF/TOF (Bruker). Microwave-assisted CNBr-mediated cleavage reactions were performed by a household microwave oven (model R-248J, 800 W, 2450 MHz) from Sharp [24]. The PEAKS software was purchased from Bioinformatics Solutions. The purification of bulk peptides was done by a preparative HPLC system from Gilson on a C_{18} reversed phase preparative column (Kromasil® from AkzoNobel, 5 μm, 250 × 30 mm).

Construction of Peptide Libraries

Comprehensive bead-based peptide libraries were synthesized using our reported method [2, 19]. For the solutions of Fmoc-isoleucine and Fmoc-glutamine, it contained 10 mol% of glycine each for discrimination of isobaric residues in the semi-automatic peptide sequencing [2, 19].

Labeling of CRP and Library Screening

For screening, CRP was first labeled using the Alexa Fluor® 647 protein labeling kit (A20173, Invitrogen) according to the supplier's protocol. 500 μl of 2 mg/mL solution of CRP was prepared in PBS and 50 μl of 1 M solution of sodium bicarbonate was added to yield a solution of pH 8.3. Then the CRP solution was transferred into a vial with the NHS-activated dye. The vial was capped and inverted a few times to fully dissolve the dye. The reaction mixture was stirred for 1 h at room temperature under dark conditions. The Alexa Fluor® 647-labeled CRP was purified by using size exclusion chromatography. Upon purification the resulting dye-labeled CRP was characterized by UV-vis spectroscopy and SDS-PAGE.

For the screen, 100 mg of library resin was transferred into an Alltech tube (8 mL, equipped with a filer) and pre-incubated in a blocking solution, comprising of 0.05% NaN_3, 0.1% Tween 20, and 0.1% BSA in PBS buffer (pH 7.4), for 1 h on a 360° shaker at 25 °C. The buffer solution was drained, and then 5 mL of 10 nM dye-labeled CRP diluted in the blocking solution was added to the swollen resin. The resulting mixture was incubated for 4 h on a 360° shaker at 25 °C. The liquid was drained and the resulting beads were washed three times with the blocking solution and three times with 0.1% Tween 20 in PBS buffer sequentially. After washing, the beads were transferred into a sample vessel of COPAS Plus (Union Biometrica) and diluted with 200 mL of PBS buffer with 0.1% Tween 20 (pH 7.4). Two-step sorting was applied to sort out positive beads [2, 19]. In the second sort, positive beads were directly sorted into a 96 titer well plate with cone-shaped wells.

CNBr-Based Cleavage of Peptides from Single Beads

A 96-well plate with sorted beads was purged by nitrogen for 15 min and then CNBr (10 μl, 0.50 M in 0.2 N HCl solution) was added into each well. After additional purging by argon/nitrogen for 15 min, the plate was sealed by a thin film and placed under microwave radiation for 1 min. The resulting solution was concentrated under centrifugal vacuum for 10 min at 45 °C and then for 50 min at 60 °C [25].

MALDI-MS and -MS/MS Analysis of Peptides Cleaved from Single Beads

To each well were added CHCA (7 μl, 0.4% solution in acetonitrile/water (70:30)) and then acetonitrile/water (7 μl, 70:30 containing 0.1% trifluoroacetic acid (v/v)). 2.5 μl of

the mixture solution was taken up to be spotted onto a 384-well MALDI plate, which was allowed to stand for 15 min to dry naturally. MS and MS/MS acquisition was conducted with ultrafleXtreme™ MALDI-TOF/TOF mass spectrometer from Bruker Daltonics.

Synthesis of Peptides for Affinity Measurements

Peptides of interest were synthesized on Rink amide resins (0.63 mmol/g) in a typical resin scale of 10 mg per sequence. With the desired sequence of peptide attained, the resin was treated in trifluoroacetic acid (95%), water (2.5%), and triisopropylsilane (2.5%) for 2 h. The cleavage cocktail was concentrated in a continuous flow of nitrogen and the crude peptides were precipitated in diethyl ether. The resulting white solid was then purified to >95% in purity by HPLC bearing a C_{18} reversed-phase preparative column. The purified peptides were used for affinity measurements *via* surface plasmon resonance (SPR).

Synthesis of Biotinylated Peptides for Dot Blot Experiments

Biotinylated peptides of interest were synthesized on Rink amide resins (0.63 mmol/g) in a typical resin scale of 10 mg per sequence. First, Fmoc-Lys(Mtt)-OH was coupled to the resin. Then, selective deprotection of 4-methyltrityl (Mtt) group was performed by reaction with TFA/TIS/DCM (3/3/94) for 2 min, 5 min and 30 min successively, using a fresh aliquot each time. At the following step, biotin-NHS in NMP with DIEA was added and the mixture was vortexed for 30 min. With all amino acids coupled, the resin was treated in trifluoroacetic acid (95%), water (2.5%) and triisopropylsilane (2.5%) for 2 h. The cleavage cocktail was concentrated in a continuous flow of nitrogen, and the crude peptides were precipitated in diethyl ether. The resulting white solid was then purified to >95% in purity by HPLC bearing a C_{18} reversed-phase preparative column.

Affinity Measurements

Affinity measurements were performed on a CM5 chip using Biacore T100 system (GE Healthcare). The instrument was primed with HBS-EP+ (GE Healthcare) buffer. Flow channel 1 (or 3) was used as a reference to subtract nonspecific binding, drift, and the bulk refractive index, while flow channel 2 (or 4) was immobilized with CRP following standard procedures. A 1:1 mixture of 0.4 M EDC and 0.1 M NHS was used to activate flow cell 2 (or 4), and 0.04 mg/mL CRP solution was injected. Deactivation of the remaining activated groups was done with a 1 M solution of ethanolamine (pH 8.5). CRP was immobilized onto the sensor chip surface by approximately 2000 response units (RU). The instrument was then primed using running buffer (HBS-EP+). Each of the hexamer peptide ligand candidates was dissolved in HBS-EP+ buffer to produce a 10 μM stock solution, which was serially diluted by a factor of 2 to produce a concentration series down to 313 nM. For a given affinity measurement, these series of peptide solutions were successively injected into the flow channel 2 (or 4) for 3 min of contact time, 5 min of dissociation time, and 3.5 min of stabilization time using a flow rate of 100 μl/min at 25 °C. Flow cell 2 (or 4) was regenerated by glycine 2.5 (GE Healthcare) after injection of each peptide solution.

Dot Blot Experiments

The affinity of the capture agents for CRP was demonstrated through dot blot experiments in 5% nonfat dry milk in TBS-T [25 mM Tris, 150 mM NaCl, 2 mM KCl, 0.1% Tween 20 (pH 7.0)]. A CRP stock solution (1 mg/mL) was prepared in PBS buffer (pH 7.4). A dilution series of CRP solution was applied to a nitrocellulose membrane, typically ranging from 1 μg to 20 ng per spot. The membrane was blocked at room temperature for 1 h in 5% nonfat milk/TBS-T, which was then washed three times with TBS-T for 10 min each. A biotinylated hexamer peptide solution was prepared at 100 nM in 0.5% nonfat milk/TBS-T and incubated over the membrane for 1 h at room temperature. After washing three times with TBS-T for 10 min, the membrane was incubated with 1:3000 streptavidin-HRP (Abcam) in 0.5% milk/TBS-T for 1 h followed by subsequent washing with TBS-T for 10 min. The resulting membrane was detected using the chemiluminescent reagents (Amersham ECL plus Western blotting detection reagents, GE Healthcare). The dot blot experiments for specificity test were conducted by the identical procedure using in-house serum proteins.

RESULTS AND DISCUSSION

Screening of a Typical Comprehensive Hexamer Peptide Library Against CRP

With an aim to identify hexamer peptide ligands that possess relatively high binding affinity to CRP as demonstrated in our previous work [2], a typical comprehensive hexamer peptide library was constructed with 18 D-amino acids as diversity elements, excluding methionine and cysteine, and screened against a target biomarker CRP conjugated with Alexa Fluor® 647. CRP is a plasma protein whose production in the mammalian biological system is rapidly increased following states of inflammation. The protein's main physiological function is to bind to damaged or apoptotic cells and this bound state activates the complementary pathway *via* interactions with C1q [26]. It is also associated with the prediction of cardiovascular disease [27, 28].

Initially, a small portion of the library beads were incubated in the presence of the dye-labeled CRP in a PBS buffer solution for 4 hours. The screening was repeated 4 times in order to acquire enough number of positive peptides for the generation of focused libraries. However, peptide sequences obtained from the collected beads with high fluorescence did not show good homology. Positively charged amino acids, *i.e.* arginine and lysine, were exclusively dominant presumably due to the nonspecific electrostatic interactions (see SI, S-Table **1**). In particular, arginine was the most dominant diversity element in the collected peptides (see SI, S-Fig. **1**). A number of previous studies [29-35] reported on strong nonspecific interactions between peptides and target macromolecules. In the work of Chen *et al.* on profiling the sequences specific to Src homology 2 (SH2) [30-32] and PDZ domains [33], protein domains tend to pick out many positive sequences that are rich in positively charged amino acids [36]. Further studies proved that the positively charged residues contributed minimally to overall binding. This phenomenon is largely

attributed to nonspecific interactions between the two objects. Although the commercially available Alexa Fluor® 647 has been used as a fluorescence probe due to its good compatibility with the excitation and emission wavelengths of the bead sorter (COPAS™), in conjunction with its near infrared (NIR) excitation wavelength that can avoid auto-fluorescence issues from TentaGel beads [2, 19], it bears multiple sulfate groups per molecule [37] that often exhibit strong tendency to interact with the peptides involving positively charged amino acids.

It should be also considered that the inherent properties of the general hexamer peptide library make it difficult to obtain meaningful positive sequences with good homology. As one drawback the hexamer peptide library is not practical in terms of screening with the entire diversity. It requires a great amount of beads, *i.e.*, 12 gram of TentaGel resin of 34 million beads, to represent all possible sequences. Meanwhile, only up to 100 mg of the library beads can be screened for one round of sorting by the bead sorter. After all, the approach with a comprehensive hexamer library requires a tremendous amount of time and labor to collect statistically meaningful data.

A breakthrough to circumvent these issues can be found in a library of a smaller size, with the identical structural features embedded between the peptides and the cleavable linker. Trimming down the number of diversity points from six to five, a pentamer library can express all possible sequences (*ca.* 1.4M) with less than 1 gram of TentaGel beads (90 μm in diameter). It allows for screening of more than 10 per cent of the entire pentamer library in a single run although the shorter peptide compromises the level of binding affinity. Besides, the exclusion of arginine in the library should be still beneficial for minimization of the electrostatic nonspecific interactions between the negatively charged dyes on the target proteins and the peptides on the beads. The "shorter" pentamer library would satisfy all these requirements.

Another drawback with the typical library structure is that it requires a tremendous amount of time and labor to collect all correct sequences from sequencing the positive peptides due to common problems associated with *de novo* sequencing [38]. Although an efficient semi-automatic peptide sequencing algorithm was developed in house [2], careful manual efforts were still required to achieve more reliable sequencing of positive peptides with enriched experience in sequencing. In particular, peptides with negatively charged residues such as glutamic acid and aspartic acid easily form sodium ion adducts in MALDI-MS resulting in difficult identification of parent mass and challenging sequencing *via* analysis of MS/MS spectra. Typically, it takes several hours to complete the sequencing of approximately 50 hexamer peptides by the semi-automatic peptide sequencing method. Besides, some parent masses from negatively charged peptides could be easily neglected in the form of sodium ion adducts.

A Comprehensive Pentamer Peptide Library Designed for Highly Efficient MS/MS-Based Sequencing

The most efficient way to solve these above issues is to devise a library of a smaller size with a facile sequencing strategy. A pentamer peptide library was designed and constructed based upon the rationale to facilitate MS/MS-based sequencing of peptides sorted out from the screenings of combinatorial peptide libraries against a target biomarker. We adopted a method from literature to enhance the ionization efficiency of peptides, which is the incorporation of positively charged functionality at C-terminus of peptides along with a spacer [20-22]. Arginine that contains guanidinium group in the residue was selected as the mass enhancer in association with easy introduction by a routine peptide coupling. Typically, arginine tends to lose ammonia (17 amu) in all fragments under MS/MS environments exhibiting a characteristic series of doublet peaks of y-ions, which greatly facilitates visual identification for sequencing. However, it is not obvious to sequence the peptides that contain more than one arginine. The MS/MS will turn out complicated due to the multiple losses of ammonia molecules in some/all fragments. One good way to avoid such a problem is to allow only one arginine in each peptide for both enhanced ionization and facile sequencing. With arginine placed at C-terminus as a mass enhancer, it needs to be opted out from the diversity region to maintain only one arginine per peptide. The exclusion of positively-charged arginine from the diversity elements will be also advantageous in a way to reduce nonspecific electrostatic interactions, which often come from the unwanted interactions between positively-charged groups in peptides and negatively-charged groups in commonly used dye molecules conjugated to target biomarkers as shown in the above screening against CRP. Amine group of lysine and imidazole group of histidine are not as serious as guanidine of arginine in terms of nonspecific interactions. Our designed library devoid of arginine in the diversity region can minimize the false positives in bead-based screenings due to reduced occurrence of nonspecific electrostatic interactions. To compensate for the excluded arginine in the comprehensive library, it can be included back in the subsequent focused libraries of significantly smaller diversity that enable easier peptide sequencing. Another method is to investigate extended positive sequences by replacing lysine(s) in the fully validated peptides with arginine(s).

With arginine introduced as an ionization promoter and excluded in the variable region of the library, a spacer should be placed to segregate the two regions to keep arginine orthogonal in the screening. A spacer also increases the overall mass of the cleaved peptide, which shifts its parent mass peak out of the matrix cluster peak-dominant mass range [20-22]. The structure illustrated in Fig. (**1**) was elaborated through a rational design in accordance with the above-mentioned factors. Two repetitive units of γ-aminobutyric acid (GABA) were inserted as a spacer. Additionally, N-terminus of the peptide chain was modified by acetyl (Ac) group to further minimize nonspecific interactions as well as to improve the stability of peptides. The addition of the GABA spacer and the arginine ionization promoter located the parent mass off the dominant peaks from the MALDI matrix clusters, α-cyano-4-hydroxy-cinnamic acid (CHCA). If necessary, ammonium phosphate solution can be treated to the MALDI-MS samples to suppress the matrix cluster peaks [2]. In addition, GABA moieties displayed distinct fragmentation properties under

MALDI-TOF/TOF environments, which helped the parent peptides effectively fragmented. With these factors combined, mass spectra could be obtained with excellent signal-to-noise ratio to easily identify parent masses. Besides signal enhancement, the suppression of sodium ion adduct formation was another great benefit.

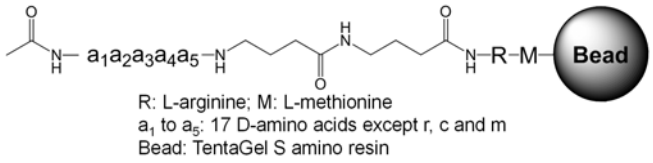


Fig. (1). Structure of a comprehensive pentamer library designed for improved sequencing.

Fig. (**2a**) illustrates an example with a peptide Ac-nfwdp (small letters indicate D-amino acids, cleaved from a single 90 μm TentaGel bead) without arginine between the spacer (two GABAs) and the cleavable linker (methionine). The matrix clusters appear dominant due to the low intensity of parent peptide in the MS spectrum. While the actual parent mass of 973 amu was not observed, only the sodium ion adduct of 995 amu was observed. The desired peak was only negligible in height presumably due to the inherent tendency of the negatively charged peptides towards weak ionization, as well as favorable formation of sodium ion adduct. With the same sequence appending arginine as well as the two-GABA spacer, Fig. (**2b**) shows one prominent peak at 1129 amu as the correct parent mass. The intensity of the parent peak in Fig. (**2b**) is approximately 10-fold higher than that in Fig. (**2a**) under the identical acquisition conditions despite containing an acidic residue. In addition, the presence of arginine effectively suppressed the formation of the sodium ion adduct. This example demonstrates that a positively charged moiety in a peptide can greatly enhance the ionization intensity as well as suppress the sodium ion adduct formation, which allows for easy identification of correct parent mass. Signal-to-noise ratio could be easily increased by 5 to 10-fold through the incorporation of arginine for most peptides including acidic residues (see SI, S-Fig. **2**). In MALDI-MS experiments, the library of a rational design turned out highly efficient to allow the parent mass easily identified by increasing its value beyond the matrix cluster region in conjunction with a significant enhancement of signal-to-noise ratio (Fig. **3**).

The incorporation of arginine next to CNBr-cleavable methionine located at C-terminus was also beneficial to enhance the ionization of fragmented peptides in MS/MS. Certain peptides that include negatively charged residues are still weakly ionized under MALDI-MS/MS conditions to correctly identify all the fragments, especially in the use of a general peptide library of vast diversity [2]. Fig. (**4**) shows an example for sequencing of a peptide (Ac-flkwf) that

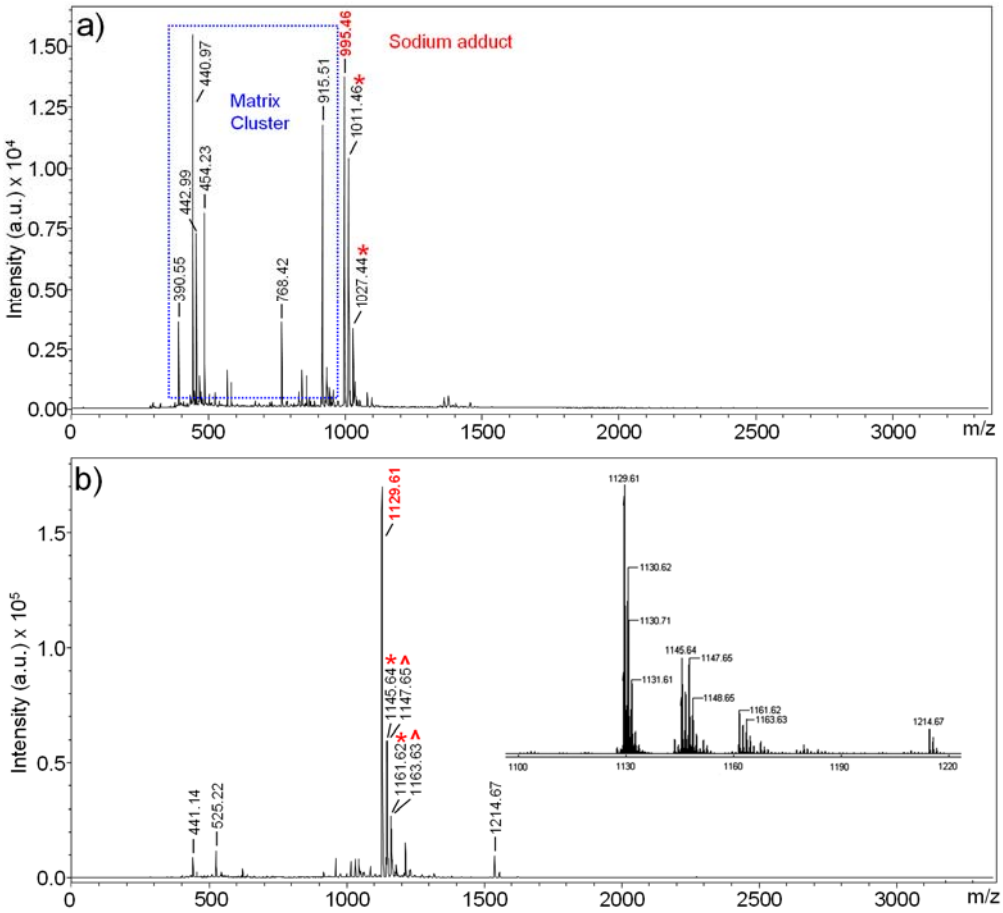


Fig. (2). Mass spectra of **a**) Ac-nfwdp-$(GABA)_2$-M* (MW = 973 amu) and **b**) Ac-nfwdp-$(GABA)_2$-R-M* (MW = 1129); M* = homoserine lactone.

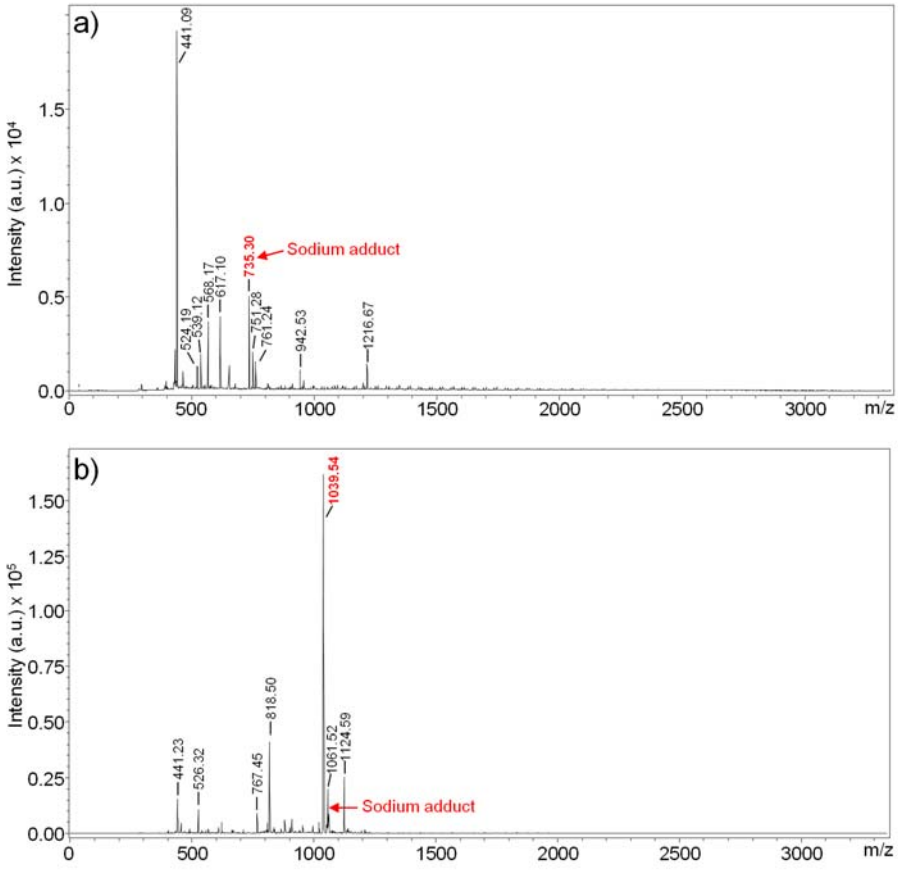


Fig. (3). Mass spectra of **a**) Ac-ddple-M* (MW = 735 amu) and **b**) Ac-ddple-$(GABA)_2$-R-M* (MW = 1039); M* = homoserine lactone.

embeds the beneficial features. The peptide was sequenced by easily identifying the characteristic series of doublets, starting from the fragment of $(GABA)_2$-R-homoserine lactone (427 amu). Sequencing of negatively charged peptides was greatly facilitated by reducing the formation of sodium ion adducts. Fig. (**5a**) shows the MS/MS of a peptide Ac-fqdye that appends arginine at C-terminus, while Fig. (**5b**) comes from the sodium ion adduct of the identical sequence that only lacks arginine at C-terminus. Without arginine as a mass enhancer, the actual parent peak $[M+H]^+$ was not at all observed in the MS (see SI, S-Fig. **3**). In the presence of arginine, the S/N ratios of MS and MS/MS peaks were typically increased up to 10-fold. Otherwise, only the sodium ion adducts could be detected in MS, which is far from being successful for sequencing in MS/MS. Another finding to highlight is that MS/MS of the sodium ion adducts did not clearly generate the monomeric immonium ions. As previously reported in our semi-automatic sequencing method [2], the immonium ions play a beneficial part to easily identify the types of amino acids that constitute the peptide to sequence. Computer programs such as PEAKS® (Bioinformatics Solutions) and flexAnalysis (Bruker) are helpful to visually identify all the y-ions and mass differences between two neighboring pairs of doublets.

To further prove that the designed library is highly efficient for rapid and reliable sequencing, several peptides were synthesized on beads with or without arginine as ionization promoter. In general, the quality of the MS and MS/MS obtained from the peptides without arginine was considerably low under the identical data acquisition conditions, whilst the peptides with arginine at C-terminus could be sequenced easily and accurately with low consumption of time and labor (see SI, S-Figs. **2**, **4**). The quality of spectra is good enough for successful sequencing of most types of peptides, even including some acidic residues which are typically problematic. Based on these results, the designed library turned out highly efficient for rapid and reliable MS/MS-based peptide sequencing, which can be utilized subsequently in high-throughput screening against target biomarkers.

For more facile isobaric differentiation, our previous method [2] was slightly modified by replacing alanine with glycine as an additive (0.1 eq.) in the coupling of isoleucine, which resulted in reduced contamination in the MS/MS from the fragmented peaks due to the larger mass difference between glycine and isoleucine (56 amu) than that between alanine and isoleucine (42 amu) [19].

Evaluation of the Designed Library in Screening Against CRP

To evaluate the designed pentamer library, screenings were carried out against Alexa Fluor® 647-labeled CRP. Reminded that we aimed to identify hexamer peptide ligands *via* two generations of screens, two or three rounds of screening exercises were sequentially performed first with a comprehensive pentamer library and then focused libraries constructed based on the first screening results. In the first

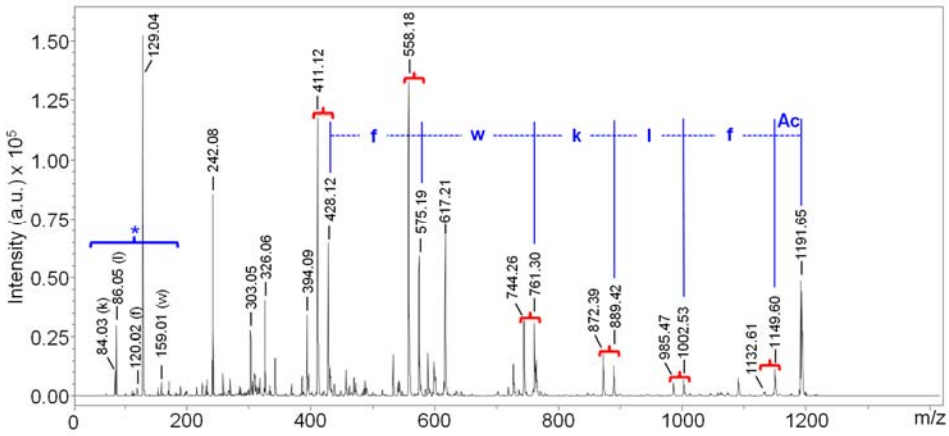


Fig. (4). Peptide sequencing *via* MS/MS of Ac-flkwf-$(GABA)_2$-R-M* by identifying doublet y ions; M* = homoserine lactone.

round screening the library beads of approximately 800 mg was synthesized to represent one copy per each sequence. About half of the library was screened to give 170 positive peptides in 3 parallel runs (see SI, S-Table **2**). The peptides cleaved from the beads were successfully analyzed by the optimized sequencing protocol. As described previously, simple visual identification of the prominent series of doublets y-ions allowed for efficient sequencing, whereas some ambiguous peptides should be confirmed by identifying the monomeric immonium ions for reliable sequencing [2]. In most cases, the identification of doublet y-ions was enough for high-throughput sequencing. The peptides from all of the collected beads could be successfully sequenced with a throughput of 50 peptides per hour.

In the initial screening of a comprehensive pentamer library, the sequencing results showed much greater

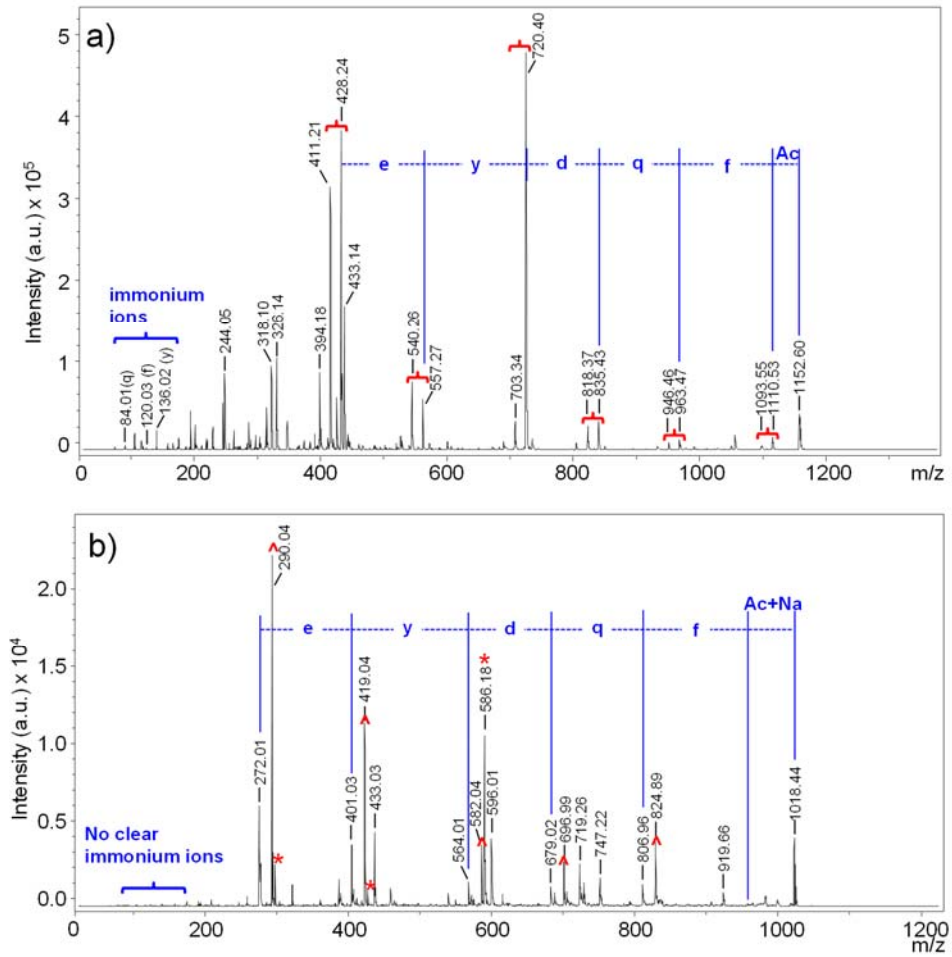
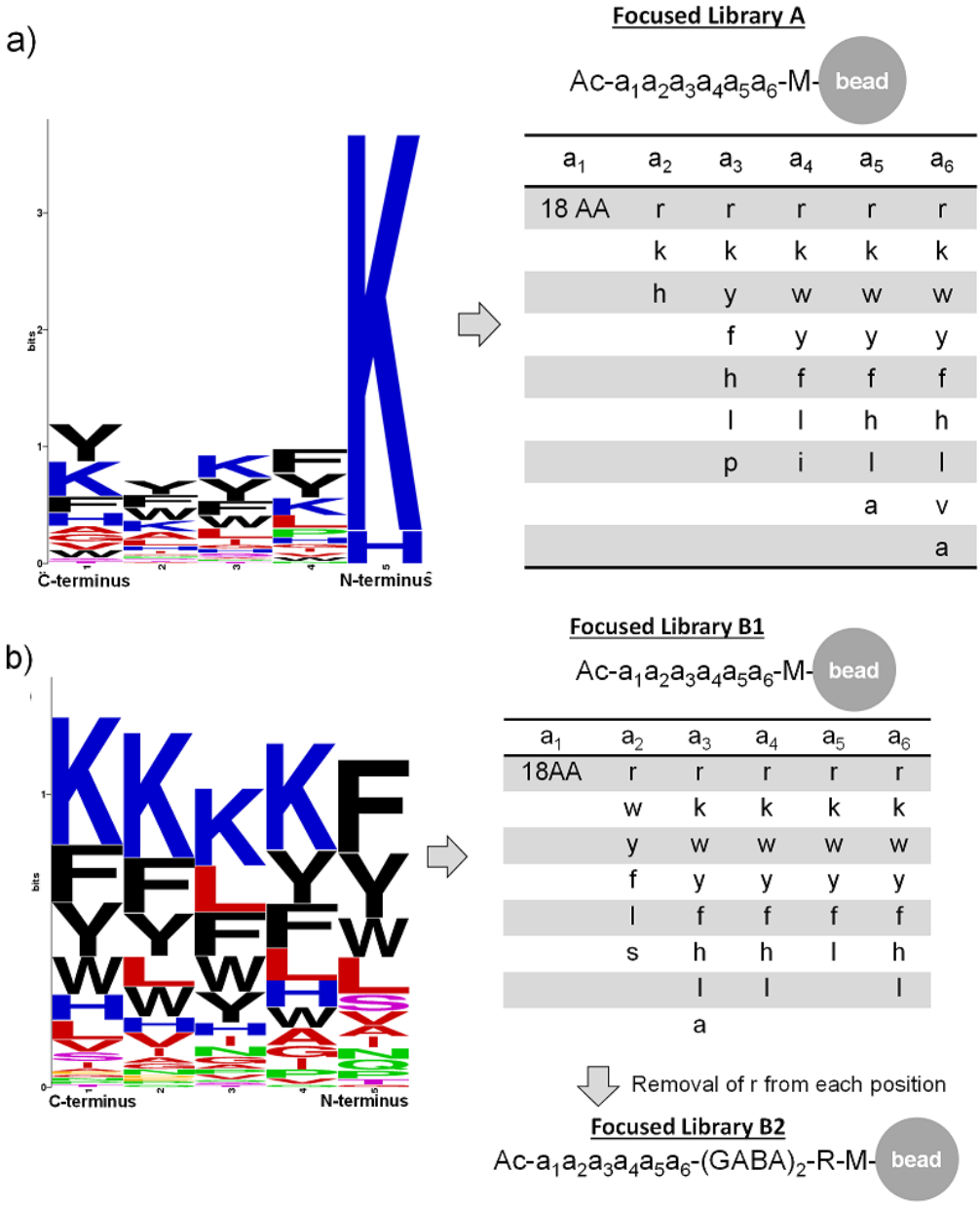


Fig. (5). MS/MS of **a**) Ac-fqdye-$(GABA)_2$-R-M* and **b**) sodium ion adduct of Ac-fqdye-$(GABA)_2$-M*; M* = homoserine lactone.

homology with several apparent sequence motifs such as wky, ykk, kfk, and kfy (see SI, S-Table **2**). For the subsequent construction of focused libraries, the positive sequences were divided into two groups based upon the nature of amino acid at N-terminus according to the hydrophilic/hydrophobic properties. Two font-histograms from the two separated groups of positive peptides were generated to figure out what amino acids were dominant at each position (Scheme **1**). Based on the occurrence of amino acids in the screening of the initial comprehensive library, two focused libraries A and B1 were proposed from each of the two groups, respectively. For enhanced binding affinity, hexamer focused libraries were constructed by adding one more diversity point at N-terminus by using the typical 18 D-amino acids. Particularly, arginine was incorporated as a diversity element to whichever positions containing k as one of dominant amino acids for the construction of the two focused libraries. 400 mg of each focused library bearing at least two sequence copies were screened. An additional hexamer focused library B2, also of the beneficial features, was generated by completely excluding arginine from the focused library B1. Details of the three focused libraries are summarized in Scheme **1**. After conducting two rounds of screening with each focused library (See SI, S-Table **3**), 11 Peptide candidates with repeatedly occurring sequence motif were short-listed for validation by surface plasmon resonance (SPR) and dot blot experiments. The significantly smaller diversity in the two focused libraries (A and B1) enabled easier peptide sequencing even though they comprise of a conventional library structure. Apparently, the focused library B2 with the designed library structure led to considerably rapid and robust peptide sequencing.



Scheme 1. Generation of focused libraries from the positive sequences containing **a**) hydrophilic amino acids (k, h) and **b**) amino acids other than k or h at N-terminus.

From SPR and dot blot results (see SI, Fig. **5**), all the 11 hexamer peptide candidates showed reasonable degree of binding to CRP. The K_D values of the peptides fall in the micro-molar to tens of micro-molar ranges determined by 1:1 binding model fitting using Biacore T100 evaluation software. Therefore, our screening strategy with a short peptide library of a novel design was successfully demonstrated for actual screening campaign against a target protein in a high-throughput fashion.

Three CRP binding candidates from the screening of the focused library B1 contain 1 or 2 lysines in each sequence. There is a possibility that lysine can be replaced with arginine to exhibit even higher binding affinity because lysine is the closest entity to arginine in terms of residual property. This speculation was validated by simple dot blot experiments, which was conducted by replacing one or both lysines in the sequence with arginines for the three candidates. The dot blot results explain that the effects of the arginine substitution on binding are ligand-dependent, as the substitution leads to either an increase, decrease or have little effects on binding (See SI, S-Fig. **6**). These results show that the arginine can be excluded to generate both comprehensive and focused libraries to facilitate peptide sequencing to stay farther from nonspecific bindings. Instead, arginine can be placed back and tested in the short-listed peptide ligand candidates at the later validation or optimization stage. Among these candidates, two representative peptides (lyfrrw and lrfrwf) showed the good responses in SPR and dot blot experiments (Fig. **6**). Especially, these two peptides displayed reasonable degree of specificity when dot blot experiments were performed with in-house serum proteins (Fig. **6b**). These initial short peptides ligands with certain degree of binding affinity and specificity can be utilized as anchor ligands for further enhancement of their binding properties by *in situ* screen or elongation [1-3].

CONCLUSION

Combinatorial peptide libraries improved for rapid and robust screenings in association with efficient MALDI-MS/MS-based peptide sequencing were investigated. The structural features include i) placement of an arginine at C-terminus as ionization promoter, ii) insertion of a spacer next

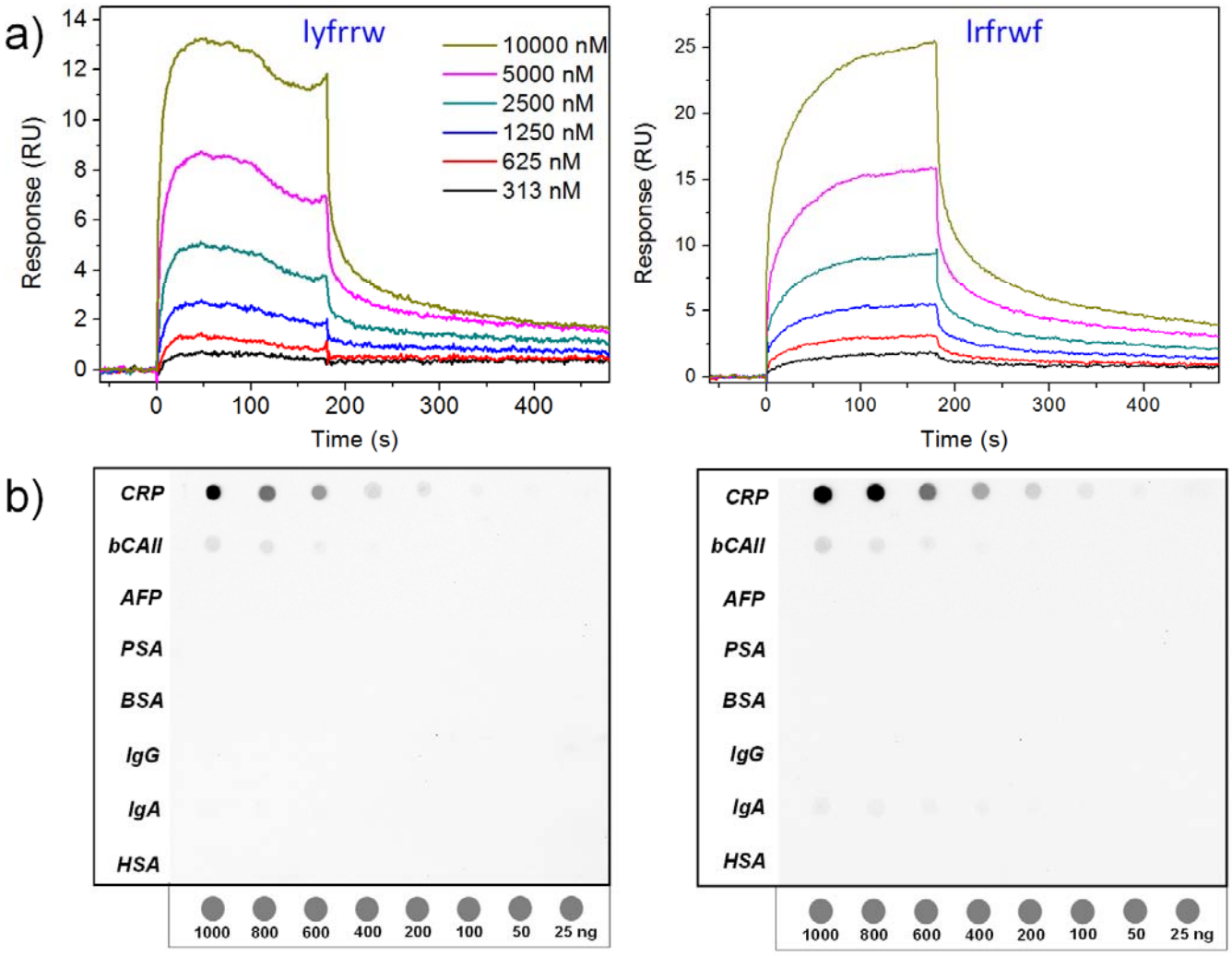


Fig. (6). a) SPR sensorgrams (immobilization level of CRP = 2,000 RU) and (**b**) dot blot results ([peptide] = 100 nM) of lyfrrw and lrfrwf, respectively.

to it, and iii) exclusion of arginine in variable peptide region of the library. Arginine as the ionization promoter also suppressed the formation of sodium ion adducts, which would jeopardize the correct pick up of the true parent mass. In addition, two repeated units of GABA as the spacer facilitated the fragmentation of peptides. The success rate of the sequencing reached near perfect in the end. Visual identification of y-ions in MS/MS was greatly enhanced due to the appearance of doublets caused by the characteristic loss of ammonia from guanidinium group of arginine as the ionization promoter, which allowed for significantly facilitated peptide sequencing. More than 50 penta- or hexamer peptides could be sequenced within one hour making our approach much more practical. It was successfully demonstrated that the efficient MS/MS-based peptide sequencing through the improved libraries was applied to a high-throughput screening campaign against CRP as a target biomarker in conjunction with our reported strategy to differentiate isobaric residues. A group of hexamer peptide ligands were successfully identified displaying high binding affinity and specificity. These high-affinity short peptide ligands can be used as initial anchor motif to further enhance the binding capabilities by adopting some existing methods such as *in situ* screen or elongation that can be performed through the MS/MS-based peptide sequencing of penta- or hexamer peptides.

CONFLICT OF INTEREST

The authors confirm that they do not have any conflicts of interest.

ACKNOWLEDGEMENTS

This work was supported by the Institute of Bioengineering and Nanotechnology (Biomedical Research Council (BMRC), Agency for Science, Technology and Research (A*STAR), Singapore), and BMRC-Science and Engineering Research Council Diagnostics Grant (A*STAR).

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Agnew, H. D.; Rohde, R. D.; Millward, S. W.; Nag, A.; Yeo, W.-S.; Hein, J. E.; Pitram, S. M.; Tariq, A. A.; Burns, V. M.; Krom, R. J.; Fokin, V. V.; Sharpless, K. B.; Heath, J. R. Iterative in situ click chemistry creates antibody-like protein-capture agents. *Angew. Chem., Int. Ed.*, **2009**, *48*, 4944-4948.

[2] Lee, S. S.; Lim, J.; Tan, S.; Cha, J.; Yeo, S. Y.; Agnew, H. D.; Heath, J. R. Accurate MALDI-TOF/TOF sequencing of one-bead-one-compound peptide libraries with application to the identification of multiligand protein affinity agents using in situ click chemistry screening. *Anal. Chem.*, **2010**, *82*, 672-679.

[3] Millward, S. W.; Henning, R. K.; Kwong, G. A.; Pitram, S.; Agnew, H. D.; Deyle, K. M.; Nag, A.; Hein, J.; Lee, S. S.; Lim, J.; Pfeilsticker, J. A.; Sharpless, K. B.; Heath, J. R. Iterative in situ click chemistry assembles a branched capture agent and allosteric inhibitor for Akt1. *J. Am. Chem. Soc.*, **2011**, *133*, 18280-18288.

[4] Wang, X.; Peng, L.; Liu, R.; Gill, S. S.; Lam, K. S. Partial alloc-deprotection approach for ladder synthesis of "one-bead one-compound" combinatorial libraries. *J. Comb. Chem.*, **2005**, *7*, 197-209.

[5] Thakkar, A.; Wavreille, A.-S.; Pei, D. Traceless capping agent for peptide sequencing by partial Edman degradation and mass spectrometry. *Anal. Chem.*, **2006**, *78*, 5935-5939.

[6] Yergey, A. L.; Coorssen, J. R.; Backlund, P. S.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L. De novo sequencing of peptides using MALDI/TOF-TOF. *J. Am. Soc. Mass. Spectrom.*, **2002**, *13*, 784-791.

[7] Krause, E.; Wenschuh, H.; Jungblut, P. R. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal. Chem.*, **1999**, *71*, 4160-4165.

[8] Valero, M. L.; Giralt, E.; Andreu, D. An investigation of residue-specific contributions to peptide desorption in MALDI-TOF mass spectrometry. *Lett. Pept. Sci.*, **1999**, *6*, 109-115.

[9] Cohen, S. L.; Chait, B. T. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.*, **1996**, *68*, 31-37.

[10] Olumee, Z.; Sadeghi, M.; Tang, X.; Vertes, A. Amino acid composition and wavelength effects in matrix-assisted laser desorption/ionization. *Rapid Commun. Mass Spectrom.*, **1995**, *9*, 744-752.

[11] Wenschuh, H.; Halada, P.; Lamer, S.; Jungblut, P.; Krause, E. The ease of peptide detection by matrix-assisted laser desorption/ionization mass spectrometry: the effect of secondary structure on signal intensity. *Rapid Commun. Mass Spectrom.*, **1998**, *12*, 115-119.

[12] Lam, K. S.; Lebl, M.; Krchnak, V. The "one-bead-one-compound" combinatorial library method. *Chem. Rev.*, **1997**, *97*, 411-448.

[13] Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature*, **1991**, *354*, 82-84.

[14] Zhou, W.; Merrick, B. A.; Khaledi, M. G.; Tomer, K. B. Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beads by matrix-assisted laser desorption/ionization mass spectrometry. *J. Am. Soc. Mass. Spectrom.*, **2000**, *11*, 273-282.

[15] Kim, J.-S.; Kim, J.-H.; Kim, H.-J. Matrix-assisted laser desorption/ionization signal enhancement of peptides by picolinamidination of amino groups. *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 495-502.

[16] Kim, J.-S.; Cui, E.; Kim, H.-J. Picolinamidination of phosphopeptides for MALDI-TOF-TOF mass spectrometric sequencing with enhanced sensitivity. *J. Am. Soc. Mass. Spectrom.*, **2009**, *20*, 1751-1758.

[17] Pashkova, A.; Moskovets, E.; Karger, B. L. Coumarin tags for improved analysis of peptides by MALDI-TOF MS and MS/MS. 1. Enhancement in MALDI MS signal intensities. *Anal. Chem.*, **2004**, *76*, 4550-4557.

[18] Pashkova, A.; Chen, H.-S.; Rejtar, T.; Zang, X.; Giese, R.; Andreev, V.; Moskovets, E.; Karger, B. L. Coumarin tags for analysis of peptides by MALDI-TOF MS and MS/MS. 2. Alexa Fluor 350 tag for increased peptide and protein identification by LC-MALDI-TOF/TOF MS. *Anal. Chem.*, **2005**, *77*, 2085-2096.

[19] Cha, J.; Lim, J.; Zheng, Y.; Tan, S.; Ang, Y. L.; Oon, J.; Ang, M. W.; Ling, J.; Bode, M.; Lee, S. S. Process automation toward ultra-high-throughput screening of combinatorial one-bead-one-compound (OBOC) peptide libraries. *J. Lab. Autom.*, **2012**, *17*, 186-200.

[20] Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. Generation and screening of combinatorial peptide libraries designed for rapid sequencing by mass spectrometry. *J. Am. Chem. Soc.*, **1995**, *117*, 3900-3906.

[21] Tornøe, C. W.; Sanderson, S. J.; Mottram, J. C.; Coombs, G. H.; Meldal, M. Combinatorial library of peptidotriazoles: identification of [1,2,3]-triazole inhibitors against a recombinant Leishmania mexicana cysteine protease. *J. Comb. Chem.*, **2004**, *6*, 312-324.

[22] Hu, B.-H.; Jones, M. R.; Messersmith, P. B. Method for screening and MALDI-TOF MS sequencing of encoded combinatorial libraries. *Anal. Chem.*, **2007**, *79*, 7275-7285.

[23] Garske, A. L.; Oliver, S. S.; Wagner, E. K.; Musselman, C. A.; LeRoy, G.; Garcia, B. A.; Kutateladze, T. G.; Denu, J. M. Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nat. Chem. Biol.*, **2010**, *6*, 283-290.

[24] Csonka, I. P.; Paizs, B.; Suhai, S. Modeling of the gas-phase ion chemistry of protonated arginine. *J. Mass Spectrom.*, **2004**, *39*, 1025-1035.

[25] Lee, S. S.; Lim, J.; Cha, J.; Tan, S.; Heath, J. R. Rapid microwave-assisted CNBr cleavage of bead-bound peptides. *J. Comb. Chem.*, **2008**, *10*, 807-809.

[26] Black, S.; Kushner, I.; Samols, D. C-reactive protein. *J. Biol. Chem.*, **2004**, *279*, 48487-48490.

[27] Ridker, P. M.; Hennekens, C. H.; Buring, J. E.; Rifai, N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *New Engl. J. Med.*, **2000**, *342*, 836-843.

[28] Ridker, P. M. High-sensitivity C-reactive protein. *Circulation*, **2001**, *103*, 1813-1818.

[29] Alluri, P. G.; Reddy, M. M.; Bachhawat-Sikder, K.; Olivos, H. J.; Kodadek, T. Isolation of protein ligands from large peptoid libraries. *J. Am. Chem. Soc.*, **2003**, *125*, 13995-14004.

[30] Beebe, K. D.; Wang, P.; Arabaci, G.; Pei, D. Determination of the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library. *Biochemistry*, **2000**, *39*, 13251-13260.

[31] Wavreille, A. S.; Garaud, M.; Zhang, Y.; Pei, D. Defining SH2 domain and PTP specificity by screening combinatorial peptide libraries. *Methods*, **2007**, *42*, 207-219.

[32] Imhof, D.; Wavreille, A. S.; May, A.; Zacharias, M.; Tridandapani, S.; Pei, D. Sequence specificity of SHP-1 and SHP-2 Src homology 2 domains. *J. Biol. Chem.*, **2006**, *281*, 20271-20282.

[33] Joo, S. H.; Pei, D. Synthesis and screening of support-bound combinatorial peptide libraries with free C-termini: determination of the sequence specificity of PDZ domains. *Biochemistry,* **2008**, *47*, 3061-3072.

[34] Samson, I.; kerremans, L.; Rozenski, J.; Samyn, B.; Van Aerschot, A.; Van Beeumen, J.; Van Aerschot, A; Herdewijn, P. Screening a random pentapeptide library, composed of 14 D-amino acids, against the COOH-terminal sequence of fructose-1,6-bisphosphate aldolase from Trypanosoma brucei. *J. Biol. Chem.*, **1997**, *272*, 11378-11383.

[35] Aggarwal, S.; Harden, J. L.; Denmeade, S. R. Synthesis and screening of a random dimeric peptide library using the one-bead-one-dimer combinatorial approach. *Bioconjugate Chem.*, **2006**, *17*, 335-340.

[36] Chen, X.; Tan, P. H.; Zhang, Y.; Pei, D. On-bead screening of combinatorial libraries: reduction of nonspecific binding by decreasing surface ligand density. *J. Comb. Chem.*, **2009**, *11*, 604-611.

[37] Panchuk-Voloshina, N.; Haugland, R. P.; Bishop-Stewart, J.; Bhalgat, M. K.; Millard, P. J.; Millard, Mao, F.; Leung, W.-Y.; Haugland, R. P. Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *J. Histochem. Cytochem.*, **1999**, *47*, 1179-1188.

[38] Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De novo sequencing of peptides by MS/MS. *Proteomics*, **2010**, *10*, 634-649.

Received: October 9, 2013 **Revised: December 31, 2013** **Accepted: January 9, 2014**