US 20030198940A1

(19) **United States**

(12) **Patent Application Publication**

Carlson et al.

(10) **Pub. No.: US 2003/0198940 A1**

(43) **Pub. Date: Oct. 23, 2003**

(54) PROTEIN CHARACTERIZATION SYSTEM

(75) Inventors: **Robert H. Carlson**, Union City, CA (US); **Ian E. Burbulis**, Kensington, CA (US); **Roger Brent**, Berkeley, CA (US)

Correspondence Address:
Kent S. Burningham, Esq.
TRASKBRITT, PC
Suite 300
230 South 500 East
Salt Lake City, UT 84102 (US)

(73) Assignee: **THE MOLECULAR SCIENCES INSTITUTE, INC.**

(21) Appl. No.: **10/444,422**

(22) Filed: **May 23, 2003**

Related U.S. Application Data

(62) Division of application No. 09/412,732, filed on Oct. 5, 1999, now Pat. No. 6,569,685.

Publication Classification

(51) **Int. Cl.^{7}** **C12Q 1/00**
(52) **U.S. Cl.** **435/4**

(57) ABSTRACT

A system and method for characterizing protein molecules. A protein molecule of interest is isolated from other types of protein molecules and modified into a one-dimensional structure. Each of a first type of amino acid residue of the protein molecule of interest is labeled with a first tag. Each of a second type of amino acid residue of the protein molecule is labeled with a second tag. The first and second tags impart to the protein molecule of interest a detectable set of characteristic ancillary properties that function as a fingerprint or characterization of the protein molecule of interest reflective of the physical structure of the protein molecule of interest as defined by the amino acid sequence of the protein molecule of interest. A library listing of the characterizations corresponding to protein molecules facilitates identification of protein molecules of interest.

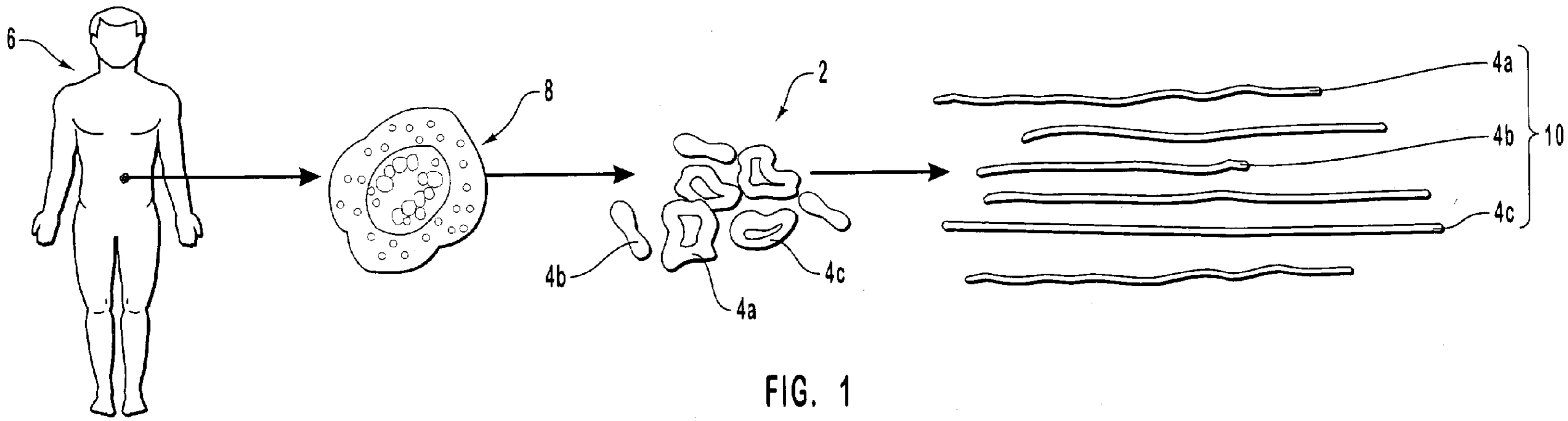

FIG. 1

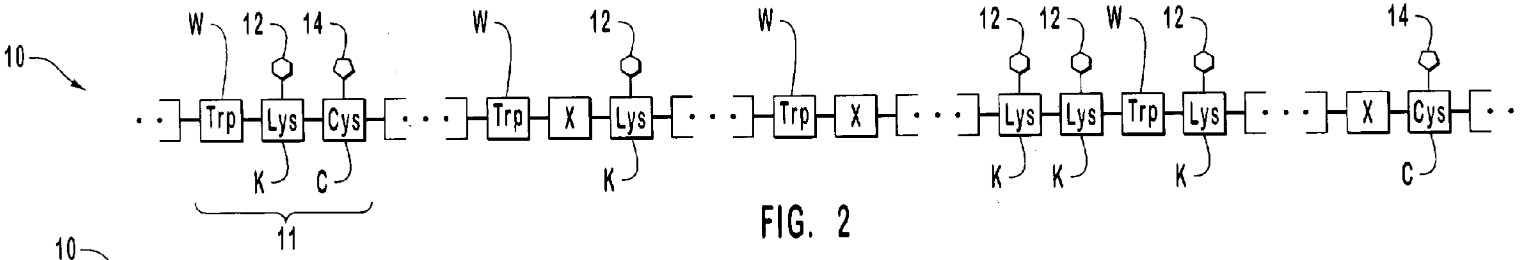

FIG. 2

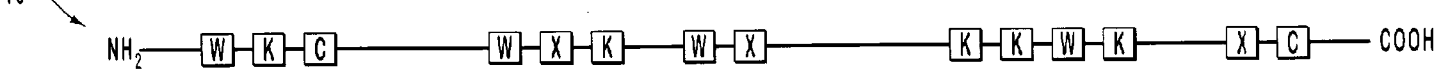

FIG. 3

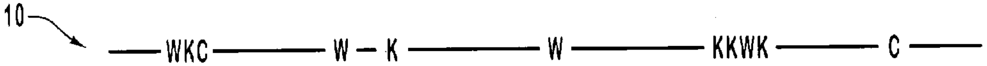

FIG. 4

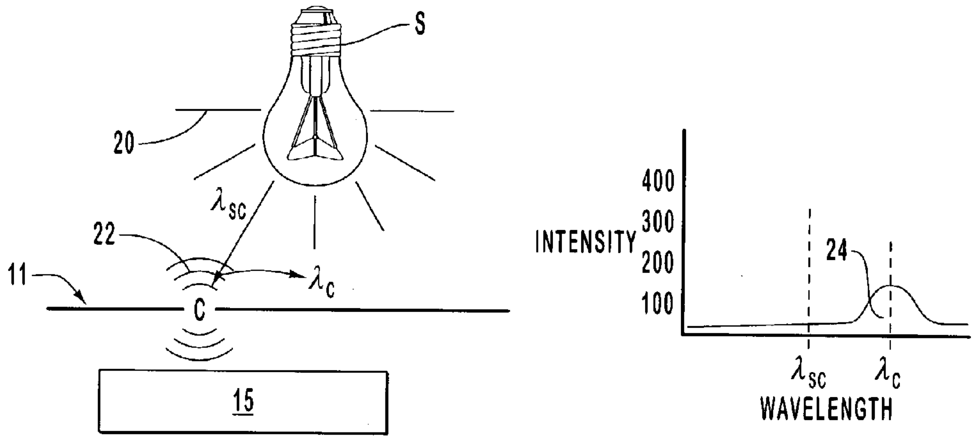

FIG. 5

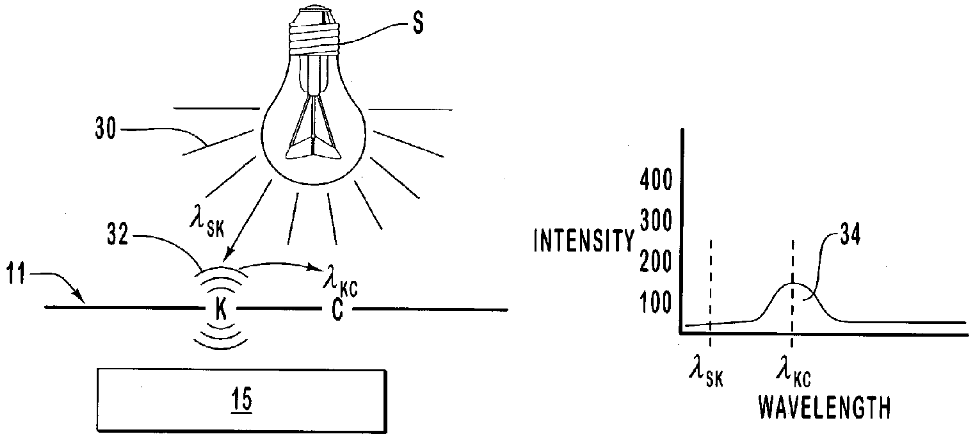

FIG. 6A

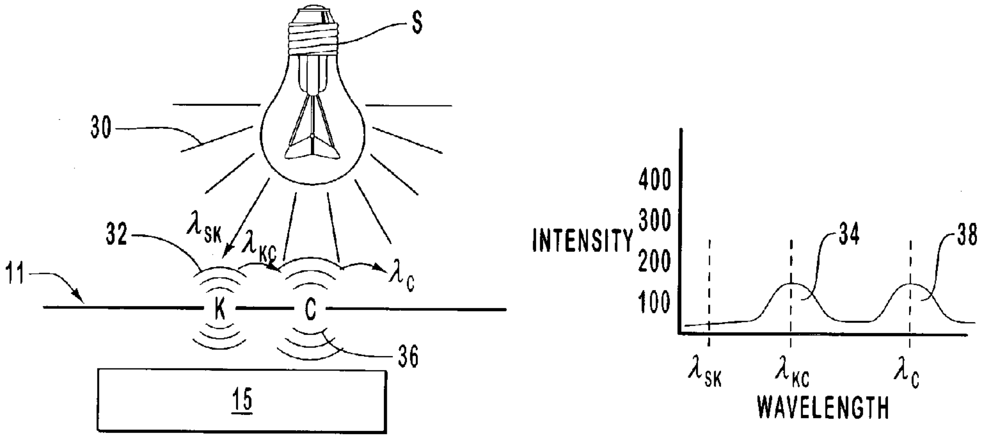

FIG. 6B

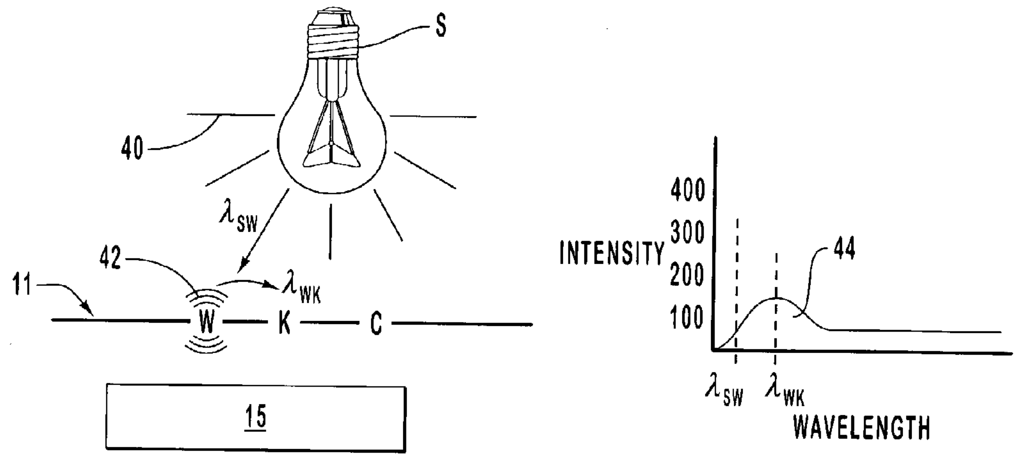

FIG. 7A

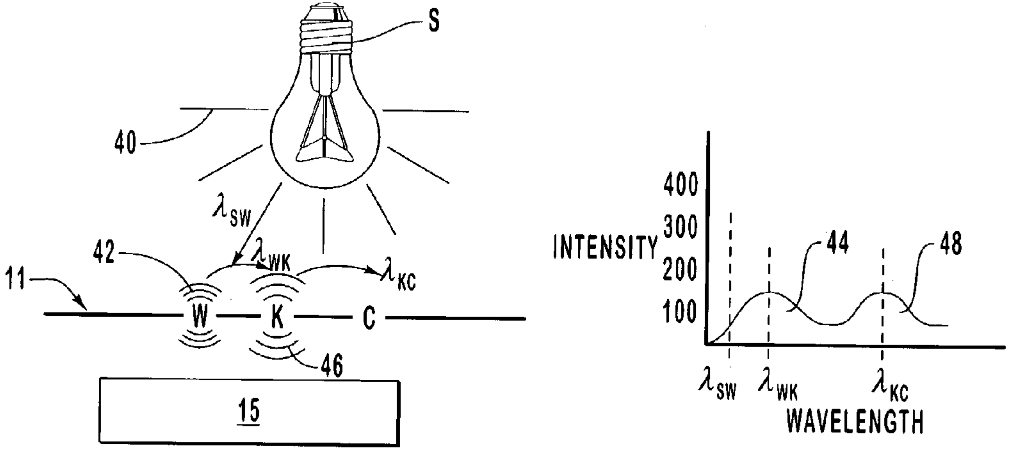

FIG. 7B

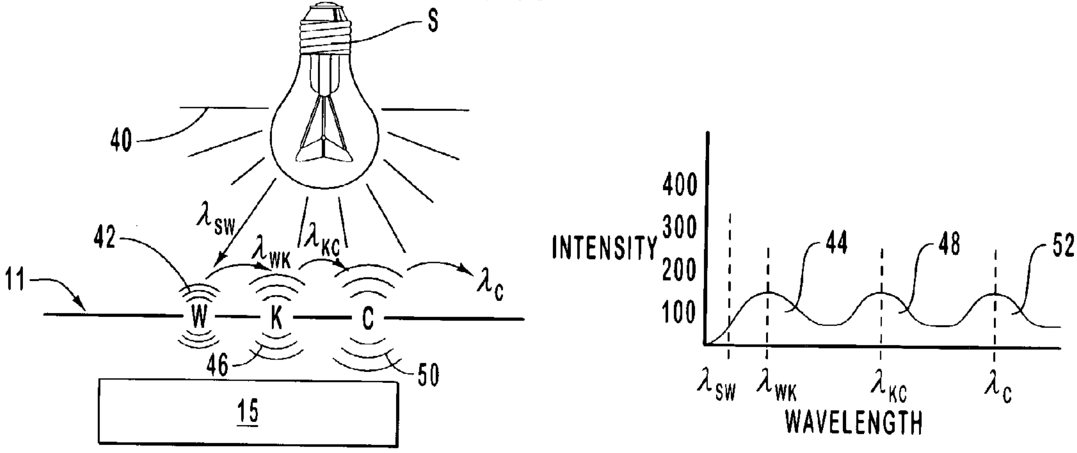

FIG. 7C

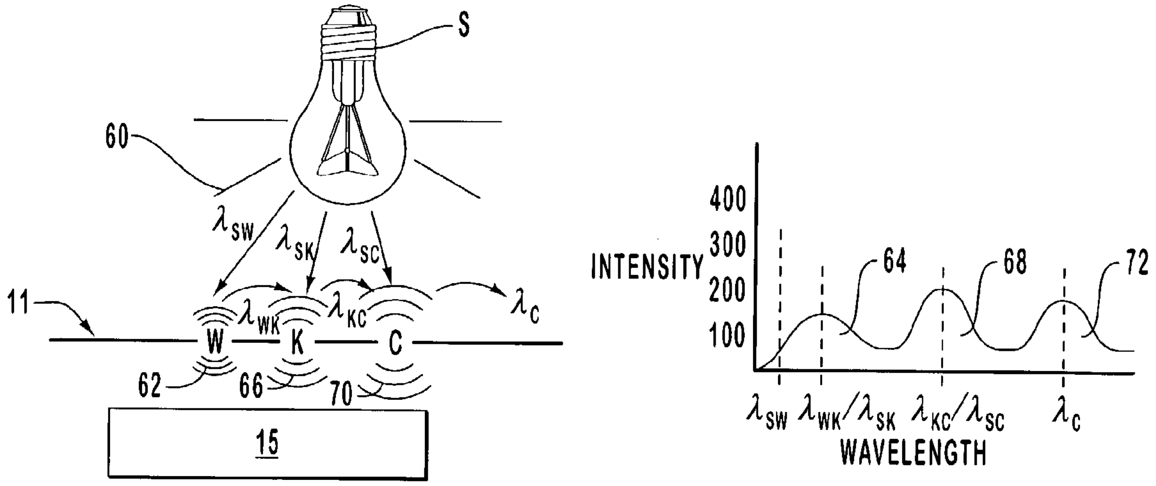
S
60
λ_{SW}
λ_{SK}
λ_{SC}
11
λ_{WK}
λ_{KC}
λ_{C}
W
K
C
62
66
70
15
INTENSITY
400
300
200
100
64
68
72
λ_{SW}
$\lambda_{WK}/\lambda_{SK}$
$\lambda_{KC}/\lambda_{SC}$
λ_{C}
WAVELENGTH

FIG. 8

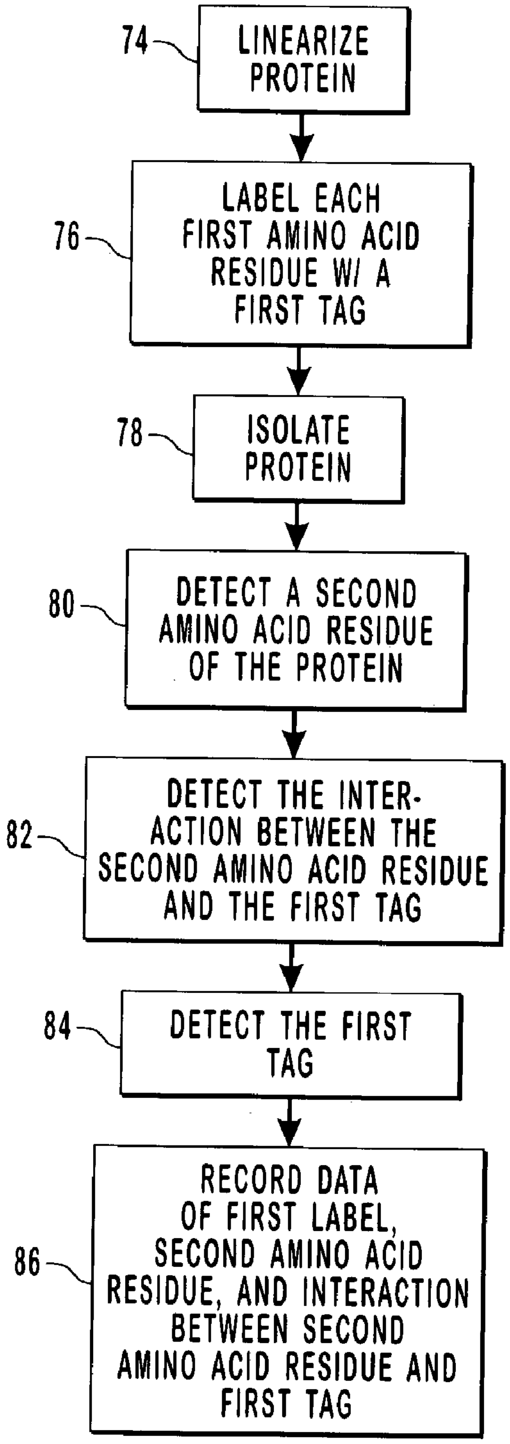
74
LINEARIZE PROTEIN
76
LABEL EACH FIRST AMINO ACID RESIDUE W/ A FIRST TAG
78
ISOLATE PROTEIN
80
DETECT A SECOND AMINO ACID RESIDUE OF THE PROTEIN
82
DETECT THE INTERACTION BETWEEN THE SECOND AMINO ACID RESIDUE AND THE FIRST TAG
84
DETECT THE FIRST TAG
86
RECORD DATA OF FIRST LABEL, SECOND AMINO ACID RESIDUE, AND INTERACTION BETWEEN SECOND AMINO ACID RESIDUE AND FIRST TAG

FIG. 9

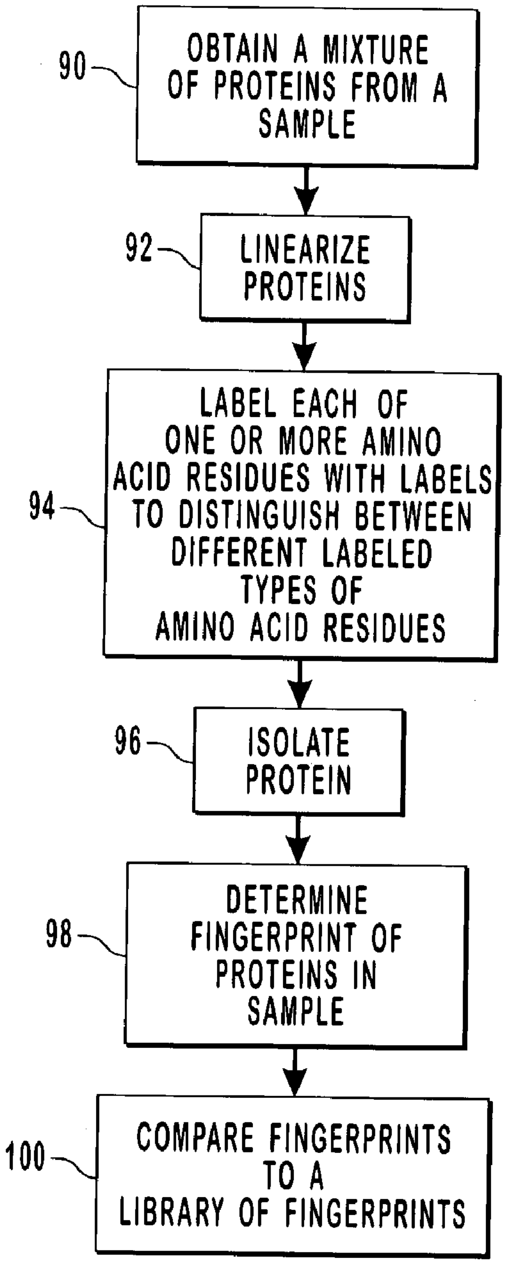
90
OBTAIN A MIXTURE OF PROTEINS FROM A SAMPLE
92
LINEARIZE PROTEINS
94
LABEL EACH OF ONE OR MORE AMINO ACID RESIDUES WITH LABELS TO DISTINGUISH BETWEEN DIFFERENT LABELED TYPES OF AMINO ACID RESIDUES
96
ISOLATE PROTEIN
98
DETERMINE FINGERPRINT OF PROTEINS IN SAMPLE
100
COMPARE FINGERPRINTS TO A LIBRARY OF FINGERPRINTS

FIG. 10

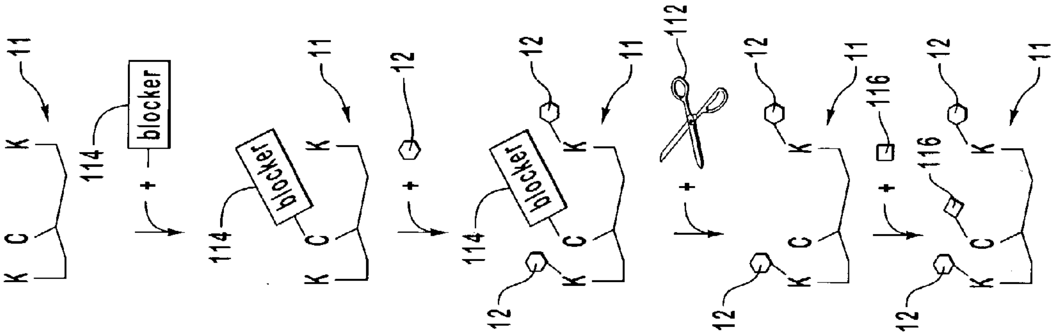

FIG. 13

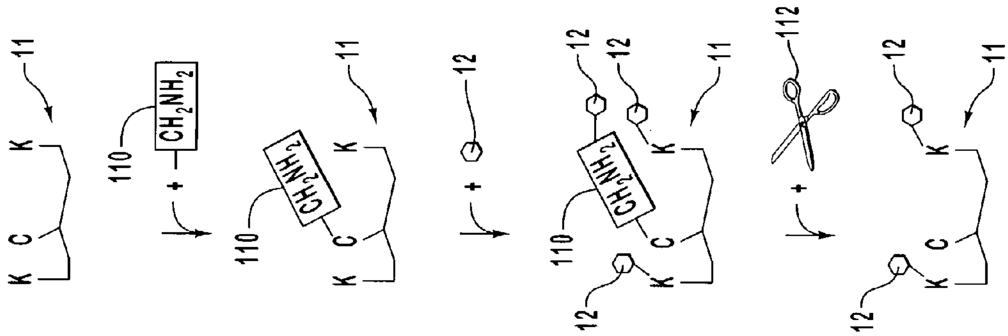

FIG. 12

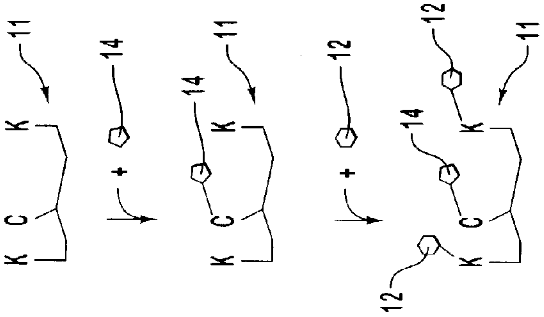

FIG. 11

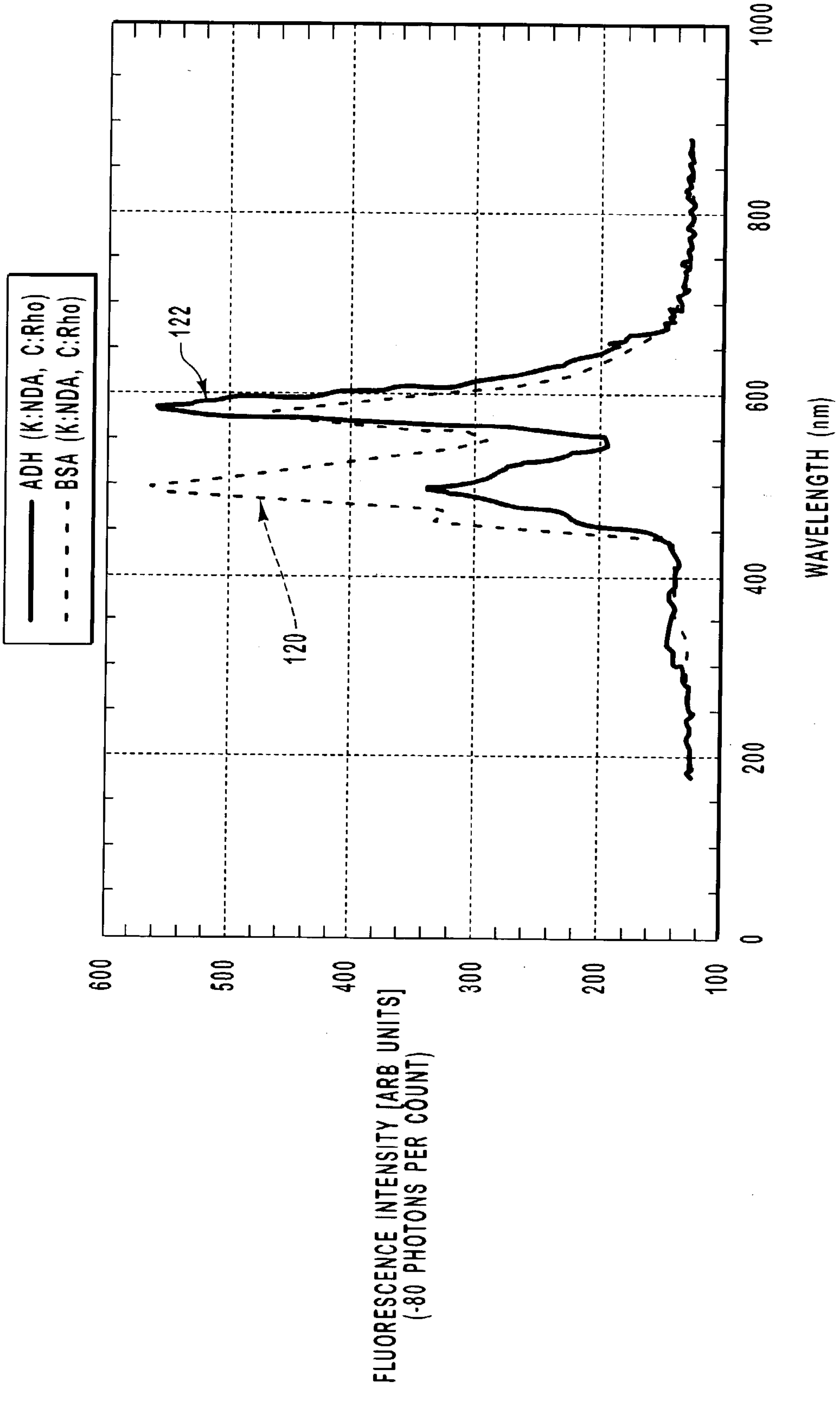
ADH (K:NDA, C:Rho)
BSA (K:NDA, C:Rho)
122
120
FLUORESCENCE INTENSITY [ARB UNITS] (~80 PHOTONS PER COUNT)
600
500
400
300
200
100
0
200
400
600
800
1000
WAVELENGTH (nm)

FIG. 14

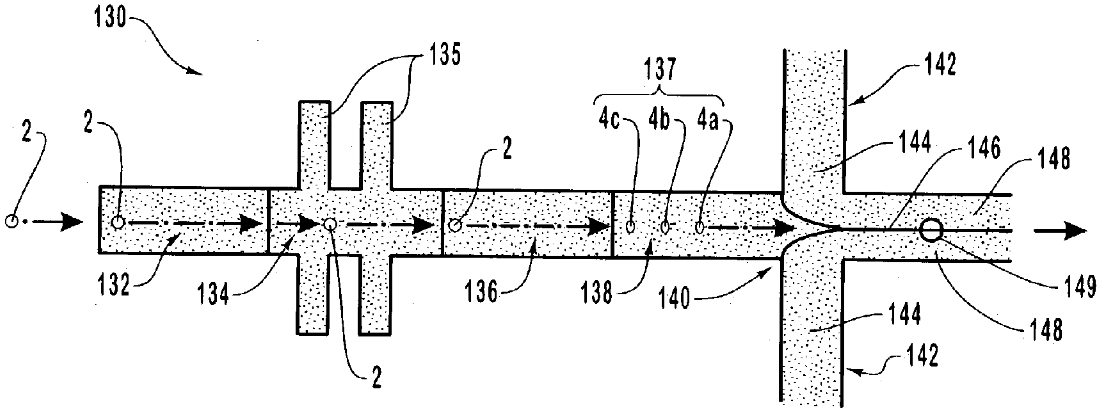
130
135
137
142
2
2
4c
4b
4a
2
144
146
148
132
134
136
138
140
149
2
144
148
142

FIG. 15

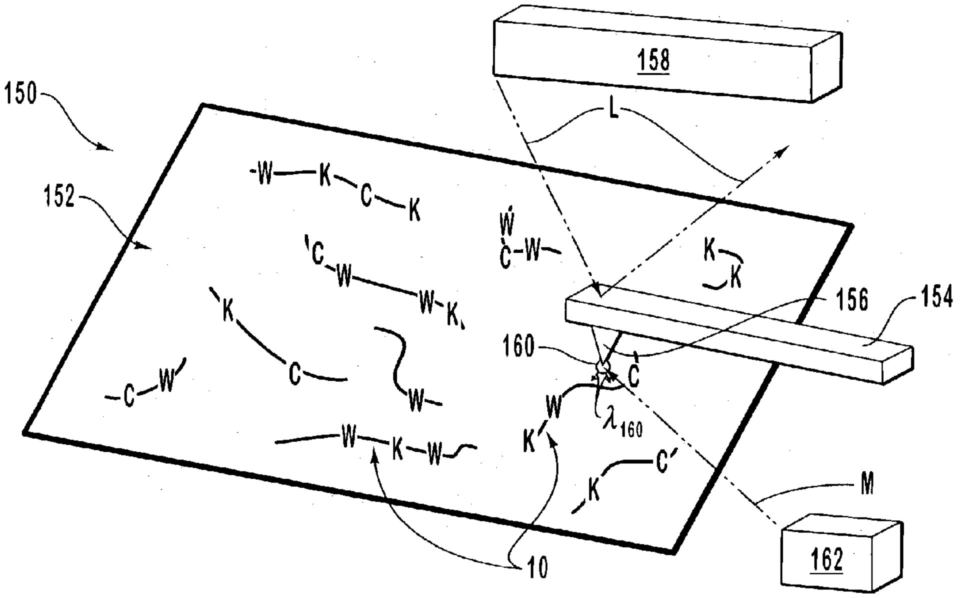
158
150
L
152
-W—K—C—K
W
C-W—
'C-W—W-K
K
K
156
154
K
160
C
C
W
λ160
–C—W
W—
—W—K—W
K
M
K—C'
10
162

FIG. 16

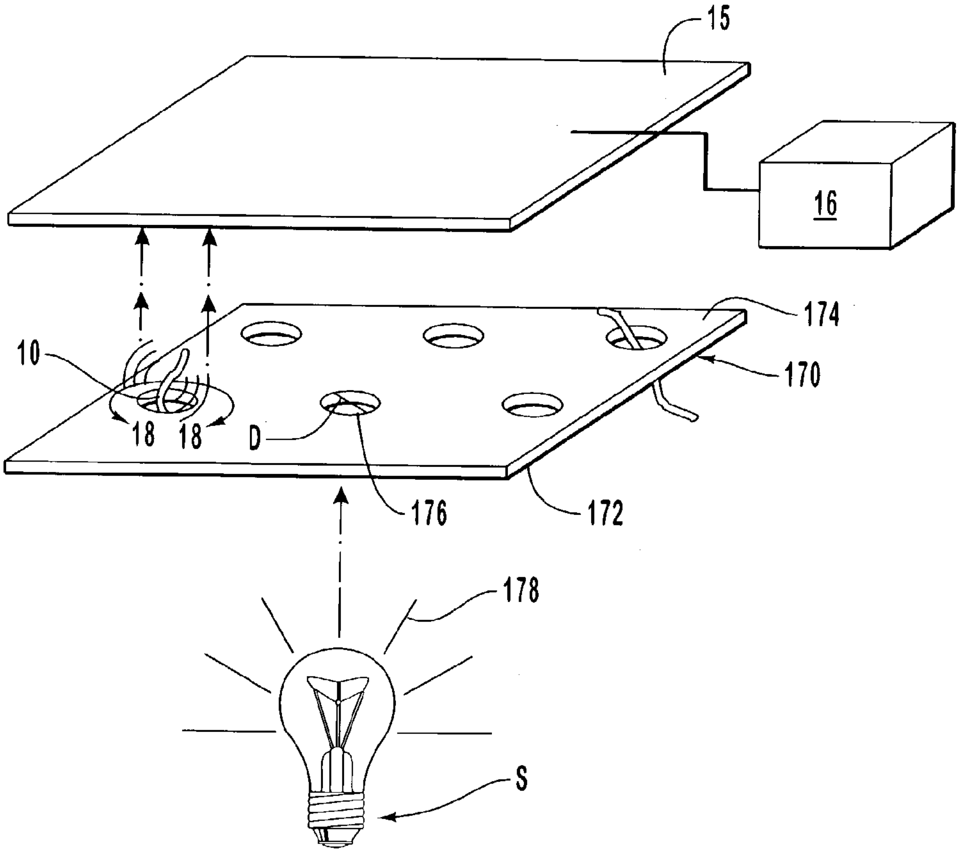
15
16
174
10
170
18
18
D
176
172
178
S

FIG. 17

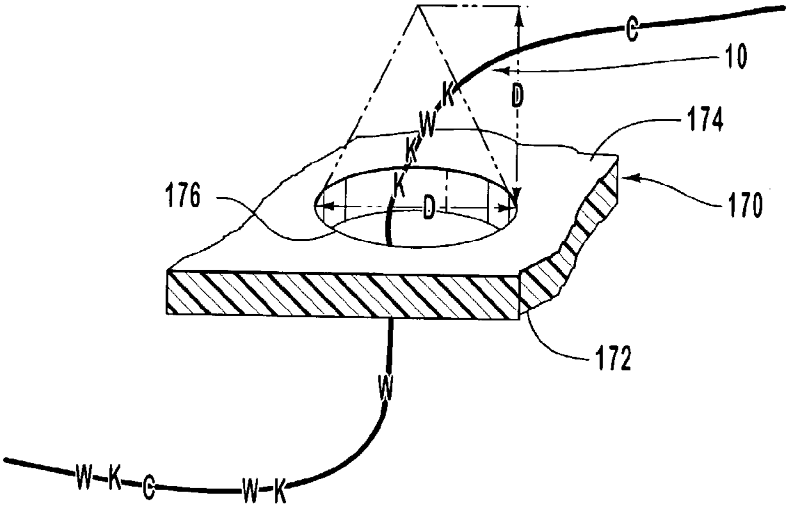
C
10
D
174
176
K
W
K
W
K
D
170
172
W
W-K-C-W-K

FIG. 18

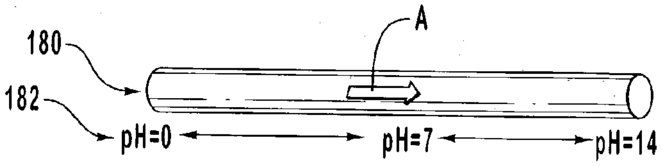

FIG. 19

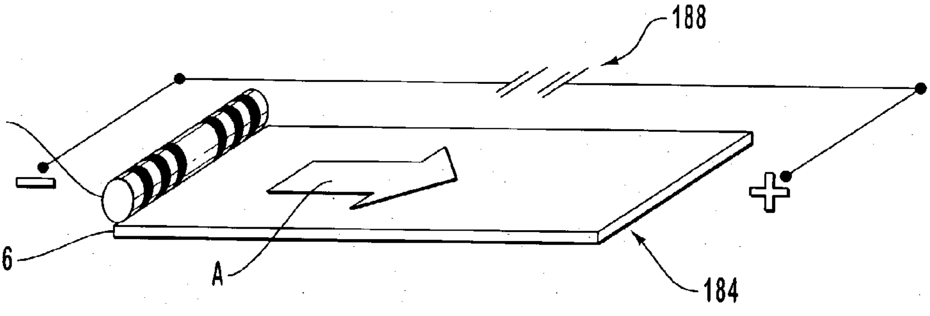

FIG. 20

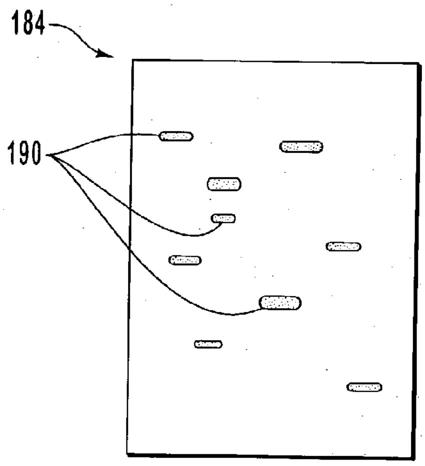

FIG. 21

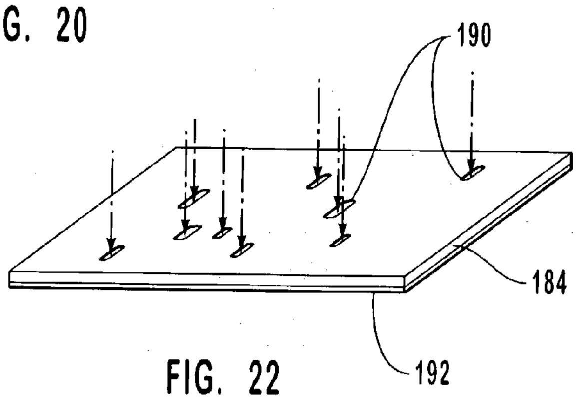

FIG. 22

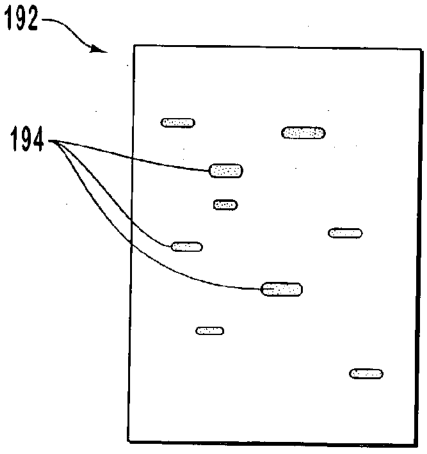

FIG. 23

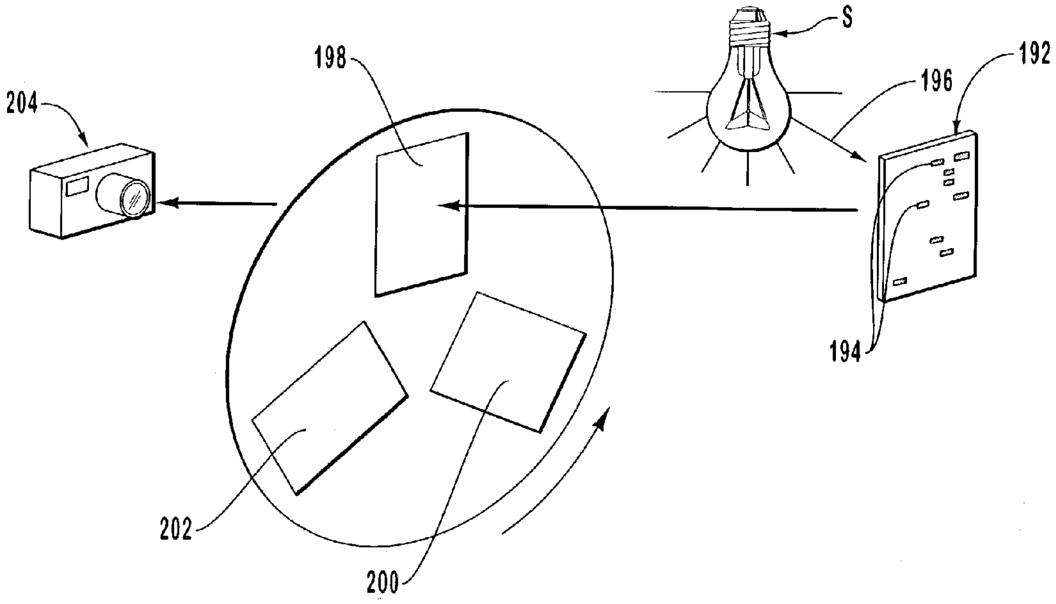

FIG. 24

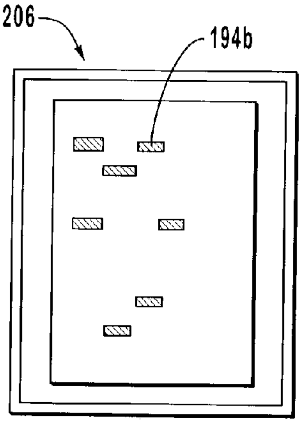

FIG. 25

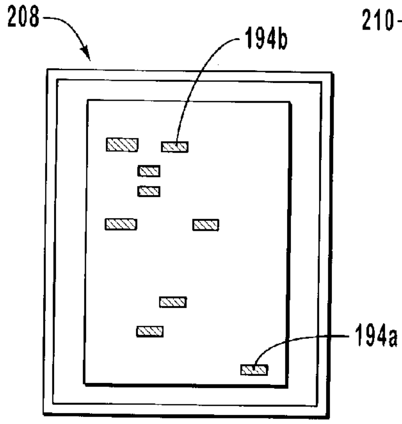

FIG. 26

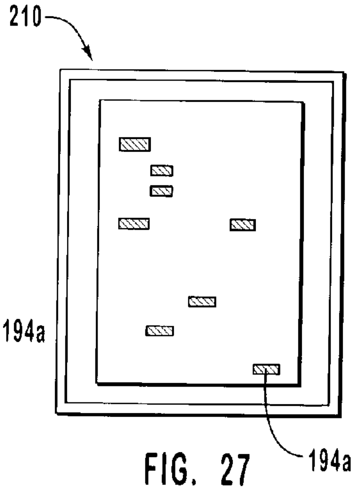

FIG. 27

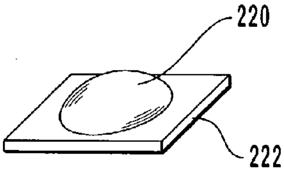

FIG. 28

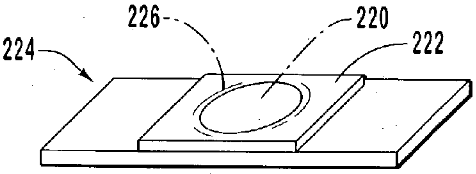

FIG. 29

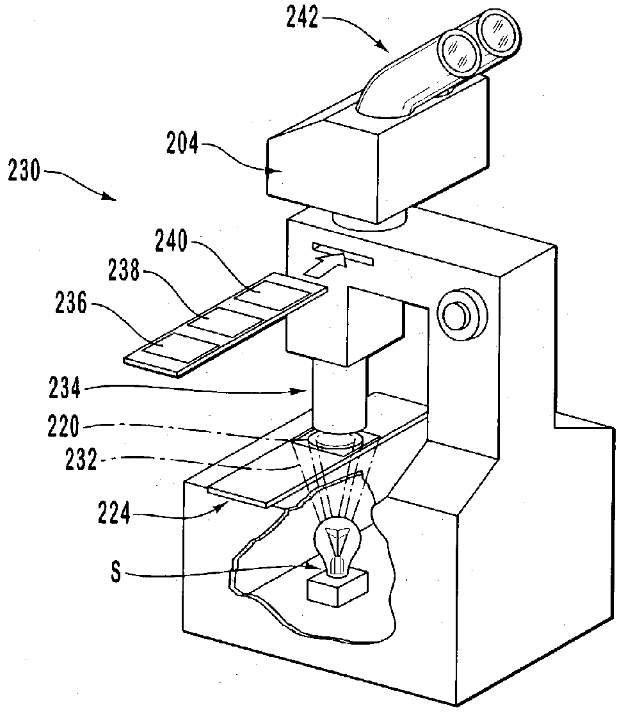

FIG. 30

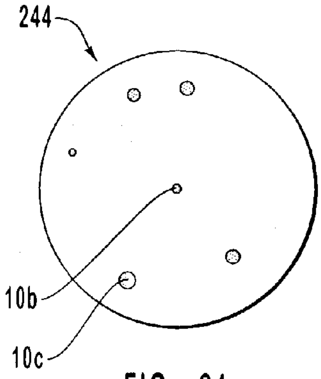

FIG. 31

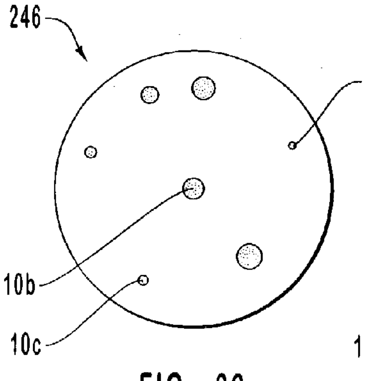

FIG. 32

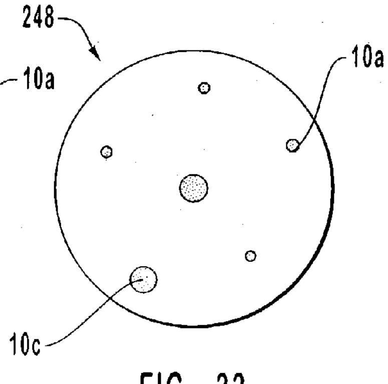

FIG. 33

PROTEIN CHARACTERIZATION SYSTEM

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a divisional application of U.S. patent application Ser. No. 412,732 that was filed on Oct. 5, 1999, and that issued as U.S. Pat. No. 6,569,685 on May 27, 2003.

BACKGROUND

[0002] 1. Field of the Invention

[0003] This invention relates to the rapid identification of protein molecules by the systematic development for each respective type of protein molecule of a set of particular, invariant, readily-detectable distinguishing characteristics, which set of characteristics will for convenience hereinafter be referred to as a fingerprint for the corresponding type of protein molecule. The invention also relates to libraries of different protein molecules and the corresponding fingerprints therefor, as well as to systems used in the identification, or fingerprinting, of protein molecules. The present invention has particular applicability to the identification of protein molecules obtained from biological samples.

[0004] 2. Background Art

[0005] There are approximately 100,000 different types of protein molecules involved in organic processes. Each protein molecule is, however, comprised of various amino acid building blocks from a group of about twenty different amino acids. Amino acids chemically connect end-to-end to form a chain that is referred to as a peptide. The amino acid building blocks in a peptide chain share as a group various of the peripheral atomic constituents of each amino acid. As a result, an amino acid in a peptide chain is not in situ a complete amino acid. Therefore, an amino acid in a peptide chain is referred to as an "amino acid residue." A peptide chain becomes a true protein molecule only when the constituent amino acid residues have been connected, when certain amino acid residues of the peptide chain have been modified by the addition to or removal of certain types of molecules from the functional chemical groups of these amino acid residues, and when the completed chain of amino acid residues assumes a particular three-dimensional structure determined by the sequence of amino acid residues and chemical modifications thereof.

[0006] Protein molecules do not naturally maintain a one-dimensional, linear arrangement. The sequence of the amino acid residues in a protein molecule causes the molecule to assume an often complex, but characteristic three-dimensional shape. A protein molecule that has been forced out of this three-dimensional shape into a one-dimensional, linear arrangement is described as having been "linearized."

[0007] Protein molecules are involved in virtually every biological process. Aberrant or mutant forms of protein molecules disrupt normal biological processes, thereby causing many types of diseases, including some cancers and inherited disorders, such as cystic fibrosis and hemophilia. The ability of a protein molecule to perform its intended function depends, in part, upon the sequence of amino acid residues of the protein molecule, modifications to particular amino acid residues of the protein molecule, and the three-dimensional structure of the protein molecule.

[0008] Alterations to the sequence of amino acid residues, to the modifications of particular amino acid residues, or to the three-dimensional structure of a protein molecule can change the way in which a protein molecule participates in biological processes. While many protein molecules and the functions thereof in biological processes are known, scientists continue the arduous task of isolating protein molecules, identifying the chemical composition and structure of each isolated protein molecule, and determining the functions of the protein molecule, as well as the consequences of changes in the structures of the protein molecule.

[0009] The sequence of the amino acid residues in a protein molecule, which imparts to the protein molecule a unique identity with a set of unique characteristics, is difficult to detect rapidly and reliably.

[0010] The identification of a protein molecule typically involves two steps: (1) purifying the protein molecule; and (2) characterizing the protein molecule.

[0011] In isolating or purifying protein molecules, a targeted protein molecule is separated from other, different types of protein molecules. Some current purification techniques are sensitive enough to purify an aberrant form of a protein molecule from normal protein molecules of the same type. Different purification techniques are based on the different characteristics of protein molecules, such as the weight of a protein molecule, the solubility of a protein molecule in water and other solvents, the reactivity of a protein molecule with various reagents, and the pH value at which the protein molecule is electrically neutral. The last is referred to as the isoelectric point of the protein molecule. Due to the large number of different types of protein molecules and because some types of protein molecules have very similar characteristics to other types of protein molecules, extremely sensitive purification processes are often required to isolate one type of protein molecule from others. The sensitivity with which similar types of protein molecules are separated from each other can be enhanced by combining different types of these purification techniques.

[0012] In some characterization processes, individual protein molecules are studied. When characterization processes that permit one to study individual protein molecules are employed, a single protein molecule in a sample can be separated or isolated from the other protein molecules in the sample by diluting the sample.

[0013] Since many purification techniques separate different types of protein molecules on the bases of the physical or chemical characteristics of the different types of protein molecules, these purification techniques may themselves reveal some information about the identity of a particular type of protein molecule. Once a particular type of protein molecule has been purified, it may be necessary to further characterize the purified protein molecule in order to identify the purified protein molecule. This is particularly true when attempting to characterize previously unidentified types of protein molecules, such as aberrant or mutant forms of a protein molecule.

[0014] Typically, protein molecules are further characterized by employing techniques that determine the weight of the protein molecule with increased sensitivity over techniques like gel electrophoresis, or by determining the sequence of amino acid residues that make up the protein molecule. One technique that is useful for performing both of these tasks is mass spectrometry.

[0015] In order to characterize a type of protein molecule by mass spectrometry, a purified type of protein molecule or a particular segment of a purified type of protein molecule is given positive and negative charges, or ionized, and made volatile in a mass spectrometer. The ionized, volatilized protein molecules or segments are then analyzed by the mass spectrometer. This produces a mass spectrum of the protein molecule or segment. The mass spectrum provides very precise information about the weight of the protein molecule or segment. Due to the precision with which a mass spectrometer determines the weight of protein molecules and segments of protein molecules, when a protein molecule or segment is analyzed, the information provided by mass spectrometry can be of use in inferring the sequence of amino acid residues in the protein molecule or segment. Mass spectrometers are also sensitive enough to provide information about modifications to particular amino acid residues of a protein molecule or segment. When a series of segments from a certain type of protein molecule are analyzed by mass spectrometry, the information about the sequences of and modifications to the amino acid residues of each segment can be combined to infer the sequence of and modifications to amino acid residues of an entire protein molecule.

[0016] Due to the sensitivity of mass spectrometry and the resulting ability to infer the sequences of the amino acid residues and modifications thereto of a particular type of protein molecule, the differences of aberrant or mutant forms of protein molecules from a normal protein molecule in amino acid residue sequences and amino acid residue modifications can also be inferred.

[0017] Nonetheless, mass spectrometry is a time-consuming process that requires expensive equipment and reagents.

SUMMARY OF THE INVENTION

[0018] It is thus a broad object of the present invention to increase the speed and efficiency with which protein molecules can be characterized.

[0019] It is also an object of the present invention to lend to a protein molecule a characteristic set of ancillary properties that are rapidly and reliably detectable.

[0020] It is a further object of the present invention to generate a listing of known protein molecules and their corresponding fingerprints as provided and determined by the method of the present invention.

[0021] Achieving the foregoing objects will fulfill further, broader objects of the present invention of improving biochemical research and healthcare.

[0022] To achieve the foregoing objects, and in accordance with the invention as embodied and broadly described herein, systems and methods for characterizing protein molecules are provided. Also provided are protein molecules having such tags attached thereto as impart the protein molecules distinguishing characteristics that are useable as fingerprints.

[0023] In one form, a system incorporating teachings of the present invention, which is capable of characterizing a protein molecule, lends to a protein molecule a characteristic set of ancillary properties that is rapidly and reliably detectable. As these ancillary properties are as uniquely identifying of the type of the protein molecule as fingerprints are reflective of the identity of a human being, the characteristic set of ancillary properties of a protein molecule function as a "fingerprint" of the protein molecule that maybe used to rapidly and reliably identify the type of the protein molecule.

[0024] A system according to teachings of the present invention has denaturation means for linearizing the protein molecule, labeling means for attaching a tag to each of a first type of amino acid residue of the protein molecule, and detector means for detecting a fingerprint of the tagged protein molecule. The fingerprint of the protein molecule has a first fingerprint constituent imparted to the protein molecule by the tags on each first type of amino acid residue in the protein molecule and a second fingerprint constituent imparted to the protein molecule by each second type of amino acid residue in the protein molecule.

[0025] A system according to teachings of the present invention may also include isolation means for separating the protein molecule from other protein molecules in a sample, as well as collation means for comparing the fingerprint of a protein molecule of interest to the fingerprints of known protein molecules listed in a library.

[0026] An example of the denaturation means is a detergent, such as sodium dodecyl sulfate (hereinafter "SDS"), which gives the entire protein molecule a negative charge and therefore pulls the protein molecule out of its three-dimensional structure. Another example of the denaturation means is β-mercaptoethanol, a chemical that breaks chemical linkages between the sulfur atoms of two amino acid residues.

[0027] A protein molecule of interest is separated from the other types of protein molecules present in a sample by way of isolation means for separating the protein molecule. Examples of isolation means that are useful in the systems and methods of the present invention include, without limitation, hydrodynamic focusing apparatus, electrophoretic gels, separation plates with apertures therethrough, and dilution systems for the sample in which the protein molecule of interest is located.

[0028] In a first example of the labeling means, a fluorescent dye is attached chemically to the amino acid residues in a protein molecule of a specific chosen type, thereby forming a tag on each amino acid residue of the specific chosen type in the protein molecule. In a second example, the labeling means is a metallic tag precursor that chemically bonds with the amino acid residues in a protein of a specific chosen type to form a tag on each amino acid residue of the specific chosen type in the protein molecule.

[0029] Of the twenty or so types of amino acid residues in protein molecules, one type of amino acid residue, known as tryptophan, self-fluoresces when exposed to electromagnetic excitation radiation of a certain range of wavelengths.

[0030] When a fluorescent dye is used as the labeling means, an example of the detector means includes electromagnetic excitation radiation of one or more excitation wavelengths or a range of excitation wavelengths that will stimulate the tryptophan amino acid residues of a protein molecule to emit radiation of a first emitted wavelength. The excitation radiation of the detector means will also cause the fluorescent dye to emit radiation of a second emitted wavelength. In this example, the detector means also includes a detector that is sensitive to the wavelengths of emitted radiation from the tryptophan amino acid residues of the protein molecule and to the fluorescent dye.

[**0031**] When the tags attached to each of the specific type of amino acid residue of the protein molecule are metallic, the detector means can include a nuclear magnetic resonance apparatus or other apparatus known in the art to be capable of detecting single metal atoms.

[**0032**] Alternatively, tags can be attached to more than one type of amino acid residue of the protein molecule. The tags on one type of amino acid residue are differentially detected from the tags on one or more other types of amino acid residues to determine different fingerprint constituents of the protein molecule.

[**0033**] According to another aspect of the invention, a listing or database is generated for use with specific protocols to identify protein molecules. This listing or database is referred to herein as a library, and includes the identities of a set of known protein molecules and information about the different fingerprint constituents of each of the known protein molecules of the listing. The different fingerprint constituents are imparted to the protein molecule by the labeling means of the system and detected by way of the detection means of the system. Collation means for comparing the fingerprint of a protein molecule of interest to the fingerprints of the known protein molecules listed in the library are then used to identify the protein molecule of interest. Typically, the function of such a collation means can be performed by a computer processor.

[**0034**] In yet another aspect, the present invention includes protein molecules that have been labeled with tags to impart fingerprint constituents to the protein molecule. Each fingerprint constituent indicates the number of a particular type of amino acid residue in a protein molecule and the relative locations of different types of amino acid residues in the protein molecule.

[**0035**] The prospect of being able to rapidly and reliably identify a type of protein molecule has utility in a wide range of research and clinical applications, such as, for example, in determining whether or not selected cells of a patient have entered early stages of cancer.

[**0036**] Additional objects and advantages of the invention will be set forth in the description which follows, and in part will be obvious from the description, or may be learned by the practice of the invention. The objects and advantages of the invention maybe realized and obtained by means of the instruments and combinations particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[**0037**] A more particular description of the invention briefly described above will be rendered by reference to a specific embodiment thereof which is illustrated in the appended drawings in order to illustrate and describe the manner in which the above-recited and other advantages and objects of the invention are obtained. Understanding that these drawings depict only a typical embodiment of the invention and are not therefore to be considered limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[**0038**] **FIG. 1** schematically illustrates steps by which a mixed sample of different types of proteins might routinely be obtained;

[**0039**] **FIG. 2** is a schematic diagram of a portion of a sequence of amino acid residues in a typical protein molecule with selected of those amino acid residues tagged;

[**0040**] **FIG. 3** is a schematic diagram of the portion of the protein molecule of **FIG. 2** with selected of the amino acid residues therein, including the tagged amino acid residues, symbolized in simplified form at a higher level of abstraction;

[**0041**] **FIG. 4** is a schematic illustration of the portion of the protein molecule illustrated in **FIG. 3** symbolized with yet enhanced simplicity at an even higher enhanced level of abstraction;

[**0042**] **FIG. 5** is a schematic illustration of a method incorporating teachings of the present invention for obtaining a first fingerprint constituent for a segment of the protein molecule depicted in **FIG. 4** using a single-stage emission produced by exposing the segment to electromagnetic radiation at a first excitation wavelength;

[**0043**] **FIGS. 6A and 6B** taken together illustrate schematically a method for obtaining a second fingerprint constituent and a third fingerprint constituent for the segment of the protein molecule depicted in **FIG. 4** using a two-stage emission produced by exposing the segment to electromagnetic radiation at a second excitation wavelength, **FIG. 6A** illustrating the initial single-stage emission in the process, and **FIG. 6B** illustrating the entirety of the two-stage emission initiated thereby;

[**0044**] FIGS. 7A-7C taken together illustrate schematically a method for obtaining fourth, fifth, and sixth fingerprint constituents for the segment of the protein molecule depicted in **FIG. 4** using a three-stage emission produced by exposing the segment to electromagnetic radiation at a third excitation wavelength, **FIG. 7A** illustrating the initial single-stage emission, **FIG. 7B** illustrating the two-stage emission caused thereby, and **FIG. 7C** illustrating the entirety of the three-stage emission;

[**0045**] **FIG. 8** is a schematic illustration of a second embodiment of a method incorporating teachings of the present invention for obtaining three fingerprint constituents for a segment of the protein molecule depicted in **FIG. 4** using a three-stage emission produced by exposing the segment to electromagnetic radiation of a broad range of excitation wavelengths;

[**0046**] **FIG. 9** is a flow chart depicting steps in a method incorporating teachings of the present invention for determining a fingerprint for a protein molecule;

[**0047**] **FIG. 10** is a flow chart depicting steps in a method incorporating teachings of the present invention for identifying a protein molecule using a fingerprint thereof determined according to the method of **FIG. 9**;

[**0048**] **FIG. 11** is a schematic diagram of steps in a first embodiment of a method for labeling more than one type of amino acid residue of a protein molecule with tags;

[**0049**] **FIG. 12** is a schematic diagram illustrating steps in a second embodiment of a method for labeling more than one type of amino acid residue of a protein molecule with tags and intermediate structures;

[**0050**] **FIG. 13** is a schematic diagram depicting steps in a third embodiment of a method for labeling more than one type of amino acid residue of a protein molecule with tags and intermediate structures;

[**0051**] **FIG. 14** is a graph depicting fingerprints imparted to two different types of protein molecules by attaching two different fluorescent tags to each of two different types of amino acid residues of the protein molecules;

[**0052**] **FIG. 15** is a schematic representation of a first embodiment of an apparatus used according to teachings of the present invention to isolate a protein molecule from other protein molecules in a sample for the purpose of facilitating the determination of a fingerprint therefor;

[**0053**] **FIG. 16** is a perspective view of a second embodiment of an apparatus used according to teachings of the present invention to isolate a protein molecule from other protein molecules in a sample for the purpose of facilitating the determination of a fingerprint therefor;

[**0054**] **FIG. 17** is a perspective view of a third embodiment of an apparatus used according to teachings of the present invention to isolate a protein molecule from other protein molecules in a sample for the purpose of facilitating the determination of a fingerprint therefor;

[**0055**] **FIG. 18** is an enlarged perspective view of a portion of the apparatus depicted in **FIG. 17**;

[**0056**] **FIG. 19** is a schematic representation of an isoelectric focusing gel used in accordance with teachings of the present invention to isolate types of protein molecules in a sample of a plurality of types of protein molecules for the purpose of identifying the types of protein molecules in the sample;

[**0057**] **FIG. 20** is a schematic representation of an electrophoretic gel used according to teachings of the present invention to refine the isolation of types of protein molecules in a sample of a plurality of types of protein molecules previously separated from each other with the isoelectric focusing gel depicted in **FIG. 19**;

[**0058**] **FIG. 21** is a schematic representation plan view of an electrophoretic gel after the various types of protein molecules in the sample have been separated from each other into respective bands in the manner illustrated in **FIGS. 19 and 20**;

[**0059**] **FIG. 22** is a schematic representation perspective view of a method for transferring the bands of the electrophoretic gel of **FIG. 21** onto a membrane;

[**0060**] **FIG. 23** is a schematic representation of the membrane of **FIG. 22** after the bands of different types of protein molecules have been transferred thereto in the manner there illustrated;

[**0061**] **FIG. 24** is a schematic representation of an embodiment of a system used according to teachings of the present invention for identifying the types of protein molecules separated into respective bands in the electrophoretic gel of **FIG. 21** and transferred to the membrane of **FIG. 23**;

[**0062**] **FIG. 25** is a schematic representation of a first photograph of the membrane of **FIG. 24** obtained by use of the system shown in **FIG. 24** and embodying data for determining a first fingerprint constituent for each of the types of proteins in the bands in the membrane of **FIG. 24**;

[**0063**] **FIG. 26** is a schematic representation of a second photograph of the membrane shown in **FIG. 24** embodying data for determining a second fingerprint constituent for each of the types of proteins in the bands in the membrane of **FIG. 24**;

[**0064**] **FIG. 27** is a schematic representation of a third photograph of the membrane shown in **FIG. 24** embodying data for determining a third fingerprint constituent for each of the types of proteins in the bands in the membrane of **FIG. 24**;

[**0065**] **FIG. 28** is a perspective view of a cover slip for a microscope slide supporting a drop of a solution containing protein molecules;

[**0066**] **FIG. 29** is a perspective view of the cover slip depicted in **FIG. 28** inverted and positioned on a microscope slide over a shallow recess therein, whereby the drop of solution hangs from the cover slip within the recess;

[**0067**] **FIG. 30** is a perspective view of an apparatus used according to teachings of the present invention to obtain photographs of a field of view of a portion of the drop of solution in **FIG. 29**; and

[**0068**] FIGS. **31**-**33** are schematic representations of a field of view of a portion of the drop of solution of **FIG. 29** obtained by use of the apparatus of **FIG. 30** using different filters, each field of view embodying data for determining fingerprints for the protein molecules appearing in that field of view.

DETAILED DESCRIPTION OF THE ILLUSTRATED EMBODIMENTS

[**0069**] **FIG. 1** depicts steps by which a mixed sample **2** of different types of protein molecules **4***a*, **4***b*, **4***c* might routinely be obtained from a living organism **6**, such as an animal, a plant, a microorganism, or the human being depicted. A biological sample **8**, such as tissue or the illustrated cell, is secured from living organism **6**. Mixed sample **2** of protein molecules **4***a*, **4***b*, **4***c* is then obtained from biological sample **8** by, for example, disrupting the cell membranes. Protein molecules **4***a*, **4***b*, **4***c* are then linearized from three-dimensional structures to one-dimensional structures, such as the linearized protein molecules **10** depicted to the right in **FIG. 1**.

[**0070**] **FIG. 2** illustrates a single linearized protein molecule **10**. Protein molecule **10** is a chain of amino acid residues that includes a number of amino acid residues K of a first type, a number of amino acid residues C of a second type, a number of tryptophan amino acid residues W, and a number of amino acid residues X of other types, many of which are not shown, but only suggested by ellipsis. A first type of tag **12** is shown chemically attached to each amino acid residue K of protein molecule **10**. A second type of tag **14** is shown chemically attached to each amino acid residue C. In **FIG. 2** no such tags are attached to tryptophan amino acid residues W or to other amino acid residues X. Tags **12** and **14** may be different types of fluorescent tags, different types of metallic tags, or different types of tags of some other detectable genre.

[**0071**] A subcombination of adjacently connected amino acid residues in protein molecule **10** is identified in **FIG. 2** as peptide **11**. In left-to-right sequence, a single amino acid residue W, an amino acid residue K carrying a tag **12**, and an amino acid residue C carrying a tag **14**. As illustrated, amino acid residue W is tryptophan, amino acid residue K is lysine, and amino acid residue C is cysteine.

[**0072**] Protein molecule **10** of **FIG. 2** is again depicted in **FIG. 3** with selected of the amino acid residues therein, including amino acid residues W, K, and C, symbolized in simplified form at a higher level of abstraction. NH_2 represents a first end, or terminus, of protein molecule **10** and COOH represents a second end of protein molecule **10**.

[**0073**] **FIG. 4** illustrates protein molecule **10** symbolized with yet enhanced simplicity at an even higher enhanced level of abstraction relative to that of **FIGS. 2 and 3**.

[**0074**] According to one aspect of the teachings of the present invention, FIGS. **5**-**7**C illustrate a method for characterizing a protein molecule on the basis of ancillary properties imparted to the protein molecule by the natural fluorescence of tryptophan amino acid residues and by fluorescent tags attached to substantially all of the lysine and cysteine amino acid residues of the protein molecule.

[**0075**] For convenience in illustrating the implementation of the disclosed protein fingerprinting technology, peptide **11** is illustrated, apart from the balance of protein molecule **10**, as a straight line in FIGS. **5**-**7**C. In the depictions of peptide **11** in these figures, only the amino acid residue W, K, or C of immediate concern to the corresponding discussion will be illustrated. As a further simplification, tag **12** on amino acid residue K and tag **14** on amino acid residue C have been omitted, as was the case in **FIGS. 3 and 4**. Nonetheless, the depictions in these figures are illustrative only, and tag **12** and tag **14** should be understood to be present, respectively, on amino acid residue K and amino acid residue C.

[**0076**] For illustrative purposes, tag **12** and tag **14** are tags that fluoresce when exposed to an appropriate respective wavelength of electromagnetic radiation. Tryptophan amino acid residue W is naturally fluorescent, meaning that amino acid residue W will fluoresce when exposed to an appropriate wavelength of electromagnetic radiation, even without the attachment thereto of any fluorescent tag, such as tag **12** or tag **14**. Therefore, no such fluorescent tag is shown on amino acid residues W in **FIG. 2** or is to be suggested in the other of the accompanying figures.

[**0077**] **FIG. 5** illustrates peptide **11** of protein molecule **10** exposed to a source S of a first primary electromagnetic excitation radiation **20** of wavelength λ_{SC}. First primary electromagnetic excitation radiation **20** stimulates the tag on amino acid residue C to fluoresce, producing emitted radiation **22** of wavelength λ_C.

[**0078**] The intensity of emitted radiation **22** is measured by a detector **15** and subjected to a spectral analysis that is reflected in the graph to the right in **FIG. 5**. That graph is characterized by a peak centered about wavelength λ_C that serves as a fingerprint constituent **24** for peptide **11** at wavelength sc.

[**0079**] Thus, as illustrated in the spectral diagram of **FIG. 5**, when a fluorescent tag is chemically attached to amino acid residue C, first primary electromagnetic excitation radiation **20** of wavelength λ_{SC} stimulates the emission of a corresponding fingerprint constituent **24**. Nonetheless, the activity reflected in **FIG. 5** is but a depiction of activity related to a single amino acid residue C in isolation from all other amino acid residues in peptide **11** or even in protein molecule **10**.

[**0080**] Different approaches are used to obtain a corresponding fingerprint constituent for the entirety of protein molecule **10** at wavelength λ_{SC}.

[**0081**] In a relatively global approach, first primary electromagnetic excitation radiation **20** is used to illuminate the entire length of protein molecule **10**. The cumulative intensity of all consequently emitted radiation is measured by a detector and subjected to an appropriate spectral analysis.

[**0082**] Alternatively, linearized protein molecule **10** is scrolled past source S and detector **15**. This results in a sequenced series of fingerprint constituents for protein molecule **10** at wavelength λ_{SC}. This scrolling process produces markedly greater information about the structure of protein molecule **10** than does the global method described previously.

[**0083**] **FIGS. 6A and 6B** depict in stages the consequence of the exposure of peptide **11** of protein molecule **10** to a source S of a second primary electromagnetic excitation radiation **30** at a wavelength ?SK that stimulates the tag on amino acid residue K to fluoresce. The process also stimulates the tag on amino acid residue C to fluoresce, albeit indirectly. Each of **FIGS. 6A and 6B** includes a graph that depicts a corresponding portion of the response spectra for peptide **11** at wavelength λ_{SK}.

[**0084**] **FIG. 6A** illustrates peptide **11** of protein molecule **10** exposed to a source S of a second primary electromagnetic excitation radiation **30** of wavelength λ_{SK}. Second Primary excitation radiation **30** causes the tag on amino acid residue K to fluoresce, producing emitted radiation **32** of wavelength λ_{KC}.

[**0085**] The intensity of emitted radiation **32** is measured by detector **15** and subjected to a spectral analysis that produces the graph to the right in **FIG. 6A**. That graph is characterized by a peak centered about wavelength λ_{KC}. This serves as a first fingerprint constituent **34** for peptide **11** at wavelength λ_{KC}.

[**0086**] Emitted radiation **32** is, however, capable of exciting the tag on amino acid residue C to fluoresce.

[**0087**] **FIG. 6B** illustrates that the exposure of peptide **11** to emitted radiation **32** excites the tag on amino acid residue C. This causes the tag on amino acid residue C to fluoresce, producing emitted radiation **36** of wavelength λ_C.

[**0088**] The intensity of emitted radiation **32** and the intensity of emitted radiation **36** are measured by detector **15** and subjected to a spectral analysis that produces the graph to the right in **FIG. 6B**. That graph is characterized not only by first fingerprint constituent **34**, but by a second peak centered about wavelength λ_C. The latter serves as a second fingerprint constituent **38** for peptide **11** at wavelength λ_{SK}.

[**0089**] For emitted radiation **32** to have the illustrated effect on the tag on amino acid residue C, the tag on amino acid residue K that produced emitted radiation **32** must be located relatively proximately along protein molecule **10** to the tag on amino acid residue C.

[**0090**] Thus, as illustrated in the spectral diagram of **FIG.** 6B, when fluorescent tags are chemically attached to two different types of amino acid residues K and C, second primary electromagnetic excitation radiation **30** of wavelength λ_{SK} that excites the tag on amino acid residue K will stimulate the emission of two corresponding additional fingerprint constituents for peptide **11**.

[**0091**] FIGS. 7A-7C illustrate in stages the consequence of the exposure of peptide **11** of protein molecule **10** to a source S of a third primary electromagnetic excitation radiation **40** at a wavelength λ_{SW} that stimulates tryptophan amino acid residue W to fluoresce. The process also indirectly stimulates the tags on amino acid residues K and C of peptide **11** to fluoresce. Each of FIGS. **7A-7C** includes a graph that depicts a corresponding portion of the response spectra for peptide **11** at wavelength λ_{SW}.

[**0092**] In **FIG. 7A** it can be seen that the exposure of peptide **11** to third primary electromagnetic excitation radiation **40** excites tryptophan amino acid residue W. This causes tryptophan amino acid residue W to fluoresce, producing emitted radiation **42** of wavelength λ_{WK}. The intensity of emitted radiation **42** is measured by detector **15** and subjected to a spectral analysis that produces the graph to the right in **FIG. 7A**. That graph is characterized by a peak centered about wavelength λ_{WK} that serves as a first fingerprint constituent **44** for peptide **11** at wavelength λ_{SW}.

[**0093**] Emitted radiation **42** of **FIG. 7A** is, however, radiation that is capable of exciting the tag on amino acid residue K to fluoresce.

[**0094**] In **FIG. 7B** it can be seen that the exposure of peptide **11** to emitted radiation **42** excites the tag on amino acid residue K. This causes the tag on amino acid residue K to fluoresce, producing emitted radiation **46** of wavelength λ_{KC}. The intensity of emitted radiation **46** is measured by detector **15** and subjected to a spectral analysis that produces the graph to the right in **FIG. 7B**. That graph is characterized, not only by first fingerprint constituent **44**, but by a second peak centered about wavelength λ_{KC}. The latter serves as a second fingerprint constituent **48** for peptide **11** at wavelength λ_{SW}.

[**0095**] For emitted radiation **42** to have the illustrated effect on the tag on amino acid residue K, tryptophan amino acid residue W that produced emitted radiation **42** must be located relatively proximately along protein molecule **10** to the tag on amino acid residue K.

[**0096**] Emitted radiation **46** of **FIG. 7B** is, however, radiation that is capable of exciting the tag on amino acid residue C to fluoresce.

[**0097**] In **FIG. 7C** it can be seen that the exposure of peptide **11** to emitted radiation **46** excites the tag on amino acid residue C. This causes the tag on amino acid residue C to fluoresce, producing emitted radiation **50** of wavelength λ_C. The intensity of emitted radiation **50** is measured by detector **15** and subjected to a spectral analysis that produces the graph to the right in **FIG. 7C**. That graph is characterized not only by first fingerprint constituent **44** and second fingerprint constituent **48**, but by a third peak centered about wavelength λ_C. The latter serves as a third fingerprint constituent **52** for peptide **11** at wavelength λ_{SW}.

[**0098**] For emitted radiation **46** to have the illustrated effect on the tag on amino acid residue C, the tag on amino acid residue K that produced emitted radiation **46** must be located relatively proximately along protein molecule **10** to the tag on amino acid residue C.

[**0099**] Thus, as illustrated in the spectral diagram of **FIG.** 7C, when fluorescent tags are chemically attached on two different types of amino acid residues K and C, third primary electromagnetic excitation radiation **40** of wavelength λ_{SW} that excites tryptophan amino acid residue W will stimulate the emission of three corresponding fingerprint constituents for peptide **11** at wavelength λ_{SW}.

[**0100**] Fingerprint constituents **24**, **34**, **38**, **44**, **48**, and **52** together comprise one possible fingerprint for peptide **11**, or by comparison for protein molecule **10**. Fingerprint constituents are obtained for additional known protein molecules and collected in a computer database. The database then serves as a library of fingerprints for a set of protein molecules with tags on amino acid residues K, C when exposed to primary excitation radiations of wavelengths λ_{SC}, λ_{SK}, and λ_{SW}.

[**0101**] An unknown protein molecule is identified by chemically attaching the same types of tags to all corresponding types of amino acid residues of that protein molecule. Fingerprint constituents of the unknown protein molecule are determined using the methodology described. The fingerprint constituents of the unknown protein molecule are compared to the entries in the library of protein fingerprints. If a matching set of protein fingerprints is located in the library, the unknown protein molecule is identified.

[**0102**] The disclosed fingerprinting method can be used to rapidly identify a plurality of unknown protein molecules in a mixture of protein molecules, such as those contained within a sample cell.

[**0103**] In a variation of the fingerprinting method illustrated in FIGS. 4-7C, shown in **FIG. 8, a** single source S that emits a broad spectrum of wavelengths of primary electromagnetic excitation radiation **60**, including wavelengths λ_{SW}, λ_{SK}, and λ_{SC}, is used to simultaneously stimulate amino acid residues W, K, and C of peptide **11** to fluoresce.

[**0104**] The process also indirectly stimulates the tags on amino acid residues K and C of peptide **11** to fluoresce. **FIG. 8** includes a graph that depicts a corresponding portion of the response spectra for peptide **11** at wavelengths λ_{SW}, λ_{SK}, and λ_{SX}.

[**0105**] In **FIG. 8** it can be seen that the exposure of peptide **1** to primary electromagnetic excitation radiation **60** excites tryptophan amino acid residue W. This causes tryptophan amino acid residue W to fluoresce, producing emitted radiation **62** of wavelength λ_{WK}.

[**0106**] Emitted radiation **62** is radiation that is capable of exciting the tag on amino acid residue K to fluoresce. In addition, wavelength λ_{SK} of primary electromagnetic excitation radiation **60** from source S causes the tag on amino acid residue K to fluoresce. When stimulated either by primary electromagnetic excitation radiation **60** or by emitted radiation **62**, the tag on amino acid residue K produces emitted radiation **66** of wavelength λ_{KC}.

[0107] Emitted radiation **66** is capable of exciting the tag on amino acid residue C to fluoresce. In addition, wavelength λ_{SC} of primary electromagnetic excitation radiation **60** from source S causes the tag on amino acid residue C to fluoresce. When stimulated either by primary electromagnetic excitation radiation **60** or by emitted radiation **66**, the tag on amino acid residue C produces emitted radiation **70** of wavelength λ_C.

[0108] The intensities of emitted radiation **62**, **66**, **70** are measured by detector **15** and subjected to a spectral analysis that produces the graph to the right in **FIG. 8**. That graph is characterized by a first peak centered about wavelength $\lambda_{WK}/\lambda_{SK}$ that serves as a first fingerprint constituent **64** for peptide **11**, by a second peak centered about wavelength $\lambda_{KC}/\lambda_{KC}$ that serves as a second fingerprint constituent **68** for peptide **11**, and by a third peak centered about wavelength λ_C that serves as a first fingerprint constituent **72** for peptide **11** when peptide **11** is exposed to primary electromagnetic excitation radiation **60** that includes wavelengths λ_{SW}, λ_{SK}, and λ_{SC}.

[0109] In another aspect of the present invention, the characterization of a protein molecule in a manner that incorporates teachings of the present invention is but a step in a method in which proteins are isolated and characterized. **FIG. 9** is a flow chart that illustrates, by way of example and not by way of limitation, one such method. The characterization steps of **FIG. 9** are broader than the characterization steps illustrated in FIGS. **5-8**. Each of the boxes of the flow chart of **FIG. 9** represents a general step of the method. The order of the boxes is not intended to be limiting, neither is any single step illustrated, as some of the steps are optional.

[0110] In box **74**, the protein molecules in a sample are exposed to one or more chemicals to convert the protein molecules from native three-dimensional structures to linear, one-dimensional structures. These chemicals are referred to herein as denaturation means for linearizing the protein molecule. When the structure of a protein molecule has been modified in this manner, the protein molecule is said to have been "linearized."

[0111] By way of example and not limitation, linearization means according to the invention can include the use of chemicals that are known to be useful in linearizing a protein molecule. These could include, without limiting the scope of the invention, ionic detergents, such as SDS, and nondetergents, such as the chaotropic salts guanadinium and urea. The chemical known as β-mercaptoethanol, which breaks chemical bonds that can form between sulfur atoms of two amino acid residues, such as methionine and cysteine amino acid residues, can also be used as linearization means.

[0112] Alternatively, it may be desirable to analyze a protein molecule in the native three-dimensional structure thereof or without disrupting chemical bonds between sulfur atoms of two amino acid residues.

[0113] In box **76**, each of a first type of amino acid residue of the protein molecule is labeled with a first tag. The amino acid residues are labeled with tags, such as fluorescent tags or metallic tags. Specific tags can be attached to specific amino acid residues by way of known chemical reactions. Alternatively, the first type of amino acid residues of the protein molecule can be labeled prior to the linearization step depicted in box **74**.

[0114] In box **78**, the protein molecule or type of protein molecule to be examined is isolated from the other protein molecules or types of protein molecules in the sample. The isolation step depicted in box **78** can occur before or after the linearization step depicted in box **74**, or before or after the labeling step depicted in box **76**.

[0115] Once the protein molecule or type of protein molecule to be examined is isolated, the protein molecule can be characterized. In box **80**, a second type of amino acid residue of the protein molecule is detected. The second type of amino acid residue can itself be detected, or some signal generated by the second type of amino acid residue can be detected. When fluorescent dyes are employed as the tags on amino acid residues of the first type, the radiation generated due to the self-fluorescence of amino acid residue W when excited by radiation, such as radiation of wavelength λ_{SW}, is detected, using, for example, the methods illustrated in **FIGS. 7A and 8**.

[0116] Next, in box **82**, an interaction between the second type of amino acid residue and the first tag on the first type of amino acid residue is detected. For example, when the first tag is a fluorescent tag as illustrated in **FIGS. 7B and 8**, the second type of amino acid residue, in this example amino acid residue W, emits radiation of a wavelength λ_{WK}, which can excite the first tag on the first type of amino acid residue, in this example amino acid residue K.

[0117] The first tag on the first type of amino acid residue of the protein molecule is then detected, as depicted by box **84** of the flow chart of **FIG. 9**. When the first tag is a fluorescent tag, the first tag can be detected by the methods illustrated in **FIGS. 6A and 8**, wherein peptide **11** is exposed to excitation radiation having a wavelength λ_{SK} that will stimulate the first tag to emit detectable radiation.

[0118] In box **86**, the data obtained from each of the steps depicted by boxes **80**, **82**, and **84** is recorded. Data can be recorded in any manner known in the art, such as manually or in a computer database.

[0119] According to another aspect, the present invention includes a method for identifying an unknown protein molecule. **FIG. 10** is a flow chart that depicts exemplary steps that may be carried out to perform the method. Each of the boxes of the flow chart of **FIG. 10** represents a general step of the method. The order in which the boxes are presented in **FIG. 10** is not meant to be limiting. Moreover, not all of the steps are required in performing the method.

[0120] In box **90**, a mixture of protein molecules is obtained from a sample. Referring again to **FIG. 1**, biological sample **8** can be obtained from living organism **6** by known processes. Protein molecules **4*a***, **4*b***, **4*c*** can then be removed from biological sample **8** by extraction processes that are known to those in the art.

[0121] In box **92** of **FIG. 10**, the protein molecules obtained from a biological sample are linearized in the same manner as described in reference to box **74** of **FIG. 9**. The linearization step of box **92** is optional, as it may be desirable to leave the protein molecule completely or partially in the natural three-dimensional configuration thereof.

[**0122**] One or more of the types of amino acid residues of the protein molecule are labeled at box **94**. The amino acid residues are labeled with known tags, such as fluorescent tags or metallic tags. Specific tags can be attached to specific amino acid residues by way of known chemical reactions. Accordingly, different fluorescent dyes can be attached to different types of amino acid residues. Alternatively or in addition, one or more of the types of amino acid residues of a protein molecule can be labeled with metallic tags.

[**0123**] As an example, **FIG. 2** illustrates each amino acid residue K as having attached thereto a first tag **12**. Each amino acid residue C shown in **FIG. 2** has a second tag **14** thereon. When different tags are used on different types of amino acid residues, the numbers and relative locations of each of the different types of amino acid residues can be distinguished from each other. FIGS. **5-8** illustrate an example of how different tags on different types of amino acid residues are used to characterize a protein molecule.

[**0124**] Different types of amino acid residues in protein molecules are labeled using various methods. A first embodiment of such a method is shown in **FIG. 11** by way of illustration and not limitation. A cysteine reactive fluorescent tag **14** is attached to each amino acid residue C of peptide **11** by known processes. A different, lysine reactive fluorescent tag **12** is attached to each amino acid residue K of peptide **11**, also by known processes.

[**0125**] **FIG. 12** depicts, by way of illustration and not limitation, a second embodiment of labeling method conducted according to the teachings of the present invention. In the second embodiment, a cysteine reactive amino group **110** having the chemical formula —CH_2NH_2 is attached to each amino acid residue C of peptide **11**. Cysteine reactive amino group **110** permits lysine reactive groups to be attached to amino acid residue C. A lysine reactive fluorescent tag **12** is then attached to each amino acid residue K and to cysteine reactive amino group **110** on each amino acid residue C of peptide **11**. The emitted radiation from tags **12** is detected to provide a first fingerprint constituent of peptide **11**. After the emitted radiation of tag **12** is detected, tags **12** attached to cysteine reactive amino groups **110** on amino acid residues C can be removed by use of a hydrolyzing reagent **112**, as known in the art. Tags **12** remaining only on amino acid residues K can then be detected to provide a second fingerprint constituent of peptide **11**.

[**0126**] A third embodiment of labeling method according to teachings of the present invention is shown in **FIG. 13** by way of illustration and not limitation. First, a cysteine reactive blocking group **114** of a type known in the art is chemically attached to each amino acid residue C of peptide **11**. A lysine reactive fluorescent tag **12** is then attached to each amino acid residue K of peptide **11**. Blocking group **114** is then removed from each amino acid residue C by use of a hydrolyzing reagent **112**, as known in the art. Next, a different fluorescent tag **116** that can react with either amino acid residue K or amino acid residue C is then chemically attached to each amino acid residue C.

[**0127**] Returning to the inventive method illustrated in **FIG. 10**, box **96** depicts the isolation of a protein molecule from other protein molecules in the sample, or of a type of protein molecule from other types of protein molecules in the sample. The isolation of a protein molecule from other protein molecules or of a type of protein molecule from other types of protein molecules can occur before or after the linearization step depicted in box **92**, or before or after the labeling step depicted in box **94**.

[**0128**] Once the desired protein molecule or type of protein molecule has been isolated, the protein molecule or type of protein molecule is characterized at box **98**. In characterizing a protein molecule or type of protein molecule, a fingerprint of the protein molecule is determined. FIGS. **5-7**B and **8** illustrate examples of a method for determining the fingerprint of a protein molecule, in which amino acid residues of the protein molecule are labeled with fluorescent tags.

[**0129**] When the fingerprint of a protein molecule of interest has been determined, that fingerprint is compared at box **100** to the fingerprints in a library of fingerprints. The library of fingerprints has a listing of known protein molecules. Each of the known protein molecules in the listing has a corresponding fingerprint that was determined by the same processes used to determine the fingerprint of the protein molecule of interest. Thus, when the fingerprint of the protein molecule of interest matches with a fingerprint in the library, the protein molecule of interest is identified.

[**0130**] **FIG. 14** is a graph that illustrates the fluorescence spectra, or fingerprints **120**, **122**, of two different protein molecules. Fingerprints **120**, **122** were obtained by the method illustrated in **FIG. 8**. Fingerprint **120** is characteristic of the protein molecule known as bovine serum albumin (hereinafter "BSA"), while fingerprint **122** corresponds to the protein molecule known as alcohol dehydrogenase (hereinafter "ADH"). Each of fingerprints **120**, **122** includes a pair of major intensity peaks, respectively at the same wavelengths.

[**0131**] Fingerprints **120**, **122** are imparted to BSA and to ADH by way of the self-fluorescence of each of the tryptophan amino acid residues, the fluorescence of napthalenedicarboxyaldehyde (hereinafter "NDA") tags on each of the lysine amino acid residues, and the fluorescence of rhodamine tags on each of the cysteine amino acid residues of BSA and of ADH.

[**0132**] The present invention includes various approaches to isolating a protein molecule of interest and to determining a fingerprint of the protein molecule of interest. Structures capable of performing each of these functions are respectively referred to as isolation means for separating the protein molecule and detector means for detecting the fingerprint of the protein molecule. FIGS. **15-33** illustrate, by way of example and not limitation, various combinations of structures for performing the functions of an isolation means and a detector means according to teachings of the present invention, thereby to isolate and determine a fingerprint of a protein molecule of interest.

[**0133**] **FIG. 15** illustrates a first embodiment of detector means that can be used according to teachings of the present invention to isolate a protein molecule from other protein molecules in a sample for the purpose of characterizing or identifying the protein molecule. The apparatus depicted in **FIG. 15** is a hydrodynamic focusing apparatus **130** that isolates individual protein molecules from each other.

[**0134**] Hydrodynamic focusing apparatus **130** has a first region **132** into which sample **2** of one or more protein molecules is introduced. From first region **132**, sample **2** flows into a second region **134** of hydrodynamic focusing apparatus **130**, where the protein molecules in sample **2** are linearized and labeled with tags by way of chemicals introduced into second region **134** by way of inlets **135**. Next, sample **2** flows into a third region **136**, where the different types of protein molecules in sample **2** are separated from each other. Third region **136** can have therein a small separation column, a microelectrophoresis gel, or other apparatus known to be capable of separating different types of protein molecules. An eluent **137** that includes the different types of protein molecules **4***a*, **4***b*, **4***c* from sample **2** flows from the separation apparatus of third region **136** into a fourth region **138** of hydrodynamic focusing apparatus **130**. Different types of protein molecules **4***a*, **4***b*, **4***c* elute separately from the separation apparatus of third region **136** into fourth region **138**.

[0135] Eluent **137**, which contains separated types of protein molecules **4***a*, **4***b*, **4***c*, flows through fourth region **138**, into a laminar flow region **140** of hydrodynamic focusing apparatus **130**. Two opposing inlets **142** communicate with laminar flow region **140** to permit the introduction of a buffer **144** into laminar flow region **140**. Buffer **144** is introduced under pressure to create a laminar flow of buffer **144** and eluent **137** as the flow paths of eluent **137** and buffer **144** merge. Due to the laminar flow of buffer **144** into eluent **137**, eluent **137** flows in a thin layer **146** between two layers **148** of buffer **144**. When buffer **144** is introduced into laminar flow region **140** under sufficient pressure, individual protein molecules **4***a*, **4***b*, **4***c* are isolated by the laminar flow of buffer **144** into eluent **137**.

[0136] As protein molecules **4***a*, **4***b*, **4***c* flow in thin layer **146** through a detection region **149** of hydrodynamic focusing apparatus **130**, a fingerprint can be determined that is imparted to each of protein molecules **4***a*, **4***b*, **4***c* in accordance with teachings of the present invention.

[0137] **FIG. 16** depicts a second embodiment of an apparatus used according to teachings of the present invention to isolate and characterize a protein molecule. The apparatus of **FIG. 16** is an atomic force microscope **150**. Individual protein molecules **10** are separately analyzed with atomic force microscope **150** due to the atomic resolution of atomic force microscope **150**.

[0138] Protein molecules **10** are diluted and placed on a support **152** with distinct protein molecules **10** separated from one another. Support **152** is formed from a material, such as mica or glass, to which protein molecules **10** will adhere and upon which protein molecules **10** will be immobilized.

[0139] Atomic force microscope **150** has a cantilever **154** with a detector tip **156**. As detector tip **156** is brought into proximity with a selected protein molecule **10**, interactions between detector tip **156** and chemical structures of selected protein molecule **10** cause cantilever **154** to vibrate. The vibrations of cantilever **154** are measured by way of a laser detection system **158** that directs a laser beam L onto cantilever **154** and detects vibrations of laser beam L as the same is reflected by cantilever **154**. The measurements are used to determine the molecular weight and length of selected protein molecule **10**.

[0140] Detector tip **156** has attached thereto a fluorescent donor molecule **160**. An excitation laser **162** directs a laser beam M toward detector tip **156** as detector tip **156** is scanned along the length of selected protein molecule **10**. Laser beam M has a wavelength that excites fluorescent donor molecule **160** to produce emitted radiation λ_{160} of a wavelength that will excite a chosen tagged amino acid residue, such as tagged amino acid residue C shown in various of protein molecules **10** on support **152**. As donor molecule **160** on detector tip **156** is brought in an excited condition into proximity with amino acid residues C with fluorescent tags, emitted radiation λ_{160} excites the tags. The excitation of the tags on each amino acid residue C in selected protein molecule **10** causes detector tip **156** to vibrate, indicating the positions of each respective of the amino acid residues C along selected protein molecule **10**.

[0141] Alternatively, the fluorescent donor molecule **160** on the detector tip **156** of the atomic force microscope **150** could be chosen so as to emit a wavelength of electromagnetic radiation that excites tryptophan amino acid residues of a protein molecule or that excites fluorescent tags on each of another type of amino acid residue of the protein molecule.

[0142] A third embodiment of an apparatus that is useful for isolating and characterizing protein molecules according to teachings of the invention is shown in **FIGS. 17 and 18**.

[0143] **FIG. 17** schematically illustrates an apparatus including a separation plate **170**, a source S of electromagnetic radiation located on and directed toward a first side **172** of separation plate **170**, and a detector **15** located on and facing an opposite, second side **174** of separation plate **170**.

[0144] Separation plate **170** is a thin, planar structure with separation apertures **176** formed therethrough. Protein molecules **10** are propelled in solution or gel through apertures **176** by way of an electric field, a technique which is referred to in the art as electrophoresis. Each aperture **176** has a diameter D sized to permit a single linear protein molecule **10** to travel therethrough. For example, diameter D can be about 1-10 nm. The speed at which protein molecules **10** travel through apertures **176** depends on the applied electric field. Adjacent apertures **176** are spaced apart from one another so as to allow for optical resolution therebetween. For example, adjacent apertures **176** can be spaced about 1, 10, or 100 μm apart from each other. Separation plate **170** is formed from a material, such as silicon or plastic, that is opaque to visible wavelengths of electromagnetic radiation and that can be micromachined or otherwise modified by known processes to fabricate apertures **176** of desired diameter and spacing.

[0145] Source S emits electromagnetic excitation radiation **178** of wavelengths that will excite tryptophan amino acid residues or fluorescent tags on each of one or more other types of amino acid residues of protein molecules **10** passing through apertures **176**. For example, the lysine and cysteine amino acid residues of protein molecules **10** could be labeled with different fluorescent tags and source S could emit a broad range of wavelengths of electromagnetic radiation to produce the spectrum depicted in **FIG. 8**.

[0146] Detector **15** is positioned to detect emitted radiation from amino acid residues of protein molecule **10** as protein molecule **10** exits specific aperture **18** among the several apertures **176** on second side **174** of separation plate **170**. Electromagnetic radiation maybe focused on detector **15** by way of an optical lens. Known apparatus that detect electromagnetic radiation, such as a charge coupled device (hereinafter "CCD") or an avalanche photodiode array, can be employed as detector **15**. Detector **15** can simultaneously detect fluorescence emitted from amino acid residues or tags on the amino acid residues of different protein molecules **10**. A processor **16** receives signals from detector **15** and generates data about the particular type of radiation detected.

[**0147**] **FIG. 18** illustrates the phenomenon of "near field excitation" that occurs as protein molecules **10** that are exposed to electromagnetic radiation from source S on first side **172** of separation plate **170** travel through aperture **176**. Diameter D is smaller than the wavelengths of excitation radiation **178** from source S and emitted radiation from stimulated amino acid residues W or fluorescent tags on amino acid residues K and C. Thus, excitation radiation and emitted radiation on first side **172** of separation plate **170** does not pass through aperture **176**. As stimulated amino acid residue W or a stimulated tag on amino acid residue K or C exits aperture **176** on second side **174** of separation plate **170**, however, amino acid residue W or the fluorescent tag remains excited by radiation from source S until leaving a substantially conical volume having a diameter D and an approximate height D. This near field excitation, the emitted radiation on second side **174** of separation plate **170**, is detected by detector **15**. The signals from detector **15** are then characterized by processor **16** as representing a specific type of amino acid residue which, when taken along with the speed at which a protein molecule travels through aperture **176**, is located at a particular position along the length of protein molecule **10**.

[**0148**] A fourth embodiment of apparatus that isolates and characterizes protein molecules in accordance with teachings of the present invention is illustrated in FIGS. **19-27**.

[**0149**] **FIG. 19** depicts an isoelectric focusing gel **180** used to isolate different types of protein molecules in a sample on the basis of the relative ratio of positively charged regions to negatively charged regions of each type of protein molecule. For each type of protein molecule, this ratio is referred to as the "isoelectric point." The process of isolating protein molecules on the basis of isoelectric points is referred to as "isoelectric focusing."

[**0150**] Prior to being introduced into isoelectric focusing gel **180**, which has a web-like or matrix structure, the protein molecules are linearized to facilitate the travel of the protein molecules through the web of isoelectric focusing gel **180**. The protein molecules can be linearized with known chemicals, such as those discussed above in reference to **FIG. 9**. When SDS, a negatively charged, or anionic, detergent is used to linearize the protein molecules, each of the protein molecules is given a net negative charge. The protein molecules can also be labeled with tags, in the same manner described above in reference to **FIG. 9**. Alternatively, the protein molecules can be labeled with tags after different types of protein molecules have been separated from each other.

[**0151**] Once the linearized protein molecules have been introduced into isoelectric focusing gel **180**, isoelectric focusing gel **180** is placed in a pH gradient **182**. Each protein molecule in isoelectric focusing gel **180** migrates in the direction of arrow A along the length of isoelectric focusing gel **180** to a pH that equals the isoelectric point of that protein molecule.

[**0152**] Isoelectric focusing can be used to isolate different types of protein molecules alone or in combination with other separation techniques. Typically, isoelectric focusing is the first step in a two-step separation process that is referred to as "two-dimensional" separation.

[**0153**] **FIG. 20** illustrates the second step of two-dimensional separation, gel electrophoresis. In gel electrophoresis, native or linearized protein molecules are introduced into an electrophoresis gel **184**. As depicted, the protein molecules that were previously separated in isoelectric focusing gel **180** are introduced into electrophoresis gel **184** at an edge **186** thereof. Like isoelectric focusing gel **180**, electrophoresis gel **184** has a web-like structure. The passageways through electrophoresis gel **184** are, however, much smaller than the passageways through isoelectric focusing gel **180** so that the linearized protein molecules traveling through electrophoresis gel **184** are separated on the basis of size.

[**0154**] As an electric field **188** is applied to electrophoresis gel **184**, the linearized protein molecules travel through electrophoresis gel **184** in the direction of arrow A. Smaller protein molecules travel more quickly than larger protein molecules through the passageways of electrophoresis gel **184**.

[**0155**] **FIG. 21** illustrates an electrophoresis gel **184** having separate bands **190** of different types of protein molecules therein. Next, as shown in **FIG. 22**, the protein molecules in bands **190** are transferred to a membrane **192** of a material such as a nitrocellulose filter or polyvinylidene difluoride, which is also referred to as "PVDF," or to another known solid support, such as a vinyl or nylon support, for further testing, a process referred to in the art as "blotting-."**FIG. 23** illustrates membrane **192** and the bands **194** of different types of protein molecules thereon. Each of bands **194** corresponds to a similarly located band **190** of electrophoresis gel **184** in **FIG. 21**.

[**0156**] The amino acid residues of the protein molecules in one or more of bands **194** can be labeled after being isolated on isoelectric focusing gel **180** and electrophoresis gel **184** and following the transfer of the protein molecules from bands **190** on electrophoresis gel **184** to membrane **192**. The different types of protein molecules separated on membrane **192** are then characterized or identified and compared with one another.

[**0157**] As depicted in **FIG. 24**, when the protein molecules are labeled with fluorescent tags, electromagnetic excitation radiation **196** of one or more appropriate wavelengths can be directed from a source S toward membrane **192**. As the tryptophan amino acid residues or different tags on one or more different types of amino residues are stimulated by excitation radiation **196**, each of the tryptophan amino acid residues or tags emit radiation of a distinct wavelength range, in a similar manner to tryptophan amino acid residue W and the tags on amino acid residues K and C depicted in **FIG. 8**. The different ranges of wavelengths of emitted radiation can be separated from each other and from excitation radiation **196** by way of different optical filters **198**, **200**, **202** that permit only specific ranges of wavelengths of emitted radiation from amino acid residues W or from tags on amino acid residues K or C to pass therethrough.

[**0158**] When one optical filter **198** is used, the intensity of emitted radiation from a single type of amino acid residue or from the tags on a single type of amino acid residue can be detected, in this case, by a camera **204**. Camera **204** can be a digital camera that creates processable signals representative of the intensities of emitted radiation of different wavelengths or an optical camera.

[**0159**] **FIGS. 25, 26**, and **27** each illustrate a picture **206**, **208**, **210** of membrane **192** taken through a single optical filter **198**, **200**, **202**, respectively. The intensity of emitted radiation of each band **194** in each wavelength range represents the number of a particular type of amino acid residue in the protein molecule isolated in that band **194** and provides a fingerprint constituent of the type of protein molecule in that band **194**.

[**0160**] As illustrated, the protein molecules of some bands **194***a* and **194***b* do not give off a particular wavelength of emitted radiation. Bands **194***a* do not appear in picture **206** of **FIG. 24** and band **194***b* does not appear in picture **210** of **FIG. 27**.

[**0161**] FIGS. **28-33** depict a fifth embodiment of apparatus for isolating and characterizing a protein molecule. In the fifth embodiment, certain amino acid residues of protein molecules in a mixed sample are labeled with fluorescent tags in the manner described in reference to **FIG. 9**. The protein molecules of the mixed sample can optionally be linearized. The protein molecules in the mixed sample are separated from one another by diluting the sample. A drop **220** of the diluted sample is then placed on a microscope cover slip **222**, as shown in **FIG. 28**.

[**0162**] A microscope slide **224** with a recess **226** formed in a surface thereof is inverted over droplet **220** to enclose droplet **220** within recess **226** between microscope slide **224** and cover slip **222**. **FIG. 29** illustrates a microscope slide **224** prepared in this manner.

[**0163**] In **FIG. 30, a** fluorescence microscope **230** is used to detect the fluorescence of the tryptophan amino acid residues and of the fluorescent tags on amino acid residues of the protein molecules in droplet **220**. Fluorescence microscope **230** has a source S of excitation radiation **232** directed toward droplet **220** held by microscope slide **224**. Source S excites tryptophan amino acid residues and fluorescent tags on one or more other amino acid residues in the same manner as that depicted in **FIG. 8**.

[**0164**] Lens **234** of fluorescence microscope **230** magnifies emitted radiation from tryptophan amino acid residues and from the fluorescent tags. The magnified emitted radiation from the tryptophan amino acid residues and the fluorescent tags on one or more other types of amino acid residues of the protein molecules in droplet **220** then passes through an optical filter **236**, **238**, **240** that permits only a specific range of wavelengths of emitted radiation from tryptophan amino acid residues or from the tags on other types of amino acid residues to pass therethrough. Optical filters **236**, **238**, and **240** screen out unwanted wavelengths of emitted radiation. When one optical filter **236** is used, the intensity of emitted radiation from a single type of amino acid residue or from the tags on a single type of amino acid residue can be detected, in this case, by a camera **204** or visually through an eyepiece **242** of fluorescence microscope **230**.

[**0165**] FIGS. **31-33** illustrate different fields of view of a portion of droplet **220** through fluorescence microscope **230**. **FIG. 31** depicts a field of view **244** through optical filter **236**. **FIG. 32** depicts a field of view **246** through optical filter **238**. **FIG. 33** depicts a field of view **248** through optical filter **240**. As illustrated, protein molecules **10***a* emit radiation of some wavelengths and can, therefore, be seen in fields of view **246** and **248**, but not in field of view **244**. The intensities of emitted radiation from other protein molecules **10***b* also differ as protein molecules **10***b* are visualized through different filters. For example, the intensity of emitted radiation of a first wavelength from protein molecule **10***b* is greater in field of view **244** of **FIG. 31** than the intensity of emitted radiation of a second wavelength from protein molecule **10***b* visualized in field of view **246** of **FIG. 32** and than the intensity of emitted radiation of a third wavelength from protein molecule **10***b* in field of view **248** of **FIG. 33**, where protein molecule **10***b* does not appear to give off any emitted radiation of the third wavelength. Protein molecule **10***c* appears in each field of view **244**, **246**, **248**.

[**0166**] The invention maybe embodied in other specific forms without departing from its spirit or essential characteristics. The described embodiments are to be considered in all respects only as illustrative and not restrictive. The scope of the invention is, therefore, indicated by the appended claims rather than by the foregoing description. All changes which come within the meaning and range of equivalency of the claims are to be embraced within their scope.

What is claimed is:

1. A system for determining a fingerprint of a protein molecule, the protein molecule including an amino acid residue of a first type and an amino acid residue of a second type, said system comprising:

(a) denaturation means for linearizing the protein molecule;

(b) labeling means for attaching a tag to the amino acid residue of the first type in the protein molecule; and

(c) detector means for determining the fingerprint of the protein molecule, the fingerprint comprising:

(i) a first fingerprint constituent imparted to the protein molecule by said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule; and

(ii) a second fingerprint constituent imparted to the protein molecule by the amino acid residue of the second type in the protein molecule.

2. A system as recited in claim 1, wherein a detergent is utilized by said denaturation means to linearize the protein molecule.

3. A system as recited in claim 2, wherein sodium dodecyl sulfate is utilized by said denaturation means to linearize the protein molecule.

4. A system as recited in claim 1, wherein a chaotropic salt is utilized by said denaturation means to linearize the protein molecule.

5. A system as recited in claim 1, wherein said tag attached to the amino acid residue of the first type by said labeling means comprises a fluorescent dye, said fluorescent dye producing an emitted radiation when stimulated.

6. A system as recited in claim 5, wherein said detector means comprises:

(a) a source of a primary excitation radiation capable of stimulating said fluorescent dye to produce said emitted radiation; and

(b) a detector sensitive to said emitted radiation from said fluorescent dye.

7. A system as recited in claim 6, wherein said detector comprises a charge coupled device.

8. A system as recited in claim 1, wherein said detector means comprises:

(a) a first source of electromagnetic radiation capable of stimulating said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule; and

(b) a second source of electromagnetic radiation capable of stimulating the amino acid residue of the second type in the protein molecule.

9. A system as recited in claim 8, wherein said detector means further comprises a charge coupled device so positioned relative to said first source and said second source as to detect:

(a) emitted radiation from said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule; and

(b) emitted radiation from the amino acid residue of the second type in the protein molecule.

10. A system as recited in claim 8, wherein said detector means further comprises a camera so positioned relative to said first source and said second source as to detect:

(a) excitation radiation from said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule; and

(b) emitted radiation from the amino acid residue of the second type in the protein molecule.

11. A system as recited in claim 8, wherein said detector means further comprises a microscope so positioned relative to said first source and said second source as to detect:

(a) excitation radiation from said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule; and

(b) emitted radiation from the amino acid residue of the second type in the protein molecule.

12. A system as recited in claim 1, wherein said detector means comprises:

(a) a source of electromagnetic radiation, said electromagnetic radiation being:

(i) capable of stimulating said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule, and

(ii) capable of stimulating the amino acid residue of the second type in the protein molecule; and

(b) a detector, said detector being:

(i) sensitive to emitted radiation from said first tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule, and

(ii) sensitive to emitted radiation from the amino acid residue of the second type in the protein molecule.

13. A system as recited in claim 1, wherein said detector means comprises an atomic force microscope.

14. A system as recited in claim 13, wherein said atomic force microscope comprises:

(a) a detector tip;

(b) a donor tag attached to said detector tip, said donor tag being capable of stimulating said tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule and said detector tip is in proximity with said tag.

15. A system as recited in claim 1, wherein said detector means comprises a nuclear magnetic resonance apparatus.

16. A system as recited in claim 15, wherein:

(a) a chemical precursor is utilized by said labeling means to attach said tag to the amino acid residue of the first type in the protein molecule; and

(b) said tag attached by said labeling means to the amino acid residue of the first type in the protein molecule is a metal tag.

17. A system as recited in claim 1, wherein said first type of amino acid residue in the protein molecule comprises an amino acid residue selected from a group consisting of cysteine and lysine.

18. A system as recited in claim 1, wherein the second type of amino acid residue in the protein molecule comprises tryptophan.

19. A system for identifying a selected protein molecule in a sample containing a plurality of protein molecules, the selected protein molecule including at least one amino acid residue of a first type and at least one amino acid residue of a second type, said system comprising:

(a) denaturation means for linearizing the protein molecules in the sample;

(b) isolation means for separating the selected protein molecule from other protein molecules in the sample;

(c) labeling means for attaching a tag to the amino acid residue of the first type in the protein molecule;

(d) detector means for determining a fingerprint for the selected protein molecule, said fingerprint comprising:

(i) a first fingerprint constituent imparted to the selected protein molecule by said first tag when said tag is attached by said labeling means to the amino acid residue of the first type in the protein molecule; and

(ii) a second fingerprint constituent imparted to the selected protein molecule by the amino acid residue of the second type in the protein molecule; and

(e) collation means for comparing said fingerprint to a library of fingerprints for known protein molecules.

20. A system as recited in claim 19, wherein said isolation means comprises a hydrodynamic focusing apparatus.

21. A system as recited in claim 19, wherein said isolation means comprises:

(a) a separation plate opaque to light; and

(b) apertures formed through said separation plate.

22. A system as recited in claim 21, wherein each of said apertures has a diameter of from about 1 nanometer to about 10 nanometers.

23. A system as recited in claim 21, wherein said separation plate is formed from a material selected from the group consisting of silicon and opaque plastics.

24. A system as recited in claim 19, wherein said isolation means comprises an electrophoresis gel.

25. A system as recited in claim 19, wherein said isolation means comprises a dilute solution.

26. A system as recited in claim 19, wherein said collation means comprises a computer processor for signals from said detector means.

27. A system for determining a fingerprint of a protein molecule having a first type of amino acid residue and a second type of amino acid residue, said system comprising:

(a) a protein molecule linearization region;

(b) a tag attachment region, said tag attachment region being in fluid association with said linearization region, in said tag attachment region a tag becoming attached to each first type of amino acid residue in the protein molecule; and

(c) a detector region in fluid association with and located downstream of said linearization region and said tag attachment region, in said detector region constituents of the fingerprint of the protein molecule being sensed, said constituents of the fingerprint of the protein molecule comprising:

(i) a first fingerprint constituent of the fingerprint of the protein molecule imparted to the protein molecule by said tag; and

(ii) a second fingerprint constituent of the fingerprint of the protein molecule imparted to the protein molecule by the second type of amino acid residue in the protein molecule.

28. A system as recited in claim 27, wherein said linearization region is located upstream of said tag attachment region.

29. A protein molecule having an identifiable fingerprint, said protein molecule comprising:

(a) a first amino acid residue, said first amino acid residue being of a first type;

(b) a first tag attached to said first amino acid residue of said protein molecule, said first tag imparting to said protein molecule a first fingerprint constituent;

(c) a second amino acid residue, said second amino acid residue being of a second type; and

(d) a second tag attached to said second amino acid residue of the protein molecule, said second tag imparting to said protein molecule a second fingerprint constituent.

30. A protein molecule as recited in claim 29, wherein said protein molecule further comprises a third amino acid residue of a third type, said third amino acid residue imparting to said protein molecule a third fingerprint constituent.

31. A protein molecule as recited in claim 30, wherein said third type of amino acid residue comprises tryptophan.

32. A protein molecule as recited in claim 29, wherein said protein molecule comprises a single said first amino acid residue.

33. A protein molecule as recited in claim 29, wherein said protein molecule comprises:

(a) a plurality of said first amino acid residues; and

(b) a plurality of said first tags, individual of said first tags being attached to corresponding of said first amino acid residues, said plurality of said first tags collectively imparting to said protein molecule said first fingerprint constituent.

34. A protein molecule as recited in claim 30, wherein said protein molecule comprises a single said third amino acid residue.

35. A protein molecule as recited in claim 30, wherein said protein molecule comprises a plurality of said third amino acid residues.

36. A fingerprint for a protein molecule, the protein molecule including at least one first amino acid residue of a first type, at least one second amino acid residue of a second type, and at least one tryptophan amino acid residue, said fingerprint comprising:

(a) a first fingerprint constituent imparted to said protein molecule by a first tag on each first amino acid residue, said first fingerprint constituent comprising emitted radiation of a first wavelength produced by said first tag when said first tag is stimulated with a first primary excitation radiation;

(b) a second fingerprint constituent imparted to said protein molecule by a second tag on each second amino acid residue, said second fingerprint constituent comprising emitted radiation of a second wavelength produced by said second tag when said second tag is stimulated with a second primary excitation radiation;

(c) a third fingerprint constituent imparted to said protein molecule by said first tag on each first amino acid residue, said third fingerprint constituent comprising a first secondary emitted radiation of said first wavelength produced by said first tag when said first tag is stimulated by said emitted radiation of said second wavelength;

(d) a fourth fingerprint constituent imparted to said protein molecule by each tryptophan amino acid residue, said fourth fingerprint constituent comprising emitted radiation of a third wavelength produced by the tryptophan amino acid residue when the said tryptophan amino acid residue is stimulated with a third primary excitation radiation;

(e) a fifth fingerprint constituent imparted to said protein molecule by said second tag on each second amino acid residue, said fifth fingerprint constituent comprising a secondary emitted radiation of said second wavelength produced by said second tag when said second tag is stimulated by said emitted radiation of said third wavelength; and

(f) a sixth fingerprint constituent imparted to said protein molecule by said first tag on each first amino acid residue, said sixth fingerprint comprising a second secondary emitted radiation of said first wavelength produced by said first tag when said first tag is stimulated by said secondary emitted radiation of said second wavelength.

37. A library of fingerprint values for a set of known protein molecules, said library comprising:

(a) a listing of the identity of each protein molecule of the set of known protein molecules;

(b) a first fingerprint constituent corresponding to each of said protein molecules of the set of known protein molecules, said first fingerprint constituent being representative of the number and sequence of a first type of amino acid residue in each of said protein molecules of the set of known protein molecules; and

(c) a second fingerprint constituent corresponding to each of said protein molecules of the set of known protein molecules, said second fingerprint constituent representative of the number and sequence of a second type of amino acid residue in each of said protein molecules of the set of known protein molecules.

38. A library as recited in claim 37, further comprising a third fingerprint constituent corresponding to each of said protein molecules of the set of known protein molecules, said third fingerprint constituent being representative of the number and sequence of a third amino acid residue in each of said protein molecules of the set of known protein molecules.

* * * * *