



Published in final edited form as:

Anal Chem. 2010 January 15; 82(2): 672–679. doi:10.1021/ac902195y.

## Accurate MALDI-TOF/TOF Sequencing of One-Bead-One-Compound Peptide Libraries with Application to the Identification of Multi-ligand Protein Affinity Agents Using In Situ Click Chemistry Screening

Su Seong Lee<sup>†</sup>, Jaehong Lim<sup>†</sup>, Sylvia Tan<sup>†</sup>, Junhoe Cha<sup>†</sup>, Shi Yun Yeo<sup>†</sup>, Heather D. Agnew<sup>‡</sup>, and James R. Heath<sup>†,‡</sup>

Su Seong Lee: sslee@ibn.a-star.edu.sg; Jaehong Lim: ; Sylvia Tan: ; Junhoe Cha: ; Shi Yun Yeo: ; Heather D. Agnew: ; James R. Heath: heath@caltech.edu

<sup>†</sup>Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, The Nanos, Singapore 138669

<sup>‡</sup>Division of Chemistry and Chemical Engineering, Caltech. MC 127-72, Pasadena, CA 91125

### Abstract

Combinatorial one-bead-one-compound (OBOC) peptide libraries are widely used for affinity screening, and the sequencing of peptides from hit beads is a key step in the process. For rapid sequencing, CNBr cleavage of the peptides from the beads, followed by *de novo* sequencing by MALDI-TOF/TOF is explored. We report on a semi-automated sequencing algorithm, and validate it through comparison against Edman degradation sequencing. The initial 44% sequencing success rate of the standard *de novo* sequencing software was improved to nearly 100%. The sequencing algorithm incorporates existing knowledge of amino acid chemistry, and a new strategy for differentiating isobaric amino acids. We tested the algorithm by using MALDI-TOF/TOF to identify a peptide biligand affinity agent against the protein bovine carbonic anhydrase II, starting from comprehensive one-bead-one-compound peptide libraries comprised of non-natural and artificial amino acid components, and using the strategy of *in situ* click/OBOC library screening.

### Introduction

Combinatorial chemical libraries can represent a broad chemical space that can be screened to identify, for example, drug candidates or biomolecular affinity agents.<sup>1</sup> Particularly useful are large one-bead-one-compound (OBOC) libraries of peptides or peptide-like molecules that are routinely generated using the split-and-mix approach.<sup>2</sup> A bottleneck for identifying hit peptides from screened OBOC libraries is the actual determination of the peptide sequences from hit beads. Edman degradation<sup>3</sup> provides the gold standard, but it is slow. The most time-efficient sequencing approach is that of Matrix-assisted laser desorption/ionization tandem mass spectrometry (MALDI-MS/MS) sequencing, but it lacks accuracy, and so is only rarely used for sequencing bead-based peptides.<sup>4</sup> This may seem surprising, considering the widespread use of MS/MS methods for sequencing peptides from proteins.<sup>5</sup> However, such sequenced peptides may be compared against a protein data base in order to rule out or add statistical weight to certain sequences. For peptides sequenced from OBOC libraries, no such data base exists, and so the *de novo* sequencing approach needs to be highly accurate.

Correspondence to: Su Seong Lee, sslee@ibn.a-star.edu.sg; James R. Heath, heath@caltech.edu.

**Supporting Information Available.** Additional information as noted in text.

Certain bead-based peptide libraries are specifically synthesized to permit MS sequencing, such as the ladder synthesis method.<sup>6</sup> Post-screening chemistries designed to generate peptide fragments, can permit partial Edman degradation/MS<sup>4</sup> sequencing. However, each of these methods has significant disadvantages. Namely, the chemistries used for library preparation are non-trivial to execute and can influence the screening results. A robust MS/MS sequencing approach that can rapidly generate accurate sequences from simply-prepared OBOC peptide libraries would be useful.

We report here on semi-automated *de novo* peptide sequencing *via* MALDI-MS/MS for application to OBOC peptide libraries. Several challenges were addressed in order to develop this algorithm, including an approach for discriminating between the isobaric amino acids K/Q and I/L. Previously reported strategies for addressing this last challenge have involved the adoption of coding tags, and are not designed for MS/MS *de novo* peptide sequencing.<sup>6a,6c</sup>

We first describe the semi-automated algorithm in detail, and we then validate the method by sequencing 5-mer peptides from single (split) beads using both Edman degradation and MALDI-TOF/TOF *de novo* sequencing. We then utilize the sequencing approach, together with the recently reported method of sequential *in situ* click/OBOC library screening,<sup>7</sup> to identify a set of biligand affinity agents against the bovine carbonic anhydrase II (bCAII) protein.

For the *in situ* click/OBOC screening method, a short primary (1°) peptide ligand is first identified using standard OBOC screening approaches. The 1° peptide is modified with a terminal acetylene group to form an anchor peptide, and prepared in bulk. To identify the peptide biligand, a comprehensive OBOC peptide library is again utilized, but each peptide contains an azide-functionalized artificial amino acid at N-terminus. That library, excess anchor peptide, and the protein target, are then incubated together. The Cu(I) catalyzed reaction between azide and acetylene groups to form a 1,2,3-triazole (Tz) linkage is well known in the literature as the CuAAC click reaction.<sup>8</sup> The basic idea behind the *in situ* click approach is that the protein itself replaces the Cu(I) catalyst,<sup>9</sup> coupling the anchor peptide with only those bead-bound peptides that bind to the protein in just the right conformation.

The *in situ* click/OBOC library screening approach can be employed sequentially, to produce, for example, triligands (from biligands), etc., with affinity and selectivity growing rapidly with ligand size. These multiligands can exhibit antibody-like binding, affinity and selectivity characteristics, coupled with chemically-designed stability.

*In situ* click/OBOC library screening only requires the sequencing of peptides from short chain (5- to 7-mer) peptides, and so is ideal for MALDI-TOF/TOF *de novo* sequencing. However, those libraries must sample broad chemical space, and hit beads must be statistically sampled. In addition, the identification of a multiligand affinity agent can involve 2 or more screens for each ligand component. Thus, an accurate and rapid sequencing strategy is required to make the approach high throughput. Our results indicate that the MALDI-TOF/TOF sequencing algorithm provides such a strategy. Three biligands against bCAII, with affinities in the 0.5-5  $\mu$ M range, are identified.

## Experimental Section

### General

N-methylpyrrolidone (NMP), diethylether, dichloromethane (DCM), and Fmoc-NH-(PEG)<sub>2</sub>-COOH (20 atoms) were purchased from Merck. Fmoc-Rink Amide MBHA Resins, natural and non-natural Fmoc-protected amino acids (Fmoc-AA's) were purchased from GL Biochem (Shanghai) Ltd. TentaGel S Amino resin was purchased from AnaSpec (Fremont, CA).  $\alpha$ -

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. MALDI-MS and MS/MS were obtained using a Bruker Autoflex II TOF/TOF. Edman sequencing of single beads was carried out on a Model 494 Procise cLC Sequencing System (Applied BioSystems, Foster City, CA). The PEAKS software was purchased from Bioinformatics Solutions Inc. (Waterloo, ON, Canada). The microwave-assisted CNBr cleavage reaction was performed by a household microwave oven (Model: R-248J, 800 W, 2450 MHz) from Sharp Inc.<sup>6</sup> The purification of bulk peptides was done by a preparative HPLC system from Gilson on a C<sub>18</sub> reversed phase preparative column (Kromasil<sup>®</sup> from AkzoNobel, 5  $\mu$ m, 250  $\times$  30 mm).

### Construction of peptide libraries

Randomized OBOC libraries of penta- to heptapeptides were synthesized using an automatic synthesizer Titan 357 (AAPPTeC) *via* standard split-and-mix methods on polyethylene glycol-grafted polystyrene beads (TentaGel S-NH<sub>2</sub>, 90  $\mu$ m, 0.29 mmol/g, 2.86  $\times$  10<sup>6</sup> beads/g). Unless otherwise specified, non-natural D-stereoisomers were used at every possible position in the peptide sequence. For all the coupling steps, a standard solid-phase peptide synthesis method with Fmoc chemistry<sup>10</sup> was used. The resin was swelled in NMP for 2 h in the Collective Vessel (CV). The coupling of Fmoc-methionine was initiated by addition of 2 equiv. of TBTU and 5 equiv. of DIEA. The coupling reaction was run for 30 min. Another 2 equiv Fmoc-methionine, 2 equiv TBTU, and 5 equiv DIEA were added, and allowed to react for 30 min ("double coupling"). Following the coupling step, the beads were thoroughly washed (4  $\times$  NMP), and treated with 20% piperidine in NMP (5 min and then 15 min with a fresh aliquot of deprotection solution). The resin was thoroughly washed (4  $\times$  NMP, 4  $\times$  DCM), and divided into multiple equal-mass aliquots for the next cycle of coupling in the Reaction Vessel (RV). With the coupling and Fmoc deprotection completed, the resins were combined in CV. The procedures were repeated until the desired length of peptide was attained. The amino acid side-chain protective groups were then removed by incubation in trifluoroacetic acid (95%), water (2.5%), and triisopropylsilane (2.5%) for 2 h. The library resin was then washed thoroughly with dichloromethane (DCM, 5  $\times$ ), methanol (MeOH, 5  $\times$ ), water (5  $\times$ ), MeOH (5  $\times$ ), DCM (5  $\times$ ), and then diethylether (5  $\times$ ). The resulting resin was dried under vacuum and stored at 4  $^{\circ}$ C.

### Synthesis of bulk peptides

Bulk synthesis of hit peptides was performed on Fmoc-Rink amide MBHA resins (0.58 mmol/g), on a typical resin scale of 0.2 g 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 diethylether. The resulting white solid then purified to >95% by HPLC on a C<sub>18</sub> reversed phase preparative column. The pure peptides were used for affinity measurements, screens, and binding assays.

**Anchor peptide N<sub>3</sub>(CH<sub>2</sub>)<sub>3</sub>C(O)-lhrywf**—The purified peptides showed >98% purity in HPLC. ESI-MS of the purified product gave peaks at *m/z* 516.5 for [M+2H]<sup>2+</sup> and 1031.7 for [M+H]<sup>+</sup>.

**Biligand rkyvfi-Tz1-lhrywf**—The purified peptides showed >98% purity in HPLC. ESI-MS of the purified product gave peaks at *m/z* 646.1 for [M+3H]<sup>3+</sup> and 968.3 for [M+2H]<sup>2+</sup>.

**Biligand lrvflk-Tz1-lhrywf**—The purified peptides showed >98% purity in HPLC. ESI-MS of the purified product gave peaks at *m/z* 629.4 for [M+3H]<sup>3+</sup> and 943.3 for [M+2H]<sup>2+</sup>.

**Biligand kwlwGI-Tz2-lhrywf**—The purified peptides showed >98% purity in HPLC. ESI-MS of the purified product gave peaks at  $m/z$  643.2 for  $[M+3H]^{3+}$  and 964.4 for  $[M+2H]^{2+}$ .

### Synthesis of bulk peptides labeled with biotin

(Supporting Information, Figure S1) On Fmoc-Rink amide MBHA resins (0.58 mmol/g), Fmoc-lysine(Mtt)-OH (Mtt = 4-methyltrityl) was introduced using a standard peptide coupling chemistry described above. The Mtt group was selectively removed in trifluoroacetic acid (3%), triisopropylsilane (3%), and DCM (94%) for 2 min, 5 min, and 30 min, successively with a fresh aliquot of deprotection solution. The biotin group was introduced by treating the resin with biotin-NHS (2 equiv.) for 30 min in the presence of 5 equiv. of DIEA in NMP. In order to improve solubility of peptides in water, polyethylene glycol moiety was introduced by treating with Fmoc-PEG2-CO<sub>2</sub>H *via* a standard peptide coupling method. 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 diethylether. The resulting white solid then purified to >90% by HPLC on a C<sub>18</sub> reversed phase column. The pure peptides were used for affinity measurements and binding assays.

**Biotinylated ifvykr**—The purified peptides showed >98% purity in HPLC. ESI-MS of the purified product gave peaks at  $m/z$  526.8 for  $[M+3H]^{3+}$  and 789.9 for  $[M+2H]^{2+}$ .

**Biotinylated biligand rkyvfi-Tz1-lhrywf**—The purified peptides showed >98% purity in HPLC. TOF MS of the purified product gave peaks at  $m/z$  2609.7 for  $[M+H]^+$ .

### CNBr cleavage of peptides from single beads

A single bead was transferred to a micro-sized vial, containing pure water (10  $\mu$ l). After addition of CNBr (10  $\mu$ l, 0.50 M in 0.2 N HCl solution) the reaction vessel was purged by argon for 15 min and placed under microwave for 1 min. After additional purging by argon for 15 min the resulting solution was concentrated under centrifugal vacuum for 10 min at 45 °C and then for 50 min at 60 °C.

### MALDI-MS and MS/MS analysis of peptides cleaved from single beads

To each vial or well were added CHCA (10  $\mu$ l, 0.5% solution in acetonitrile/water (70:30)) and then acetonitrile/water (10  $\mu$ l, 70:30 containing 0.1% trifluoroacetic acid (v/v)). 2  $\mu$ l of the mixture solution was taken up to be spotted onto a 384-well MALDI plate, which was stood for 15 min as drying naturally.

### Library screening

For screening, the bCAII protein was first labeled using the Alexa Fluor<sup>®</sup> 647 protein labeling kit (A20173, Invitrogen), according to the supplier's protocol. First, a 2 mg/ml solution of bCAII was dissolved in 0.1 M sodium bicarbonate (pH ~ 8.3). 0.5 ml of this bCAII solution was transferred into the vial of the reactive dye. The vial was capped and inverted a few times to fully dissolved dye. The reaction mixture was stirred for 1 hr at room temperature under dark conditions. The Alexa Fluor<sup>®</sup> 647-labeled bCAII was purified from the mixture using the size exclusion purification resin in the labeling kit. Purified and labeled bCAII was characterized by UV-Vis spectroscopy and SDS-PAGE. For the screen, 200 mg of library resin was transferred into an 8-ml Alltech vessel and pre-incubated in a blocking solution, 0.05% NaN<sub>3</sub>, 0.1% Tween 20 and 0.1% BSA in PBS buffer (pH 7.4) for 1 hr on 360-degree shaker at 25°C. The buffer solution was drained by vacuum and then 5 ml of 10 nM dye-labeled bCAII diluted in blocking solution was added to the swollen resin. The resulting mixture was incubated for 15 - 18 hr on 360-degree shaker at 25°C. The liquid was drained by vacuum and

non-specifically bound proteins were eliminated by washing 3 times with blocking solution, 3 times with 0.1% Tween 20 in PBS buffer, sequentially. Lastly, the resin was washed 6 times with PBS buffer. After stringent washing, 200 mg of the assayed library resin was transferred into sample vessel of COPAS Plus (Union Biometrica)<sup>11</sup> and diluted with 200 ml of PBS buffer (pH=7.4). We applied two-step sorting (Supporting Information, Figure S2). In the second sorting, positive beads were directly sorted into a 96 titer well plate with conical-shaped wells.

### In situ screen

5 ml of 10nM Alexa Fluor<sup>®</sup> 647-labeled bCAII diluted in blocking solution (PBS buffer, pH 7.4), 0.1% Tween 20, 0.1% bovine serum albumin (BSA), 0.05% NaN<sub>3</sub>) was pre-incubated with anchor peptide, N<sub>3</sub>(CH<sub>2</sub>)<sub>3</sub>C(O)-lhrywf. The anchor ligand was supplied at  $\geq 2000$ -fold excess of the protein. The pre-incubation was done in an 8-ml Alltech vessel for 2 hr on a 360-degree shaker at 25 °C. 200 mg of the OBOC peptide library, coupled with 4-pentynoic acid at N-terminus, was transferred into another 8-ml Alltech vessel and incubated with the blocking solution for 1 hr on a 360-degree shaker at 25°C. It was then washed with 3  $\times$  5 ml of the blocking solution, drained by vacuum. The pre-incubated solution of 10nM Alexa Fluor<sup>®</sup> 647-labeled bCAII and the anchor peptide were then added to the swollen resin. The resulting mixture was incubated for 18 hr on a 360-degree shaker at 25 °C. Non-specifically bound proteins were eliminated by washing 3 times with blocking solution and further incubation with 5 ml blocking solution for 2 hr. Lastly, the mixture was drained by vacuum and washed 6 times with 0.1% Tween 20 in PBS buffer. After stringent washing, 200 mg of the assayed library resin was transferred into the sample vessel of a COPAS Plus (Union Biometrica) automated bead sorter,<sup>11</sup> and diluted with 200 ml of PBS buffer (pH=7.4). A two-step sorting (see Supporting Information) was applied. In the second sorting, positive beads were directly sorted into a 96 titer well plate with conical-shaped wells. Sequencing of these beads was performed with MALDI-TOF/TOF and the semi-automated algorithm. The 8 most abundant amino acid at each position were then chosen to generate a focused hexamer library i.e. P- $\times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ -m(100%)-TentaGel (P = 5-pentynoic acid). For this library, the amino acids at each position  $\times 6 \times 1$  were:  $\times 6$  (v,i,l,f,k,t,w,r);  $\times 5$  (f,l,r,k,v,i,n,a);  $\times 4$  (f,y,l,k,r,v,i,w);  $\times 3$  (f,w,l,k,v,r,y,a);  $\times 2$  (k,r,l,f,v,w,s,p); and,  $\times 1$  (k,r,w,f,s,y,l,i). A second *in situ* screening was repeated with the focused hexamer library and anchor peptide, and positive beads were sorted, collected and sequenced.

### Affinity measurements

Affinity measurements were performed using Biacore T100 system and research grade CM5 sensor chips (GE Healthcare). The instrument was primed with HBS-EP+ (GE Healthcare) buffer. Flow cell 1 (or 3) was used as a reference to subtract nonspecific binding, drift, and bulk refractive index while flow cell 2 (or 4) was immobilized with bCAII following standard procedures. 1:1 mixture of 0.4 M EDC and 0.1 M NHS was used to activate flow cell 2 (or 4) and 0.1mg/ml bCAII solution was injected. Blocking of the remaining activated groups was done with 1M solution of ethanolamine (pH 8.5). bCAII was immobilized onto the sensor chip surface by approximately 6000 response unit (RU). The instrument was then primed using running buffer (HBS-EP+). Two biligands identified through this work, and the biligand reported previously<sup>12</sup> were dissolved in HBS-EP+ buffer to produce 5  $\mu$ M peptide stock solutions for each peptide, which were serially diluted by a factor of 2 to produce a concentration series down to 2 nM. For a given affinity measurement, these series of peptide solutions successively were injected to flow cell 2 (or 4) for 3 min of contact time, 5 min of dissociation time, and 3.5 min of stabilization time using flow rate of 100  $\mu$ L/min at 25 °C. Flow cell 2 (or 4) was regenerated by Glycine 2.5 (GE Healthcare) after injection of each peptide solution. Data processing and kinetic analysis were performed using Biacore T100 Evaluation Software (version 2.0.1, Biacore). Two different statistical fit routines, supplied by the vendor software package, were employed to extract affinity constant ( $K_D$ ) values: the 'two-

state reaction' fit, and the 'heterogeneous ligand' fit.  $K_D$  values by the both fitting methods were similar to  $K_D$  value (3  $\mu\text{M}$ ) for the reported biligand<sup>12</sup> (see Supporting Information, S-Fig. 10).

### Dot blot selectivity/sensitivity assays in serum

The affinity of the biligand capture agents for bCAII was demonstrated through the use of dot blot experiments in 5% non-fat dry milk in TBS-T [25 mM Tris, 150 mM NaCl, 2 mM KCl, 0.5% Tween20 (pH 8.0)]. bCAII solution was prepared as 10 mg/ml stocks in PBS buffer (pH 7.4). A dilution series of bCAII solution was applied to a nitrocellulose membrane, typically ranging from 20  $\mu\text{g}$  to 2 ng per spot. The membrane was blocked at room temperature for 2 hr in 5% non-fat milk/TBS-T. The membrane was then washed with TBS-T. The biotinylated biligand solution was prepared at 1  $\mu\text{M}$  in 5% non-fat milk/TBS-T and incubated over the membrane 2 hr at room temperature. For the comparison, 1:4000 biotinylated bCAII-antibody (Abcam, Cambridge, MA) was prepared in 5% non-fat milk/TBS-T and incubated over the membrane 2 hr at room temperature. After 3 times washing with TBS-T for 10 min, 1:3000 Streptavidin-HRP (Abcam, Cambridge, MA) prepared in 0.5% milk/TBS-T was added to the membrane and incubated for 2 hr. After 3 times washing with TBS-T for 10 min, the membrane was treated to chemiluminescent reagents (Amersham ECL Plus™ Western Blotting Detection Reagents, GE Healthcare) and then immediately developed on film.

## Results and Discussion

### Semi-automatic peptide sequencing

All OBOC peptide libraries utilized here were prepared using the double coupling method to ensure very high purity OBOC peptide libraries. **Library 1** was a 5-mer library that was comprehensive in the D-stereoisomers of all natural (L) amino acids except cysteine and methionine. We refer to these D-stereoisomers by the lower-case of the standard, single letter abbreviation (i.e. W = L-tryptophan, w = D-tryptophan). For all MALDI-MS/MS sequencing, peptides from single beads (or half-beads) were cleaved using optimized CNBr cleavage conditions.<sup>13</sup>

To establish an initial point of reference, 48 beads were randomly selected from **Library 1**, halved, and then sequenced by both Edman degradation and MALDI-MS/MS using commercial sequencing software (PEAKS; Bioinformatics Solutions Inc., Waterloo, ON, Canada). For *de novo* MS/MS sequencing, the cleaved peptide is initially interrogated using MALDI-MS. Mass peaks originating from the matrix background are subtracted, and the highest intensity remaining mass peak is chosen as the parent ion for MS/MS analysis. The commercial software rank-orders several potential sequences from the MS/MS analysis of the selected parent ion. The highest scoring candidate that contained 5 amino acids, including homoserine lactone, matched the sequence generated by Edman degradation analysis of the other half of the same bead with a 44% success rate. (Supporting Information, Table S1).

This low sequencing success rate originated from 3 general influences.

1. Incorrect assignment of the parent ion. This could be attributed to the formation of a  $\text{Na}^+$  adduct to the parent ion, (Supporting Information, Figure S3) or to the cyclization of N-terminal amino acids to pyroglutamate, such as glutamic acid and glutamine.
2. Mass-modification of amino acids through chemical processes associated with the peptide cleavage chemistry or the laser desorption steps. This led to incorrect identification of constituent amino acids.
3. Incorrect assignment of the isobaric amino acids K/Q and I/L.

The accuracy of the parent ion selection could be increased in two ways. First, sodium or potassium adducts of the parent ion could be greatly reduced by washing the sample spot on the MALDI plate with 10 mM ammonium monobasic phosphate in water<sup>14</sup> (Supporting Information, Figure S4). Second, the cyclization of N-terminal amino acids to pyroglutamate could be identified by the presence of typically lower-intensity mass peaks shifted relative to the parent ion, i.e., -17 amu from glutamine and -18 amu from glutamic acid (Supporting Information, Figure S5). Finally, calibration of the instrument with mass standards for every 8 samples improved measurement accuracy of both parent and fragment ion masses.

Once the parent ion in the MS spectrum was successfully selected and its MS/MS spectrum was obtained, accurate sequence information from PEAKS was obtained by taking into account the known number of amino acids in a sequence and the presence of a homoserine lactone at C-terminus. In addition, specific amino acids in the MS/MS spectrum were identified by accounting for specific amino acid chemistries. For example, the presence of tryptophan (w) in a sequence was confirmed by ion peaks from peptides with oxidized w<sup>15,16</sup> in the MS spectrum. This was useful for differentiating w from the isobaric amino acid combinations of d-a, e-g, and v-s. The oxidation of w occurred during the CNBr cleavage reaction, although the amount of the oxidized product could be reduced by argon purging. In addition, several amino acids such as r, k, q and w consistently show unique low mass fragment peaks other than the immonium ions<sup>17-19</sup> (Supporting Information, Figure S6 to Figure S8). These and other considerations that led to an improved sequencing algorithm are included in the Supporting Information.

A third challenge for obtaining accurate peptide sequences involved discriminating between the isobaric amino acid pairs k/q and l/i. We developed a simple strategy in which, during the library preparation a small percent of a low molecular weight amino acid was substituted when either q or i was coupled onto the beads (Scheme 1).<sup>6a,6c,6d</sup> This percentage was chosen to permit identification by MALDI-MS/MS, but to be sufficiently low so as not to influence the library screening results. For q, 6 – 10% of g was added, and for i, 7 – 10% of a was added. This bead chemistry was validated by Edman degradation sequencing. A sequence with a q exhibits a lower intensity mass peak at -71 amu relative to the parent ion. For i, a lower intensity peak at -42 amu is observed. If a sequence contains both q and k, the exact position of q can be obtained by analyzing the MS/MS of the additional smaller peak. The approach for discriminating between i and l is analogous. This simple strategy was successfully applied to the sequencing of hit beads from **Library 2** (Supporting Information, Table S5 and S6).

Scheme 2 shows the overall procedure for the semi-automatic peptide sequencing. We applied this approach towards successfully sequencing the (halved) beads which were unsuccessfully sequenced using the standard commercial algorithm of PEAKS (Table 1). (Supporting Information, Table S1) An example, demonstrated in Scheme 1, shows the semi-automatic sequencing of the peptide lweykm\* which was unsuccessfully sequenced using the standard software.

To further validate the semi-automatic sequencing approach, 28 randomly selected beads from **Library 1** were halved and then sequenced by both of MALDI-TOF/TOF and Edman degradation. The semi-automatic approach exhibited a nearly 100% sequencing success rate except isobaric amino acid differentiation, since the relevant chemistry was not incorporated into the preparation of **Library 1** (see Supporting Information, Table S2). Three non-matching sequences were correct in amino acid identification, but not sequence order.

### Identification of a Peptide Affinity Agent Against bCAII

The sequencing algorithm was validated by applying it towards identifying a biligand affinity agent against the protein bCAII, using the *in situ* click/OBOC library screening approach. We

began by identifying a 1° ligand *via* two generations of screens against first a comprehensive and then a focused 6-mer OBOC peptide library. The hit beads from the comprehensive 5-mer peptide library were analyzed using the semi-automated MALDI-TOF/TOF method, which resulted in great reproducibility as well as a high homology to sequences separately identified *via* the Edman degradation method (Figure 1). That 1° ligand modified with 4-pentynoic acid at N-terminus then served as the anchor ligand for two similar generations of *in situ* click/OBOC screens designed to identify a 2° ligand, and thus a biligand. For all screens, a COPAS Plus Automated Bead Sorter was used to sort hit beads, MALDI-TOF/TOF data were collected from individual hit beads, and the sequencing algorithm (including the strategy for differentiating isobaric amino acids) was applied to identify hit sequences. For first generation screens, the sequences guided the development of a focused library. For second generation screens, they provided the identity of 1° or 2° ligands. The final biligand candidates then were tested as affinity agents against bCAII using surface plasmon resonance (SPR) affinity measurements, as well as dot blot experiments (Figure 2). The conditions of the screens, and the relevant OBOC library descriptions are provided in Table 2.

For the second generation 1° ligand screen (2gen\_1°L) using a focused OBOC library, common sequence motifs such as lhry and wrw were observed, including 10 hit sequences of lhryw<sub>x</sub> (x = r, f, k or y). (Supporting Information, Table S4) The peptide lhrywf was chosen as the 1° ligand, and modified with 4-azidobutanoic acid [N<sub>3</sub>(CH<sub>2</sub>)<sub>3</sub>CO<sub>2</sub>H] at N-terminus to form the anchor peptide for use in the biligand screens.

The 2gen\_2°L (focused library *in situ*) screen yielded very selective positive sequences representing xfy(k/r)(k/r), kxfv(x/y) and x(r/k)x(w/f)x(r/k) (x = hydrophobic amino acids such as l, i, v and f). (Supporting Information, Table S5 and S6) These sequences are similar to those reported recently from similar screening procedures, but using Edman degradation-based sequencing.<sup>12</sup> Based upon this screen, two triazole coupled peptide biligands were synthesized in bulk, and their binding affinities against bCAII was characterized using SPR (Supporting Information, Figure S10). Note that the best of these biligands exhibits an affinity for bCAII in the 0.5 – 5 μM, which is as good or better than the biligands identified previously through the use of *in situ* click/OBOC library screening and Edman degradation sequencing.<sup>12</sup> Compared to the previous work, the MALDI TOF/TOF sequencing approach was both significantly more efficient and permitted a deeper statistical sampling of all the screens carried out here. For example, with Edman degradation, sequencing 25 hit beads takes several days, while the MS/MS method reported here can accurately sequence 100 – 200 beads in a few hours, thus permitting a much broader exploration of the results of an affinity screen. The binding affinity was confirmed by dot blot experiments (Figure 2).

## Conclusions

We have demonstrated an efficient algorithm for the robust and accurate use of MALDI-TOF/TOF for the sequencing of peptides cleaved from single TentaGel beads. An initial 44% sequencing success rate by a *de novo* sequencing software PEAKS was improved nearly up to 100% by a semi-automated sequencing algorithm that was combined with a new strategy for differentiating isobaric amino acids, and with graphical functions of PEAKS. The MS/MS approach was validated in two ways. First, single beads, randomly chosen from a comprehensive OBOC 5-mer peptide library, were split in half and sequenced using both the MS/MS approach, and Edman degradation. Excellent agreement was found between the two methods. Second, the sequencing approach was utilized to successfully identify two biligand affinity agents against bCAII through the approach of *in situ* click/OBOC screening. The sequencing approach and algorithm reported here should provide a valuable tool for making the approach of sequential *in situ* click/OBOC screening a high throughput approach towards the identification of high affinity, high selectivity protein capture agents.



## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

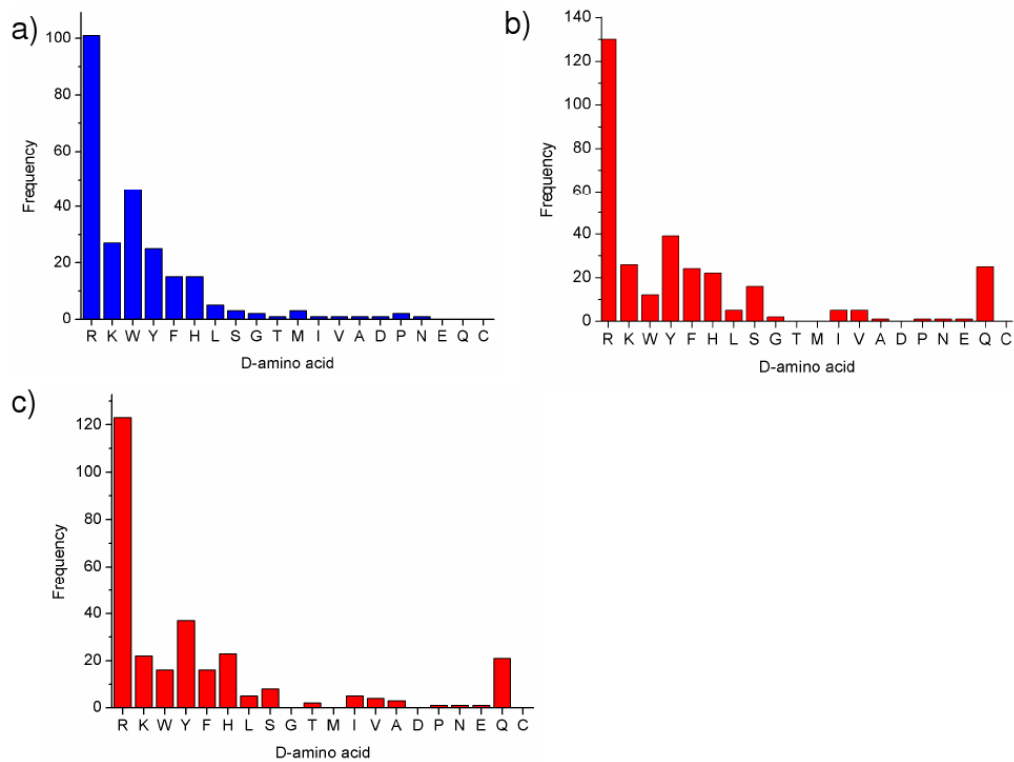
## Acknowledgments

This work was supported by the Institute of Bioengineering and Nanotechnology (Biomedical Research Council, Agency for Science, Technology and Research, Singapore), with additional support (JRH) from the National Cancer Institute Grant No. 5U54 CA119347 (J.R.H., P.I.) and funding from the Grand Duchy of Luxembourg.

## References

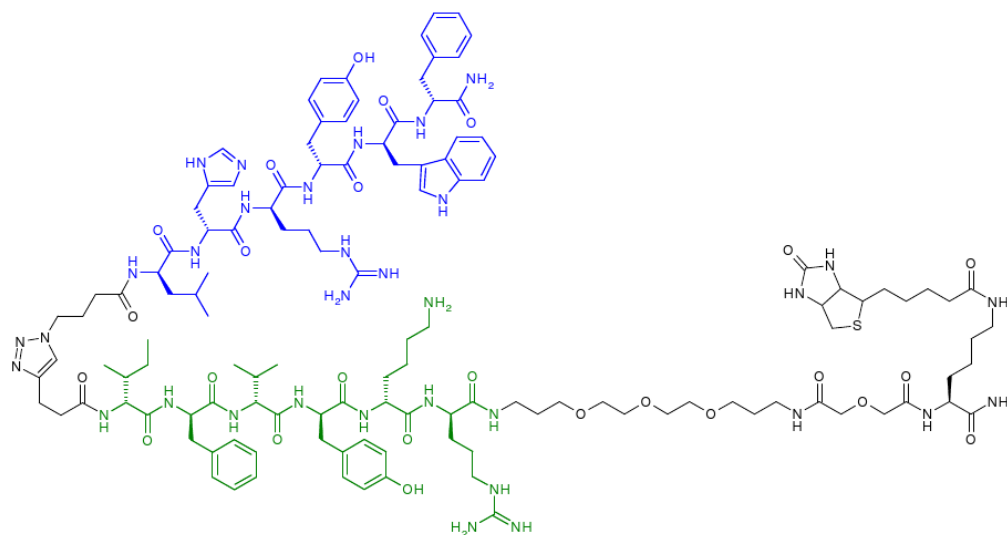
1. a) Furka A, Sebestyén F, Asgedom M, Dibo G. *Int J Pept Protein Res* 1991;37:487–493. [PubMed: 1917305] b) Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, Cuervo JH. *Nature* 1991;354:84–86. [PubMed: 1719428]
2. a) Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ. *Nature* 1991;354:82–84. [PubMed: 1944576] b) Lam KS, Lebl M, Krchňák V. *Chem Rev* 1997;97:411–448. [PubMed: 11848877]
3. a) Edman P. *Acta Chem Scand* 1950;4:283–293. b) Niall HD. *Methods Enzymol* 1973;27:942–1010. [PubMed: 4773306]
4. a) Chait BT, Wang R, Beavis RC, Kent SBH. *Science* 1993;262:89–92. [PubMed: 8211132] b) Wang P, Arabaci G, Pei D. *J Comb Chem* 2001;3:251–254. [PubMed: 11350247] c) Sweeney MC, Pei D. *J Comb Chem* 2003;5:218–222. [PubMed: 12739936] d) Thakkar A, Wavreile AS, Pei D. *Anal Chem* 2006;16:5935–5939. [PubMed: 16906744] e) Joo SH, Xiao Q, Ling Y, Gopishetty B, Pei D. *J Am Chem Soc* 2006;128:13000–13009. [PubMed: 17002397]
5. Pevtsov S, Fedulova I, Mirzaei H, Buck C, Zhang X. *J Proteome Res* 2006;5:3018–3028. [PubMed: 17081053]
6. a) Youngquist RS, Fuentes GR, Lacey MP, Keough T. *J Am Chem Soc* 1995;117:3900–3906. b) Buchardt J, Schiødt CB, Korg-Jensen C, Delaissé JM, Foged NT, Meldal M. *J Comb Chem* 2000;2:624–638. [PubMed: 11138549] c) Wang X, Peng L, Liu R, Gill SS, Lam KS. *J Comb Chem* 2005;7:197–209. [PubMed: 15762747] d) Hu BH, Jones MR, Messersmith PB. *Anal Chem* 2007;79:7275–7285. [PubMed: 17713965] e) Kim YG, Shin DS, Kim EM, Park HY, Lee CS, Kim JH, Lee BS, Lee YS, Kim BG. *Angew Chem Int Ed* 2007;46:5408–5411.
7. Agnew HD, Rohde RD, Millward SW, Nag A, Yeo WS, Hein JE, Pitram SM, Tariq AA, Burns VM, Krom RJ, Fokin VV, Sharpless KB, Heath JR. *Angew Chem Int Ed* 2009;48:1–5.
8. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. *Angew Chem* 2002;114:2708–2711. *Angew Chem Int Ed* 2002;41:2596–2599.
9. a) Lewis WG, Green LG, Grynszpan F, Radić Z, Carlier PR, Taylor P, Finn MG, Sharpless KB. *Angew Chem* 2002;114:1095–1099. *Angew Chem Int Ed* 2002;41:1053–1057. b) Manetsch R, Krasinski A, Radić Z, Raushel J, Taylor P, Sharpless KB, Kolb HC. *J Am Chem Soc* 2004;126:12809–12818. [PubMed: 15469276] c) Bourne Y, Kolb HC, Radić Z, Sharpless KB, Taylor P, Marchot P. *Proc Natl Acad Sci U S A* 2004;101:1449–1454. [PubMed: 14757816] d) Mocharla VP, Colasson B, Lee LV, Röper S, Sharpless KB, Wong CH, Kolb HC. *Angew Chem* 2005;117:118–122. *Angew Chem Int Ed* 2005;44:116–120. e) Whiting M, Muldoon J, Lin YC, Silverman SM, Lindstrom W, Olson AJ, Kolb HC, Finn MG, Sharpless KB, Elder JH, Fokin VV. *Angew Chem* 2006;118:1463–1467. *Angew Chem Int Ed* 2006;45:1435–1439.
10. Fields GB, Noble RL. *Int J Pept Protein Res* 1990;35:161–214. [PubMed: 2191922]
11. a) Burns AR, Kwok TCY, Howard A, Houston E, Johanson K, Chan A, Cutler SR, McCourt P, Roy PJ. *Nature protocols* 2006;1:1906–1914. b) Tornøe CW, Sanderson SJ, Mottram JC, Coombs GH, Meldal M. *J Comb Chem* 2004;6:312–324. [PubMed: 15132590] c) Garske AL, Denu JM. *Biochemistry* 2006;45:94–101. [PubMed: 16388584] d) Hu BH, Jones MR, Messersmith PB. *Anal Chem* 2007;79:7275–7285. [PubMed: 17713965] e) Kodadek T, Bachhawat-Sikder K. *Mol Biosyst* 2006;2:25–35. [PubMed: 16880920]
12. Gehrig PM, Hunziker PE, Zahariev S, Pongor S. *J Am Soc Mass Spectrom* 2004;15:142–149. [PubMed: 14766281]

13. Lee SS, Lim J, Cha J, Tan S, Heath JR. *J Comb Chem* 2008;10:807–809. [PubMed: 18811218]
14. Smirnov IP, Zhu S, Taylor T, Huang Y, Ross P, Papayanopoulos IA, Martin SA, Pappin DJ. *Anal Chem* 2004;76:2958–2965. [PubMed: 15144210]
15. Kraft P, Mills J, Dratz E. *Anal Biochem* 2001;292:76–86. [PubMed: 11319820]
16. Ablonczy, Zs; Knapp, DR.; Darrow, R.; Organisciak, DT.; Crouch, RK. *Mol Vision* 2000;6:109–115.
17. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens JP, Morla A, Bouchu D. *Rapid Commun Mass Spectrom* 2003;17:1297–1311. [PubMed: 12811753]
18. Papayannopoulos IA. *Mass Spectrom Rev* 1995;14:49–73.
19. Falick AM, Hines WM, Medzihradzky KF, Baldwin MA, Gidson BW. *J Am Soc Mass Spectrom* 1993;4:882–893.

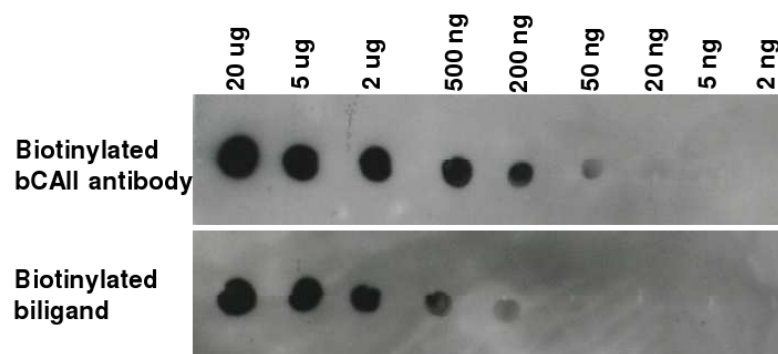
**Figure 1.**

Histograms of amino acids identified from sequencing hit beads using various approaches. Frequency on y-axis indicates the number of each D-amino acids occurring in the sequencing results. Each histogram represents the results from the screening of 200 mg of 5-mer comprehensive library beads against bCAII. a) 51 Manually picked hit beads, sequenced by Edman degradation. b) 63 Automatically sorted hit beads sequenced by the semi-automatic MS/MS method. c) 58 automatically sorted hit beads sequenced by the semi-automatic MS/MS method.

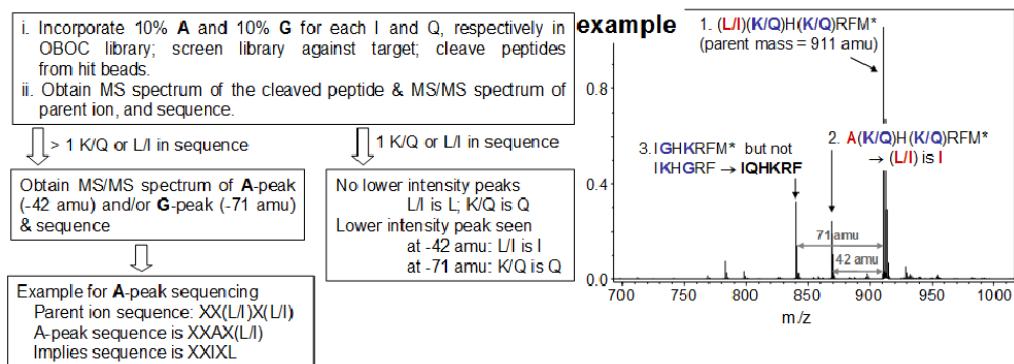
a)



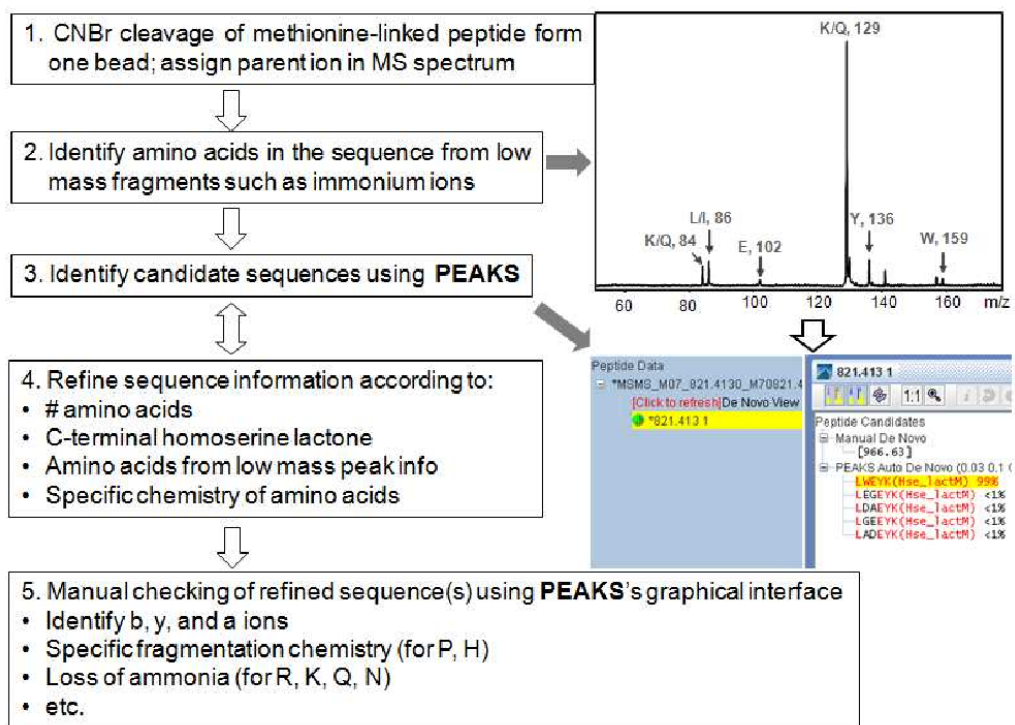
b)

**Figure 2.**

a) Structure of the biotinylated biligand (blue: 1° ligand, green: 2° ligand). b) Dot blot experiments by the biligand (the amount of bCAII: 20  $\mu$ g to 2 ng, biotinylated bCAII antibody from Abcam).

**Scheme 1.**

Procedure of differentiation of isobaric amino acids and one example (Supporting Information, Figure S9 for MS/MS data and sequencing).

**Scheme 2.**

Procedure for the semi-automatic sequencing of a peptide cleaved from a single bead through the use of MALDI-TOF/TOF measurements in combination with PEAKS *de novo* sequencing software.

**Table 1**

Peptide sequencing by the semi-automatic method in combination of PEAKS software. [( ) = sequence with the highest score just by PEAKS]. The chemical strategy for differentiating between the isobaric amino acids k/q and l/i was not included here.

	Observed mass	Sequences by MALDI-TOF/TOF	Calculated mass	Sequences by Edman degradation
1	676	qvaylm* (qvaylm*)	676	kvyim
2	644	wldagm* (adlwgm*)	644	wldagm
3	743	drqnqm* (rdqnqm*)	743	drqnkm
4	730	hpkhem* (htfnem*)	730	hpkhem
5	614	pktegm* (kptegm*)	614	pktegm
6	821	lweykm* (qwnykm*)	821	lweykm
7	672	lstRim* (lstrvm*)	672	lstrim
8	873	hyewrm* (hyerwm*)	873	hyewrm
9	612	tpklam* (talkpm*)	612	tpklam
10	676	gwadem* (wgadem*)	676	gwadem
11	640	pddpnm* (vldpnm*)	640	pddpnm

**Table 2**

The conditions of the screens along with the relevant OBOC library.

Screen	[bCAII]	Time; Temp	OBOC library (all methionine linked)	%Hits; # Hits sequenced
1gen_1°L	10 nM bCAII	15 hr; 25°C	5-mer library comprehensive in all d- stereoisomers of the natural amino acids except m,c.	121 hits; ~ 0.02%
2gen_1°L	10 nM bCAII	15 hr; 25°C	6-mer library comprehensive in r, y, k, q, h, f, w, s, l, i, v.	60 hits; ~ 0.02%
1gen_2°L	1 nM bCAII + 20 μM lhryw with 4- azidobutanoic acid coupled at N-terminus	18 hr; 25°C	6-mer library comprehensive in all d- stereoisomers of the natural amino acids except m,c; 5- pentynoic acid coupled at N-terminus	146 hits; ~ 0.024%
2gen_2oL	1 nM bCAII + 20 μM lhryw with 4-azidobutanoic acid coupled at N- terminus	18 hr; 25°C	P-×6×5×4×3×2×1-m (100%)-TentaGel (P = 5-pentynoic acid; ×6 = v,i,l,f,k,t,w,r; ×5 = f,l,r,k,v,i,n,a; ×4 = f,y,l,k,r,v,i,w; ×3 = f,w,l,k,v,r,y,a; ×2 = k,r,l,f,v,w,s,p; ×1 = k,r,w,f,s,y,l,i.	58 hits; ~ 0.01%