

DIFFERENTIATION OF ISOBARIC AMINO ACIDS AND OTHER SPECIES

FIELD OF INVENTION

This invention relates generally to differentiating isobaric species. In some
5 embodiments, isobaric amino acids are differentiated.

BACKGROUND OF INVENTION

Split-and-mix synthesis approaches have been used to produce peptide libraries on
small beads, with each bead containing an individual peptide molecule. Such libraries are
referred to as one-bead-one-compound (OBOC) libraries. A common use of OBOC
10 libraries is to identify molecules from the libraries that perform some function of interest.
As an example, an OBOC library may be used to identify a molecule (i.e., a peptide) that
binds to a particular protein by screening the library for beads that are associated with the
protein ("hit" beads). The hit beads can be separated from the rest of the library, and the
identity of the peptide on a particular hit bead can be determined using a peptide
15 sequencing strategy.

Peptide sequencing using mass spectrometry is complicated by the existence of
amino acids with similar masses (e.g., isobaric amino acids) that cannot be differentiated
by mass spectrometry. Thus, a need exists for methods for differentiating isobaric amino
acids.

SUMMARY OF INVENTION

This invention relates generally to methods for differentiating isobaric species. In
some embodiments, isobaric amino acids are differentiated. The subject matter of the
present invention involves, in some cases, interrelated products, alternative solutions to a
particular problem, and/or a plurality of different uses of one or more systems and/or
25 articles.

In one aspect, the invention is directed to a method. In one set of embodiments, the
method includes an act of growing amino acid sequences on a surface. In some cases, at
least one of the amino acids within the amino acid sequence is an isobaric amino acid. In
certain embodiments, when the isobaric amino acid is added to a medium exposed to the
30 sequences for addition thereto, the medium containing the isobaric amino acid further
contains an entity, incorporatable in an amino acid sequence, that has a molecular weight
different than the isobaric amino acid.

In another set of embodiments, the method includes an act of distinguishing, using mass spectrometry, a first amino acid sequence from a second amino acid sequence having the same number of amino acid residues as the first sequence. In some embodiments, the second amino acid sequence is identical to the first amino acid sequence except at one or more locations where the first amino acid sequence contains a first, isobaric amino acid. In some cases, the second amino acid sequence contains a second amino acid that has a molecular weight different than the first amino acid.

According to yet another set of embodiments, the method includes an act of exposing an amino acid sequence, attached to a surface, to a solution containing an isobaric amino acid. In some embodiments, the isobaric amino acid is a member of an isobaric group consisting of a first amino acid species and a second amino acid species. In certain instances, when the isobaric amino acid incorporated into the amino acid sequence is the first amino acid species, the solution further comprises an entity, incorporatable in an amino acid sequence, that has a molecular weight different than the isobaric amino acid. In some cases, when the isobaric amino acid is incorporated into the amino acid sequence is the second amino acid species, the solution is substantially free of the incorporatable entity.

The invention, in another aspect, is directed to a composition. In one set of embodiments, the composition includes a first amino acid sequence and a second amino acid sequence having the same number of amino acid residues as the first sequence. In some cases, the second amino acid sequence is identical to the first amino acid sequence except at one or more locations where the first amino acid sequence contains a first, isobaric amino acid. In certain instances, the second amino acid sequence contains a second amino acid that has a molecular weight different than the first amino acid.

The invention, in yet another aspect, is directed to an article. According to one set of embodiments, the article includes a surface to which is attached a first amino acid sequence and a second amino acid sequence, each via a cleavable moiety. In some embodiments, the second amino acid sequence has the same number of amino acid residues as the first sequence. In certain instances, the second amino acid sequence is identical to the first amino acid sequence except at one or more locations where the first amino acid sequence contains a first, isobaric amino acid. In some instances, the second

amino acid sequence contains a second amino acid with a molecular weight different from the first amino acid.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

BRIEF DESCRIPTION OF DRAWINGS

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

FIG. 1A shows an analysis method in accordance with an embodiment of the invention;

FIG. 1B shows an analysis method in accordance with another embodiment of the invention;

FIG. 1C shows a logic tree analysis flow chart in accordance with one embodiment of the invention;

FIG. 2 shows a sequencing flowchart in accordance with another embodiment of the invention;

FIG. 3 shows a table of hit peptides and the corresponding parent peak masses in accordance with an embodiment of the invention;

FIG. 4A shows a mass spectrum of a hit peptide containing isobaric amino acids in accordance with another embodiment of the invention;

FIG. 4B shows a MALDI-TOF MS/MS spectrum of the parent peak from a hit peptide containing isobaric amino acids in accordance with an embodiment of the invention;

FIG. 4C shows a MALDI-TOF MS/MS spectrum of a mass peak 71 Da less than the parent peak in accordance with another embodiment of the invention;

FIG. 5 shows a table of hit sequences generated from the screening of an OBOC hexamer peptide library with bCAII in accordance with one embodiment of the invention;

FIG. 6 shows a mass spectrum of a hit peptide containing isobaric amino acids in accordance with another embodiment of the invention;

FIG. 7 shows HPLC traces from Edman degradation sequencing of a peptide in accordance with an embodiment of the invention; and

FIG. 8 shows an SPR sensorgram for the binding of HIRYKR-CONH₂ to bCAII in accordance with an embodiment of the invention.

DETAILED DESCRIPTION

Techniques for differentiating isobaric species are described. In one aspect, an isobaric species is substituted with a tagging species that can be identified using mass spectrometry. The isobaric species may be a subunit of a first polymer having a defined sequence, e.g., the isobaric species may be an amino acid in a protein or a peptide sequence. A tagging species may be substituted for the isobaric species in a second polymer having an otherwise identical sequence as the first polymer. For instance, the second polymer may have the same number of sequences as the first polymer, and substantially the same sequence of subunits, with a few exceptions such as the tagging species for the isobaric species. In some embodiments, the first polymer and the second polymer are prepared essentially simultaneously in the same reaction vessel. A polymer (such as a protein) having a defined subunit sequence and containing an isobaric species or a tagging species may be analyzed by mass spectrometry to determine the sequence of the polymer.

In one aspect, the invention allows for the determination of unknown entities, such as proteins or peptides. For example, the unknown entities may be peptide sequences, some of which interact with a protein target of interest and some of which do not. The unknown entities may be exposed to a protein target of interest and by separating the

unknown entities that bind to or otherwise interact with the protein target of interest from sequences that do not and analyzing the binding or interacting unknown entities (e.g., using mass spectrometry), information regarding the unknown entity may be obtained. However, in some cases, some of the unknown entities may contain isobaric species, such as isobaric amino acids, which cannot be adequately distinguished using certain techniques, such as certain mass spectrometry techniques.

Isobaric species are species having different molecular structures, but having a mass that cannot be distinguished. For instance, the isobaric species may be structural isomers of each other, or contain different atoms that nevertheless sum to the same atomic mass. An isobaric group member refers to a composition that cannot be distinguished from other members of the isobaric group (e.g., two amino acids having the same mass). Examples of isobaric groups include leucine/isoleucine and glutamine/lysine. An isobaric group member may be any type of composition. For example, the isobaric group member may be a single molecule (i.e., an amino acid, a nucleoside, a sugar, etc.) or a group of molecules (i.e., a dimer, trimer, tetramer, etc.).

In some cases, isobaric species may not be distinguishable from other members of the isobaric group after undergoing a chemical transformation. That is, the isobaric species may not be sufficiently distinguishable after undergoing a chemical transformation. For example, two members of an isobaric group may or may not be isobaric (i.e., distinguishable by mass) when incorporated into a polymer; however, upon analysis by a technique such as mass spectrometry, one or more of the isobaric group members may undergo a chemical transformation (i.e., fragmentation, ionization, etc.) such that they cannot be distinguished from one another on the basis of mass. It should be understood that in some embodiments, some species that may be isobaric when analyzed by a first technique, may not necessarily be isobaric when analyzed by a second technique.

A non-limiting example will now be described. FIG. 1A shows a particle 100 having a probe entity 110 and a tagging entity 130 attached to the surface of the particle via a linker 150. It should be understood that although particles are described in FIG. 1A, this is by way of example only, and in other embodiments, other systems may be used, e.g., the probe entity and tagging entity may be in solution, attached to a planar substrate, or the like. Probe entity 110 comprises a sequence of subunits chosen from a pool of

subunits and includes an isobaric group member 120 (i.e., isobaric group member 120 has the same mass as one or more other subunits in the pool of subunits). Tagging entity 130 comprises an identical sequence of subunits as probe entity 110, except that the isobaric group member is substituted with a tagging species 140 having a mass different than that of the isobaric group member 120. The tagging species codes for the presence of the isobaric group member 120 in the probe entity 110. The probe entity and tagging entity may be cleaved from the particle, e.g., via linker 150, as described in more detail below, to form a mixture that is subsequently analyzed by mass spectrometry (or other techniques) to produce a mass spectrum 180. Since the probe entity and tagging entity differ only by the presence of the isobaric group member and tagging species, respectively, the mass spectrum of the mixture shows a parent peak 182 for the probe entity that differs in mass from the parent peak 184 for the tagging entity by an amount generally equal to the difference in mass between the isobaric group member and tagging species. Thus, in this example, the presence of two peaks having this mass difference indicates the presence of an isobaric group member in the probe entity. The parent peak 182 for the probe entity is then fragmented and mass analyzed to determine the sequence of the probe entity (mass spectrum 181). Spectrum 181 shows the fragments of peptide 160 in order of increasing mass from left to right. Generally, the difference in mass between the peaks indicates the mass of the subunits in the probe entity and allows the subunits to be identified, e.g., using known or readily available techniques.

In another non-limiting example, a probe entity containing two or more isobaric group members can be sequenced using tandem mass spectrometry. Referring now to FIG. 1B, a particle 200 is shown having a probe entity 210 and a tagging entity 230 attached to the surface of the particle via a linker 250. Probe entity 210 comprises a sequence of subunits chosen from a pool of subunits and includes two isobaric group members 216 and 222 having the same mass but differing in chemical structure. Tagging entity 230 comprises an identical sequence of subunits except that the isobaric group member 216 is substituted with a tagging species 240 having a mass different than that of the isobaric group members 216 and 222. The tagging species codes for the presence of the isobaric group member 216 in the probe entity 210. The probe entity and tagging entity may be cleaved from the particle to form a mixture that is subsequently analyzed by

mass spectrometry to produce a mass spectrum 280. The mass spectrum of the mixture shows a parent peak 282 for the probe entity that differs in mass from the parent peak 284 for the tagging entity by an amount equal to the difference in mass between the isobaric group member and tagging species. Sequencing of the parent peak 282 (i.e., probe entity 5 210) by fragmentation (spectrum 281) indicates the positions of the two isobaric group members but does not distinguish them from each other. Sequencing of the parent peak 284 (i.e., tagging entity 230) by fragmentation (spectrum 283) reveals that position 2 has a different mass than the isobaric group members indicating that isobaric group member 216 is in position 2 and isobaric group member 222 is in position 4.

10 Accordingly, based on the differences in mass spectrometry data and knowledge of the particles and their probe and tagging entities, a polymer, such as a peptide or protein, may be analyzed using mass spectroscopy or other techniques to determine sequences containing, or suspected of containing, isobaric species such as isobaric amino acids. Thus, after a subset of particles and their probes are collected, e.g., upon interaction or 15 reaction with an unknown entity, the probes may be analyzed to determine the sequences of the probes, and thus, information (e.g., sequence information) regarding the unknown entity.

Thus, in some instances, particles that are at least suspected of containing isobaric species may be analyzed as follows. One non-limiting example of a suitable technique 20 with reference to a logic tree analysis flow chart used to analyze mass spectrometry data is provided in FIG. 1C. In this example, the probe entity and the tagging entity are peptides, but it should be understood that other probe entities and tagging entities may also be analyzed in analogous fashion. The flow chart of this example begins with the acquisition of a mass spectrum (MS spectrum) for an isolated peptide sample. The mass spectrum is 25 analyzed for the parent peak of the probe entity and for any additional peaks (i.e., a tagging entity parent peak). In some embodiments, the probe entities and tagging entities are members of a library of entities for which all possible masses are known. Thus, the parent peak of the probe entity can be identified, for example, by comparison to a list of possible masses for probe entities. Likewise the parent peak of a tagging entity can also be 30 identified. If no peak is present in addition to the probe entity parent peak, then the parent peak may be sequenced to determine the identity of the parent peak (i.e., the parent peak of

the peptide may be fragmented to determine the amino acid sequence). Since no tagging entity was present, this indicates that the probe entity either (1) contained no isobaric amino acids or (2) contained an isobaric amino acid that was not coded by a tagging species. For example, for the isobaric amino acid pair leucine and isoleucine, leucine may be coded by alanine and isoleucine may not be coded. Thus, if the mass spectrum of a peptide sample does not have a second parent peak corresponding to a tagging entity, then the probe entity contains isoleucine or does not contain an isobaric amino acid. Conversely, if the mass spectrum reveals a second parent peak corresponding to a tagging entity, then the probe entity contains leucine.

Still referring to FIG. 1C, if a second parent peak in addition to the probing entity parent peak is present, this indicates the presence of a tagging species in the second parent peak corresponding to an isobaric amino acid in the probe entity. The first parent peak is then sequenced. If more than one isobaric amino acid is not present in the probe entity, then the sequence of the probe entity can be obtained solely from sequencing the parent mass of the probe entity. However, if more than one isobaric amino acid is present in the probe entity, then the tagging entity is sequenced to determine the position of the tagging species (i.e., the tagged amino acid). In some cases, this may be repeated if yet more isobaric amino acids are present in the probe entity.

Any suitable techniques may be used to determine the mass of an entity, for example, mass spectrometry, mass sedimentation, or the like. Any mass spectrometry technique may be used in embodiments involving mass spectrometry. For example, techniques such as MALDI, ESI, FAB, etc. may be used to ionize a species. The mass of an ionized analyte may be determined using analyzers such as time-of-flight (TOF), quadrupole, Fourier transform ion cyclotron resonance, orbitrap, sector field, etc. In some embodiments, an ion may be fragmented to produce secondary ions. The secondary ions may also be fragmented, for example using a technique such as tandem mass spectrometry. For instance, methods such as collision-induced dissociation, electron capture dissociation, electron transfer dissociation, infrared multiphoton dissociation, blackbody infrared radiative dissociation, and the like may be used to fragment ions.

A probe entity may be any composition that can contain an isobaric group member. For example, a probe entity may contain subunits. In some embodiments, the probe entity

may be a peptide and the subunits may be amino acid residues of the peptide. The peptide may comprise any amino acid, including alanine, cysteine, aspartic acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, arparagine, pyrrollysine, proline, glutamine, arginine, serine, threonine, selenocysteine, valine, tryptophan, tyrosine, etc. Other amino acids, including non-naturally-occurring amino acids may also be used in some cases, and in some embodiments the non-naturally-occurring amino acids may also be isobaric, e.g., with other non-naturally-occurring amino acids and/or with naturally occurring amino acids. Thus, the peptide may be modified using techniques known in the art. For example, the peptide may be modified at the N-terminus and/or the C-terminus. The side groups of the amino acids may also be modified. In some cases, the backbone of the peptide may be modified. In some embodiments, the peptide may be linked to another polymer, for example a biological polymer such as a nucleic acid and/or a polysaccharide. The peptide may also be linked to other polymers, for example polyesters (e.g., polylactide, polyglycolide, PLGA, polycaprolactone, etc.), polyamides, polyurethanes, polyanhydrides, etc. and combinations thereof.

An “amino acid” is given its ordinary meaning as used in the field of biochemistry. An isolated amino acid typically, but not always (for example, as in the case of proline) has a general structure $\text{NH}_2\text{-CHR-COOH}$. R may be any suitable moiety; for example, R may be a hydrogen atom, a methyl group, or an isopropyl group. A series of isolated amino acids may be connected to form a peptide or a protein by reaction of the -NH_2 of one amino acid with the -COOH of another amino acid to form a peptide bond (-CO-NH-). In such cases, each of the R groups on the peptide or protein can be referred to as an amino acid residue. The amino acid may be one of the 20 amino acids commonly found in nature (the “natural amino acids”), or an unnatural amino acid, i.e., an amino acid that is not one of the natural amino acids. Non-limiting examples of unnatural amino acids include alloisoleucine, allothreonine, homophenylalanine, homoserine, homocysteine, 5-hydroxylysine, 4-hydroxyproline, 4-carboxyglutamic acid, cysteic acid, cyclohexylalanine, ethylglycine, norleucine, norvaline, 3-aminobutyric acid, beta-amino acids (e.g., beta-alanine), *N*-methylated amino acids such as *N*-methylglycine, *N*-methylalanine, *N*-methylvaline, *N*-methyllleucine, *N*-methylisoleucine, *N*-methylnorleucin, *N*-methyl-2-

aminobutyric acid, *N*-methyl-2-aminopentanoic acid, etc., as well as the D-isomers of the natural amino acids.

The tagging entity may be identical to a probe entity, or in other instances, the tagging entity may differ from the probe entity in one or more places. For example, a
5 tagging entity may comprise the same subunits arranged in the same order but differing from a probe entity in that one or more of the corresponding subunits is substituted with a tagging species. For example, a peptidic probe entity may contain a first amino acid from an isobaric group containing the first amino acid and a second amino acid, and a tagging
10 species having a distinguishable mass from the isobaric group members may be substituted for the first amino acid (i.e., the tagging species “codes” for the first amino acid). In some embodiments, a first tagging species may be substituted for a first amino acid and a second tagging species may be substituted for a second amino acid.

A tagging species may be any composition that can distinguish a member of an isobaric group (e.g., by a difference in mass). In some embodiments, the tagging species
15 may be a composition incorporatable into a polymer. For example, the tagging species may be an amino acid, a nucleoside, a saccharide, and the like. In some embodiments, a tagging species may be a composition that may be substituted for a subunit in a polymer but does not belong to the same chemical group as the other subunits in the polymer. For example, the tagging species may have an amino group and a carboxyl group such that it
20 can be incorporated into a peptide but may not be an amino acid.

In some cases, a probe entity and a tagging entity may be attached to a surface, e.g., the surface of a particle, the surface of a planar substrate, etc. The entities may be attached by any suitable method. For example, the surface may have a functional group attached thereto to which a probe entity and/or tagging entity may be connected. An entity
25 may be connected to a surface using a covalent bond, an ionic bond, van der Waals forces, hydrophobic interactions, etc. Chemical bonds such as esters, amides, ethers, anhydrides, urethanes, etc. may be used to attach an entity to a surface. In some embodiments, a metal-sulfur bond (i.e., gold-sulfur) may be used to attach an entity to a surface.

In some embodiments, a cleavable linker may be used to connect an entity to a
30 surface (i.e., a particle). Such linkers are known to those skilled in the art, for example in solid-phase peptide synthesis and solid-phase oligonucleotide synthesis. In some

instances, a methionine residue may be used as a linker. A methionine linker may be cleaved by a reagent such as CNBr, which cleaves peptide bonds at the C-terminus of methionine residues. Examples of other types of linkers that may be used include acid-cleavable linkers, base-cleavable linkers, photo-cleavable linkers, and redox-cleavable
5 linkers such as those mediated by periodate, 2,3-Dichloro-5,6-Dicyanobenzoquinone (DDQ), cerium (IV) ammonium nitrate (CAN), etc. The cleavable linker may be susceptible to cleavage under conditions that are essentially benign to the entity. In other embodiments, the entity may be cleaved at one or more places in addition to being cleaved at the linker.

10 In some aspects, a probe entity and a tagging entity may be synthesized essentially simultaneously. For example, in the instance of peptide entities, solid-phase synthesis techniques known to those skilled in the art may be used to synthesize a probe entity and a tagging entity essentially simultaneously by sequentially adding amino acid residues to a growing peptide chain in the same reaction vessel. When it is desired to add a specific
15 non-isobaric amino acid to a peptide chain, the peptide chain may be exposed to a reaction mixture containing essentially one non-isobaric amino acid. By contrast, to add an isobaric amino acid to a peptide chain that has been assigned a tagging species (i.e., another amino acid having a different mass), the peptide may be exposed to a reaction mixture containing a combination of the isobaric amino acid and the tagging species. The
20 ratio of isobaric amino acid to tagging species may be any suitable ratio, e.g., less than about 100:1, less than about 10:1, less than about 5:1, less than about 2:1, less than about 1:1, less than about 1:2, less than about 1:5, less than about 1:10, less than about 1:100, etc.

In some instances, a library of probe entities and tagging entities may be prepared.
25 In some instances the library of entities may be attached to particles (i.e., beads). For example, a one-bead-one-compound (OBOC) library may be generated in one embodiment. An OBOC library generally contains a plurality of beads where each bead is attached to one or more copies of a single compound (i.e., a probe entity). For example, a first bead may be attached to a first peptide having a first sequence and a second bead may
30 be attached to a second peptide having a second sequence, which is typically different from the first sequence. In this example, it should be understood that each of the beads

also may be attached to a tagging entity corresponding to the probe entity attached to the bead. A peptide OBOC library may be generated according to split-and-pool synthesis techniques known to those skilled in the art.

In some embodiments, it may be desired to know which, if any, of the library members interact with a target of interest (i.e., a protein). The library members may be contacted with a suitable target of interest, and members that bind to the target of interest may be enriched using techniques known in the art, such as fluorescence sorting. This process may be repeated one or more times in order to narrow the number of "hits" (i.e., library members that bind to the target). The isolated hits may then be subjected to the methods described above to determine the identity of the hits. Targets of interest include biopolymers such as proteins, nucleic acids, and polysaccharides, and specific compositions such as

The following examples are provided for illustration purposes and are not intended to be limiting.

15

EXAMPLE 1

This example demonstrates synthesis of an OBOC hexamer peptide library. "Split-and-mix" synthesis of a peptide library was performed using an automatic synthesizer (Titan 357, AAPPTEC). TentaGel S Amino beads (2.2 g, 90 micron diameter, loading of NH_2 : 0.24 mmol/g) were swelled in NMP (27 ml) for 2 hr in a Collective Vessel (CV). For incorporation of 100% methionine as a linker, fmoc-Met (2 equiv, 0.2 M solution in NMP) was added to the CV, as well as TBTU (2 equiv, 0.2 M solution in NMP), and DIEA (5 equiv, 0.5 M solution in NMP), after draining the solvent. The resulting mixture was vortexed for 30 min. The solution was drained and the coupling step was repeated by using fresh reagent solutions. The resulting beads were thoroughly washed by NMP (27 ml x 4). Next, 20% piperidine in NMP (27 ml) was added and the CV was vortexed for 5 min. The liquid was drained and a fresh solution of 20% piperidine in NMP (27 ml) was added, and the CV was vortexed for another 15 min. The resulting beads were thoroughly washed by NMP (27 ml x 4) and DCM (27 ml x 4), followed by distribution equally into 11 Reaction Vessels (RV). One of the 11 selected Fmoc-protected D-amino acids as diversity elements (F, H, K, L, R, S, V, W, Y, 9:1 mixture of Q and G, and 9:1 mixture of I and A, 3 equiv each), TBTU (3 equiv) and DIEA (7.5 equiv) were added to each RV. The

RV was then vortexed for 30 min. After draining the solution, the coupling step was repeated. The resulting beads in each RV were washed by NMP (2 ml $\times$ 4). Again, 20% piperidine in NMP (2 ml $\times$ 4) was added to each RV, which was vortexed for 15 min. The liquid was drained and a fresh solution of 20% Piperidine in NMP (2 ml $\times$ 4) was added with vortexing for another 30 min. Beads in each RV were thoroughly washed by NMP (2 ml $\times$ 4) and DCM (2 ml $\times$ 4), which were combined into the CV. The overall split, coupling, deprotection, and mix processes were repeated until the beads appended a hexamer, excluding the initial methionine. The beads were transferred to a 50 ml reactor equipped with a filter. The protective groups in the residues were removed by shaking in TFA-water-TIS (27 ml, 94:3:3, v/v) for 2 h. The liquid was drained and the resulting beads were thoroughly washed by DCM (27 ml $\times$ 3), methanol (27 ml $\times$ 3), water (27 ml $\times$ 3), methanol (27 ml $\times$ 3), DCM (27 ml $\times$ 3), and diethylether (27 ml), successively, and then dried under reduced pressure for 24 h.

EXAMPLE 2

This example demonstrates library screening. Alexa Fluor® 647 protein labeling kit (A20173, Invitrogen) was chosen as the reactive dye for labeling bovine carbonic anhydrase (bCAII) following the supplier's protocol. Briefly, 2 mg/mL of bCAII in 0.1 M sodium bicarbonate (pH $\sim$ 8.3) was transferred 0.5 mL of bCAII solution into the vial of the reactive dye, and the vial was capped and inverted several times to dissolve the dye completely. The reaction mixture was stirred for 1 hr at room temperature under dark conditions. The Alexa Fluor® 647 labeled bCAII (bCAII-A647) was purified from the mixture using the size exclusion purification resin in the kit. Purified and labeled bCAII-A647 was characterized by NanoDrop (Thermo Scientific) and gel documentation (Typhoon) after SDS-PAGE. Dried library resin (200 mg) was transferred into an 8 mL Alltech vessel and pre-incubated in blocking solution, 0.05% NaN₃, 0.1% Tween 20 and 0.1% BSA in PBS buffer (pH 7.4) for 1 hr on a 360° shaker at room temperature. The buffer solution was drained by vacuum and then 5 mL of 50 nM dye labeled bCAII diluted in blocking solution was added to the swollen resin. The resulting mixture was incubated at on a 360° rotating thermostat shaker for 15 hr at 25 °C. The liquid was drained by vacuum and non-specifically bound proteins were eliminated by washing 3 times with blocking solution and 7 times with 0.1% Tween 20 in PBS, sequentially. After stringent

washing, 200 mg of the assayed library resin was transferred into sample vessel of COPAS Plus (Union Biometrica) and diluted with 200 mL of 0.1% Tween 20 in PBS buffer.

Hit beads were sorted into cone-shaped wells of a 96 well plate using a COPAS Plus. Gating and sorting regions were defined for sorting beads on the COPAS Plus. The gating region was defined for beads based on their time-of-flight (TOF) and red fluorescence, and the sorting region was chosen for selecting beads based on their time-of-flight (TOF) and uniformity red fluorescence after incubation with bCAII-A647. A two step sorting strategy was used for rapid and robust sorting. The first sorting was to purify beads in a high concentration regime (>1000 beads/mL). During the first stage sorting, 200 mg (c.a. 300,000 beads) of assayed library beads in PBS were sorted using deionized water as the sheath solution. The beads in the sample cup were passed through the flow cell and focused hydrodynamically at a rate of >100 objects/s. The TOF, red fluorescence, and red fluorescence peak height of the beads were detected using a red diode laser ($\lambda=635$ nm). Due to the auto-fluorescence of the beads, the argon ion laser was intentionally turned off to minimize bleed-through effect. Less than 5,000 beads (less than 1.7% of the total beads) were collected in the first sorting. For the second step sorting, the sorted beads from the first step were thoroughly washed with deionized water and transferred to a sample cup in the COPAS Plus and diluted with 100 mL of deionized water. The beads in the sample cup were passed through the flow cell at a rate of <5 objects/s. Each hit bead was directly sorted into a 96 well plate with cone-shaped wells.

EXAMPLE 3

This example demonstrates MALDI mass spectrometry of a peptide. CNBr (10 microliters, 0.50 M in 0.2 N HCl solution) was added into each well of the 96 well plate in Example 2, and the plate was purged with argon and microwaved for 1 min. The resulting solution was concentrated under centrifugal vacuum for 2 h.

To each vial or well was added CHCA (10 microliters, 0.5% solution in acetonitrile/water (70:30)) and then acetonitrile/water (10 microliters, 70:30 containing 0.1% trifluoroacetic acid (v/v)). The sample was centrifuged for 2 min and 2 microliters were withdrawn and spotted onto a 384-well MALDI-MS plate, which was allowed to stand for 15 min to allow evaporation of the sample.

EXAMPLE 4

This example demonstrates sequencing of peptides containing isobaric amino acids using mass spectrometry. The hit peptides from the screening of the OBOC hexamer peptide library against bCAII in Example 2 were successfully sequenced by using the isobaric differentiation approach. As indicated in Example 1, glycine (G) was used as a tag for Q to differentiate between K and Q, and alanine (A) was used as a tag for I to differentiate between I and L. FIG. 2 illustrates the scheme used to obtain the full peptide sequence. An MS spectrum obtained using MALDI-TOF/TOF was first examined to identify any additional mass peaks besides the parent mass peak. In this example, a peak 42 amu smaller than the parent peak indicates that I is present in the sequence, and a peak 71 amu smaller than the parent mass peak indicates that Q is present in the sequence. The sequence of the parent peak is obtained by analysis of its MS/MS spectrum. If the sequence contains only one L/I or one K/Q, then the L/I will be I, while K/Q will be Q. However, if the parent peptide contains more than one K/Q or one L/I, then the additional mass peaks are subjected to MS/MS fragmentation to identify the position of the isobaric amino acid I or Q.

The sequencing scheme of FIG. 2 was applied to the hit peptides from screening of the hexamer library against bCAII as shown in FIG. 3. The hit peptides refer to the peptides that have binding effect with the targeted protein. The peptides from 59 hit beads were first cleaved from the bead before being subjected to mass analysis.

FIGs. 4A-4C illustrate an example of the sequencing of a hit peptide containing isobaric amino acids. In FIG. 4A, the MS spectrum of a released peptide from a single bead showed two additional small peaks in addition to the parent peak. The parent peak with the highest intensity had an m/z value of 911. The other two smaller peaks had a difference of m/z -42 amu and -71 amu from the parent mass. As indicated in the analysis shown in FIG. 2, the presence of the peak with the m/z value of 869 signifies that I is present in the sequence, while the presence of the peak with the m/z value of 840 indicates that Q is present in the sequence. From the MS/MS spectrum of FIG. 4B, the sequence of the parent peptide could be determined by *de novo* sequencing. The peptide sequence could be determined from either a top-down or a bottom-up approach by finding the b- and the y-type ions respectively yielding (I/L)(Q/K)H(Q/K)RF. Since there is I in the peptide sequence as indicated by the -42 amu peak, the peptide sequence should be

I(Q/K)H(Q/K)RF. The -42 amu mass peak does not have to undergo MS/MS fragmentation as there is only one amino acid from the I/L isobaric pair present in the parent peptide sequence. Because there is no -142 amu peak for IGHGRF, the peptide should have one Q and one K. In order to identify the position of Q in the sequence, the mass peak of 840 was subjected to MS/MS analysis. FIG. 4C shows sequencing from the MS/MS spectrum of the -71 amu peak ($m/z = 840$). The presence of G in the peptide sequence confirms the presence of Q in that position. Consequently, the final sequence should be IQHKRF (SEQ ID NO: 1).

EXAMPLE 5

This example demonstrates sequencing of peptides from a 10% cleavable OBOC peptide library, where the peptides contain isobaric amino acids. Hit peptides obtained from screening an OBOC hexamer peptide library against bCAII where only 10% of the peptides have a cleavable methionine linker group can be successfully sequenced with isobaric differentiation. Having only 10% of the peptides on a bead contain a linker group has important consequences because the linker group may influence the screening result. FIG. 5 shows 8 hit sequences from the screening.

An example of a mass spectrum of a hit peptide from the brightest bead is illustrated in FIG. 6. The sequence of the peptide (parent $m/z = 955$) as determined by MS/MS analysis was H(I/L)RY(Q/K)R. A peak at m/z 931 that is 42 amu smaller than the parent peak indicated that I was present. The sequence was therefore read as HIRY(Q/K)R. Since there was no peak 71 amu smaller than the parent peak, Q is not present. As a result, the final peptide sequence is HIRYKR (SEQ ID NO: 2).

This peptide sequence of HIRYKR can be validated by Edman degradation of the remaining 90% of the peptides on the recovered bead that do not contain a cleavable linker (FIG. 7). From FIG. 7, the type and position of the amino acid present in the peptide can be identified by HPLC. Each individual HPLC spectrum was compared to amino acid standards by matching the retention time. HPLC analysis indicated that the first amino acid in the sequence is H and that the second amino acid in the sequence is I. For the second position in the peptide, a smaller peak corresponding to A is seen in the spectrum in addition to the peak corresponding to I, since 10% A was added as a tag at the coupling

of I. Therefore, the peptide sequence generated with Edman degradation matches the sequence as obtained by MALDI-TOF/TOF.

The 'hit' peptide sequence of HIRYKR was re-synthesized so that the affinity of the peptide to the protein could be determined by surface plasmon resonance (SPR). An RV well in a Titan 357 automated synthesizer (AAPTEC) was used as a reactor. Rink Amide resins (200 mg, the loading of NH_2 : 0.29 mmol/g) were swelled to equilibrium in NMP (5 ml) for 2 h and then solvent was drained. Fmoc-d-Arg(Pbf)-OH (2 equiv, 0.2 M solution in NMP), TBTU (2 equiv, 0.2 M solution in NMP), and DIEA (5 equiv, 0.2 M solution in NMP) were added successively, and the resulting beads were vortexed for 30 min. The coupling step was repeated using fresh portions of reagent solutions after the liquid was drained. The resulting beads were thoroughly washed by NMP (3 ml x 4). Next, 20% piperidine in NMP (5 ml, v/v) was added, and the RV was vortexed for 5 min. The liquid was drained, a fresh solution of 20% piperidine in NMP (3 ml, v/v) was added, and the RV was vortexed for another 15 min. The resulting beads were thoroughly washed by NMP (3 ml x 4) and DCM (3 ml x 4). The coupling of the remaining amino acids was performed as described above using fmoc-d-Lys(Trt)-OH, fmoc-d-Tyr(t-Bu)-OH, fmoc-d-Arg(Pbf)-OH, fmoc-Ile-OH, and fmoc-d-His(Trt)-OH, sequentially. The beads were transferred to an 8 ml reactor equipped with a filter and incubated in 3 ml of trifluoroacetic acid (TFA)/water/triisopropylsilane (TIS) (94/3/3, v/v/v) at room temperature for 2 h. The cleavage solution was collected with another portion of TFA (3 ml) used to rinse the beads. The combined solution was nearly evaporated by using a centrifugal concentrator GeneVac EZ2 Plus, and diethylether (*ca.* 5 ml) was added to decant the crude peptide. The final purification was carried out using preparative HPLC to yield the HIRYKR peptide having a carboxamide C-terminus as a white solid (3.5 mg, purity: >99%).

Affinity measurements were performed using a Biacore 3000 system (GE Healthcare) equipped with research grade CM5 sensor chips (GE Healthcare). The CM5 sensor chip was preconditioned after docking and priming the instrument 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. Using a flow rate of 20 microliters/min, the sensor chip's four flow cells were activated at 25 °C with a 7 min injection of amine-coupling kit (GE Healthcare) (1:1 mixture of 0.4 M EDC and 0.1 M NHS). To create a

bCAII surface, 1 mg of bCAII was dissolved in 1 mL of 10 mM sodium acetate (pH 5.0) and injected across flow cell 2 only. Approximately 1000 response units (RU) of bCAII were immobilized onto this surface. Finally, 1.0 M ethanolamine (pH 8.5) was injected across the four flow cells for 4 min to deactivate the surfaces. After bCAII
5 immobilization, the instrument was primed three times using running buffer (1x PBS + 3% DMSO). HIRYKR peptide (26.13 micrograms) was dissolved in 18 microliters DMSO and diluted with HBS-EP buffer to produce a 50 micromolar peptide stock solution. The peptide stock solution was serially diluted by a factor of 2 to a concentration of 390 nM. The concentration series of peptide solutions were injected into flow cell 2 (or 4) for 6 min
10 using a flow rate of 30 microliters/min at 25 °C. The 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 Biaevaluation software (version 4.1, Biacore). The SPR data was plotted as a function of resonance units (RUs) versus time (a sensorgram) (FIG. 8). The curves were fitted with a 1:1 Langmuir binding curve (one analyte to one
15 ligand binding), and the dissociation constant (K_D) was determined to be in the micromolar range. This experiment further validates the sequencing result that the hit peptide HIRYKR has a significant binding affinity to the protein bCAII.

While several embodiments of the present invention have been described and
20 illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials,
25 and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described
30 herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents

thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems,
5 articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

10 The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are
15 conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A
20 and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, “or” should be understood to
25 have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer
30 to the inclusion of exactly one element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e.

“one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase “at least one,” in
5 reference to a list of one or more elements, should be understood to mean at least one
element selected from any one or more of the elements in the list of elements, but not
necessarily including at least one of each and every element specifically listed within the
list of elements and not excluding any combinations of elements in the list of elements.
This definition also allows that elements may optionally be present other than the elements
10 specifically identified within the list of elements to which the phrase “at least one” refers,
whether related or unrelated to those elements specifically identified. Thus, as a non-
limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or,
equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one,
optionally including more than one, A, with no B present (and optionally including
15 elements other than B); in another embodiment, to at least one, optionally including more
than one, B, with no A present (and optionally including elements other than A); in yet
another embodiment, to at least one, optionally including more than one, A, and at least
one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any
20 methods claimed herein that include more than one step or act, the order of the steps or
acts of the method is not necessarily limited to the order in which the steps or acts of the
method are recited.

In the claims, as well as in the specification above, all transitional phrases such as
“comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,”
25 “composed of,” and the like are to be understood to be open-ended, i.e., to mean including
but not limited to. Only the transitional phrases “consisting of” and “consisting essentially
of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the
United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

Sequence Listing
SEQUENCE LISTING

<110> Agency for Science, Technology and Research
Heath, James R
Lee, Su S
Lim, Jaehong
Cha, Junhoe
Tan, Sylvia
Yeo, Shi Y

<120> Differentiation of Isobaric Amino Acids and Other Species

<130> S1507.70018W000

<160> 2

<170> PatentIn version 3.5

<210> 1

<211> 6

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic Peptide

<400> 1

Ile Gln His Lys Arg Phe
1 5

<210> 2

<211> 6

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic Peptide

<400> 2

His Ile Arg Tyr Lys Arg
1 5

CLAIMS

1. A method, comprising:
growing amino acid sequences on a surface, wherein at least one of the amino acids within the amino acid sequence is an isobaric amino acid, and wherein, when the isobaric amino acid is added to a medium exposed to the sequences for addition thereto, the medium containing the isobaric amino acid further contains an entity, incorporatable in an amino acid sequence, that has a molecular weight different than the isobaric amino acid, wherein the amino acid sequences on the surface are formed from a plurality of amino acids that define amino acid sequences having the same length.
2. The method of claim 1, wherein at least some of the amino acid sequences are attached to the surface by a cleavable linker.
3. The method of any one of claims 1 or 2, further comprising analyzing the amino acid sequence by mass spectrometry.
4. The method of claim 3, wherein the mass spectrometry comprises tandem mass spectrometry.
5. The method of claim 3, wherein the mass spectrometry comprises ESI.
6. The method of claim 3, wherein the mass spectrometry comprises MALDI-TOF/TOF.
7. The method of any one of claims 3-6, wherein analyzing the amino acid sequences comprises sequencing the amino acid sequences.
8. The method of any one of claims 1-7, wherein isobaric amino acid is an unnatural amino acid.
9. The method of any one of claims 1-8, wherein the entity comprises an unnatural amino acid.

10. A composition, consisting of:
 - a substrate having attached thereto amino acid sequences, the amino acid sequences consisting of a first amino acid sequence and a second amino acid sequence, each defined by a plurality of amino acids forming the respective sequences, the second amino acid sequence having the same number of amino acid residues as the first sequence, wherein the second amino acid sequence is identical to the first amino acid sequence except at one or more locations where the first amino acid sequence contains a first, isobaric amino acid, and the second amino acid sequence contains at a corresponding position, a second amino acid structurally different from the first amino acid and having a molecular weight different than the first amino acid.
11. The composition of claim 10, further comprising a particle comprising the first amino acid sequence and the second amino acid sequence.
12. The composition of claim 11, wherein the first amino acid sequence and the second amino acid sequence are each attached to the particle by a cleavable linker.
13. The composition of claim 12, wherein the cleavable linker is a methionine residue.
14. The composition of any one of claims 10-13, wherein the isobaric amino acid is leucine or isoleucine.
15. The composition of any one of claims 10-14, wherein the isobaric amino acid is lysine or glutamine.
16. The composition of any one of claims 10-15, wherein the isobaric amino acid is an unnatural amino acid.
17. The composition of any one of claims 10-16, wherein the second amino acid is an unnatural amino acid.
18. A surface to which is attached amino acid sequences, the amino acid sequences consisting of a first amino acid sequence and a second amino acid

sequence, each via a cleavable moiety, the first amino acid sequence and the second amino acid sequence each defined by a plurality of amino acids forming the respective sequences, the second amino acid sequence having the same number of amino acid residues as the first sequence, wherein the second amino acid sequence is identical to the first amino acid sequence except at one or more locations where the first amino acid sequence contains a first, isobaric amino acid, and the second amino acid sequence contains at a corresponding position, a second amino acid with a molecular weight different from the first amino acid.

19. The surface of claim 18, wherein the cleavable moiety is a methionine residue.
20. The surface of any one of claims 18 or 19, wherein the isobaric amino acid is leucine or isoleucine.
21. The surface of any one of claims 18 or 19, wherein the isobaric amino acid is lysine or glutamine.
22. The surface of any one of claims 18-21, wherein the isobaric amino acid is an unnatural amino acid.
23. The surface of any one of claims 18-22, wherein the second amino acid is an unnatural amino acid.
24. A method comprising:
 - distinguishing, using mass spectrometry, a first amino acid sequence from a second amino acid sequence having the same number of amino acid residues as the first sequence, the first amino acid sequence and the second amino acid sequence each being defined by a plurality of amino acids forming the respective sequences, wherein the second amino acid sequence is identical to the first amino acid sequence except at one or more locations where the first amino acid sequence contains a first, isobaric amino acid, the second amino acid sequence contains at a corresponding position, a second amino acid that has a molecular weight different than the first amino acid, and wherein the first and second amino acid sequences are attached to a same substrate.

25. The method of claim 24, the method further comprising essentially simultaneously cleaving the first and second amino acid sequences from a common particle prior to subjecting the first and second amino acid sequences to mass spectrometry.
26. The method of claim 25, wherein the mass spectrometry comprises tandem mass spectrometry.
27. The method of claim 25, wherein the mass spectrometry comprises ESI.
28. The method of claim 25, wherein mass spectrometry comprises MALDI-TOF/TOF.
29. The method of any one of claims 24-28, wherein the second amino acid sequence has a parent peak m/z value different than the parent peak m/z value of the first amino acid sequence.
30. The method of any one of claims 24-29, wherein mass spectrometry comprises tandem mass spectrometry.
31. The method of any one of claims 24-30, wherein the isobaric amino acid is an unnatural amino acid.
32. The method of any one of claims 24-31, wherein the second amino acid is an unnatural amino acid.
33. A method, comprising:
 - exposing an amino acid sequence, attached to a surface, to a solution containing an isobaric amino acid, the isobaric amino acid being a member of an isobaric group consisting of a first amino acid species and a second amino acid species, wherein, when the isobaric amino acid incorporated into the amino acid sequence is the first amino acid species, the solution further comprises an entity, incorporatable in an amino acid sequence, that has a molecular weight different than the isobaric amino acid, and when the isobaric amino acid incorporated into the amino acid sequence is the second amino acid species, the solution is substantially free of the incorporatable entity, wherein the amino acid sequences on the surface are formed

from a plurality of amino acids that define amino acid sequences having the same length.

34. The method of claim 33, wherein at least some of the amino acid sequences are attached to the surface by a cleavable linker.
35. The method of any one of claims 33 or 34, further comprising analyzing the amino acid sequence by mass spectrometry.
36. The method of claim 35, wherein the mass spectrometry comprises tandem mass spectrometry.
37. The method of claim 35, wherein the mass spectrometry comprises ESI.
38. The method of claim 35, wherein the mass spectrometry comprises MALDI-TOF/TOF.
39. The method of any one of claims 35-38, wherein analyzing the amino acid sequences comprises sequencing the amino acid sequences.
40. The method of any one of claims 33-39, wherein isobaric amino acid is an unnatural amino acid.
41. The method of any one of claims 33-40, wherein the entity comprises an unnatural amino acid.