(12) STANDARD PATENT (11) Application No. **AU 2004292226 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Proteome epitope tags and methods of use thereof in protein modification analysis

(51) International Patent Classification(s)
C12Q 1/26 (2006.01) ***G01N 33/68*** (2006.01)
C12Q 1/34 (2006.01) ***G06F 19/00*** (2006.01)
C12Q 1/48 (2006.01)

(21) Application No: **2004292226** (22) Date of Filing: **2004.11.15**

(87) WIPO No: **WO05/050223**

(30) Priority Data

(31) Number	(32) Date	(33) Country
10/712,425	**2003.11.13**	**US**

(43) Publication Date: **2005.06.02**
(44) Accepted Journal Date: **2008.06.26**

(71) Applicant(s)
Epitome Biosystems Inc.

(72) Inventor(s)
Meng, Xun;Lee, Frank D.;Livingston, David

(74) Agent / Attorney
Shelston IP, Level 21 60 Margaret Street, Sydney, NSW, 2000

(56) Related Art
US 2002/137119
AU 2003302118

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 June 2005 (02.06.2005)

PCT

(10) International Publication Number
WO 2005/050223 A3

(51) International Patent Classification[7]: G06F 19/00, G01N 33/68

(21) International Application Number:
PCT/US2004/038283

(22) International Filing Date:
15 November 2004 (15.11.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/712,425 13 November 2003 (13.11.2003) US

(71) Applicant *(for all designated States except US)*: **EPITOME BIOSYSTEMS INC.** [US/US]; 100 Beaver Street, Waltham, MA 02453 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEE, Frank, D.** [US/US]; 6 Norfolk Road, Chestnut Hill, MA 02467 (US). **MENG, Xun** [CN/US]; 11 George Street, Newton, MA 02458 (US). **LIVINGSTON, David** [US/US]; 206 Rumstick Road, Barrington, RI 02806 (US).

(74) Agents: **VINCENT, Matthew, P.** et al.; Ropes & Gray_LLP, One International Place, Boston, MA 02110-2624 (US).

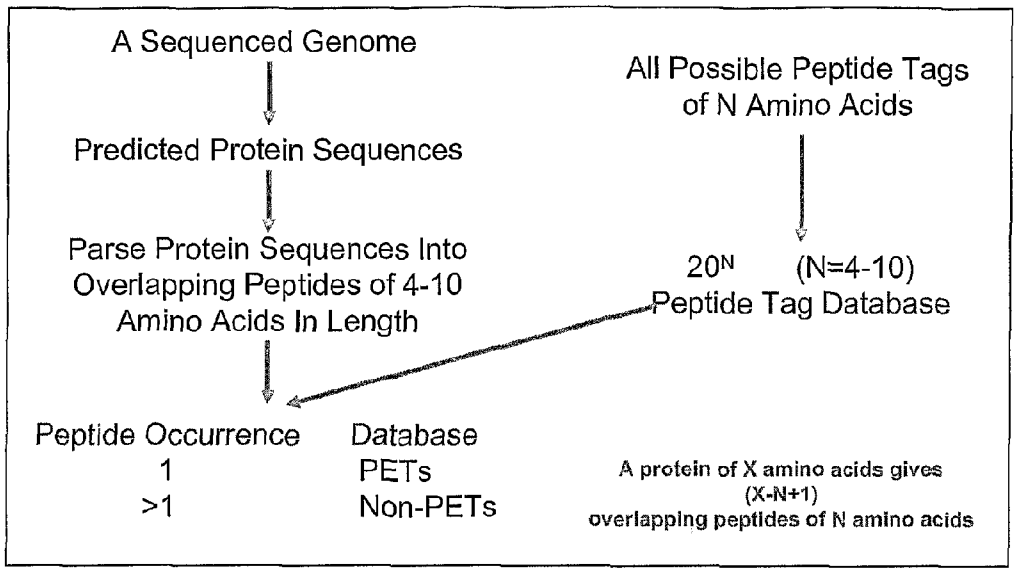
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: PROTEOME EPITOPE TAGS AND METHODS OF USE THEREOF IN PROTEIN MODIFICATION ANALYSIS



(57) Abstract: Disclosed are methods for reliably detecting the presence of proteins, especially proteins with various post-translational modifications (phosphorylation, glycosylation, methylation, acetylation, etc.) in a sample by the use of one or more capture agents that recognize and interact with recognition sequences uniquely characteristic of a set of proteins (Proteome Epitope Tags, or PETs) in the sample. Arrays comprising these capture agents or PETs are also provided.

WO 2005/050223 A3

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
9 February 2006

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PROTEOME EPITOPE TAGS AND METHODS OF USE THEREOF IN PROTEIN MODIFICATION ANALYSIS

Background of the Invention

Genomic studies are now approaching "industrial" speed and scale, thanks to advances in gene sequencing and the increasing availability of high-throughput methods for studying genes, the proteins they encode, and the pathways in which they are involved. The development of DNA microarrays has enabled massively parallel studies of gene expression as well as genomic DNA variations.

DNA microarrays have shown promise in advanced medical diagnostics. More specifically, several groups have shown that when the gene expression patterns of normal and diseased tissues are compared at the whole genome level, patterns of expression characteristic of the particular disease state can be observed. Bittner et al., (2000) Nature 406:536–540; Clark et al., (2000) Nature 406:532–535; Huang et al., (2001) Science 294:870–875; and Hughes et al., (2000) Cell 102:109–126. For example, tissue samples from patients with malignant forms of prostate cancer display a recognizably different pattern of mRNA expression to tissue samples from patients with a milder form of the disease. C.f., Dhanasekaran et al., (2001) Nature 412 (2001), pp. 822–826.

However, as James Watson pointed out recently proteins are really the "actors in biology" (*"A Cast of Thousands" Nature Biotechnology March 2003).* A more attractive approach would be to monitor key proteins directly. These might be biomarkers identified by DNA microarray analysis. In this case, the assay required might be relatively simple, examining only 5–10 proteins. Another approach would be to use an assay that detects hundreds or thousands of protein features, such as for the direct analysis of blood, sputum or urine samples, etc. It is reasonable to believe that the body would react in a specific way to a particular disease state and produce a distinct "biosignature" in a complex data set, such as the levels of 500 proteins in the blood. One could imagine that in the future a single blood test could be used to diagnose most conditions.

The motivation for the development of large-scale protein detection assays as

basic research tools is different to that for their development for medical diagnostics. The utility of biosignatures is one aspect researchers desire in order to understand the molecular basis of cellular response to a particular genetic, physiological or environmental stimulus. DNA microarrays do a good job in this role, but detection of proteins would allow for more accurate determination of protein levels and, more importantly, could be designed to quantitate the presence of different splice variants or isoforms. These events, to which DNA microarrays are largely or completely blind, often have pronounced effects on protein activities.

This has sparked great interest in the development of devices such as protein-detecting microarrays (PDMs) to allow similar experiments to be done at the protein level, particularly in the development of devices capable of monitoring the levels of hundreds or thousands of proteins simultaneously.

Prior to the present invention, PDMs that even approach the complexity of DNA microarrays do not exist. There are several problems with the current approaches to massively parallel, *e.g.*, cell-wide or proteome wide, protein detection. First, reagent generation is difficult: One needs to first isolate every individual target protein in order to isolate a detection agent against every protein in an organism and then develop detection agents against the purified protein. Since the number of proteins in the human organism is currently estimated to be about 30,000 this requires a lot of time (years) and resources. Furthermore, detection agents against native proteins have less defined specificity since it is a difficult task to know which part of the proteins the detection agents recognize. This problem causes considerable cross-reactivity of when multiple detection agents are arrayed together, making large-scale protein detection array difficult to construct. Second, current methods achieve poor coverage of all possible proteins in an organism. These methods typically include only the soluble proteins in biological samples. They often fail to distinguish splice variants, which are now appreciated as being ubiquitous. They exclude a large number of proteins that are bound in organellar and cellular membranes or are insoluble when the sample is processed for detection. Third, current methods are not general to all proteins or to all types of biological samples. Proteins vary quite widely in their chemical character. Groups of proteins require different processing conditions in order to keep them stably solubilized for

detection. Any one condition may not suit all the proteins. Further, biological samples vary in their chemical character. Individual cells considered identical express different proteins over the course of their generation and ultimate death. Physiological fluids like urine and blood serum are relatively simple, but biopsy tissue samples are very complex. Different protocols need to be used to process each type of sample and achieve maximal solubilization and stabilization of proteins.

Current detection methods are either not effective over all proteins uniformly or cannot be highly multiplexed to enable simultaneous detection of a large number of proteins (*e.g.*, > 5,000). Optical detection methods would be most cost effective but suffer from lack of uniformity over different proteins. Proteins in a sample have to be labeled with dye molecules and the different chemical character of proteins leads to inconsistency in efficiency of labeling. Labels may also interfere with the interactions between the detection agents and the analyte protein leading to further errors in quantitation. Non-optical detection methods have been developed but are quite expensive in instrumentation and are very difficult to multiplex for parallel detection of even moderately large samples (*e.g.*, > 100 samples).

Another problem with current technologies is that they are burdened by intracellular life processes involving a complex web of protein complex formation, multiple enzymatic reactions altering protein structure, and protein conformational changes. These processes can mask or expose binding sites known to be present in a sample. For example, prostate specific antigen (PSA) is known to exist in serum in multiple forms including free (unbound) forms, *e.g.*, pro-PSA, BPSA (BPH-associated free PSA), and complexed forms, *e.g.*, PSA-ACT, PSA-A2M (PSA-alpha$_2$-macroglobulin), and PSA-API (PSA-alpha$_1$-protease inhibitor) (see Stephan C. *et al.* (2002) *Urology* 59:2-8). Similarly, Cyclin E is known to exist not only as a full length 50 kD protein, but also in five other low molecular weight forms ranging in size from 34 to 49 kD. In fact, the low molecular weight forms of cyclin E are believed to be more sensitive markers for breast cancer than the full length protein (see Keyomarsi K. *et al.* (2002) *N. Eng. J. Med.* 347(20):1566-1575).

Sample collection and handling prior to a detection assay may also affect the nature of proteins that are present in a sample and, thus, the ability to detect these

proteins. As indicated by Evans M. J. *et al.* (2001) *Clinical Biochemistry* 34:107-112 and Zhang D. J. *et al.* (1998) *Clinical Chemistry* 44(6):1325-1333, standardizing immunoassays is difficult due to the variability in sample handling and protein stability in plasma or serum. For example, PSA sample handling, such as sample freezing, affects the stability and the relative levels of the different forms of PSA in the sample (Leinonen J, Stenman UH (2000) *Tumour Biol.* 21(1):46-53).

Finally, current technologies are burdened by the presence of autoantibodies which affect the outcome of immunoassays in unpredictable ways, *e.g.*, by leading to analytical errors (Fitzmaurice T. F. *et al.* (1998) *Clinical Chemistry* 44(10):2212-2214).

These problems prompted the question whether it is even possible to standardize immunoassays for hetergenous protein antigens. (Stenman U-H. (2001) Immunoassay Standardization: Is it possible? Who is responsible? Who is capable? Clinical Chemistry 47 (5) 815-820). Thus, a great need exists in the art for efficient and simple methods of parallel detection of proteins that are expressed in a biological sample and, particularly, for methods that can overcome the imprecisions caused by the complexity of protein chemistry and for methods which can detect all or a majority of the proteins expressed in a given cell type at a given time, or for proteome-wide detection and quantitation of proteins expressed in biological samples.

2004292226 05 Jun 2008

Summary of the Invention

According to a broad aspect, the present invention provides a method for detecting the presence or absence of a post-translational modification at a location on a target protein within a sample, comprising:

a. computationally analysing an amino acid sequence of said target protein to identify one or more potential sites for said post-translational modification;

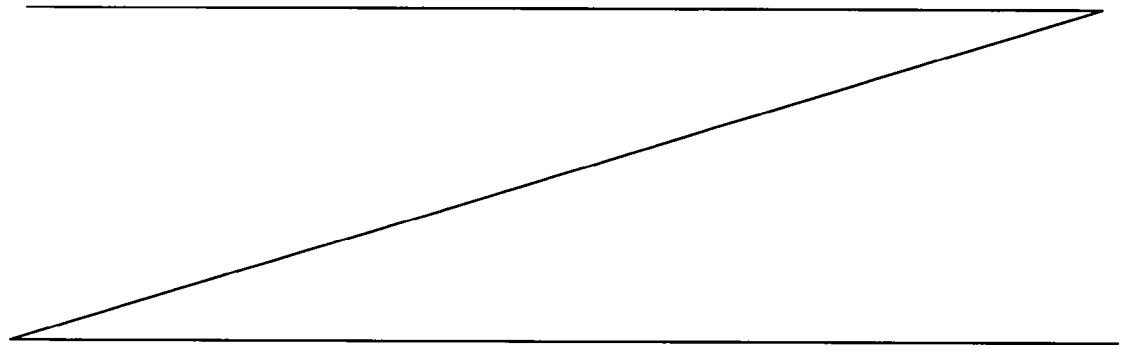
b. computationally identifying an amino acid sequence of at least one fragment of said target protein, said fragment predictably resulting from a treatment of said target protein within said sample, and said fragment comprising at least one of said potential post-translational modification sites and, separate therefrom, a PET (proteome epitope tag) unique to said fragment within said sample;

c. generating a capture agent that specifically binds said PET separate from said post-translational modification on said fragment, and immobilizing said capture agent to a support;

d. subjecting said sample to the treatment to produce said fragment, rendering said fragment soluble in solution, and contacting said fragment with said capture agent to bind said fragment, at said PET, to said capture agent;

e. detecting, on said fragment bound to said capture agent, the presence or absence of said post-translational modification by using a secondary capture agent specific for said post-translational modification separate from said PET on said fragment, wherein said secondary capture agent is labelled by a detectable moiety.

NEXT PAGE: PAGE 5e

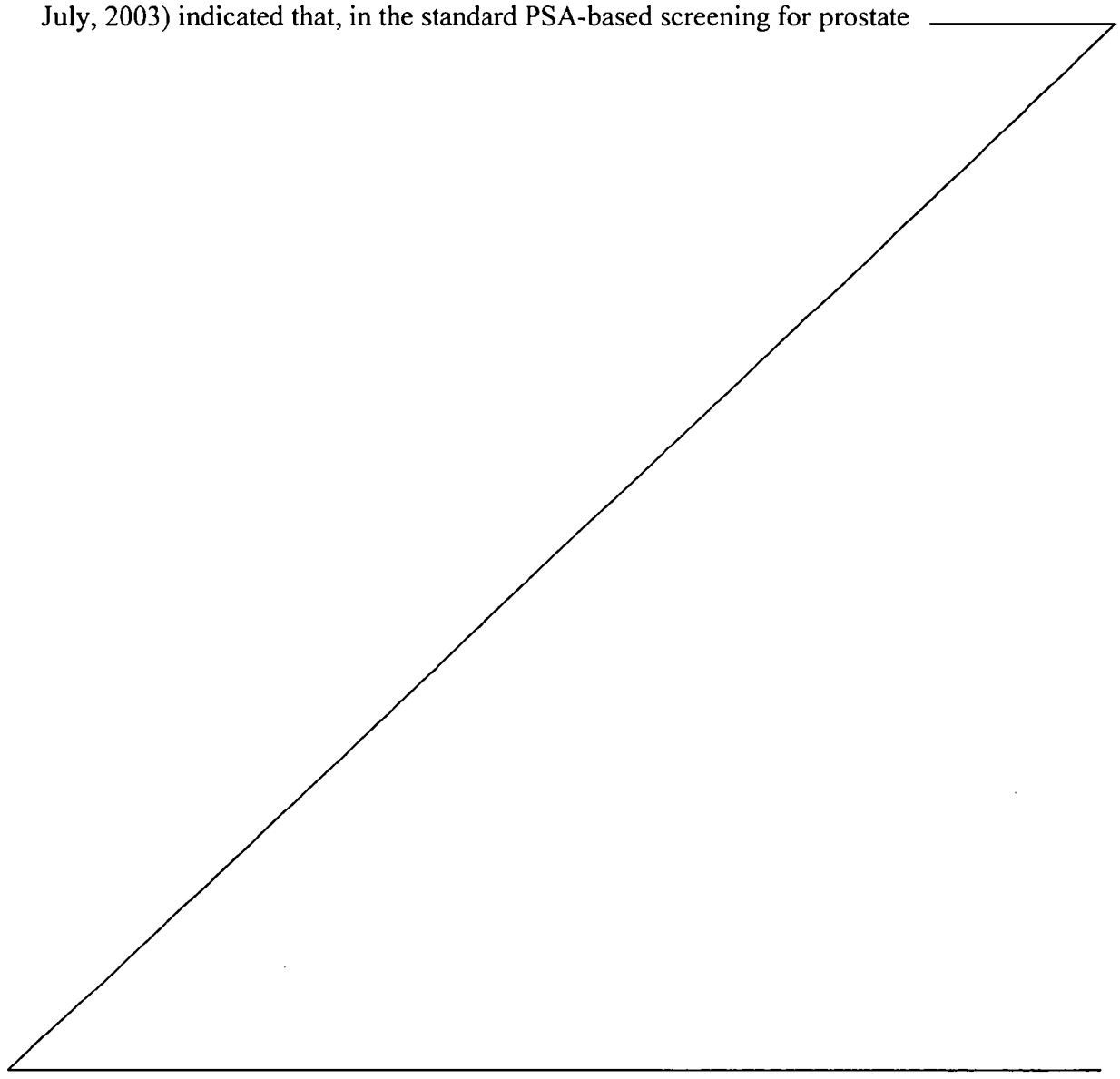


Unless the context clearly requires otherwise, throughout the description and the claims, the words "comprise", "comprising", and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in the sense of "including, but not limited to".

The present invention is directed to methods and reagents for reproducible protein detection and quantitation, *e.g.*, parallel detection and quantitation, in complex biological samples. Salient features to certain embodiments of the present invention reduce the complexity of reagent generation, achieve greater coverage of all protein classes in an organism, greatly simplify the sample processing and analyte stabilisation process, and enable effective and reliable parallel detection, *e.g.* by optical or other automated detection methods, and quantitation of proteins and/or post-translationally modified forms, and, enable multiplexing of standardized capture agents for proteins with minimal cross-reactivity and well-defined specificity for large-scale, proteome-wide protein detection.

Embodiments of the present invention also overcome the imprecisions in detection methods caused by: the existence of proteins in multiple forms in a sample (*e.g.* various post-translationally modified forms of various complexed or aggregated forms); the variability in sample handling and protein stability in a sample, such as plasma or serum; and the presence of autoantibodies in samples. In certain embodiments, using a targeted fragmentation protocol, the methods of the present invention assure that a binding site on a protein of interest, which may have been masked due to one of the foregoing reasons, is made available to interact with a capture agent. In other embodiments, the sample proteins are subjected to conditions in which they are denatured, and optionally are alkylated, so as to render buried (or otherwise cryptic) PET moieties accessible to solvent and interaction with capture agents. As a result, the present invention allows for detection methods having increased sensitivity and more accurate protein quantitation capabilities. This advantage of the present invention will be particularly useful in, for example, protein marker-type disease detection assays (*e.g.* PSA or Cyclin E based assays) as it will allow for an improvement in the predictive value, sensitivity, and reproducibility of these assays. The present invention can standardize detection and measurement assays for all proteins from all samples.

For example, a recent study by Punglia et al. (*N. Engl. J. Med.* 349(4): 335-42, July, 2003) indicated that, in the standard PSA-based screening for prostate



cancer, if the threshold PSA value for undergoing biopsy were set at 4.1 ng per milliliter, 82 percent of cancers in younger men and 65 percent of cancers in older men would be missed. Thus a lower threshold level of PSA for recommending prostate biopsy, particularly in younger men, may improve the clinical value of the PSA test. However, at lower detection limits, background can become a significant issue. It would be immensly advantageous if the sensitivity / selectivity of the assay can be improved by, for example, the method of the instant invention.

In a specific embodiment, the invention provides a method to detect and quantitate the presence of specific modified polypeptides in a sample. In a general sense, the invention provides a method to identify a URS or PET uniquely associated with a modification site on a peptide fragment, which PET can then be captured and detected / quantitated by specific capture agents. The method applies to virtually all kinds of post-translational modifications, including but are not limited to phosphorylation, glycosylation, etc., as long as the modification can be reliably detected, for example, by phospho-antibodies. The method also applies to the detection of alternative splicing forms of otherwise identical proteins.

The present invention is based, at least in part, on the realization that exploitation of unique recognition sequences (URSs) or Proteome Epitope Tags (PETs) present within individual proteins can enable reproducible detection and quantitation of individual proteins in parallel in a milieu of proteins in a biological sample. As a result of this PET-based approach, the methods of the invention detect specific proteins in a manner that does not require preservation of the whole protein, nor even its native tertiary structure, for analysis. Moreover, the methods of the invention are suitable for the detection of most or all proteins in a sample, including insoluble proteins such as cell membrane bound and organelle membrane bound proteins.

The present invention is also based, at least in part, on the realization that PETs can serve as Proteome Epitope Tags characteristic of a specific organism's proteome and can enable the recognition and detection of a specific organism.

The present invention is also based, at least in part, on the realization that high-affinity agents (such as antibodies) with predefined specificity can be generated

for defined, short length peptides and when antibodies recognize protein or peptide epitopes, only 4-6 (on average) amino acids are critical. See, for example, Lerner RA (1984) Advances In Immunology. 36:1-45.

The present invention is also based, at least in part, on the realization that by denaturing (including thermo- and/or chemical- denaturation) and/or fragmenting (such as by protease digestion including digestion by thermo-protease) all proteins in a sample to produce a soluble set of protein analytes, e.g., in which even otherwise buried PETs including PETs in protein complexes / aggregates are solvent accessible, the subject method provides a reproducible and accurate (intra-assay and inter-assay) measurement of proteins.

The present invention is also based, at least in part, on the realization that protein modifications associated with PETs on a fragmented peptide can be readily detected and quantitated by isolating the associated PET followed by detection / quantitation of the modification.

Accordingly, in one aspect, the present invention provides a method for globally detecting the presence of a protein(s) (*e.g.*, membrane bound protein(s)) in an organism's proteome. The method includes providing a sample which has been denatured and/or fragmented to generate a collection of soluble polypeptide analytes; contacting the polypeptide analytes with a plurality of capture agents (*e.g.*, capture agents immobilized on a solid support such as an array) under conditions such that interaction of the capture agents with corresponding unique recognition sequences occurs, thereby globally detecting the presence of protein(s) in an organism's proteome.

The method is suitable for use in, for example, diagnosis (*e.g.*, clinical diagnosis or environmental diagnosis), drug discovery, protein sequencing or protein profiling. In one embodiment, at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of an organism's proteome is detectable from arrayed capture agents.

The capture agent may be a protein, a peptide, an antibody, *e.g.*, a single chain antibody, an artificial protein, an RNA or DNA aptamer, an allosteric ribozyme, a small molecule or electronic means of capturing a PET.

The sample to be tested (*e.g.*, a human, yeast, mouse, C. elegans, Drosophila melanogaster or Arabidopsis thaliana sample, such whole cell lysate) may be fragmented by the use of a proteolytic agent. The proteolytic agent can be any agent, which is capable of predictably cleaving polypeptides between specific amino acid residues (i.e., the proteolytic cleavage pattern). The predictability of cleavage allows a computer to generate fragmentation patterns in sillico, which will greatly aid the process of searching PETs unique to a sample.

According to one embodiment of this aspect of the present invention a proteolytic agent is a proteolytic enzyme. Examples of proteolytic enzymes, include but are not limited to trypsin, calpain, carboxypeptidase, chymotrypsin, V8 protease, pepsin, papain, subtilisin, thrombin, elastase, gluc-C, endo lys-C or proteinase K, caspase-1, caspase-2, caspase-3, caspase-4, caspase-5, caspase-6, caspase-7, caspase-8, MetAP-2, adenovirus protease, HIV protease and the like.

The following table summarizes the result of analyzing pentamer PETs in the human proteome using different proteases. A total of 23,446 sequences are tagged before protease digestion.

Protease	Cleavage Site	Fragment Length	Tagged Proteins
Chymotrypsin	after W,F,Y	12.7	21,990
S.A. V-8 E specific	after E	13.7	23,120
Post-Proline Cleaving Enzyme	after P	15.7	23,009
Trypsin	after K, R	8.5	22,408

According to another embodiment of this aspect of the present invention a proteolytic agent is a proteolytic chemical such as cyanogen bromide and 2-nitro-5-thiocyanobenzoate. In still other embodiments, the proteins of the test sample can be fragmented by physical shearing; by sonication, or some combination of these or other treatment steps.

An important feature for certain embodiments, particularly when analyzing complex samples, is to develop a fragmentation protocol that is known to reproducibly generate peptides, preferably soluble peptides, which serve as the unique recognition sequences. The collection of polypeptide analytes generated from the fragmentation may be 5-30, 5-20, 5-10, 10-20, 20-30, or 10-30 amino acids long, or longer. Ranges intermediate to the above recited values, *e.g.*, 7-15 or 15-25 are also intended to be part of this invention. For example, ranges using a combination of any of the above recited values as upper and/or lower limits are intended to be included.

The unique recognition sequence may be a linear sequence or a non-contiguous sequence and may be 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, or 30 amino acids in length. In certain embodiments, the unique recognition sequence is selected from the group consisting of SEQ ID NOs:1-546 or a sub-collection thereof.

In one embodiment, the protein(s) being detected is characteristic of a pathogenic organism, *e.g.*, anthrax, small pox, cholera toxin, Staphylococcus aureus α-toxin, Shiga toxin, cytotoxic necrotizing factor type 1, Escherichia coli heat-stable toxin, botulinum toxins, or tetanus neurotoxins.

In another aspect, the present invention provides a method for detecting the presence of a protein, preferably simultaneous or parallel detection of multiple proteins, in a sample. The method includes providing a sample which has been denatured and/or fragmented to generate a collection of soluble polypeptide analytes; providing an array comprising a support having a plurality of discrete regions to which are bound a plurality of capture agents, wherein each of the capture agents is bound to a different discrete region and wherein each of the capture agents is able to recognize and interact with a unique recognition sequence within a protein; contacting the array of capture agents with the polypeptide analytes; and determining which discrete regions show specific binding to the sample, thereby detecting the presence of a protein in a sample.

To further illustrate, the present invention provides a packaged protein detection array. Such arrays may include an addressable array having a plurality of

features, each feature independently including a discrete type of capture agent that selectively interacts with a unique recognition sequence (URS) or PET of an analyte protein, e.g., under conditions in which the analyte protein is a soluble protein produced by proteolysis and/or denaturation. The features of the array are disposed in a pattern or with a label to provide the identity of interactions between analytes and the capture agents, e.g., to ascertain the identity and/or quantity of a protein occurring in the sample. The packaged array may also include instructions for (i) contacting the addressable array with a sample containing polypeptide analytes produced by denaturation and/or cleavage of proteins at amide backbone positions; (ii) detecting interaction of said polypeptide analytes with said capture agent moieties; (iii) and determining the identity of polypeptide analytes, or native proteins from which they are derived, based on interaction with capture agent moieties.

In yet a further aspect, the present invention provides a method for detecting the presence of a protein in a sample by providing a sample which has been denatured and/or fragmented to generate a collection of soluble polypeptide analytes; contacting the sample with a plurality of capture agents, wherein each of the capture agents is able to recognize and interact with a unique recognition sequence within a protein, under conditions such that the presence of a protein in the sample is detected.

In another aspect, the present invention provides a method for detecting the presence of a protein in a sample by providing an array of capture agents comprising a support having a plurality of discrete regions (features) to which are bound a plurality of capture agents, wherein each of the capture agents is bound to a different discrete region and wherein the plurality of capture agents are capable of interacting with at least 50% of an organism's proteome; contacting the array with the sample; and determining which discrete regions show specific binding to the sample, thereby detecting the presence of a protein in the sample.

In a further aspect, the present invention provides a method for globally detecting the presence of a protein(s) in an organism's proteome by providing a sample comprising the protein and contacting the sample with a plurality of capture

agents under conditions such that interaction of the capture agents with corresponding unique recognition sequences occurs, thereby globally detecting the presence of protein(s) in an organism's proteome.

In another aspect, the present invention provides a plurality of capture agents, wherein the plurality of capture agents are capable of interacting with at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of an organism's proteome and wherein each of the capture agents is able to recognize and interact with a unique recognition sequence within a protein.

In yet another aspect, the present invention provides an array of capture agents, which includes a support having a plurality of discrete regions to which are bound a plurality of capture agents (, *e.g.*, at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000 or 13000 different capture agents), wherein each of the capture agents is bound to a different discrete region and wherein each of the capture agents is able to recognize and interact with a unique recognition sequence within a protein. The capture agents may be attached to the support, *e.g.*, via a linker, at a density of 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 or 1000 capture agents/cm^2. In one embodiment, each of the discrete regions is physically separated from each of the other discrete regions.

The capture agent array can be produced on any suitable solid surface, including silicon, plastic, glass, polymer, such as cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene, ceramic, photoresist or rubber surface. Preferably, the silicon surface is a silicon dioxide or a silicon nitride surface. Also preferably, the array is made in a chip format. The solid surfaces may be in the form of tubes, beads, discs. silicon chips, microplates, polyvinylidene difluoride (PVDF) membrane, nitrocellulose membrane, nylon membrane, other purous membrane, non-porous membrane, e.g., plastic, polymer, perspex, silicon, amongst others, a plurality of polymeric pins, or a plurality of microtitre wells, or any other surface suitable for immobilizing proteins and/or conducting an immunoassay or other binding assay.

The capture agent may be a protein, a peptide, an antibody, *e.g.*, a single chain antibody, an artificial protein, an RNA or DNA aptamer, an allosteric ribozyme or a small molecule.

In a further aspect, the present invention provides a composition comprising a plurality of isolated unique recognition sequences, wherein the unique recognition sequences are derived from at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of an organism's proteome. In one embodiment, each of the unique recognition sequences is derived from a different protein.

In another aspect, the present invention provides a method for preparing an array of capture agents. The method includes providing a plurality of isolated unique recognition sequences, the plurality of unique recognition sequences derived from at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of an organism's proteome; generating a plurality of capture agents capable of binding the plurality of unique recognition sequences; and attaching the plurality of capture agents to a support having a plurality of discrete regions, wherein each of the capture agents is bound to a different discrete region, thereby preparing an array of capture agents.

In one fundamental aspect, the invention provides an apparatus for detecting simultaneously the presence of plural specific proteins in a multi-protein sample, e.g., a body fluid sample or a cell sample produced by lysing a natural tissue sample or microorganism sample. The apparatus comprises a plurality of immobilized capture agents for contact with the sample and which include at least a subset of agents which respectively bind specifically with individual unique recognition sequences, and means for detecting binding events between respective capture agents and the unique recognition sequences, *e.g.*, probes for detecting the presence and/or concentration of unique recognition sequences bound to the capture agents. The unique recognition sequences are selected such that the presence of each sequence is unambiguously indicative of the presence in the sample (before it is fragmented) of a target protein from which it was derived. Each sample is treated with a set proteolytic protocol so that the unique recognition sequences are generated reproducibly. Optionally, the means for detecting binding events may include means for detecting data indicative of the amount of bound unique recognition sequence. This permits assessment of the relative quantity of at least two target proteins in said sample.

The invention also provides methods for simultaneously detecting the presence of plural specific proteins in a multi-protein sample. The method comprises denaturing and/or fragmenting proteins in a sample using a predetermined protocol to generate plural unique recognition sequences, the presence of which in the sample are indicative unambiguously of the presence of target proteins from which they were derived. At least a portion of the Recognition Sequences in the sample are contacted with plural capture agents which bind specifically to at least a portion of the unique recognition sequences. Detection of binding events to particular unique recognition sequences indicate the presence of target proteins corresponding to those sequences.

In another aspect, the present invention provides methods for improving the reproducibility of protein binding assays conducted on biological samples. The improvement enables detecting the presence of the target protein with greater effective sensitivity, or quantitating the protein more reliably (*i.e.*, reducing standard deviation). The methods include: (1) treating the sample using a pre-determined protocol which A) inhibits masking of the target protein caused by target protein-protein non covalent or covalent complexation or aggregation, target protein degradation or denaturing, target protein post-translational modification, or environmentally induced alteration in target protein tertiary structure, and B) fragments the target protein to, thereby, produce at least one peptide epitope (*i.e.*, a PET) whose concentration is directly proportional to the true concentration of the target protein in the sample; (2) contacting the so treated sample with a capture agent for the PET under suitable binding conditions, and (3) detecting binding events qualitatively or quantitatively.

For certain embodiments of the subject assay, the capture agents that are made available according to the teachings herein can be used to develop multiplex assays having increased sensitivity, dynamic range and/or recovery rates relative to, for example ELISA and other immunoassays. Such improved performance characteristics can include one or more of the following: a regression coefficient (R2) of 0.95 or greater for a reference standard, e.g., a comparable control sample, more preferably an R2 greater than 0.97, 0.99 or even 0.995; an average recovery rate of at least 50 percent, and more preferably at least 60, 75, 80 or even 90 percent;

a average positive predictive value for the occurrence of proteins in a sample of at least 90 percent, more preferably at least 95, 98 or even 99 percent; an average diagnostic sensitivity (DSN) for the occurrence of proteins in a sample of 99 percent or higher, more preferably at least 99.5 or even 99.8 percent; an average diagnostic specificity (DSP) for the occurrence of proteins in a sample of 99 percent or higher, more preferably at least 99.5 or even 99.8 percent.

Another aspect of the invention provides a method for detecting the presence of a post-translational modification on a target protein within a sample, comprising: (1) computationally analyzing amino acid sequence of said target protein to identify one or more candidate site for said post-translational modification; (2) computationally identifying the amino acid sequence of one or more fragment of said target protein, said fragment predictably results from a treatment of said target protein within said sample, and said fragment encompasses said potential post-translational modification site and a PET (proteome epitope tag) unique to said fragment within said sample; (3) generating a capture agent that specifically binds said PET, and immobilizing said capture agent to a support; (4) subjecting said sample to said treatment to render said fragment soluble in solution, and contacting said sample after said treatment to said capture agent; (5) detecting, on said fragment bound to said capture agent, the presence or absence of said post-translational modification.

In one embodiment, said post-translational modification is acetylation, amidation, deamidation, prenylation, formylation, glycosylation, hydroxylation, methylation, myristoylation, phosphorylation, ubiquitination, ribosylation or sulphation.

In one embodiment, said post-translational modification is phosphorylation on tyrosine, serine or threonine.

In one embodiment, said step of computationally analyzing amino acid sequences includes a Nearest-Neighbor Analysis that identifies said PET based on criteria that also include one or more of pI, charge, steric, solubility, hydrophobicity, polarity and solvent exposed area.

In one embodiment, the method further comprises determining the specificity

of said capture agent generated in (3) against one or more nearest neighbor(s), if any, of said PET.

In one embodiment, peptide competition assay is used in determining the specificity of said capture agent generated in (3) against said nearest neighbor(s) of said PET.

In one embodiment, said step of computationally analyzing amino acid sequences includes a solubility analysis that identifies said PET that are predicted to have at least a threshold solubility under a designated solution condition.

In one embodiment, the length of said PET is selected from 5-10 amino acids, 10-15 amino acids, 15-20 amino acids, 20-25 amino acids, 25-30 amino acids, or 30-40 amino acids.

In one embodiment, said capture agent is a full-length antibody, or a functional antibody fragment selected from: an Fab fragment, an F(ab')2 fragment, an Fd fragment, an Fv fragment, a dAb fragment, an isolated complementarity determining region (CDR), a single chain antibody (scFv), or derivative thereof.

In one embodiment, said capture agent is nucleotides; nucleic acids; PNA (peptide nucleic acids); proteins; peptides; carbohydrates; artificial polymers; or small organic molecules.

In one embodiment, said capture agent is aptamers, scaffolded peptides, or small organic molecules.

In one embodiment, said treatment is denaturation and/or fragmentation of said sample by a protease, a chemical agent, physical shearing, or sonication.

In one embodiment, said denaturation is thermo-denaturation or chemical denaturation.

In one embodiment, said thermo-denaturation is followed by or concurrent with proteolysis using thermo-stable proteases.

In one embodiment, said thermo-denaturation comprises two or more cycles of thermo-denaturation followed by protease digestion.

In one embodiment, said fragmentation is carried out by a protease selected

from trypsin, chymotrypsin, pepsin, papain, carboxypeptidase, calpain, subtilisin, gluc-C, endo lys-C, or proteinase K.

In one embodiment, said sample is a body fluid selected from: saliva, mucous, sweat, whole blood, serum, urine, amniotic fluid, genital fluid, fecal material, marrow, plasma, spinal fluid, pericardial fluid, gastric fluid, abdominal fluid, peritoneal fluid, pleural fluid, synovial fluid, cyst fluid, cerebrospinal fluid, lung lavage fluid, lymphatic fluid, tears, prostatitc fluid, extraction from other body parts, or secretion from other glands; or from supernatant, whole cell lysate, or cell fraction obtained by lysis and fractionation of cellular material, extract or fraction of cells obtained directly from a biological entity or cells grown in an artificial environment.

In one embodiment, said sample is obtained from human, mouse, rat, frog (Xenopus), fish (zebra fish), fly (Drosophila melanogaster), nematode (C. elegans), fission or budding yeast, or plant (Arabidopsis thaliana).

In one embodiment, said sample is produced by treatment of membrane bound proteins.

In one embodiment, said treatment is carried out under conditions to preserve said post-translational modification.

In one embodiment, said PET and said candidate site for said post-translational modification do not overlap.

In one embodiment, said capture agent is optimized for selectivity for said PET under denaturing conditions.

In one embodiment, step (5) is effectuated by using a secondary capture agent specific for said post-translational modification, wherein said secondary capture agent is labeled by a detectable moiety selected from: an enzyme, a fluorescent label, a stainable dye, a chemilumninescent compound, a colloidal particle, a radioactive isotope, a near-infrared dye, a DNA dendrimer, a water-soluble quantum dot, a latex bead, a selenium particle, or a europium nanoparticle.

In one embodiment, said post-translational modification is phosphorylation, and said secondary capture agent is a labeled secondary antibody specific for

phosphorylated tyrosine, phosphorylated serine, or phosphorylated threonine.

In one embodiment, said secondary antibody is labeled by an enzyme or a fluorescent group.

In one embodiment, said enzyme is HRP (horse radish peroxidase).

In one embodiment, said post-translational modification is phosphorylation, and said secondary capture agent is a fluoresent dye that specifically stains phosphoamino acids.

In one embodiment, said fluoresent dye is Pro-Q Diamond dye.

In one embodiment, said post-translational modification is glycosylation, and said labeled secondary capture agent is a labeled lectin specific for one or more sugar moieties attached to the glycosylation site.

In one embodiment, said post-translational modification is ubiquitination, and said labeled secondary capture agent is a labeled secondary antibody specific for ubiquitin.

In one embodiment, said sample contains billion molar excess of unrelated proteins or fragments thereof relative to said fragment.

In one embodiment, the method further comprises qantitating the amount of said fragment bound to said capture agent.

In one embodiment, step (3) is effectuated by immunizing an animal with an antigen comprising said PET sequence.

In one embodiment, the N- or C-terminus, or both, of said PET sequence are blocked to eliminate free N- or C-terminus, or both.

In one embodiment, the N- or C-terminus of said PET sequence are blocked by fusing the PET sequence to a heterologous carrier polypeptide, or blocked by a small chemical group.

In one embodiment, said carrier is KLH or BSA.

Another aspect of the invention provides an array of capture agents for identifying all potential substrates of a kinase within a proteome, comprising a plurality of capture agents, each immobilized on a distinct addressable location on

solid support, each of said capture agents specifically binds a PET uniquely associated with a peptide fragment that predictably results from a treatment of all proteins within said proteome, wherein said peptide fragment encompasses one or more potential phosphorylation sites of said kinase.

In one embodiment, said solid support is beads or an array device in a manner that encodes the identity of said capture agents disposed thereon.

In one embodiment, said array includes 100 or more different capture agents.

In one embodiment, said array device includes a diffractive grating surface.

In one embodiment, said capture agents are antibodies or antigen binding portions thereof, and said array is an arrayed ELISA.

In one embodiment, said array device is a surface plasmon resonance array.

In one embodiment, said beads are encoded as a virtual array.

Another aspect of the invention provides a method of identifying, in a sample, potential substrates of a kinase, comprising: (1) computationally analyzing amino acid sequences of all proteins in a proteome to identify all candidate phosphorylation sites for said kinase; (2) computationally identifying all peptide fragments encompassing one or more said candidate phosphorylation sites, said fragments predictably result from a treatment of all proteins within said proteome; (3) for each said fragments identified in (2), identifying one PET unique to said fragment within said sample; (4) obtaining capture agents specific for each PET identified in (3), respectively, and immobilizing said capture agents to generate the array of the subject invention; (5) contacting said array of capture agents with a sample of said proteome subjected to said treatment, and (6) detecting the presence of phosphorylated residues within any fragments bound to said capture agents, if any, wherein the presence of phosphorylated residues within a specific fragment bound to a specific capture agent is indicative that the protein, from which said specific fragment is derived from, is a substrate of said kinase.

In one embodiment, said proteome is a human proteome.

In one embodiment, said candidate phosphorylation sites are predicted based on the consensus sequence of phosphorylation by said kinase.

In one embodiment, said consensus sequence is obtained from a phosphorylation site database.

In one embodiment, said sample is pre-treated by an agent that is a known agonist of said kinase, or a known agonist of the signaling pathway to which said kinase belongs.

In one embodiment, said treatment is carried out under conditions to preserve phosphorylation.

In one embodiment, the method further comprises verifying phosphorylation of said identified substrate by said kinase in vitro or in vivo.

In one embodiment, said proteome and said kinase are from the same organism.

In one embodiment, step (6) is effectuated by using a labeled secondary capture agent specific for phosphorylated residues.

Another aspect of the invention provides an array of capture agents for identifying all potential substrates of an enzyme catalyzing post-translational modification within a proteome, comprising a plurality of capture agents, each immobilized on a distinct addressable location on solid support, each of said capture agents specifically binds a PET uniquely associated with a peptide fragment that predictably results from a treatment of all proteins within said proteome, wherein said peptide fragment encompasses one or more potential post-translational modification sites of said enzyme.

Another aspect of the invention provides a method of identifying, in a sample, potential substrates of an enzyme that catalyze a post-translational modification selected from acetylation, amidation, deamidation, prenylation, formylation, glycosylation, hydroxylation, methylation, myristoylation, phosphorylation, ubiquitination, ribosylation or sulphation, comprising: (1) computationally analyzing amino acid sequences of all proteins in a proteome to identify all candidate post-translational modification sites for said enzyme; (2) computationally identifying all peptide fragments encompassing one or more said candidate post-translational modification sites, said fragments predictably result

from a treatment of all proteins within said proteome; (3) for each said fragments identified in (2), identifying one PET unique to said fragment within said sample; (4) obtaining capture agents specific for each PET identified in (3), respectively, and immobilizing said capture agents in the array of the subject invention; (5) contacting said array of capture agents with a sample of said proteome subjected to said treatment, and (6) detecting the presence of residues with said post-translational modification within any fragments bound to said capture agents, if any, wherein the presence of residues with said post-translational modification within a specific fragment bound to a specific capture agent is indicative that the protein, from which said specific fragment is derived from, is a substrate of said enzyme.

Another aspect of the invention provides an array of capture agents for determining which, if any, of a selected number of signal transduction pathways within a proteome is activated or inhibited in response to a stimulation, comprising: a plurality of capture agents, each immobilized on a distinct addressable location on solid support, each of said capture agents specifically binds a unique PET associated with a peptide fragment that predictably results from a treatment of one or more key proteins of said signal transduction pathways, said peptide fragment encompasses one or more sites predictably post-translationally modified upon activation or inhibition of said pathway; wherein each of said signal transduction pathways is represented by one or more said key proteins.

In one embodiment, said signal transduction pathways are immune pathways activated by IL-4, IL-13, or Token-like receptor; seven-transmembrane receptor pathways activated by adrenergic, PAC1 receptor, Dictyostelium discoideum cAMP chemotaxis, Wnt/Ca^{2+}/cGMP, or G Protein-independent seven transmembrane receptor; circadian rhythm pathway of murine or Drosophila; insulin pathway; FAS pathway; TNF pathway; G-Protein coupled receptor pathways; integrin pathways; mitogen-activated protein kinase pathways of MAPK, JNK, or p38; estrogen receptor pathway; phosphoinositide 3-kinase pathway; Transforming Growth Factor-β (TGF-β) pathway; B Cell antigen receptor pathway; Jak-STAT pathway; STAT3 pathway; T Cell signal transduction pathway; Type 1 Interferon (α/β) pathway; jasmonate biochemical pathway; or jasmonate signaling pathway.

In one embodiment, said proteome is that of human, mouse, rat, frog (Xenopus), fish (zebra fish), fly (Drosophila melanogaster), nematode (C. elegans), fission or budding yeast, or plant (Arabidopsis thaliana).

In one embodiment, said post-translational modification is phosphorylation on a tyrosine, a serine, or a threonine residue.

In one embodiment, said stimulation is treatment of cells by a growth factor, a cytokine, a hormone, a steroid, a lipid, an antigen, a small molecule (Ca^{2+}, cAMP, cGMP), an osmotic shock, a heat or cold shock, a pH change, a change in ionic strength, a mechanical force, a viral or bacterial infection, or an attachment or detachment from a neighboring cell or a surface with or without a coated protein.

In one embodiment, activation or inhibition of at least one of said signal transduction pathways is manifested by a type of post-translational modification different from those of other signal transduction pathways.

In one embodiment, at least 3, 5, 10, 20, 50, 100, 200, 500, or 1000 signaling pathways are represented.

In one embodiment, signaling pathways of at least two different organisms are represented.

In one embodiment, similar signaling pathways of different organisms are represented.

In one embodiment, all capture agents are specific for proteins belonging to the same signal transduction pathway, and wherein all proteins of said signal transduction pathway that are predictably post-translationally modified are represented.

In one embodiment, one or more of said key proteins are post-translationally modified upon activation or inhibition of at least two of said signal transduction pathways. In this embodiment, the status of post-translational modification of these key proteins may indicate cross-talk between different, or even seemingly irrelavent, signaling pathways, since signals converge to these key proteins from many different pathways.

In one embodiment, the array further includes instructions for: (1) denaturing

and/or fragmentation of a sample containing polypeptide analytes, in a way compatible with the array; (2) detecting interaction of said polypeptide analytes or fragments thereof with said capture agents.

In one embodiment, the instructions further includes one or more of: data for calibration procedures and preparation procedures, and statistical data on performance characteristics of the capture agents.

In one embodiment, the array has a recovery rate of at least 50 percent.

In one embodiment, the array has an overall positive predictive value for occurrence of proteins in said sample of at least 90 percent.

In one embodiment, the array has an overall diagnostic sensitivity (DSN) for occurrence of proteins in said sample of 99 percent or higher.

In one embodiment, said array comprises at least 1,000 or 10,000 different capture agents bound to said support.

In one embodiment, said capture agents are bound to said support at a density of 100 capture agents /cm^2.

In one embodiment, the array further includes one or more labeled reference peptides including PET portions that bind to said capture agents, wherein said binding of said capture agents with said polypeptide analytes is detected by a competitive binding assay with said reference peptides.

In one embodiment, the addressable array is collection of beads, each of which comprises a discrete species of capture agent and one or more labels which identify the bead.

Another aspect of the invention provides a method of using the array of the subject invention for determining which, if any, of a selected number of signal transduction pathways within a sample from a proteome is activated or inhibited in response to a stimulation, comprising: (1) subjecting said sample to said stimulation; (2) subjecting said sample to the treatment of the subject invention to render said peptide fragment of the subject invention soluble in solution; (3) contacting said sample after said treatment to the array of the subject invention; (4) detecting the presence, and/or quantitate the amount of post-translationally modified residues

within any fragments bound to said capture agents, if any, wherein a change in the presence and/or amount of post-translationally modified residues within a specific fragment bound to a specific capture agent on said array, after said stimulation, is indicative that the signal transduction pathway represented by said specific fragment is activated or inhibited.

In one embodiment, said stimulation is effectuated by a candidate analog of a drug, and wherein activation or inhibition of a specific signal transduction pathway is monitored.

In one embodiment, said specific signal transduction pathway is one that is affected by said drug.

In one embodiment, the method further comprises comparing the degree of activation / inhibition of said specific signal transduction pathway by said analog and said drug.

In one embodiment, said specific signal transduction pathway is one that mediates a side effect of said drug.

Another aspect of the invention provides a business method for a biotechnology or pharmaceutical business, the method comprising: (i) identifying, using the method of the subject invention, one or more substrates for an enzyme catalyzing a post-translational modification; (ii) optionally, verifying the post-translational modification of said substrates by said enzyme; (iii) licensing to a third party the right to manufacture, or explore the use of said substrate as a target of said enzyme.

Another aspect of the invention provides a business method for providing protein detection arrays for identifying substrates of a post-translational modification enzyme, the method comprising: (i) identifying, within a proteome, one or more protein(s) or fragments thereof that have at least one site for said potential post-translational modification; (ii) identifying one or more PETs for each of one or more protein(s) or fragments thereof identified in (i); (iii) generating one or more capture agent(s) for each of said PETs identified in (ii), each of said capture agent(s) specifically bind one of said PETs for which said capture agent(s) is generated; (iv) fabricating arrays of capture agent(s) generated in (iii), wherein each of said capture

agents is bound to a different discrete region or address of said solid support; (v) packaging said arrays of capture agent(s) in (iv) for use in diagnostic and/or research experimentation.

In one embodiment, the business method further comprises marketing said arrays of capture agent(s).

In one embodiment, the business method further comprises distributing said arrays of capture agent(s).

Another aspect of the invention provides a composition comprising a plurality of capture agents, wherein said plurality of capture agents are, collectively, capable of specifically interacting with all potential substrates of a post-translational modification enzyme within an organism's proteome, and wherein each of said capture agents is able to recognize and interact with only one PET within said potential substrate or fragment thereof containing the post-translational modification site.

In one embodiment, said capture agents are selected from the group consisting of: nucleotides; nucleic acids; PNA (peptide nucleic acids); proteins; peptides; carbohydrates; artificial polymers; and small organic molecules.

In one embodiment, said capture agents are antibodies, or antigen binding fragments thereof.

In one embodiment, said capture agent is a full-length antibody, or a functional antibody fragment selected from: an Fab fragment, an F(ab')2 fragment, an Fd fragment, an Fv fragment, a dAb fragment, an isolated complementarity determining region (CDR), a single chain antibody (scFv), or derivative thereof.

In one embodiment, each of said capture agents is a single chain antibody.

Another aspect of the invention provides a business method for generating arrays of capture agents for marketing in research and development, the method comprising: (1) identifying one or more protein(s), a post-translational modification of which protein(s) represent the activation of at least one signal transduction pathway within an organism; (2) identifying one or more PETs for each of said protein(s), or fragment thereof containing at least one site for said post-translational

modification; (3) generating one or more capture agent(s) for each of said PETs identified in (2), each of said capture agent(s) specifically bind one of said PETs for which said capture agent(s) is generated; (4) fabricating arrays of capture agent(s) generated in (3) on solid support, wherein each of said capture agents is bound to a different discrete region of said solid support; (5) packaging said arrays of capture agent(s) in (4) for diagnosis and/or research use in commercial and/or academic laboratories.

In one embodiment, the business method further comprises marketing said arrays of capture agent(s) in (4) or said packaged arrays of capture agent(s) in (5) to potential customers and/or distributors.

In one embodiment, the business method further comprises distributing said arrays of capture agent(s) in (4) or said packaged arrays of capture agent(s) in (5) to customers and/or distributors.

Another aspect of the invention provides a business method for generating arrays of capture agents for marketing in research and development, the method comprising: (1) identifying one or more protein(s), a post-translational modification of which protein(s) represent the activation of at least one signal transduction pathway within an organism; (2) identifying one or more PETs for each of said protein(s), or fragment thereof containing at least one site for said post-translational modification; (3) licensing to a third party the right to manufacture or use said one or more PET(s) identified in (2).

Another aspect of the invention provides a method of immunizing a host animal against a disease condition associated with the presence or overexpression of a protein, comprising: (1) computationally analyzing the amino acid sequence of said protein to identify one or more PET(s) unique to said protein within the proteome of said host animal; (2) administering said one or more PET(s) identified in (1) to said host animal as an immunogen.

In one embodiment, said one or more PET(s) is administered to said host animal in a formulation designed to enhance the immune response of said host animal.

In one embodiment, said formulation comprises liposomes with or without

additional adjuvants selected from: lipopolysaccharide (LPS), lipid A, muramyl dipeptide (MDP), glucan or cytokine.

In one embodiment, said cytokine is an interleukin, an interferon, or an colony stimulating factor.

In one embodiment, said formulation comprises a viral or bacterial vector encoding said one or more PET(s).

In one embodiment, said protein is from an organism different from the host animal.

In one embodiment, said protein is from a tumor cell, an infectious agent or a parasitic agent.

In one embodiment, said infectious agent is SARS virus.

Another aspect of the invention provides a method of generating antibodies specific for a marker protein for use in immunohistochemistry, the method comprising computationally analyzing the amino acid sequence of said marker protein to identify one or more PET(s) unique to said marker protein, wherein said PET(s) is located on the surface of said marker protein.

In one embodiment, said PET(s) excludes residues known to form cross-links under the fixation condition to be used in immunohistochemistry.

Another aspect of the invention provides a method for simultaneous unambiguous detection / quantification of a family of related proteins in a sample, comprising: (1) computationally analyzing amino acid sequences for said family of related proteins expected to be present in a sample of proteins, and identifying a common PET sequence unique to the said family of proteins; (2) generating a capture agent that selectively and specifically binds said common PET; (3) contacting said sample with said capture agent identified in (2); and (4) detecting the presence and/or measuring the amount of proteins bound to said capture agent, thereby simultaneously detecting / quantifying said family of related proteins in said sample.

In one embodiment, said family of related proteins are denatured and digested by protease or chemical agents prior to step (3).

In one embodiment, the method further comprises identifying at least one PET unique to each member of said family of related proteins to facilitate detection / quantification of said each member.

In one embodiment, said family of related proteins comprises a family of related kinases or cytokines.

In one embodiment, said sample is a body fluid selected from: saliva, mucous, sweat, whole blood, serum, urine, amniotic fluid, genital fluid, fecal material, marrow, plasma, spinal fluid, pericardial fluid, gastric fluid, abdominal fluid, peritoneal fluid, pleural fluid, synovial fluid, cyst fluid, cerebrospinal fluid, lung lavage fluid, lymphatic fluid, tears, prostatitc fluid, extraction from other body parts, or secretion from other glands; or from supernatant, whole cell lysate, or cell fraction obtained by lysis and fractionation of cellular material, extract or fraction of cells obtained directly from a biological entity or cells grown in an artificial environment.

Another aspect of the invention provides a method of processing a sample for use in PET-associated detection / quantitation of a target protein therein, the method comprising denaturing all proteins of said sample, and/or fragmenting all proteins of said sample by a protease, a chemical agent, physical shearing, or sonication.

In one embodiment, said denaturation is thermo-denaturation or chemical denaturation.

In one embodiment, said thermo-denaturation is followed by or concurrent with proteolysis using thermo-stable proteases.

In one embodiment, said thermo-denaturation comprises two or more cycles of thermo-denaturation followed by protease digestion.

In one embodiment, each of said two or more cycles of thermo-denaturation is carried out by denaturing at about 90°C followed by protease digestion at about 50°C.

In one embodiment, wherein said fragmentation is carried out by a protease selected from trypsin, chymotrypsin, pepsin, papain, carboxypeptidase, calpain,

subtilisin, gluc-C, endo lys-C, or proteinase K.

In one embodiment, said sample is a body fluid selected from: saliva, mucous, sweat, whole blood, serum, urine, amniotic fluid, genital fluid, fecal material, marrow, plasma, spinal fluid, pericardial fluid, gastric fluid, abdominal fluid, peritoneal fluid, pleural fluid, synovial fluid, cyst fluid, cerebrospinal fluid, lung lavage fluid, lymphatic fluid, tears, prostatitc fluid, extraction from other body parts, or secretion from other glands; or from supernatant, whole cell lysate, or cell fraction obtained by lysis and fractionation of cellular material, extract or fraction of cells obtained directly from a biological entity or cells grown in an artificial environment.

In one embodiment, said target protein forms or tends to form complexes or aggregates with other proteins within said sample.

In one embodiment, said target protein is a TGF-beta protein.

Another aspect of the invention provides a SARS virus-specific PET amino acid sequence as listed in Table SARS.

Another aspect of the invention provides a method of generating antibodies specific for a PET sequence, the method comprising: (1) administering to an animal a peptide immunogen comprising said PET sequence; (2) screening for antibodies specific for said PET sequence using a peptide fragment comprising said PET sequence, said peptide fragment predictably results from a treatment of a protein comprising said PET sequence.

In one embodiment, said peptide immunogen consists essentially of said PET sequence.

In one embodiment, the N- or C-terminus, or both, of said PET sequence are blocked to eliminate free N- or C-terminus, or both.

In one embodiment, more than one peptide immunogens, each comprising a PET sequence, are adminitered to said animal.

In one embodiment, said more than one peptide immunogens encompasses PET sequences derived from different proteins.

In one embodiment, said peptide immunogen comprises more than one PET sequences.

In one embodiment, said more than one PET sequences are linked by short linker sequences.

In one embodiment, said more than one PET sequences are derived from different proteins.

It is also contemplated that all embodiments of the invention, including those specifically described for different aspects of the invention, can be combined with any other embodiments of the invention as appropriate.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the sequence of the Interleukin-8 receptor A and the pentamer unique recognition sequences (URS) or PETs within this sequence.

Figure 2 depicts the sequence of the Histamine H1 receptor and the pentamer unique recognition sequences (URS) or PETs within this sequence that are not destroyed by trypsin digestion.

Figure 3 is an alternative format for the parallel detection of PET from a complex sample. In this type of "virtual array" each of many different beads displays a capture agent directed against a different PET. Each different bead is color-coded by covalent linkage of two dyes (dye1 and dye2) at a characteristic ratio. Only two different beads are shown for clarity. Upon application of the sample, the capture agent binds a cognate PET, if present in the sample. Then a mixture of secondary binding ligands (in this case labeled PET peptides) conjugated to a third fluorescent tag is applied to the mixture of beads. The beads can then be analyzed using flow cytometry other detection method that can resolve, on a bead-by-bead basis, the ratio of dye1 and dye2 and thus identify the PET captured on the bead, while the fluorescence intensity of dye3 is read to quantitate the amount of

labeled PET on the bead (which will in inversely reflect the analyte PET level).

Figure 4 illustrates the result of extraction of intracellular and membrane proteins. Top Panel: M: Protein Size Marker; H-S: HELA-Supernatant; H-P: HELA-Pellet; M-S: MOLT4-Supernatant; M-P: MOLT4-Pellet. Bottom panel shows that >90% of the proteins are solublized. Briefly, cells were washed in PBS, then suspended (5 x 10^6 cells/ml) in a buffer with 0.5% Triton X-100 and homogenized in a Dounce homogenizer (30 strokes). The homogenized cells were centrifuged to separate the soluble portion and the pellet, which were both loaded to the gel.

Figure 5 illustrates the process for PET-specific antibody generation.

Figure 6 illustrates a general scheme of sample preparation prior to its use in the methods of the instant invention. The left side shows the process for chemical denaturation followed by protease digestion, the right side illustrates the preferred thermo-denaturation and fragmentation. Although the most commonly used protease trypsin is depicted in this illustration, any other suitable proteases described in the instant application may be used. The process is simple, robust & reproducible, and is generally applicable to main sample types including serum, cell lysates and tissues.

Figure 7 provides an illustrative example of serum sample pre-treatment using either the thermo-denaturation or the chemical denaturation as described in Figure 6.

Figure 8 shows the result of thermo-denaturation and chemical denaturation of serum proteins and cell lysates (MOLT4 and Hela cells).

Figure 9 illustrates the structure of mature TGF-beta dimer, and one complex form of mature TGF-beta with LAP and LTBP.

Figure 10 depicts PET-based array for (AKT) kinase substrate identification.

Figure 11 illustrates a general approach to identify all PETs of a given length in an organism with sequenced genome or a sample with known proteome. Although in this illustrative figure, the protein sequences are parsed into overlapping peptides of 4-10 amino acids in length to identify PETs of 4-10 amino acids, the same scheme is to be used for PETs of any other lengths.

Figure 12 lists the results of searching the whole human proteome (a total of

29,076 proteins, which correspond to about 12 million 4-10 overlapping peptides) for PETs, and the number of PETs identified for each N between 4-10.

Figure 13 shows the result of percentage of human proteins that have at least one PET(s).

Figure 14 provides further data resulting from tryptic digest of the human proteome.

Figure 15 illustrates a schematic drawing of fluorescence sandwich immunoassay for specific capture and quantitation of a targeted peptide in a complex peptide mixture, and results of readout fluorescent signal detected by the secondary antibody.

Figure 16 illustrates the sandwich assay used to detect a tagged-human PSA protein.

Figure 17 illustrates the PETs and their nearest neighbors for the detection of phospho-peptides in SHIP-2 and ABL.

Figure 18 illustrates a general approach to use the sandwich assay for detecting N proteins with N+1 PET-specific antibodies.

Figure 19 illustrates the common PETs and kinase-specific PETs useful for the detection of related kinases.

Figure 20 shows two SARS-specific PETs and their nearest neighbors in both the human proteome and the related Coronaviruses.

Figure 21 shows a design for the PET-based assay for standardized serum TGF-beta measurement.

Figure 22 is a schematic drawing showing the general principal of detecting PET-associated protein modification using sandwich assay.

Figure 23 is a schematic diagram of one embodiment of the detection of post-translational modification (e.g., phosphorylation or glycosylation). A target peptide is digested by a protease, such as Trypsin to yield smaller, PET-containing fragments. One of the fragments (PTP2) also contains at least one modification of interest. Once the fragments are isolated by capture agents on a support, the presence

of phosphorylation can be detected by, for example, HRP-conjugated anti-phospho-amino acid antibodies; and the presence of sugar modification can be detected by, for example, lectin.

Figure 24 illustrates that PET-specific antibodies are highly specific for the PET antigen and do not bind the nearest neighbors of the PET antigen.

Detailed Description of the Invention

The present invention provides methods, reagents and systems for detecting, *e.g.*, globally detecting, the presence of a protein or a panel of proteins, especially protein with a specific type of modification (phosphorylation, glycosylation, alternative splicing, mutation, etc.) in a sample. In certain embodiments, the method may be used to quantitate the level of expression or post-translational modification of one or more proteins in the sample. The method includes providing a sample which has, preferably, been fragmented and/or denatured to generate a collection of peptides, and contacting the sample with a plurality of capture agents, wherein each of the capture agents is able to recognize and interact with a unique recognition sequence (URS) or PET characteristic of a specific protein or modified state. Through detection and deconvolution of binding data, the presence and/or amount of a protein in the sample is determined.

In the first step, a biological sample is obtained. The biological sample as used herein refers to any body sample such as blood (serum or plasma), sputum, ascites fluids, pleural effusions, urine, biopsy specimens, isolated cells and/or cell membrane preparation (see Figure 4). Methods of obtaining tissue biopsies and body fluids from mammals are well known in the art.

Retrieved biological samples can be further solubilized using detergent-based or detergent free (i.e., sonication) methods, depending on the biological specimen and the nature of the examined polypeptide (i.e., secreted, membrane anchored or intracellular soluble polypeptide).

In certain embodiment, the sample may be denatured by detergent-free methods, such as thermo-denaturation. This is especially useful in applications

where detergent needs to be removed or is preferably removed in future analysis.

In certain embodiments, the solubilized biological sample is contacted with one or more proteolytic agents. Digestion is effected under effective conditions and for a period of time sufficient to ensure complete digestion of the diagnosed polypeptide(s). Agents that are capable of digesting a biological sample under moderate conditions in terms of temperature and buffer stringency are preferred. Measures are taken not to allow non-specific sample digestion, thus the quantity of the digesting agent, reaction mixture conditions (i.e., salinity and acidity), digestion time and temperature are carefully selected. At the end of incubation time proteolytic activity is terminated to avoid non-specific proteolytic activity, which may evolve from elongated digestion period, and to avoid further proteolysis of other peptide-based molecules (i.e., protein-derived capture agents), which are added to the mixture in following steps.

If the sample is thermo-denatured, protease active at high temperatures, such as those isolated from thermophilic bacteria, can be used after the denaturation.

In the next method step the rendered biological sample is contacted with one or more capture agents, which are capable of discriminately binding one or more protein analytes through interaction via PET binding, and the products of such binding interactions examined and, as necessary, deconvolved, in order to identify and/or quantitate proteins found in the sample.

The present invention is based, at least in part, on the realization that unique recognition sequences (URSs) or PETs, which can be identified by computational analysis, can characterize individual proteins in a given sample, e.g., identify a particular protein from amongst others and/or identify a particular post-translationally modified form of a protein. The use of agents that bind PETs can be exploitated for the detection and quantitation of individual proteins from a milieu of several or many proteins in a biological sample. The subject method can be used to assess the status of proteins or protein modifications in, for example, bodily fluids, cell or tissue samples, cell lystates, cell membranes, etc. In certain embodiments, the method utilizes a set of capture agents which discriminate between splice variants, allelic variants and/or point mutations (e.g., altered amino acid sequences arising

from single nucleotide polymorphisms).

As a result of the sample preparation, namely denaturation and/or proteolysis, the subject method can be used to detect specific proteins / modifications in a manner that does not require the homogeneity of the target protein for analysis and is relatively refractory to small but otherwise significant differences between samples. The methods of the invention are suitable for the detection of all or any selected subset of all proteins in a sample, including cell membrane bound and organelle membrane bound proteins.

In certain embodiments, the detection step(s) of the method are not sensitive to post-translational modifications of the native protein; while in other embodiments, the preparation steps are designed to preserve a post-translational modification of interest, and the detection step(s) use a set of capture agents able to discriminate between modified and unmodified forms of the protein. Exemplary post-translational modifications that the subject method can be used to detect and quantitate include acetylation, amidation, deamidation, prenylation (such as farnesylation or geranylation), formylation, glycosylation, hydroxylation, methylation, myristoylation, phosphorylation, ubiquitination, ribosylation and sulphation. In one specific embodiment, the phosphorylation to be assessed is phosphorylation on tyrosine, serine, threonine or histidine residue. In another specific embodiment, the addition of a hydrophobic group to be assessed is the addition of a fatty acid, e.g., myristate or palmitate, or addition of a glycosyl-phosphatidyl inositol anchor. In certain embodiment, the present method can be used to assess protein modification profile of a particular disease or disorder, such as infection, neoplasm (neoplasia), cancer, an immune system disease or disorder, a metabolism disease or disorder, a muscle and bone disease or disorder, a nervous system disease or disorder, a signal disease or disorder, or a transporter disease or disorder.

As used herein, the term "unique recognition sequence," "URS," "Proteome Epitope Tag," or "PET" is intended to mean an amino acid sequence that, when detected in a particular sample, unambiguously indicates that the protein from which it was derived is present in the sample. For instance, a PET is selected such that its

presence in a sample, as indicated by detection of an authentic binding event with a capture agent designed to selectively bind with the sequence, necessarily means that the protein which comprises the sequence is present in the sample. A useful PET must present a binding surface that is solvent accessible when a protein mixture is denatured and/or fragmented, and must bind with significant specificity to a selected capture agent with minimal cross reactivity. A unique recognition sequence is present within the protein from which it is derived and in no other protein that may be present in the sample, cell type, or species under investigation. Moreover, a PET will preferably not have any closely related sequence, such as determined by a nearest neighbor analysis, among the other proteins that may be present in the sample. A PET can be derived from a surface region of a protein, buried regions, splice junctions, or post translationally modified regions.

Perhaps the ideal PET is a peptide sequence which is present in only one protein in the proteome of a species. But a peptide comprising a PET useful in a human sample may in fact be present within the structure of proteins of other organisms. A PET useful in an adult cell sample is "unique" to that sample even though it may be present in the structure of other different proteins of the same organism at other times in its life, such as during embryology, or is present in other tissues or cell types different from the sample under investigation. A PET may be unique even though the same amino acid sequence is present in the sample from a different protein provided one or more of its amino acids are derivatized, and a binder can be developed which resolves the peptides.

When referring herein to "uniqueness" with respect to a PET, the reference is always made in relation to the foregoing. Thus, within the human genome, a PET may be an amino acid sequence that is truly unique to the protein from which it is derived. Alternatively, it may be unique just to the sample from which it is derived, but the same amino acid sequence may be present in, for example, the murine genome. Likewise, when referring to a sample which may contain proteins from multiple different organism, uniqueness refers to the ability to unambiguously identify and discriminate between proteins from the different organisms, such as being from a host or from a pathogen.

Thus, a PET may be present within more than one protein in the species, provided it is unique to the sample from which it is derived. For example, a PET may be an amino acid sequence that is unique to: a certain cell type, *e.g.*, a liver, brain, heart, kidney or muscle cell; a certain biological sample, *e.g.*, a plasma, urine, amniotic fluid, genital fluid, marrow, spinal fluid, or pericardial fluid sample; a certain biological pathway, *e.g.*, a G-protein coupled receptor signaling pathway or a tumor necrosis factor (TNF) signaling pathway.

In this sense, the instant invention provides a method to identify application-specific PETs, depending on the type of proteins present in a given sample. This information may be readily obtained from a variety of sources. For example, when the whole genome of an organism is concerned, the sequenced genome provides each and every protein sequences that can be encoded by this genome, sometimes even including hypothetical proteins. This "virtually translated proteome" obtained from the sequenced genome is expected to be the most comprehensive in terms of representing all proteins in the sample. Alternatively, the type of transcribed mRNA species within a sample may also provide useful information as to what type of proteins may be present within the sample. The mRNA species present may be identified by DNA microarrays, SNP analysis, or any other suitable RNA analysis tools available in the art of molecular biology. An added advantage of RNA analysis is that it may also provide information such as alternative splicing and mutations. Finally, direct protein analysis using techniques such as mass spectrometry may help to identify the presence of specific post-translation modifications and mutations, which may aid the design of specific PETs for specific applications.

The PET may be found in the native protein from which it is derived as a contiguous or as a non-contiguous amino acid sequence. It typically will comprise a portion of the sequence of a larger peptide or protein, recognizable by a capture agent either on the surface of an intact or partially degraded or digested protein, or on a fragment of the protein produced by a predetermined fragmentation protocol. The PET may be 5, 6, 7, 8, 9, 10, 11, 12, 13,14, 15, 16, 17, 18, 19 or 20 amino acid residues in length. In a preferred embodiment, the PET is 6, 7, 8, 9 or 10 amino acid residues, preferably 8 amino acids in length.

The term "discriminate", as in "capture agents able to discriminate between", refers to a relative difference in the binding of a capture agent to its intended protein analyte and background binding to other proteins (or compounds) present in the sample. In particular, a capture agent can discriminate between two different species of proteins (or species of modifications) if the difference in binding constants is such that a statistically significant difference in binding is produced under the assay protocols and detection sensitivities. In preferred embodiments, the capture agent will have a discriminating index (D.I.) of at least 0.5, and even more preferably at least 0.1, 0.001, or even 0.0001, wherein D.I. is defined as $K_d(a)/K_d(b)$, $K_d(a)$ being the dissociation constant for the intended analyte, $K_d(b)$ is the dissociation constant for any other protein (or modified form as the case may be) present in sample.

As used herein, the term "capture agent" includes any agent which is capable of binding to a protein that includes a unique recognition sequence, *e.g.*, with at least detectable selectivity. A capture agent is capable of specifically interacting with (directly or indirectly), or binding to (directly or indirectly) a unique recognition sequence. The capture agent is preferably able to produce a signal that may be detected. In a preferred embodiment, the capture agent is an antibody or a fragment thereof, such as a single chain antibody, or a peptide selected from a displayed library. In other embodiments, the capture agent may be an artificial protein, an RNA or DNA aptamer, an allosteric ribozyme or a small molecule. In other embodiments, the capture agent may allow for electronic (*e.g.*, computer-based or information-based) recognition of a unique recognition sequence. In one embodiment, the capture agent is an agent that is not naturally found in a cell.

As used herein, the term "globally detecting" includes detecting at least 40% of the proteins in the sample. In a preferred embodiment, the term "globally detecting" includes detecting at least 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of the proteins in the sample. Ranges intermediate to the above recited values, *e.g.*, 50%-70% or 75%-95%, are also intended to be part of this invention. For example, ranges using a combination of any of the above recited values as upper and/or lower limits are intended to be included.

As used herein, the term "proteome" refers to the complete set of chemically

distinct proteins found in an organism.

As used herein, the term "organism" includes any living organism including animals, *e.g.*, avians, insects, mammals such as humans, mice, rats, monkeys, or rabbits; microorganisms such as bacteria, yeast, and fungi, *e.g.*, *Escherichia coli*, *Campylobacter*, *Listeria*, *Legionella*, *Staphylococcus*, *Streptococcus*, *Salmonella*, *Bordatella*, *Pneumococcus*, *Rhizobium*, *Chlamydia*, *Rickettsia*, *Streptomyces*, *Mycoplasma*, *Helicobacter pylori*, *Chlamydia pneumoniae*, *Coxiella burnetii*, *Bacillus Anthracis*, and *Neisseria*; protozoa, *e.g.*, Trypanosoma brucei; viruses, *e.g.*, human immunodeficiency virus, rhinoviruses, rotavirus, influenza virus, Ebola virus, simian immunodeficiency virus, feline leukemia virus, respiratory syncytial virus, herpesvirus, pox virus, polio virus, parvoviruses, Kaposi's Sarcoma-Associated Herpesvirus (KSHV), adeno-associated virus (AAV), Sindbis virus, Lassa virus, West Nile virus, enteroviruses, such as 23 Coxsackie A viruses, 6 Coxsackie B viruses, and 28 echoviruses, Epstein-Barr virus, caliciviruses, astroviruses, and Norwalk virus; fungi, *e.g.*, *Rhizopus, neurospora*, yeast, *or puccinia*; tapeworms, *e.g.*, Echinococcus granulosus, E. multilocularis, E. vogeli and E. oligarthrus; and plants, *e.g.*, Arabidopsis thaliana, rice, wheat, maize, tomato, alfalfa, oilseed rape, soybean, cotton, sunflower or canola.

As used herein, "sample" refers to anything which may contain a protein analyte. The sample may be a biological sample, such as a biological fluid or a biological tissue. Examples of biological fluids include urine, blood, plasma, serum, saliva, semen, stool, sputum, cerebral spinal fluid, tears, mucus, amniotic fluid or the like. Biological tissues are aggregates of cells, usually of a particular kind together with their intercellular substance that form one of the structural materials of a human, animal, plant, bacterial, fungal or viral structure, including connective, epithelium, muscle and nerve tissues. Examples of biological tissues also include organs, tumors, lymph nodes, arteries and individual cell(s). The sample may also be a mixture of target protein containing molecules prepared in vitro.

As used herein, "a comparable control sample" refers to a control sample that is only different in one or more defined aspects relative to a test sample, and the present methods, kits or arrays are used to identify the effects, if any, of these

defined difference(s) between the test sample and the control sample, e.g., on the amounts and types of proteins expressed and/or on the protein modification profile. For example, the control biosample can be derived from physiological normal conditions and/or can be subjected to different physical, chemical, physiological or drug treatments, or can be derived from different biological stages, etc.

"Predictably result from a treatment" means that a peptide fragment can be reliably generated by certain treatments, such as site specific protease digestion or chemical fragmentation. Since the digestion sites are quite specific, the peptide fragment generated by specific treatments can be reliably predicted in silico.

A report by MacBeath and Schreiber (*Science* 289 (2000), pp. 1760–1763) in 2000 established that proteins could be printed and assayed in a microarray format, and thereby had a large role in renewing the excitement for the prospect of a protein chip. Shortly after this, Snyder and co-workers reported the preparation of a protein chip comprising nearly 6000 yeast gene products and used this chip to identify new classes of calmodulin- and phospholipid-binding proteins (Zhu *et al.*, *Science* 293 (2001), pp. 2101–2105). The proteins were generated by cloning the open reading frames and overproducing each of the proteins as glutathione-S-transferase-(GST) and His-tagged fusions. The fusions were used to facilitate the purification of each protein and the His-tagged family were also used in the immobilization of proteins. This and other references in the art established that microarrays containing thousands of proteins could be prepared and used to discover binding interactions. They also reported that proteins immobilized by way of the His tag – and therefore uniformly oriented at the surface – gave superior signals to proteins randomly attached to aldehyde surfaces.

Related work has addressed the construction of antibody arrays (de Wildt *et al.*, Antibody arrays for high-throughput screening of antibody–antigen interactions. *Nat. Biotechnol.* **18** (2000), pp. 989–994; Haab, B.B. *et al.* (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2, RESEARCH0004.1–RESEARCH0004.13). Specifically, in an early landmark report, de Wildt and Tomlinson immobilized phage libraries presenting scFv antibody fragments on filter

paper to select antibodies for specific antigens in complex mixtures (supra). The use of arrays for this purpose greatly increased the throughput when evaluating antibodies, allowing nearly 20,000 unique clones to be screened in one cycle. Brown and co-workers extended this concept to create molecularly defined arrays wherein antibodies were directly attached to aldehyde-modified glass. They printed 115 commercially available antibodies and analyzed their interactions with cognate antigens with semi-quantitative results (supra). Kingsmore and co-workers used an analogous approach to prepare arrays of antibodies recognizing 75 distinct cytokines and, using the rolling-circle amplification strategy (Lizardi *et al.*, Mutation detection and single molecule counting using isothermal rolling circle amplification. *Nat. Genet.* **19** (1998), pp. 225–233), could measure cytokines at femtomolar concentrations (Schweitzer *et al.*, Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nat. Biotechnol.* **20** (2002), pp. 359–365).

These examples demonstrate the many important roles that protein chips can play, and give evidence for the widespread activity in fabrication of these tools. The following subsections describes in further detail about various aspects of the invention.

I. Type of Capture Agents

In certain preferred embodiments, the capture agents used should be capable of selective affinity reactions with PET moieties. Generally, such ineraction will be non-covalent in nature, though the present invention also contemplates the use of capture reagents that become covalently linked to the PET.

Examples of capture agents which can be used include, but are not limited to: nucleotides; nucleic acids including oligonucleotides, double stranded or single stranded nucleic acids (linear or circular), nucleic acid aptamers and ribozymes; PNA (peptide nucleic acids); proteins, including antibodies (such as monoclonal or recombinantly engineered antibodies or antibody fragments), T cell receptor and MHC complexes, lectins and scaffolded peptides; peptides; other naturally occurring polymers such as carbohydrates; artificial polymers, including plastibodies; small organic molecules such as drugs, metabolites and natural products; and the like.

In certain embodiments, the capture agents are immobilized, permanently or

reversibly, on a solid support such as a bead, chip, or slide. When employed to analyze a complex mixture of proteins, the immobilized capture agent are arrayed and/or otherwise labeled for deconvolution of the binding data to yield identity of the capture agent (and therefore of the protein to which it binds) and (optionally) to quantitate binding. Alternatively, the capture agents can be provided free in solution (soluble), and other methods can be used for deconvolving PET binding in parallel.

In one embodiment, the capture agents are conjugated with a reporter molecule such as a fluorescent molecule or an enzyme, and used to detect the presence of bound PET on a substrate (such as a chip or bead), in for example, a "sandwich" type assay in which one capture agent is immobilized on a support to capture a PET, while a second, labeled capture agent also specific for the captured PET may be added to detect /quantitate the captured PET. In this embodiment, the peptide fragment contains two unique, non-overlapping PETs, one recognized by the immobilized the capture agent, the other recognized by the labled detecting capture agent. In a related embodiment, one PET unique to the peptide fragment can be used in conjunction with a common PET shared among several protein family members. The spacial arrangement of these two PET is such that binding by one capture agent will not substancially affect the binidng by the other capture agent. In addition, the length of the peptide fragment is such that it encompasses two PETs properly spaced from each other. Preferably, peptide fragments is at least about 15 residues for sandwich assay. In other embodiments a labeled-PET peptide is used in a competitive binding assay to determine the amount of unlabeled PET (from the sample) binds to the capture agent. In this embodiment, the peptide fragment need only be long enough to encompass one PET, so peptides as short as 5-8 residues may be suitable.

Generally, the sandwich assay tend to be more (e.g., about 10, 100, or 1000 fold more) sensitive than the competitive binding assay.

An important advantage of the invention is that useful capture agents can be identified and/or synthesized even in the absence of a sample of the protein to be detected. With the completion of the whole genome in a number of organisms, such as human, fly (Drosophila melanogaster) and nematode (C. elegans), PET of a given

length or combination thereof can be identified for any single given protein in a certain organism, and capture agents for any of these proteins of interest can then be made without ever cloning and expressing the full length protein.

In addition, the suitability of any PET to serve as an antigen or target of a capture agent can be further checked against other available information. For example, since amino acid sequence of many proteins can now be inferred from available genomic data, sequence from the structure of the proteins unique to the sample can be determined by computer aided searching, and the location of the peptide in the protein, and whether it will be accessible in the intact protein, can be determined. Once a suitable PET peptide is found, it can be synthesized using known techniques. With a sample of the PET in hand, an agent that interacts with the peptide such as an antibody or peptidic binder, can be raised against it or panned from a library. In this situation, care must be taken to assure that any chosen fragmentation protocol for the sample does not restrict the protein in a way that destroys or masks the PET. This can be determined theoretically and/or experimentally, and the process can be repeated until the selected PET is reliably retrieved by a capture agent(s).

The PET set selected according to the teachings of the present invention can be used to generate peptides either through enzymatic cleavage of the protein from which they were generated and selection of peptides, or preferably through peptide synthesis methods.

Proteolytically cleaved peptides can be separated by chromatographic or electrophoretic procedures and purified and renatured via well known prior art methods.

Synthetic peptides can be prepared by classical methods known in the art, for example, by using standard solid phase techniques. The standard methods include exclusive solid phase synthesis, partial solid phase synthesis methods, fragment condensation, classical solution synthesis, and even by recombinant DNA technology. See, e.g., Merrifield, J. Am. Chem. Soc., 85:2149 (1963), incorporated herein by reference. Solid phase peptide synthesis procedures are well known in the art and further described by John Morrow Stewart and Janis Dillaha Young, Solid

Phase Peptide Syntheses (2nd Ed., Pierce Chemical Company, 1984).

Synthetic peptides can be purified by preparative high performance liquid chromatography [Creighton T. (1983) Proteins, structures and molecular principles. WH Freeman and Co. N.Y.] and the composition of which can be confirmed via amino acid sequencing.

In addition, other additives such as stabilizers, buffers, blockers and the like may also be provided with the capture agent.

A. Antibodies

In one embodiment, the capture agent is an antibody or an antibody-like molecule (collectively "antibody"). Thus an antibody useful as capture agent may be a full length antibody or a fragment thereof, which includes an "antigen-binding portion" of an antibody. The term "antigen-binding portion," as used herein, refers to one or more fragments of an antibody that retain the ability to specifically bind to an antigen. It has been shown that the antigen-binding function of an antibody can be performed by fragments of a full-length antibody. Examples of binding fragments encompassed within the term "antigen-binding portion" of an antibody include (i) a Fab fragment, a monovalent fragment consisting of the V_L, V_H, C_L and C_{H1} domains; (ii) a $F(ab')_2$ fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the V_H and C_{H1} domains; (iv) a Fv fragment consisting of the V_L and V_H domains of a single arm of an antibody, (v) a dAb fragment (Ward et al., (1989) *Nature* 341:544-546), which consists of a V_H domain; and (vi) an isolated complementarity determining region (CDR). Furthermore, although the two domains of the Fv fragment, V_L and V_H, are coded for by separate genes, they can be joined, using recombinant methods, by a synthetic linker that enables them to be made as a single protein chain in which the V_L and V_H regions pair to form monovalent molecules (known as single chain Fv (scFv); see, *e.g.*, Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883; and Osbourn *et al.* 1998, *Nature Biotechnology* 16: 778). Such single chain antibodies are also intended to be

encompassed within the term "antigen-binding portion" of an antibody. Any V_H and V_L sequences of specific scFv can be linked to human immunoglobulin constant region cDNA or genomic sequences, in order to generate expression vectors encoding complete IgG molecules or other isotypes. V_H and V_L can also be used in the generation of Fab , Fv or other fragments of immunoglobulins using either protein chemistry or recombinant DNA technology. Other forms of single chain antibodies, such as diabodies are also encompassed. Diabodies are bivalent, bispecific antibodies in which V_H and V_L domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites (see, *e.g.*, Holliger, P., *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6444-6448; Poljak, R. J., *et al.* (1994) *Structure* 2:1121-1123).

Still further, an antibody or antigen-binding portion thereof may be part of a larger immunoadhesion molecule, formed by covalent or noncovalent association of the antibody or antibody portion with one or more other proteins or peptides. Examples of such immunoadhesion molecules include use of the streptavidin core region to make a tetrameric scFv molecule (Kipriyanov, S.M., et al. (1995) Human Antibodies and Hybridomas 6:93-101) and use of a cysteine residue, a marker peptide and a C-terminal polyhistidine tag to make bivalent and biotinylated scFv molecules (Kipriyanov, S.M., et al. (1994) Mol. Immunol. 31:1047-1058). Antibody portions, such as Fab and $F(ab')_2$ fragments, can be prepared from whole antibodies using conventional techniques, such as papain or pepsin digestion, respectively, of whole antibodies. Moreover, antibodies, antibody portions and immunoadhesion molecules can be obtained using standard recombinant DNA techniques.

Antibodies may be polyclonal or monoclonal. The terms "monoclonal antibodies" and "monoclonal antibody composition," as used herein, refer to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of an antigen, whereas the term "polyclonal antibodies" and "polyclonal antibody composition" refer to a population of antibody molecules that contain multiple species of antigen binding sites capable of interacting with a particular antigen. A monoclonal antibody

composition, typically displays a single binding affinity for a particular antigen with which it immunoreacts.

Any art-recognized methods can be used to generate an PET-directed antibody. For example, a PET (alone or linked to a hapten) can be used to immunize a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal or vertebrate). For example, the methods described in U.S. Patent Nos. 5,422,110; 5,837,268; 5,708,155; 5,723,129;and 5,849,531 (the contents of each of which are incorporated herein by reference) can be used. The immunogenic preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with a PET induces a polyclonal anti-PET antibody response. The anti-PET antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized PET.

The antibody molecules directed against a PET can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-PET antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare, *e.g.*, monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), or the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a PET immunogen as described above, and the culture supernatants

of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a PET.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-PET monoclonal antibody (see, *e.g.*, G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind a PET, *e.g.*, using a standard ELISA assay.

In addition, automated screening of antibody or scaffold libraries against arrays of target proteins / PETs will be the most rapid way of developing thousands of reagents that can be used for protein expression profiling. Furthermore, polyclonal antisera, hybridomas or selection from library systems may also be used to quickly generate the necessary capture agents. A high-throughput process for antibody isolation is described by Hayhurst and Georgiou in *Curr Opin Chem Biol* 5(6):683-9, December 2001 (incorporated by reference).

The PET antigens used for the generation of PET-specific antibodies are preferably blocked at either the N- or C-terminal end, most preferably at both ends (see Figure 5) to generate neutral groups, since antibodies raised against peptides with non-neutralized ends may not be functional for the methods of the invention. The PET antigens can be most easily synthesized using standard molecular biology or chemical methods, for example, with a peptide synthesizer. The terminals can be blocked with NH2- or COO- groups as appropriate, or any other blocking agents to eliminate free ends. In a preferred embodiment, one end (either N- or C-terminus) of the PET will be conjugated with a carrier protein such as KLH or BSA to facilitate antibody generation. KLH represents Keyhole-limpet hemocyanin, an oxygen carrying copper protein found in the keyhole-limpet (*Megathura crenulata*), a primitive mollusk sea snail. KLH has a complex molecular arrangement and contains a diverse antigenic structure and elicits a strong nonspecific immune response in host animals. Therefore, when small peptides (which may not be very immunogenic) are used as immunogens, they are preferably conjugated to KLH or other carrier proteins (BSA) for enhanced immune responses in the host animal. The resulting antibodies can be affinity purified using a polypeptide corresponding to the PET-containing tryptic peptide of interest (see Figure 5).

Blocking the ends of PET in antibody generation may be advantageous, since in many (if not most) cases, the selected PETs are contained within larger (tryptic) fragments. In these cases, the PET-specific antibodies are required to bind PETs in the middle of a peptide fragment. Therefore, blocking both the C- and N-terminus of the PETs best simulates the antibody binding of peptide fragments in a digested sample. Similarly, if the selected PET sequence happens to be at the N- or C-terminal end of a target fragment, then only the other end of the immunogen needs to be blocked, preferably by a carrier such as KLH or BSA..

Figure 24 below shows that PET-specific antibodies are highly specific and have high affinity for their respective PET-antigens.

When generating PET-specific antibodies, preferably monoclonal antibodies, a peptide immunogen comprising essentially of the target PET sequence may be administered to an animal according to standard antibody generation protocol for

short peptide antigens. In one embodiment, the short peptide antigen may be conjugated with a carrier such as KLH. However, when screening for antibodies specific for the PET sequence, it is preferred that the parental peptide fragments containing the PET sequence (such as the fragment resulting from trypsin digestion) is used. This ensures that the identified antibodies will be not only specific for the original PET sequence, but also able to recognize the PET peptide fragment for which the antibody is designed. Optionally, the specificity of the identified antibody can be further verified by reacting with the original immunogen such as the end-blocked PET sequence itself.

In certain embodiments, several different immunogens for different PET sequences may be simultaneously administered to the same animal, so that different antibodies may be generated in one animal. Obviously, for each immunogen, a separate screen would be needed to identify antibodies specific for the immunogen.

In an alternative embodiment, different PETs may be linked together in a single, longer immunogen for administration to an animal. The linker sequence can be flexible linkers such as GS, GSSSS or repeats thereof (such as three-peats).

In both embodiments described above, the different immunogens may be from the same or different organisms or proteomes. These methods are all potential means of reducing costs in antibody generation. An unexpected advantage of using linked PET sequences as immunogen is that longer immunogens may at certain situations produce higher affinity antibodies than those produced using short PET sequences.

(i) PET-Specific Antibody Knowledge Database

The instant invention also provides an antibody knowledge database, which provides various important information pertaining to these antibodies. A specific subset of the antibodies will be PET-specific antibodies, which are either generated de novo based on the criteria set forth in the instant application, or generated by others in the prior art, which happens to recognize certain PETs.

Information to be included in the knowledge database can be quite

comprehensive. Such knowledge may be further classified as public or proprietary. Examples of public information may include: target protein name, antibody source, catalog number, potential applications, etc. Exemplary proprietary information includes parental tryptic fragments in one or more organisms or specific samples, immunogen peptide sequences and whether or not they are PETs, affinity for the target PET, degree of cross-reactivity with other related epitopes (such as the closest nearest neighbors), and usefulness for various PET assays.

To this end, such information about 1000 anti-peptide antibodies are already collected / generated in the knowledge database. Among them, about 128 antibodies are deemed compatible for trypsin digested samples. Certain commercially available antibodies, the immunogen and the PET sequences they happen to contain, and the nearest neighbors of these PETs are listed below.

Commercial Anti-PET Antibodies

Protein	PTP (Immunogen/PET underlined)	Nearest Neighbors
Anti-Cyclin F	TASP**TSSVDGGL**GALP.K	**SASI**DGGL; **SSSS**DGGL; TGSVDGG**A**; **ESSS**DGGL
Anti-phospho SHC (Tyr239)	FAGMPITLTVS**TSSLNLMA**ADCK	**IS**TA**S**LNL; **IS**TSSLN**V**; VS**L**SSLNL; **MD**TSSLNL
Anti-phospho-PP2A (Tyr307)	EEE**ADINQLTE**EFF.K	ADLNQLT**Q**; **R**DINQL**S**E; ADF**N**QL**A**E; ADIN**M**VTE
Anti-Cdk8	ATS**QQPPQYSH**QTHR	QEPPQYSH; QQ**Q**PQ**F**SH; QQPPQ**HSK**; QQPPQ**QQ**H

B. Proteins and peptides

Other methods for generating the capture agents of the present invention include phage-display technology described in, for example, Dower *et al.*, WO 91/17271, McCafferty *et al.*, WO 92/01047, Herzig *et al.*, US 5,877,218, Winter *et al.*, US 5,871,907, Winter *et al.*, US 5,858,657, Holliger *et al.*, US 5,837,242, Johnson *et al.*, US 5,733,743 and Hoogenboom *et al.*, US 5,565,332 (the contents of each of which are incorporated by reference). In these methods, libraries of phage are produced in which members display different antibodies, antibody binding sites, or peptides on their outer surfaces. Antibodies are usually displayed as Fv or Fab fragments. Phage displaying sequences with a desired specificity are selected by affinity enrichment to a specific PET.

Methods such as yeast display and *in vitro* ribosome display may also be used to generate the capture agents of the present invention. The foregoing methods are described in, for example, Methods in Enzymology Vol 328 -Part C: Protein-protein interactions & Genomics and Bradbury A. (2001) *Nature Biotechnology* 19:528-529, the contents of each of which are incorporated herein by reference.

In a related embodiment, proteins or polypeptides may also act as capture agents of the present invention. These peptide capture agents also specifically bind to an given PET, and can be identified, for example, using phage display screening against an immobilized PET, or using any other art-recognized methods. Once identified, the peptidic capture agents may be prepared by any of the well known methods for preparing peptidic sequences. For example, the peptidic capture agents may be produced in prokaryotic or eukaryotic host cells by expression of polynucleotides encoding the particular peptide sequence. Alternatively, such peptidic capture agents may be synthesized by chemical methods. Methods for expression of heterologous peptides in recombinant hosts, chemical synthesis of peptides, and *in vitro* translation are well known in the art and are described further in Maniatis *et al.*, Molecular Cloning: A Laboratory Manual (1989), 2nd Ed., Cold Spring Harbor, N.Y.; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego,

Calif.; Merrifield, J. (1969) *J. Am. Chem. Soc.* 91:501; Chaiken, I. M. (1981) *CRC Crit. Rev. Biochem.* 11:255; Kaiser *et al.* (1989) *Science* 243:187; Merrifield, B. (1986) *Science* 232:342; Kent, S. B. H. (1988) *Ann. Rev. Biochem.* 57:957; and Offord, R. E. (1980) *Semisynthetic Proteins*, Wiley Publishing, which are incorporated herein in their entirety by reference).

The peptidic capture agents may also be prepared by any suitable method for chemical peptide synthesis, including solution-phase and solid-phase chemical synthesis. Preferably, the peptides are synthesized on a solid support. Methods for chemically synthesizing peptides are well known in the art (see, *e.g.*, Bodansky, M. *Principles of Peptide Synthesis*, Springer Verlag, Berlin (1993) and Grant, G.A (ed.). *Synthetic Peptides: A User's Guide*, W.H. Freeman and Company, New York (1992). Automated peptide synthesizers useful to make the peptidic capture agents are commercially available.

C. Scaffolded peptides

An alternative approach to generating capture agents for use in the present invention makes use of antibodies are scaffolded peptides, e.g., peptides displayed on the surface of a protein. The idea is that restricting the degrees of freedom of a peptide by incorporating it into a surface-exposed protein loop could reduce the entropic cost of binding to a target protein, resulting in higher affinity. Thioredoxin, fibronectin, avian pancreatic polypeptide (aPP) and albumin, as examples, are small, stable proteins with surface loops that will tolerate a great deal of sequence variation. To identify scaffolded peptides that selectively bind a target PET, libraries of chimeric proteins can be generated in which random peptides are used to replace the native loop sequence, and through a process of affinity maturation, those which selectively bind a PET of interest are identified.

D. Simple peptides and peptidomimetic compounds

Peptides are also attractive candidates for capture agents because they

combine advantages of small molecules and proteins. Large, diverse libraries can be made either biologically or synthetically, and the "hits" obtained in binding screens against PET moieties can be made synthetically in large quantities.

Peptide-like oligomers (Soth et al. (1997) Curr. Opin. Chem. Biol. 1:120–129) such as peptoids (Figliozzi et al., (1996) Methods Enzymol. 267:437–447) can also be used as capture reagents, and can have certain advantages over peptides. They are impervious to proteases and their synthesis can be simpler and cheaper than that of peptides, particularly if one considers the use of functionality that is not found in the 20 common amino acids.

E. *Nucleic* acids

In another embodiment, aptamers binding specifically to a PET may also be used as capture agents. As used herein, the term "aptamer," e.g., RNA aptamer or DNA aptamer, includes single-stranded oligonucleotides that bind specifically to a target molecule. Aptamers are selected, for example, by employing an in vitro evolution protocol called systematic evolution of ligands by exponential enrichment. Aptamers bind tightly and specifically to target molecules; most aptamers to proteins bind with a K_d (equilibrium dissociation constant) in the range of 1 pM to 1 nM. Aptamers and methods of preparing them are described in, for example, E.N. Brody et al. (1999) Mol. Diagn. 4:381-388, the contents of which are incorporated herein by reference.

In one embodiment, the subject aptamers can be generated using SELEX, a method for generating very high affinity receptors that are composed of nucleic acids instead of proteins. See, for example,. Brody et al. (1999) Mol. Diagn. 4:381-388. SELEX offers a completely in vitro combinatorial chemistry alternative to traditional protein-based antibody technology. Similar to phage display, SELEX is advantageous in terms of obviating animal hosts, reducing production time and labor, and simplifying purification involved in generating specific binding agents to a particular target PET.

To further illustrate, SELEX can be performed by synthesizing a random

oligonucleotide library, e.g., of greater than 20 bases in length, which is flanked by known primer sequences. Synthesis of the random region can be achieved by mixing all four nucleotides at each position in the sequence. Thus, the diversity of the random sequence is maximally 4^n, where n is the length of the sequence, minus the frequency of palindromes and symmetric sequences. The greater degree of diversity conferred by SELEX affords greater opportunity to select for oligonuclotides that form 3-dimensional binding sites. Selection of high affinity oligonucleotides is achieved by exposing a random SELEX library to an immobilized target PET. Sequences, which bind readily without washing away, are retained and amplified by the PCR, for subsequent rounds of SELEX consisting of alternating affinity selection and PCR amplification of bound nucleic acid sequences. Four to five rounds of SELEX are typically sufficient to produce a high affinity set of aptamers.

Therefore, hundreds to thousands of aptamers can be made in an economically feasible fashion. Blood and urine can be analyzed on aptamer chips that capture and quantitate proteins. SELEX has also been adapted to the use of 5-bromo (5-Br) and 5-iodo (5-I) deoxyuridine residues. These halogenated bases can be specifically cross-linked to proteins. Selection pressure during *in vitro* evolution can be applied for both binding specificity and specific photo-cross-linkability. These are sufficiently independent parameters to allow one reagent, a photo-cross-linkable aptamer, to substitute for two reagents, the capture antibody and the detection antibody, in a typical sandwich array. After a cycle of binding, washing, cross-linking, and detergent washing, proteins will be specifically and covalently linked to their cognate aptamers. Because no other proteins are present on the chips, protein-specific stain will now show a meaningful array of pixels on the chip. Combined with learning algorithms and retrospective studies, this technique should lead to a robust yet simple diagnostic chip.

In yet another related embodiment, a capture agent may be an allosteric ribozyme. The term "allosteric ribozymes," as used herein, includes single-stranded oligonucleotides that perform catalysis when triggered with a variety of effectors, *e.g.*, nucleotides, second messengers, enzyme cofactors, pharmaceutical agents, proteins, and oligonucleotides. Allosteric ribozymes and methods for preparing them are described in, for example, S. Seetharaman *et al.* (2001) *Nature Biotechnol.* 19:

336-341, the contents of which are incorporated herein by reference. According to Seetharaman *et al.*, a prototype biosensor array has been assembled from engineered RNA molecular switches that undergo ribozyme-mediated self-cleavage when triggered by specific effectors. Each type of switch is prepared with a 5'-thiotriphosphate moiety that permits immobilization on gold to form individually addressable pixels. The ribozymes comprising each pixel become active only when presented with their corresponding effector, such that each type of switch serves as a specific analyte sensor. An addressed array created with seven different RNA switches was used to report the status of targets in complex mixtures containing metal ion, enzyme cofactor, metabolite, and drug analytes. The RNA switch array also was used to determine the phenotypes of *Escherichia coli* strains for adenylate cyclase function by detecting naturally produced 3',5'- cyclic adenosine monophosphate (cAMP) in bacterial culture media.

F. Plastibodies

In certain embodiments the subject capture agent is a plastibody. The term "plastibody" refers to polymers imprinted with selected template molecules. See, for example, Bruggemann (2002) Adv Biochem Eng Biotechnol 76:127-63; and Haupt et al. (1998) Trends Biotech. 16:468-475. The plastibody principle is based on molecular imprinting, namely, a recognition site that can be generated by stereoregular display of pendant functional groups that are grafted to the sidechains of a polymeric chain to thereby mimic the binding site of, for example, an antibody.

G. Chimeric binding agents derived from two low-affinity ligands

Still another strategy for generating suitable capture agents is to link two or more modest-affinity ligands and generate high affinity capture agent. Given the appropriate linker, such chimeric compounds can exhibit affinities that approach the product of the affinities for the two individual ligands for the PET. To illustrate, a collection of compounds is screened at high concentrations for weak interactors of a target PET. The compounds that do not compete with one another are then identified

and a library of chimeric compounds is made with linkers of different length. This library is then screened for binding to the PET at much lower concentrations to identify high affinity binders. Such a technique may also be applied to peptides or any other type of modest-affinity PET-binding compound.

H. Labels for Capture Agents

The capture agents of the present invention may be modified to enable detection using techniques known to one of ordinary skill in the art, such as fluorescent, radioactive, chromatic, optical, and other physical or chemical labels, as described herein below.

I. Miscellaneous

In addition, for any given PET, multiple capture agents belonging to each of the above described categories of capture agents may be available. These multiple capture agents may have different properties, such as affinity / avidity / specificity for the PET. Different affinities are useful in covering the wide dynamic ranges of expression which some proteins can exhibit. Depending on specific use, in any given array of capture agents, different types / amounts of capture agents may be present on a single chip / array to achieve optimal overall performance.

In a preferred embodiment, capture agents are raised against PETs that are located on the surface of the protein of interest, *e.g.*, hydrophilic regions. PETs that are located on the surface of the protein of interest may be identified using any of the well known software available in the art. For example, the Naccess program may be used.

Naccess is a program that calculates the accessible area of a molecule from a PDB (Protein Data Bank) format file. It can calculate the atomic and residue accessibilities for both proteins and nucleic acids. Naccess calculates the atomic accessible area when a probe is rolled around the Van der Waal's surface of a macromolecule. Such three-dimensional co-ordinate sets are available from the PDB

at the Brookhaven National laboratory. The program uses the Lee & Richards (1971) *J. Mol. Biol.*, 55, 379-400 method, whereby a probe of given radius is rolled around the surface of the molecule, and the path traced out by its center is the accessible surface.

The solvent accessibility method described in Boger, J., Emini, E.A. & Schmidt, A., Surface probability profile-An heuristic approach to the selection of synthetic peptide antigens, Reports on the Sixth International Congress in Immunology (Toronto) 1986 p.250 also may be used to identify PETs that are located on the surface of the protein of interest. The package MOLMOL (Koradi, R. *et al.* (1996) *J. Mol. Graph.* 14:51–55) and Eisenhaber's ASC method (Eisenhaber and Argos (1993) *J. Comput. Chem.* 14:1272–1280; Eisenhaber *et al.* (1995) *J. Comput. Chem.* 16:273–284) may also be used.

In another embodiment, capture agents are raised that are designed to bind with peptides generated by digestion of intact proteins rather than with accessible peptidic surface regions on the proteins. In this embodiment, it is preferred to employ a fragmentation protocol which reproducibly generates all of the PETs in the sample under study.

II. Tools Comprising Capture Agents (Arrays, etc.)

In certain embodiments, to construct arrays, *e.g.*, high-density arrays, of capture agents for efficient screening of complex chemical or biological samples or large numbers of compounds, the capture agents need to be immobilized onto a solid support (*e.g.*, a planar support or a bead). A variety of methods are known in the art for attaching biological molecules to solid supports. See, generally, Affinity Techniques, Enzyme Purification: Part B, Meth. Enz. 34 (ed. W. B. Jakoby and M. Wilchek, Acad. Press, N.Y. 1974) and Immobilized Biochemicals and Affinity Chromatography, Adv. Exp. Med. Biol. 42 (ed. R. Dunlap, Plenum Press, N.Y. 1974). The following are a few considerations when constructing arrays.

A. Formats and surfaces consideration

Protein arrays have been designed as a miniaturisation of familiar immunoassay methods such as ELISA and dot blotting, often utilizing fluorescent readout, and facilitated by robotics and high throughput detection systems to enable multiple assays to be carried out in parallel. Common physical supports include glass slides, silicon, microwells, nitrocellulose or PVDF membranes, and magnetic and other microbeads. While microdrops of protein delivered onto planar surfaces are widely used, related alternative architectures include CD centrifugation devices based on developments in microfluidics [Gyros] and specialized chip designs, such as engineered microchannels in a plate [The Living Chip™, Biotrove] and tiny 3D posts on a silicon surface [Zyomyx]. Particles in suspension can also be used as the basis of arrays, providing they are coded for identification; systems include color coding for microbeads [Luminex, Bio-Rad] and semiconductor nanocrystals [QDots™, Quantum Dots], and barcoding for beads [UltraPlex™, Smartbeads] and multimetal microrods [Nanobarcodes™ particles, Surromed]. Beads can also be assembled into planar arrays on semiconductor chips [LEAPS technology, BioArray Solutions].

B. Immobilisation considerations

The variables in immobilization of proteins such as antibodies include both the coupling reagent and the nature of the surface being coupled to. Ideally, the immobilization method used should be reproducible, applicable to proteins of different properties (size, hydrophilic, hydrophobic), amenable to high throughput and automation, and compatible with retention of fully functional protein activity. Orientation of the surface-bound protein is recognized as an important factor in presenting it to ligand or substrate in an active state; for capture arrays the most efficient binding results are obtained with orientated capture reagents, which generally requires site-specific labeling of the protein.

The properties of a good protein array support surface are that it should be chemically stable before and after the coupling procedures, allow good spot morphology, display minimal nonspecific binding, not contribute a background in detection systems, and be compatible with different detection systems.

Both covalent and noncovalent methods of protein immobilization are used and have various pros and cons. Passive adsorption to surfaces is methodologically simple, but allows little quantitative or orientational control; it may or may not alter the functional properties of the protein, and reproducibility and efficiency are variable. Covalent coupling methods provide a stable linkage, can be applied to a range of proteins and have good reproducibility; however, orientation may be variable, chemical dramatization may alter the function of the protein and requires a stable interactive surface. Biological capture methods utilizing a tag on the protein provide a stable linkage and bind the protein specifically and in reproducible orientation, but the biological reagent must first be immobilized adequately and the array may require special handling and have variable stability.

Several immobilization chemistries and tags have been described for fabrication of protein arrays. Substrates for covalent attachment include glass slides coated with amino- or aldehyde-containing silane reagents [Telechem]. In the Versalinx™ system [Prolinx], reversible covalent coupling is achieved by interaction between the protein derivatized with phenyldiboronic acid, and salicylhydroxamic acid immobilized on the support surface. This also has low background binding and low intrinsic fluorescence and allows the immobilized proteins to retain function. Noncovalent binding of unmodified protein occurs within porous structures such as HydroGel™ [PerkinElmer], based on a 3-dimensional polyacrylamide gel; this substrate is reported to give a particularly low background on glass microarrays, with a high capacity and retention of protein function. Widely used biological capture methods are through biotin / streptavidin or hexahistidine / Ni interactions, having modified the protein appropriately. Biotin may be conjugated to a poly-lysine backbone immobilized on a surface such as titanium dioxide [Zyomyx] or tantalum pentoxide [Zeptosens].

Arenkov *et al.*, for example, have described a way to immobilize proteins while preserving their function by using microfabricated polyacrylamide gel pads to capture proteins, and then accelerating diffusion through the matrix by microelectrophoresis (Arenkov *et al.* (2000), Anal Biochem 278(2):123-31). The patent literature also describes a number of different methods for attaching biological molecules to solid supports. For example, U.S. Patent No. 4,282,287

describes a method for modifying a polymer surface through the successive application of multiple layers of biotin, avidin, and extenders. U.S. Patent No. 4,562,157 describes a technique for attaching biochemical ligands to surfaces by attachment to a photochemically reactive arylazide. U.S. Patent No. 4,681,870 describes a method for introducing free amino or carboxyl groups onto a silica matrix, in which the groups may subsequently be covalently linked to a protein in the presence of a carbodiimide. In addition, U.S. Patent No. 4,762,881 describes a method for attaching a polypeptide chain to a solid substrate by incorporating a light-sensitive unnatural amino acid group into the polypeptide chain and exposing the product to low-energy ultraviolet light.

The surface of the support is chosen to possess, or is chemically derivatized to possess, at least one reactive chemical group that can be used for further attachment chemistry. There may be optional flexible adapter molecules interposed between the support and the capture agents. In one embodiment, the capture agents are physically adsorbed onto the support.

In certain embodiments of the invention, a capture agent is immobilized on a support in ways that separate the capture agent's PET binding site region and the region where it is linked to the support. In a preferred embodiment, the capture agent is engineered to form a covalent bond between one of its termini to an adapter molecule on the support. Such a covalent bond may be formed through a Schiff-base linkage, a linkage generated by a Michael addition, or a thioether linkage.

In order to allow attachment by an adapter or directly by a capture agent, the surface of the substrate may require preparation to create suitable reactive groups. Such reactive groups could include simple chemical moieties such as amino, hydroxyl, carboxyl, carboxylate, aldehyde, ester, amide, amine, nitrile, sulfonyl, phosphoryl, or similarly chemically reactive groups. Alternatively, reactive groups may comprise more complex moieties that include, but are not limited to, sulfo-N-hydroxysuccinimide, nitrilotriacetic acid, activated hydroxyl, haloacetyl (e.g., bromoacetyl, iodoacetyl), activated carboxyl, hydrazide, epoxy, aziridine, sulfonylchloride, trifluoromethyldiaziridine, pyridyldisulfide, N-acyl-imidazole, imidazolecarbamate, succinimidylcarbonate, arylazide, anhydride, diazoacetate,

benzophenone, isothiocyanate, isocyanate, imidoester, fluorobenzene, biotin and avidin. Techniques of placing such reactive groups on a substrate by mechanical, physical, electrical or chemical means are well known in the art, such as described by U.S. Pat. No. 4,681,870, incorporated herein by reference.

Once the initial preparation of reactive groups on the substrate is completed (if necessary), adapter molecules optionally may be added to the surface of the substrate to make it suitable for further attachment chemistry. Such adapters covalently join the reactive groups already on the substrate and the capture agents to be immobilized, having a backbone of chemical bonds forming a continuous connection between the reactive groups on the substrate and the capture agents, and having a plurality of freely rotating bonds along that backbone. Substrate adapters may be selected from any suitable class of compounds and may comprise polymers or copolymers of organic acids, aldehydes, alcohols, thiols, amines and the like. For example, polymers or copolymers of hydroxy-, amino-, or di-carboxylic acids, such as glycolic acid, lactic acid, sebacic acid, or sarcosine may be employed. Alternatively, polymers or copolymers of saturated or unsaturated hydrocarbons such as ethylene glycol, propylene glycol, saccharides, and the like may be employed. Preferably, the substrate adapter should be of an appropriate length to allow the capture agent, which is to be attached, to interact freely with molecules in a sample solution and to form effective binding. The substrate adapters may be either branched or unbranched, but this and other structural attributes of the adapter should not interfere stereochemically with relevant functions of the capture agents, such as a PET interaction. Protection groups, known to those skilled in the art, may be used to prevent the adapter's end groups from undesired or premature reactions. For instance, U.S. Pat. No. 5,412,087, incorporated herein by reference, describes the use of photo-removable protection groups on a adapter's thiol group.

To preserve the binding affinity of a capture agent, it is preferred that the capture agent be modified so that it binds to the support substrate at a region separate from the region responsible for interacting with it's ligand, *i.e.*, the PET.

Methods of coupling the capture agent to the reactive end groups on the surface of the substrate or on the adapter include reactions that form linkage such as

thioether bonds, disulfide bonds, amide bonds, carbamate bonds, urea linkages, ester bonds, carbonate bonds, ether bonds, hydrazone linkages, Schiff-base linkages, and noncovalent linkages mediated by, for example, ionic or hydrophobic interactions. The form of reaction will depend, of course, upon the available reactive groups on both the substrate/adapter and capture agent.

C. Array fabrication consideration

Preferably, the immobilized capture agents are arranged in an array on a solid support, such as a silicon-based chip or glass slide. One or more capture agents designed to detect the presence (and optionally the concentration) of a given known protein (one previously recognized as existing) is immobilized at each of a plurality of cells / regions in the array. Thus, a signal at a particular cell / region indicates the presence of a known protein in the sample, and the identity of the protein is revealed by the position of the cell. Alternatively, capture agents for one or a plurality of PET are immobilized on beads, which optionally are labeled to identify their intended target analyte, or are distributed in an array such as a microwell plate.

In one embodiment, the microarray is high density, with a density over about 100, preferably over about 1000, 1500, 2000, 3000, 4000, 5000 and further preferably over about 9000, 10000, 11000, 12000 or 13000 spots per cm^2, formed by attaching capture agents onto a support surface which has been functionalized to create a high density of reactive groups or which has been functionalized by the addition of a high density of adapters bearing reactive groups. In another embodiment, the microarray comprises a relatively small number of capture agents, e.g., 10 to 50, selected to detect in a sample various combinations of specific proteins which generate patterns probative of disease diagnosis, cell type determination, pathogen identification, etc.

Although the characteristics of the substrate or support may vary depending upon the intended use, the shape, material and surface modification of the substrates must be considered. Although it is preferred that the substrate have at least one surface which is substantially planar or flat, it may also include indentations,

protuberances, steps, ridges, terraces and the like and may have any geometric form (*e.g.*, cylindrical, conical, spherical, concave surface, convex surface, string, or a combination of any of these). Suitable substrate materials include, but are not limited to, glasses, ceramics, plastics, metals, alloys, carbon, papers, agarose, silica, quartz, cellulose, polyacrylamide, polyamide, and gelatin, as well as other polymer supports, other solid-material supports, or flexible membrane supports. Polymers that may be used as substrates include, but are not limited to: polystyrene; poly(tetra)fluoroethylene (PTFE); polyvinylidenedifluoride; polycarbonate; polymethylmethacrylate; polyvinylethylene; polyethyleneimine; polyoxymethylene (POM); polyvinylphenol; polylactides; polymethacrylimide (PMI); polyalkenesulfone (PAS); polypropylene; polyethylene; polyhydroxyethylmethacrylate (HEMA); polydimethylsiloxane; polyacrylamide; polyimide; and various block co-polymers. The substrate can also comprise a combination of materials, whether water-permeable or not, in multi-layer configurations. A preferred embodiment of the substrate is a plain 2.5 cm x 7.5 cm glass slide with surface Si-OH functionalities.

Array fabrication methods include robotic contact printing, ink-jetting, piezoelectric spotting and photolithography. A number of commercial arrayers are available [e.g. Packard Biosience] as well as manual equipment [V & P Scientific]. Bacterial colonies can be robotically gridded onto PVDF membranes for induction of protein expression in situ.

At the limit of spot size and density are nanoarrays, with spots on the nanometer spatial scale, enabling thousands of reactions to be performed on a single chip less than 1mm square. BioForce Laboratories have developed nanoarrays with 1521 protein spots in 85sq microns, equivalent to 25 million spots per sq cm, at the limit for optical detection; their readout methods are fluorescence and atomic force microscopy (AFM).

A microfluidics system for automated sample incubation with arrays on glass slides and washing has been codeveloped by NextGen and PerkinElmer Lifesciences.

For example, capture agent microarrays may be produced by a number of

means, including "spotting" wherein small amounts of the reactants are dispensed to particular positions on the surface of the substrate. Methods for spotting include, but are not limited to, microfluidics printing, microstamping (see, *e.g.*, U.S. Pat. No. 5,515,131, U.S. Pat. No. 5,731,152, Martin, B.D. *et al.* (1998), *Langmuir* 14: 3971-3975 and Haab, BB *et al.* (2001) *Genome Biol* 2 and MacBeath, G. *et al.* (2000) *Science* 289: 1760-1763), microcontact printing (see, *e.g.*, PCT Publication WO 96/29629), inkjet head printing (Roda, A. *et al.* (2000) *BioTechniques* 28: 492-496, and Silzel, J.W. *et al.* (1998) *Clin Chem* 44: 2036-2043), microfluidic direct application (Rowe, C.A. *et al.* (1999) *Anal Chem* 71: 433-439 and Bernard, A. *et al.* (2001), *Anal Chem* 73: 8-12) and electrospray deposition (Morozov, V.N. *et al.* (1999) *Anal Chem* 71: 1415-1420 and Moerman R. *et al.* (2001) *Anal Chem* 73: 2183-2189). Generally, the dispensing device includes calibrating means for controlling the amount of sample deposition, and may also include a structure for moving and positioning the sample in relation to the support surface. The volume of fluid to be dispensed per capture agent in an array varies with the intended use of the array, and available equipment. Preferably, a volume formed by one dispensation is less than 100 nL, more preferably less than 10 nL, and most preferably about 1nL. The size of the resultant spots will vary as well, and in preferred embodiments these spots are less than 20,000 μm in diameter, more preferably less than 2,000 μm in diameter, and most preferably about 150-200 μm in diameter (to yield about 1600 spots per square centimeter). Solutions of blocking agents may be applied to the microarrays to prevent non-specific binding by reactive groups that have not bound to a capture agent. Solutions of bovine serum albumin (BSA), casein, or nonfat milk, for example, may be used as blocking agents to reduce background binding in subsequent assays.

In preferred embodiments, high-precision, contact-printing robots are used to pick up small volumes of dissolved capture agents from the wells of a microtiter plate and to repetitively deliver approximately 1 nL of the solutions to defined locations on the surfaces of substrates, such as chemically-derivatized glass microscope slides. Examples of such robots include the GMS 417 Arrayer, commercially available from Affymetrix of Santa Clara, CA, and a split pin arrayer constructed according to instructions downloadable from the Brown lab website at

http://cmgm.stanford.edu/pbrown. This results in the formation of microscopic spots of compounds on the slides. It will be appreciated by one of ordinary skill in the art, however, that the current invention is not limited to the delivery of 1 nL volumes of solution, to the use of particular robotic devices, or to the use of chemically derivatized glass slides, and that alternative means of delivery can be used that are capable of delivering picoliter or smaller volumes. Hence, in addition to a high precision array robot, other means for delivering the compounds can be used, including, but not limited to, ink jet printers, piezoelectric printers, and small volume pipetting robots.

In one embodiment, the compositions, *e.g.*, microarrays or beads, comprising the capture agents of the present invention may also comprise other components, *e.g.*, molecules that recognize and bind specific peptides, metabolites, drugs or drug candidates, RNA, DNA, lipids, and the like. Thus, an array of capture agents only some of which bind a PET can comprise an embodiment of the invention.

As an alternative to planar microarrays, bead-based assays combined with fluorescence-activated cell sorting (FACS) have been developed to perform multiplexed immunoassays. Fluorescence-activated cell sorting has been routinely used in diagnostics for more than 20 years. Using mAbs, cell surface markers are identified on normal and neoplastic cell populations enabling the classification of various forms of leukemia or disease monitoring (recently reviewed by Herzenberg et al. *Immunol Today* **21** (2000), pp. 383–390).

Bead-based assay systems employ microspheres as solid support for the capture molecules instead of a planar substrate, which is conventionally used for microarray assays. In each individual immunoassay, the capture agent is coupled to a distinct type of microsphere. The reaction takes place on the surface of the microspheres. The individual microspheres are color-coded by a uniform and distinct mixture of red and orange fluorescent dyes. After coupling to the appropriate capture molecule, the different color-coded bead sets can be pooled and the immunoassay is performed in a single reaction vial. Product formation of the PET targets with their respective capture agents on the different bead types can be detected with a fluorescence-based reporter system. The signal intensities are

measured in a flow cytometer, which is able to quantify the amount of captured targets on each individual bead. Each bead type and thus each immobilized target is identified using the color code measured by a second fluorescence signal. This allows the multiplexed quantification of multiple targets from a single sample. Sensitivity, reliability and accuracy are similar to those observed with standard microtiter ELISA procedures. Color-coded microspheres can be used to perform up to a hundred different assay types simultaneously (LabMAP system, Laboratory Muliple Analyte Profiling, Luminex, Austin, TX, USA). For example, microsphere-based systems have been used to simultaneously quantify cytokines or autoantibodies from biological samples (Carson and Vignali, *J Immunol Methods* **227** (1999), pp. 41–52; Chen et al., *Clin Chem* **45** (1999), pp. 1693–1694; Fulton et al., *Clin Chem* **43** (1997), pp. 1749–1756). Bellisario et al. (*Early Hum Dev* **64** (2001), pp. 21–25) have used this technology to simultaneously measure antibodies to three HIV-1 antigens from newborn dried blood-spot specimens.

Bead-based systems have several advantages. As the capture molecules are coupled to distinct microspheres, each individual coupling event can be perfectly analyzed. Thus, only quality-controlled beads can be pooled for multiplexed immunoassays. Furthermore, if an additional parameter has to be included into the assay, one must only add a new type of loaded bead. No washing steps are required when performing the assay. The sample is incubated with the different bead types together with fluorescently labeled detection antibodies. After formation of the sandwich immuno-complex, only the fluorophores that are definitely bound to the surface of the microspheres are counted in the flow cytometer.

D. Related non-array formats

An alternative to an array of capture agents is one made through the so-called "molecular imprinting" technology, in which peptides (e.g. selected PETs) are used as templates to generate structurally complementary, sequence-specific cavities in a polymerisable matrix; the cavities can then specifically capture (digested) proteins which have the appropriate primary amino acid sequence [ProteinPrint™, Aspira Biosystems]. To illustrate, a chosen PET can be synthesized, and a universal matrix

of polymerizable monomers is allowed to self assemble around the peptide and crosslinked into place. The PET, or template, is then removed, leaving behind a cavity complementary in shape and functionality. The cavities can be formed on a film, discrete sites of an array or the surface of beads. When a sample of fragmented proteins is exposed to the capture agent, the polymer will selectively retain the target protein containing the PET and exclude all others. After the washing, only the bound PET-containing peptides remain. Common staining and tagging procedures, or any of the non-labeling techniques described below can be used to detect expression levels and/or post translational modifications. Alternatively, the captured peptides can be eluted for further analysis such as mass spectrometry analysis. See WO 01/61354 A1, WO 01/61355 A1, and related applications / patents.

Another methodology which can be used diagnostically and in expression profiling is the ProteinChip® array [Ciphergen], in which solid phase chromatographic surfaces bind proteins with similar characteristics of charge or hydrophobicity from mixtures such as plasma or tumor extracts, and SELDI-TOF mass spectrometry is used to detection the retained proteins. The ProteinChip® is credited with the ability to identify novel disease markers. However, this technology differs from the protein arrays under discussion here since, in general, it does not involve immobilization of individual proteins for detection of specific ligand interactions.

E. Single Assay Format

PET-specific affinity capture agents can also be used in a single assay format. For example, such agents can be used to develop a better assay for detecting circulating agents, such as PSA, by providing increased sensitivity, dynamic range and/or recovery rate. For instance, the single assays can have functional performance characteristics which exceed traditional ELISA and other immunoassays, such as one or more of the following: a regression coefficient (R2) of 0.95 or greater for a reference standard, e.g., a comparable control sample, more preferably an R2 greater than 0.97, 0.99 or even 0.995; a recovery rate of at least 50 percent, and more preferably at least 60, 75, 80 or even 90 percent; a positive predictive value for

occurrence of the protein in a sample of at least 90 percent, more preferably at least 95, 98 or even 99 percent; a diagnostic sensitivity (DSN) for occurrence of the protein in a sample of 99 percent or higher, more preferably at least 99.5 or even 99.8 percent; a diagnostic specificity (DSP) for occurrence of the protein in a sample of 99 percent or higher, more preferably at least 99.5 or even 99.8 percent.

III. Methods of Detecting Binding Events

The capture agents of the invention, as well as compositions, *e.g.*, microarrays or beads, comprising these capture agents have a wide range of applications in the health care industry, *e.g.*, in therapy, in clinical diagnostics, in *in vivo* imaging or in drug discovery. The capture agents of the present invention also have industrial and environmental applications, *e.g.*, in environmental diagnostics, industrial diagnostics, food safety, toxicology, catalysis of reactions, or high-throughput screening; as well as applications in the agricultural industry and in basic research, *e.g.*, protein sequencing.

The capture agents of the present invention are a powerful analytical tool that enables a user to detect a specific protein, or group of proteins of interest present within complex samples. In addition, the invention allow for efficient and rapid analysis of samples; sample conservation and direct sample comparison. The invention enables "multi-parametric" analysis of protein samples. As used herein, a "multi-parametric" analysis of a protein sample is intended to include an analysis of a protein sample based on a plurality of parameters. For example, a protein sample may be contacted with a plurality of PETs, each of the PETs being able to detect a different protein within the sample. Based on the combination and, preferably the relative concentration, of the proteins detected in the sample the skilled artisan would be able to determine the identity of a sample, diagnose a disease or pre-disposition to a disease, or determine the stage of a disease

The capture agents of the present invention may be used in any method suitable for detection of a protein or a polypeptide, such as, for example, in immunoprecipitations, immunocytochemistry, Western Blots or nuclear magnetic resonance spectroscopy (NMR).

To detect the presence of a protein that interacts with a capture agent, a variety of art known methods may be used. The protein to be detected may be labeled with a detectable label, and the amount of bound label directly measured. The term "label" is used herein in a broad sense to refer to agents that are capable of providing a detectable signal, either directly or through interaction with one or more additional members of a signal producing system. Labels that are directly detectable and may find use in the present invention include, for example, fluorescent labels such as fluorescein, rhodamine, BODIPY, cyanine dyes (e.g. from Amersham Pharmacia), Alexa dyes (e.g. from Molecular Probes, Inc.), fluorescent dye phosphoramidites, beads, chemilumninescent compounds, colloidal particles, and the like. Suitable fluorescent dyes are known in the art, including fluoresceinisothiocyanate (FITC); rhodamine and rhodamine derivatives; Texas Red; phycoerythrin; allophycocyanin; 6-carboxyfluorescein (6-FAM); 2',7'-dimethoxy-41,51-dichloro carboxyfluorescein (JOE); 6-carboxy-X-rhodamine (ROX); 6-carboxy-21,41,71,4,7-hexachlorofluorescein (HEX); 5-carboxyfluorescein (5-FAM); N,N,N1,N'-tetramethyl carboxyrhodamine (TAMRA); sulfonated rhodamine; Cy3; Cy5, etc. Radioactive isotopes, such as ^{35}S, ^{32}P, ^{3}H, ^{125}I, etc., and the like can also be used for labeling. In addition, labels may also include near-infrared dyes (Wang *et al., Anal. Chem.*, 72:5907-5917 (2000), upconverting phosphors (Hampl *et al., Anal. Biochem.*, 288:176-187 (2001), DNA dendrimers (Stears *et al., Physiol. Genomics* 3: 93-99 (2000), quantum dots (Bruchez *et al., Science* 281:2013-2016 (1998), latex beads (Okana *et al., Anal. Biochem.* 202:120-125 (1992), selenium particles (Stimpson *et al., Proc. Natl. Acad. Sci.* 92:6379-6383 (1995), and europium nanoparticles (Harma *et al., Clin. Chem.* 47:561-568 (2001). The label is one that preferably does not provide a variable signal, but instead provides a constant and reproducible signal over a given period of time.

A very useful labeling agent is water-soluble quantum dots, or so-called "functionalized nanocrystals" or "semiconductor nanocrystals" as described in U.S. Pat. No. 6,114,038. Generally, quantum dots can be prepared which result in relative monodispersity (e.g., the diameter of the core varying approximately less than 10% between quantum dots in the preparation), as has been described previously (Bawendi et al., 1993, J. Am. Chem. Soc. 115:8706). Examples of quantum dots are

known in the art to have a core selected from the group consisting of CdSe, CdS, and CdTe (collectively referred to as "CdX")(see, e.g., Norris et al., 1996, Physical Review B. 53:16338-16346; Nirmal et al., 1996, Nature 383:802-804; Empedocles et al., 1996, Physical Review Letters 77:3873-3876; Murray et al., 1996, Science 270: 1355-1338; Effros et al., 1996, Physical Review B. 54:4843-4856; Sacra et al., 1996, J. Chem. Phys. 103:5236-5245; Murakoshi et al., 1998, J. Colloid Interface Sci. 203:225-228; Optical Materials and Engineering News, 1995, Vol. 5, No. 12; and Murray et al., 1993, J. Am. Chem. Soc. 115:8706-8714; the disclosures of which are hereby incorporated by reference).

CdX quantum dots have been passivated with an inorganic coating ("shell") uniformly deposited thereon. Passivating the surface of the core quantum dot can result in an increase in the quantum yield of the luminescence emission, depending on the nature of the inorganic coating. The shell which is used to passivate the quantum dot is preferably comprised of YZ wherein Y is Cd or Zn, and Z is S, or Se. Quantum dots having a CdX core and a YZ shell have been described in the art (see, e.g., Danek et al., 1996, Chem. Mater. 8:173-179; Dabbousi et al., 1997, J. Phys. Chem. B 101:9463; Rodriguez-Viejo et al., 1997, Appl. Phys. Lett. 70:2132-2134; Peng et al., 1997, J. Am. Chem. Soc. 119:7019-7029; 1996, Phys. Review B. 53:16338-16346; the disclosures of which are hereby incorporated by reference). However, the above described quantum dots, passivated using an inorganic shell, have only been soluble in organic, non-polar (or weakly polar) solvents. To make quantum dots useful in biological applications, it is desirable that the quantum dots are water-soluble. "Water-soluble" is used herein to mean sufficiently soluble or suspendable in an aqueous-based solution, such as in water or water-based solutions or buffer solutions, including those used in biological or molecular detection systems as known by those skilled in the art.

U.S. Pat. No. 6,114,038 provides a composition comprising functionalized nanocrystals for use in non-isotopic detection systems. The composition comprises quantum dots (capped with a layer of a capping compound) that are water-soluble and functionalized by operably linking, in a successive manner, one or more additional compounds. In a preferred embodiment, the one or more additional compounds form successive layers over the nanocrystal. More particularly, the

functionalized nanocrystals comprise quantum dots capped with the capping compound, and have at least a diaminocarboxylic acid which is operatively linked to the capping compound. Thus, the functionalized nanocrystals may have a first layer comprising the capping compound, and a second layer comprising a diaminocarboxylic acid; and may further comprise one or more successive layers including a layer of amino acid, a layer of affinity ligand, or multiple layers comprising a combination thereof. The composition comprises a class of quantum dots that can be excited with a single wavelength of light resulting in detectable luminescence emissions of high quantum yield and with discrete luminescence peaks. Such functionalized nanocrystal may be used to label capture agents of the instant invention for their use in the detection and/or quantitation of the binding events.

U.S. Pat. No. 6,326,144 describes quantum dots (QDs) having a characteristic spectral emission, which is tunable to a desired energy by selection of the particle size of the quantum dot. For example, a 2 nanometer quantum dot emits green light, while a 5 nanometer quantum dot emits red light. The emission spectra of quantum dots have linewidths as narrow as 25-30 nm depending on the size heterogeneity of the sample, and lineshapes that are symmetric, gaussian or nearly gaussian with an absence of a tailing region. The combination of tunability, narrow linewidths, and symmetric emission spectra without a tailing region provides for high resolution of multiply-sized quantum dots within a system and enables researchers to examine simultaneously a variety of biological moieties tagged with QDs. In addition, the range of excitation wavelengths of the nanocrystal quantum dots is broad and can be higher in energy than the emission wavelengths of all available quantum dots. Consequently, this allows the simultaneous excitation of all quantum dots in a system with a single light source, usually in the ultraviolet or blue region of the spectrum. QDs are also more robust than conventional organic fluorescent dyes and are more resistant to photobleaching than the organic dyes. The robustness of the QD also alleviates the problem of contamination of the degradation products of the organic dyes in the system being examined. These QDs can be used for labeling capture agents of protein, nucleic acid, and other biological molecules in nature. Cadmium Selenide quantum dot nanocrystals are available from Quantum

Dot Corporation of Hayward, California.

Alternatively, the sample to be tested is not labeled, but a second stage labeled reagent is added in order to detect the presence or quantitate the amount of protein in the sample. Such "sandwich based" methods of detection have the disadvantage that two capture agents must be developed for each protein, one to capture the PET and one to label it once captured. Such methods have the advantage that they are characterized by an inherently improved signal to noise ratio as they exploit two binding reactions at different points on a peptide, thus the presence and/or concentration of the protein can be measured with more accuracy and precision because of the increased signal to noise ratio.

In yet another embodiment, the subject capture array can be a "virtual arrays". For example, a virtual array can be generated in which antibodies or other capture agents are immobilized on beads whose identity, with respect to the particular PET it is specific for as a consequence to the associated capture agent, is encoded by a particular ratio of two or more covalently attached dyes. Mixtures of encoded PET-beads are added to a sample, resulting in capture of the PET entities recognized by the immobilized capture agents.

To quantitate the captured species, a sandwich assay with fluorescently labeled antibodies that bind the captured PET, or a competitive binding assay with a fluorescently labeled ligand for the capture agent, are added to the mix. In one embodiment, the labeled ligand is a labeled PET that competes with the analyte PET for binding to the capture agent. The beads are then introduced into an instrument, such as a flow cytometer, that reads the intensity of the various fluorescence signals on each bead, and the identity of the bead can be determined by measuring the ratio of the dyes (Figure 3). This technology is relatively fast and efficient, and can be adapted by researchers to monitor almost any set of PET of interest.

In another embodiment, an array of capture agents are embedded in a matrix suitable for ionization (such as described in Fung et al. (2001) Curr. Opin. Biotechnol. 12:65–69). After application of the sample and removal of unbound molecules (by washing), the retained PET proteins are analyzed by mass spectrometry. In some instances, further proteolytic digestion of the bound species

with trypsin may be required before ionization, particularly if electrospray is the means for ionizing the peptides.

All the above named reagents may be used to label the capture agents. Preferably, the capture agent to be labeled is combined with an activated dye that reacts with a group present on the protein to be detected, *e.g.*, amine groups, thiol groups, or aldehyde groups.

The label may also be a covalently bound enzyme capable of providing a detectable product signal after addition of suitable substrate. Examples of suitable enzymes for use in the present invention include horseradish peroxidase, alkaline phosphatase, malate dehydrogenase and the like.

Enzyme-Linked Immunosorbent Assay (ELISA) may also be used for detection of a protein that interacts with a capture agent. In an ELISA, the indicator molecule is covalently coupled to an enzyme and may be quantified by determining with a spectrophotometer the initial rate at which the enzyme converts a clear substrate to a correlated product. Methods for performing ELISA are well known in the art and described in, for example, Perlmann, H. and Perlmann, P. (1994). Enzyme-Linked Immunosorbent Assay. In: Cell Biology: A Laboratory Handbook. San Diego, CA, Academic Press, Inc., 322-328; Crowther, J.R. (1995). Methods in Molecular Biology, Vol. 42-ELISA: Theory and Practice. Humana Press, Totowa, NJ.; and Harlow, E. and Lane, D. (1988). Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 553-612, the contents of each of which are incorporated by reference. Sandwich (capture) ELISA may also be used to detect a protein that interacts with two capture agents. The two capture agents may be able to specifically interact with two PETs that are present on the same peptide (*e.g.*, the peptide which has been generated by fragmentation of the sample of interest, as described above). Alternatively, the two capture agents may be able to specifically interact with one PET and one non-unique amino acid sequence, both present on the same peptide (*e.g.*, the peptide which has been generated by fragmentation of the sample of interest, as described above). Sandwich ELISAs for the quantitation of proteins of interest are especially valuable when the concentration of the protein in the sample is low and/or the protein of interest is present in a

sample that contains high concentrations of contaminating proteins.

A fully-automated, microarray-based approach for high-throughput, ELISAs was described by Mendoza et al. (*BioTechniques* **27**:778-780,782-786,788, 1999). This system consisted of an optically flat glass plate with 96 wells separated by a Teflon mask. More than a hundred capture molecules were immobilized in each well. Sample incubation, washing and fluorescence-based detection were performed with an automated liquid pipettor. The microarrays were quantitatively imaged with a scanning charge-coupled device (CCD) detector. Thus, the feasibility of multiplex detection of arrayed antigens in a high-throughput fashion using marker antigens could be successfully demonstrated. In addition, Silzel et al. (*Clin Chem* **44** pp. 2036–2043, 1998) could demonstrate that multiple IgG subclasses can be detected simultaneously using microarray technology. Wiese et al. (*Clin Chem* **47** pp. 1451–1457, 2001) were able to measure prostate-specific antigen (PSA), -(1)-antichymotrypsin-bound PSA and interleukin-6 in a microarray format. Arenkov et al. (supra) carried out microarray sandwich immunoassays and direct antigen or antibody detection experiments using a modified polyacrylamide gel as substrate for immobilized capture molecules.

Most of the microarray assay formats described in the art rely on chemiluminescence- or fluorescence-based detection methods. A further improvement with regard to sensitivity involves the application of fluorescent labels and waveguide technology. A fluorescence-based array immunosensor was developed by Rowe et al. (*Anal Chem* 71 (1999), pp. 433–439; and *Biosens Bioelectron* **15** (2000), pp. 579–589) and applied for the simultaneous detection of clinical analytes using the sandwich immunoassay format. Biotinylated capture antibodies were immobilized on avidin-coated waveguides using a flow-chamber module system. Discrete regions of capture molecules were vertically arranged on the surface of the waveguide. Samples of interest were incubated to allow the targets to bind to their capture molecules. Captured targets were then visualized with appropriate fluorescently labeled detection molecules. This array immunosensor was shown to be appropriate for the detection and measurement of targets at physiologically relevant concentrations in a variety of clinical samples.

A further increase in the sensitivity using waveguide technology was achieved with the development of the planar waveguide technology (Duveneck et al., *Sens Actuators B* **B38** (1997), pp. 88–95). Thin-film waveguides are generated from a high-refractive material such as Ta_2O_5 that is deposited on a transparent substrate. Laser light of desired wavelength is coupled to the planar waveguide by means of diffractive grating. The light propagates in the planar waveguide and an area of more than a square centimeter can be homogeneously illuminated. At the surface, the propagating light generates a so-called evanescent field. This extends into the solution and activates only fluorophores that are bound to the surface. Fluorophores in the surrounding solution are not excited. Close to the surface, the excitation field intensities can be a hundred times higher than those achieved with standard confocal excitation. A CCD camera is used to identify signals simultaneously across the entire area of the planar waveguide. Thus, the immobilization of the capture molecules in a microarray format on the planar waveguide allows the performance of highly sensitive miniaturized and parallelized immunoassays. This system was successfully employed to detect interleukin-6 at concentrations as low as 40 fM and has the additional advantage that the assay can be performed without washing steps that are usually required to remove unbound detection molecules (Weinberger et al., *Pharmacogenomics* 1 (2000), pp. 395–416).

Alternative strategies pursued to increase sensitivity are based on signal amplification procedures. For example, immunoRCA (immuno rolling circle amplification) involves an oligonucleotide primer that is covalently attached to a detection molecule (such as a second capture agent in a sandwich-type assay format). Using circular DNA as template, which is complementary to the attached oligonucleotide, DNA polymerase will extend the attached oligonucleotide and generate a long DNA molecule consisting of hundreds of copies of the circular DNA, which remains attached to the detection molecule. The incorporation of thousands of fluorescently labeled nucleotides will generate a strong signal. Schweitzer et al. (*Proc Natl Acad Sci USA* **97** (2000), pp. 10113–10119) have evaluated this detection technology for use in microarray-based assays. Sandwich immunoassays for hulgE and prostate-specific antigens were performed in a microarray format. The antigens could be detected at femtomolar concentrations and

it was possible to score single, specifically captured antigens by counting discrete fluorescent signals that arose from the individual antibody–antigen complexes. The authors demonstrated that immunoassays employing rolling circle DNA amplification are a versatile platform for the ultra-sensitive detection of antigens and thus are well suited for use in protein microarray technology.

Radioimmunoassays (RIA) may also be used for detection of a protein that interacts with a capture agent. In a RIA, the indicator molecule is labeled with a radioisotope and it may be quantified by counting radioactive decay events in a scintillation counter. Methods for performing direct or competitive RIA are well known in the art and described in, for example, Cell Biology: A Laboratory Handbook. San Diego, CA, Academic Press, Inc., the contents of which are incorporated herein by reference.

Other immunoassays commonly used to quantitate the levels of proteins in cell samples, and are well-known in the art, can be adapted for use in the instant invention. The invention is not limited to a particular assay procedure, and therefore is intended to include both homogeneous and heterogeneous procedures. Exemplary other immunoassays which can be conducted according to the invention include fluorescence polarization immunoassay (FPIA), fluorescence immunoassay (FIA), enzyme immunoassay (EIA), nephelometric inhibition immunoassay (NIA). An indicator moiety, or label group, can be attached to the subject antibodies and is selected so as to meet the needs of various uses of the method which are often dictated by the availability of assay equipment and compatible immunoassay procedures. General techniques to be used in performing the various immunoassays noted above are known to those of ordinary skill in the art. In one embodiment, the determination of protein level in a biological sample may be performed by a microarray analysis (protein chip).

In several other embodiments, detection of the presence of a protein that interacts with a capture agent may be achieved without labeling. For example, determining the ability of a protein to bind to a capture agent can be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.*

(1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore).

In another embodiment, a biosensor with a special diffractive grating surface may be used to detect / quantitate binding between non-labeled PET-containing peptides in a treated (digested) biological sample and immobilized capture agents at the surface of the biosensor. Details of the technology is described in more detail in B. Cunningham, P. Li, B. Lin, J. Pepper, "Colorimetric resonant reflection as a direct biochemical assay technique," Sensors and Actuators B, Volume 81, p. 316-328, Jan 5 2002, and in PCT No. WO 02/061429 A2 and US 2003/0032039. Briefly, a guided mode resonant phenomenon is used to produce an optical structure that, when illuminated with collimated white light, is designed to reflect only a single wavelength (color). When molecules are attached to the surface of the biosensor, the reflected wavelength (color) is shifted due to the change of the optical path of light that is coupled into the grating. By linking receptor molecules to the grating surface, complementary binding molecules can be detected / quantitated without the use of any kind of fluorescent probe or particle label. The spectral shifts may be analyzed to determine the expression data provided, and to indicate the presence or absence of a particular indication.

The biosensor typically comprises: a two-dimensional grating comprised of a material having a high refractive index, a substrate layer that supports the two-dimensional grating, and one or more detection probes immobilized on the surface of the two-dimensional grating opposite of the substrate layer. When the biosensor is illuminated a resonant grating effect is produced on the reflected radiation spectrum. The depth and period of the two-dimensional grating are less than the wavelength of the resonant grating effect.

A narrow band of optical wavelengths can be reflected from the biosensor when it is illuminated with a broad band of optical wavelengths. The substrate can comprise glass, plastic or epoxy. The two-dimensional grating can comprise a material selected from the group consisting of zinc sulfide, titanium dioxide, tantalum oxide, and silicon nitride.

The substrate and two-dimensional grating can optionally comprise a single unit. The surface of the single unit comprising the two-dimensional grating is coated with a material having a high refractive index, and the one or more detection probes are immobilized on the surface of the material having a high refractive index opposite of the single unit. The single unit can be comprised of a material selected from the group consisting of glass, plastic, and epoxy.

The biosensor can optionally comprise a cover layer on the surface of the two-dimensional grating opposite of the substrate layer. The one or more detection probes are immobilized on the surface of the cover layer opposite of the two-dimensional grating. The cover layer can comprise a material that has a lower refractive index than the high refractive index material of the two-dimensional grating. For example, a cover layer can comprise glass, epoxy, and plastic.

A two-dimensional grating can be comprised of a repeating pattern of shapes selected from the group consisting of lines, squares, circles, ellipses, triangles, trapezoids, sinusoidal waves, ovals, rectangles, and hexagons. The repeating pattern of shapes can be arranged in a linear grid, i.e., a grid of parallel lines, a rectangular grid, or a hexagonal grid. The two-dimensional grating can have a period of about 0.01 microns to about I micron and a depth of about 0.01 microns to about 1 micron.

To illustrate, biochemical interactions occurring on a surface of a calorimetric resonant optical biosensor embedded into a surface of a microarray slide, microtiter plate or other device, can be directly detected and measured on the sensor's surface without the use of fluorescent tags or calorimetric labels. The sensor surface contains an optical structure that, when illuminated with collimated white light, is designed to reflect only a narrow band of wavelengths (color). The narrow wavelength is described as a wavelength "peak." The "peak wavelength value" (PWV) changes when biological material is deposited or removed from the sensor surface, such as when binding occurs. Such binding-induced change of PWV can be measured using a measurement instrument disclosed in US2003/0032039.

In one embodiment, the instrument illuminates the biosensor surface by directing a collimated white light on to the sensor structure. The illuminated light may take the form of a spot of collimated light. Alternatively, the light is generated

in the form of a fan beam. The instrument collects light reflected from the illuminated biosensor surface. The instrument may gather this reflected light from multiple locations on the biosensor surface simultaneously. The instrument can include a plurality of illumination probes that direct the light to a discrete number of positions across the biosensor surface. The instrument measures the Peak Wavelength Values (PWVs) of separate locations within the biosensor-embedded microtiter plate using a spectrometer. In one embodiment, the spectrometer is a single-point spectrometer. Alternatively, an imaging spectrometer is used. The spectrometer can produce a PWV image map of the sensor surface. In one embodiment, the measuring instrument spatially resolves PWV images with less than 200 micron resolution.

In one embodiment, a subwavelength structured surface (SWS) may be used to create a sharp optical resonant reflection at a particular wavelength that can be used to track with high sensitivity the interaction of biological materials, such as specific binding substances or binding partners or both. A colormetric resonant diffractive grating surface acts as a surface binding platform for specific binding substances (such as immobilized capture agents of the instant invention). SWS is an unconventional type of diffractive optic that can mimic the effect of thin-film coatings. (Peng & Morris, "Resonant scattering from two-dimensional gratings," J. Opt. Soc. Am. A, Vol. 13, No. 5, p. 993, May; Magnusson, & Wang, "New principle for optical filters," Appl. Phys. Lett., 61, No. 9, p. 1022, August, 1992; Peng & Morris, "Experimental demonstration of resonant anomalies in diffraction from two-dimensional gratings," Optics Letters, Vol. 21, No. 8, p. 549, April, 1996). A SWS structure contains a surface-relief, two-dimensional grating in which the grating period is small compared to the wavelength of incident light so that no diffractive orders other than the reflected and transmitted zeroth orders are allowed to propagate. A SWS surface narrowband filter can comprise a two-dimensional grating sandwiched between a substrate layer and a cover layer that fills the grating grooves. Optionally, a cover layer is not used. When the effective index of refraction of the grating region is greater than the substrate or the cover layer, a waveguide is created. When a filter is designed accordingly, incident light passes into the waveguide region. A two-dimensional grating structure selectively couples light at a

narrow band of wavelengths into the waveguide. The light propagates only a short distance (on the order of 10-100 micrometers), undergoes scattering, and couples with the forward- and backward-propagating zeroth-order light. This sensitive coupling condition can produce a resonant grating effect on the reflected radiation spectrum, resulting in a narrow band of reflected or transmitted wavelengths (colors). The depth and period of the two-dimensional grating are less than the wavelength of the resonant grating effect.

The reflected or transmitted color of this structure can be modulated by the addition of molecules such as capture agents or their PET-containing binding partners or both, to the upper surface of the cover layer or the two-dimensional grating surface. The added molecules increase the optical path length of incident radiation through the structure, and thus modify the wavelength (color) at which maximum reflectance or transmittance will occur. Thus in one embodiment, a biosensor, when illuminated with white light, is designed to reflect only a single wavelength. When specific binding substances are attached to the surface of the biosensor, the reflected wavelength (color) is shifted due to the change of the optical path of light that is coupled into the grating. By linking specific binding substances to a biosensor surface, complementary binding partner molecules can be detected without the use of any kind of fluorescent probe or particle label. The detection technique is capable of resolving changes of, for example, about 0.1 nm thickness of protein binding, and can be performed with the biosensor surface either immersed in fluid or dried. This PWV change can be detected by a detection system consists of, for example, a light source that illuminates a small spot of a biosensor at normal incidence through, for example, a fiber optic probe. A spectrometer collects the reflected light through, for example, a second fiber optic probe also at normal incidence. Because no physical contact occurs between the excitation/detection system and the biosensor surface, no special coupling prisms are required. The biosensor can, therefore, be adapted to a commonly used assay platform including, for example, microtiter plates and microarray slides. A spectrometer reading can be performed in several milliseconds, thus it is possible to efficiently measure a large number of molecular interactions taking place in parallel upon a biosensor surface, and to monitor reaction kinetics in real time.

Various embodiments, variations of the biosensor described above can be found in US2003/0032039, incorporated herein by reference in its entirety.

One or more specific capture agents may be immobilized on the two-dimensional grating or cover layer, if present. Immobilization may occur by any of the above described methods. Suitable capture agents can be, for example, a nucleic acid, polypeptide, antigen, polyclonal antibody, monoclonal antibody, single chain antibody (scFv), F(ab) fragment, F(ab')2 fragment, Fv fragment, small organic molecule, even cell, virus, or bacteria. A biological sample can be obtained and/or deribed from, for example, blood, plasma, serum, gastrointestinal secretions, homogenates of tissues or tumors, synovial fluid, feces, saliva, sputum, cyst fluid, amniotic fluid, cerebrospinal fluid, peritoneal fluid, lung lavage fluid, semen, lymphatic fluid, tears, or prostatitc fluid. Preferably, one or more specific capture agents are arranged in a microarray of distinct locations on a biosensor. A microarray of capture agents comprises one or more specific capture agents on a surface of a biosensor such that a biosensor surface contains a plurality of distinct locations, each with a different capture agent or with a different amount of a specific capture agent. For example, an array can comprise 1, 10, 100, 1,000, 10,000, or 100,000 distinct locations. A biosensor surface with a large number of distinct locations is called a microarray because one or more specific capture agents are typically laid out in a regular grid pattern in x-y coordinates. However, a microarray can comprise one or more specific capture agents laid out in a regular or irregular pattern.

A microarray spot can range from about 50 to about 500 microns in diameter. Alternatively, a microarray spot can range from about 150 to about 200 microns in diameter. One or more specific capture agents can be bound to their specific PET-containing binding partners.

In one biosensor embodiment, a microarray on a biosensor is created by placing microdroplets of one or more specific capture agents onto, for example, an x-y grid of locations on a two-dimensional grating or cover layer surface. When the biosensor is exposed to a test sample comprising one or more PET binding partners, the binding partners will be preferentially attracted to distinct locations on the

microarray that comprise capture agents that have high affinity for the PET binding partners. Some of the distinct locations will gather binding partners onto their surface, while other locations will not. Thus a specific capture agent specifically binds to its PET binding partner, but does not substantially bind other PET binding partners added to the surface of a biosensor. In an alternative embodiment, a nucleic acid microarray (such as an aptamer array) is provided, in which each distinct location within the array contains a different aptamer capture agent. By application of specific capture agents with a microarray spotter onto a biosensor, specific binding substance densities of 10,000 specific binding substances/in^2 can be obtained. By focusing an illumination beam of a fiber optic probe to interrogate a single microarray location, a biosensor can be used as a label-free microarray readout system.

For the detection of PET binding partners at concentrations of less than about 0.1 ng/ml, one may amplify and transduce binding partners bound to a biosensor into an additional layer on the biosensor surface. The increased mass deposited on the biosensor can be detected as a consequence of increased optical path length. By incorporating greater mass onto a biosensor surface, an optical density of binding partners on the surface is also increased, thus rendering a greater resonant wavelength shift than would occur without the added mass. The addition of mass can be accomplished, for example, enzymatically, through a "sandwich" assay, or by direct application of mass (such as a second capture agent specific for the PET peptide) to the biosensor surface in the form of appropriately conjugated beads or polymers of various size and composition. Since the capture agents are PET-specific, multiple capture agents of different types and specificity can be added together to the captured PETs. This principle has been exploited for other types of optical biosensors to demonstrate sensitivity increases over 1500× beyond sensitivity limits achieved without mass amplification. See, e.g., Jenison et al., "Interference-based detection of nucleic acid targets on optically coated silicon," Nature Biotechnology, 19: 62-65, 2001.

In an alternative embodiment, a biosensor comprises volume surface-relief volume diffractive structures (a SRVD biosensor). SRVD biosensors have a surface that reflects predominantly at a particular narrow band of optical wavelengths when

illuminated with a broad band of optical wavelengths. Where specific capture agents and/or PET binding partners are immobilized on a SRVD biosensor, the reflected wavelength of light is shifted. One-dimensional surfaces, such as thin film interference filters and Bragg reflectors, can select a narrow range of reflected or transmitted wavelengths from a broadband excitation source. However, the deposition of additional material, such as specific capture agents and/or PET binding partners onto their upper surface results only in a change in the resonance linewidth, rather than the resonance wavelength. In contrast, SRVD biosensors have the ability to alter the reflected wavelength with the addition of material, such as specific capture agents and/or binding partners to the surface.

A SRVD biosensor comprises a sheet material having a first and second surface. The first surface of the sheet material defines relief volume diffraction structures. Sheet material can comprise, for example, plastic, glass, semiconductor wafer, or metal film. A relief volume diffractive structure can be, for example, a two-dimensional grating, as described above, or a three-dimensional surface-relief volume diffractive grating. The depth and period of relief volume diffraction structures are less than the resonance wavelength of light reflected from a biosensor. A three-dimensional surface-relief volume diffractive grating can be, for example, a three-dimensional phase-quantized terraced surface relief pattern whose groove pattern resembles a stepped pyramid. When such a grating is illuminated by a beam of broadband radiation, light will be coherently reflected from the equally spaced terraces at a wavelength given by twice the step spacing times the index of refraction of the surrounding medium. Light of a given wavelength is resonantly diffracted or reflected from the steps that are a half-wavelength apart, and with a bandwidth that is inversely proportional to the number of steps. The reflected or diffracted color can be controlled by the deposition of a dielectric layer so that a new wavelength is selected, depending on the index of refraction of the coating.

A stepped-phase structure can be produced first in photoresist by coherently exposing a thin photoresist film to three laser beams, as described previously. See e.g., Cowen, "The recording and large scale replication of crossed holographic grating arrays using multiple beam interferometry," in International Conference on the Application, Theory, and Fabrication of Periodic Structures, Diffraction

Gratings, and Moire Phenomena II, Lerner, ed., Proc. Soc. Photo-Opt. Instrum. Eng., 503, 120-129, 1984; Cowen, "Holographic honeycomb microlens," Opt. Eng. 24, 796-802 (1985); Cowen & Slafer, "The recording and replication of holographic micropatterns for the ordering of photographic emulsion grains in film systems," J Imaging Sci. 31, 100-107, 1987. The nonlinear etching characteristics of photoresist are used to develop the exposed film to create a three-dimensional relief pattern. The photoresist structure is then replicated using standard embossing procedures. For example, a thin silver film may be deposited over the photoresist structure to form a conducting layer upon which a thick film of nickel can be electroplated. The nickel "master" plate is then used to emboss directly into a plastic film, such as vinyl, that has been softened by heating or solvent. A theory describing the design and fabrication of three-dimensional phase-quantized terraced surface relief pattern that resemble stepped pyramids is described: Cowen, "Aztec surface-relief volume diffractive structure," J. Opt. Soc. Am. A, 7:1529 (1990). An example of a three-dimensional phase-quantized terraced surface relief pattern may be a pattern that resembles a stepped pyramid. Each inverted pyramid is approximately 1 micron in diameter. Preferably, each inverted pyramid can be about 0.5 to about 5 microns diameter, including for example, about 1 micron. The pyramid structures can be close-packed so that a typical microarray spot with a diameter of 150-200 microns can incorporate several hundred stepped pyramid structures. The relief volume diffraction structures have a period of about 0.1 to about 1 micron and a depth of about 0.1 to about 1 micron.

One or more specific binding substances, as described above, are immobilized on the reflective material of a SRVD biosensor. One or more specific binding substances can be arranged in microarray of distinct locations, as described above, on the reflective material.

A SRVD biosensor reflects light predominantly at a first single optical wavelength when illuminated with a broad band of optical wavelengths, and reflects light at a second single optical wavelength when one or more specific binding substances are immobilized on the reflective surface. The reflection at the second optical wavelength results from optical interference. A SRVD biosensor also reflects light at a third single optical wavelength when the one or more specific capture

agents are bound to their respective PET binding partners, due to optical interference. Readout of the reflected color can be performed serially by focusing a microscope objective onto individual microarray spots and reading the reflected spectrum with the aid of a spectrograph or imaging spectrometer, or in parallel by, for example, projecting the reflected image of the microarray onto an imaging spectrometer incorporating a high resolution color CCD camera.

A SRVD biosensor can be manufactured by, for example, producing a metal master plate, and stamping a relief volume diffractive structure into, for example, a plastic material like vinyl. After stamping, the surface is made reflective by blanket deposition of, for example, a thin metal film such as gold, silver, or aluminum. Compared to MEMS-based biosensors that rely upon photolithography, etching, and wafer bonding procedures, the manufacture of a SRVD biosensor is very inexpensive.

A SWS or SRVD biosensor embodiment can comprise an inner surface. In one preferred embodiment, such an inner surface is a bottom surface of a liquid-containing vessel. A liquid-containing vessel can be, for example, a microtiter plate well, a test tube, a petri dish, or a microfluidic channel. In one embodiment, a SWS or SRVD biosensor is incorporated into a microtiter plate. For example, a SWS biosensor or SRVD biosensor can be incorporated into the bottom surface of a microtiter plate by assembling the walls of the reaction vessels over the resonant reflection surface, so that each reaction "spot" can be exposed to a distinct test sample. Therefore, each individual microtiter plate well can act as a separate reaction vessel. Separate chemical reactions can, therefore, occur within adjacent wells without intermixing reaction fluids and chemically distinct test solutions can be applied to individual wells.

This technology is useful in applications where large numbers of biomolecular interactions are measured in parallel, particularly when molecular labels would alter or inhibit the functionality of the molecules under study. High-throughput screening of pharmaceutical compound libraries with protein targets, and microarray screening of protein-protein interactions for proteomics are examples of applications that require the sensitivity and throughput afforded by the compositions

and methods of the invention.

Unlike surface plasmon resonance, resonant mirrors, and waveguide biosensors, the described compositions and methods enable many thousands of individual binding reactions to take place simultaneously upon the biosensor surface. This technology is useful in applications where large numbers of biomolecular interactions are measured in parallel (such as in an array), particularly when molecular labels alter or inhibit the functionality of the molecules under study. These biosensors are especially suited for high-throughput screening of pharmaceutical compound libraries with protein targets, and microarray screening of protein-protein interactions for proteomics. A biosensor of the invention can be manufactured, for example, in large areas using a plastic embossing process, and thus can be inexpensively incorporated into common disposable laboratory assay platforms such as microtiter plates and microarray slides.

Other similar biosensors may also be used in the instant invention. Numerous biosensors have been developed to detect a variety of biomolecular complexes including oligonucleotides, antibody-antigen interactions, hormone-receptor interactions, and enzyme-substrate interactions. In general, these biosensors consist of two components: a highly specific recognition element and a transducer that converts the molecular recognition event into a quantifiable signal. Signal transduction has been accomplished by many methods, including fluorescence, interferometry (Jenison et al., "Interference-based detection of nucleic acid targets on optically coated silicon," Nature Biotechnology, 19, p. 62-65; Lin et al., "A porous silicon-based optical interferometric biosensor," Science, 278, p. 840-843, 1997), and gravimetry (A. Cunningham, Bioanalytical Sensors, John Wiley & Sons (1998)). Of the optically-based transduction methods, direct methods that do not require labeling of analytes with fluorescent compounds are of interest due to the relative assay simplicity and ability to study the interaction of small molecules and proteins that are not readily labeled.

These direct optical methods include surface plasmon resonance (SPR) (Jordan & Corn, "Surface Plasmon Resonance Imaging Measurements of Electrostatic Biopolymer Adsorption onto Chemically Modified Gold Surfaces,"

Anal. Chem., 69:1449-1456 (1997); plasmom-resonant particles (PRPs) (Schultz *et al., Proc. Nat. Acad. Sci.,* 97: 996-1001 (2000); grating couplers (Morhard et al., "Innnobilization of antibodies in micropattems for cell detection by optical diffraction," Sensors and Actuators B, 70, p. 232-242, 2000); ellipsometry (Jin et al., "A biosensor concept based on imaging ellipsometry for visualization of biomolecular interactions," Analytical Biochemistry, 232, p. 69-72, 1995), evanascent wave devices (Huber et al., "Direct optical immunosensing (sensitivity and selectivity)," Sensors and Actuators B, 6, p.122.126, 1992), resonance light scattering (Bao *et al., Anal. Chem.*, 74:1792-1797 (2002), and reflectometry (Brecht & Gauglitz, "Optical probes and transducers," Biosensors and Bioelectronics, 10, p. 923-936, 1995). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules. Theoretically predicted detection limits of these detection methods have been determined and experimentally confirmed to be feasible down to diagnostically relevant concentration ranges.

Surface plasmon resonance (SPR) has been successfully incorporated into an immunosensor format for the simple, rapid, and nonlabeled assay of various biochemical analytes. Proteins, complex conjugates, toxins, allergens, drugs, and pesticides can be determined directly using either natural antibodies or synthetic receptors with high sensitivity and selectivity as the sensing element. Immunosensors are capable of real-time monitoring of the antigen–antibody reaction. A wide range of molecules can be detected with lower limits ranging between 10^{-9} and 10^{-13} mol/L. Several successful commercial developments of SPR immunosensors are available and their web pages are rich in technical information. Wayne *et al.* (*Methods* **22**: 77-91, 2000) reviewed and highlighted many recent developments in SPR-based immunoassay, functionalizations of the gold surface, novel receptors in molecular recognition, and advanced techniques for sensitivity enhancement.

Utilization of the optical phenomenon surface plasmon resonance (SPR) has seen extensive growth since its initial observation by Wood in 1902 (*Phil. Mag.* **4** (1902), pp. 396–402). SPR is a simple and direct sensing technique that can be used to probe refractive index (η) changes that occur in the very close vicinity of a thin

metal film surface (Otto *Z. Phys.* **216** (1968), p. 398). The sensing mechanism exploits the properties of an evanescent field generated at the site of total internal reflection. This field penetrates into the metal film, with exponentially decreasing amplitude from the glass–metal interface. Surface plasmons, which oscillate and propagate along the upper surface of the metal film, absorb some of the plane-polarized light energy from this evanescent field to change the total internal reflection light intensity I_r. A plot of I_r versus incidence (or reflection) angle θ produces an angular intensity profile that exhibits a sharp dip. The exact location of the dip minimum (or the SPR angle θ_r) can be determined by using a polynomial algorithm to fit the I_r signals from a few diodes close to the minimum. The binding of molecules on the upper metal surface causes a change in η of the surface medium that can be observed as a shift in θ_r.

The potential of SPR for biosensor purposeswas realized in 1982–1983 by Liedberg et al., who adsorbed an immunoglobulin G (IgG) antibody overlayer on the gold sensing film, resulting in the subsequent selective binding and detection of IgG (Nylander et al., *Sens. Actuators* **3** (1982), pp. 79–84; Liedberg et al., *Sens. Actuators* **4** (1983), pp. 229–304). The principles of SPR as a biosensing technique have been reviewed previously (Daniels et al., *Sens. Actuators* **15** (1988), pp. 11–18; VanderNoot and Lai, *Spectroscopy* **6** (1991), pp. 28–33; Lundström *Biosens. Bioelectron.* **9** (1994), pp. 725–736; Liedberg et al., *Biosens. Bioelectron.* **10** (1995); Morgan et al., *Clin. Chem.* **42** (1996), pp. 193–209; Tapuchi et al., *S. Afr. J. Chem.* **49** (1996), pp. 8–25). Applications of SPR to biosensing were demonstrated for a wide range of molecules, from virus particles to sex hormone-binding globulin and syphilis. Most importantly, SPR has an inherent advantage over other types of biosensors in its versatility and capability of monitoring binding interactions without the need for fluorescence or radioisotope labeling of the biomolecules. This approach has also shown promise in the real-time determination of concentration, kinetic constant, and binding specificity of individual biomolecular interaction steps. Antibody–antigen interactions, peptide/protein–protein interactions, DNA hybridization conditions, biocompatibility studies of polymers, biomolecule–cell receptor interactions, and DNA/receptor–ligand interactions can all be analyzed (Pathak and Savelkoul, *Immunol. Today* **18** (1997), pp. 464–467). Commercially, the

use of SPR-based immunoassay has been promoted by companies such as Biacore (Uppsala, Sweden) (Jönsson et al., *Ann. Biol. Clin.* **51** (1993), pp. 19–26), Windsor Scientific (U.K.) (WWW URL for Windsor Scientific IBIS Biosensor), Quantech (Minnesota) (WWW URL for Quantech), and Texas Instruments (Dallas, TX) (WWW URL for Texas Instruments).

In yet another embodiment, a fluorescent polymer superquenching-based bioassays as disclosed in WO 02/074997 may be used for detecting binding of the unlabeled PET to its capture agents. In this embodiment, a capture agent that is specific for both a target PET peptide and a chemical moiety is used. The chemical moiety includes (a) a recognition element for the capture agent, (b) a fluorescent property-altering element, and (c) a tethering element linking the recognition element and the property-altering element. A composition comprising a fluorescent polymer and the capture agent are co-located on a support. When the chemical moiety is bound to the capture agent, the property-altering element of the chemical moiety is sufficiently close to the fluorescent polymer to alter (quench) the fluorescence emitted by the polymer. When an analyte sample is introduced, the target PET peptide, if present, binds to the capture agent, thereby displacing the chemical moiety from the receptor, resulting in de-quenching and an increase of detected fluorescence. Assays for detecting the presence of a target biological agent are also disclosed in the application.

In another related embodiment, the binding event between the capture agents and the PET can be detected by using a water-soluble luminescent quantum dot as described in US2003/0008414A1. In one embodiment, a water-soluble luminescent semiconductor quantum dot comprises a core, a cap and a hydrophilic attachment group. The "core" is a nanoparticle-sized semiconductor. While any core of the IIB-VIB, IIIB-VB or IVB-IVB semiconductors can be used in this context, the core must be such that, upon combination with a cap, a luminescent quantum dot results. A IIB-VIB semiconductor is a compound that contains at least one element from Group IEB and at least one element from Group VIB of the periodic table, and so on. Preferably, the core is a IIB-VIB, IIIB-VB or IVB-IVB semiconductor that ranges in size from about 1 nm to about 10 nm. The core is more preferably a IIB-VIB semiconductor and ranges in size from about 2 nm to about 5 nm. Most

preferably, the core is CdS or CdSe. In this regard, CdSe is especially preferred as the core, in particular at a size of about 4.2 nm.

The "cap" is a semiconductor that differs from the semiconductor of the core and binds to the core, thereby forming a surface layer on the core. The cap must be such that, upon combination with a given semiconductor core, results in a luminescent quantum dot. The cap should passivate the core by having a higher band gap than the core. In this regard, the cap is preferably a IIB-VIB semiconductor of high band gap. More preferably, the cap is ZnS or CdS. Most preferably, the cap is ZnS. In particular, the cap is preferably ZnS when the core is CdSe or CdS and the cap is preferably CdS when the core is CdSe.

The "attachment group" as that term is used herein refers to any organic group that can be attached, such as by any stable physical or chemical association, to the surface of the cap of the luminescent semiconductor quantum dot and can render the quantum dot water-soluble without rendering the quantum dot no longer luminescent. Accordingly, the attachment group comprises a hydrophilic moiety. Preferably, the attachment group enables the hydrophilic quantum dot to remain in solution for at least about one hour, one day, one week, or one month. Desirably, the attachment group is attached to the cap by covalent bonding and is attached to the cap in such a manner that the hydrophilic moiety is exposed. Preferably, the hydrophilic attachment group is attached to the quantum dot via a sulfur atom. More preferably, the hydrophilic attachment group is an organic group comprising a sulfur atom and at least one hydrophilic attachment group. Suitable hydrophilic attachment groups include, for example, a carboxylic acid or salt thereof, a sulfonic acid or salt thereof, a sulfamic acid or salt thereof, an amino substituent, a quaternary ammonium salt, and a hydroxy. The organic group of the hydrophilic attachment group of the present invention is preferably a C1-C6 alkyl group or an aryl group, more preferably a C1-C6 alkyl group, even more preferably a C1-C3 alkyl group. Therefore, in a preferred embodiment, the attachment group of the present invention is a thiol carboxylic acid or thiol alcohol. More preferably, the attachment group is a thiol carboxylic acid. Most preferably, the attachment group is mercaptoacetic acid.

Accordingly, a preferred embodiment of a water-soluble luminescent

semiconductor quantum dot is one that comprises a CdSe core of about 4.2 nm in size, a ZnS cap and an attachment group. Another preferred embodiment of a watersoluble luminescent semiconductor quantum dot is one that comprises a CdSe core, a ZnS cap and the attachment group mercaptoacetic acid. An especially preferred water-soluble luminescent semiconductor quantum dot comprises a CdSe core of about 4.2 nm, a ZnS cap of about 1 nm and a mercaptoacetic acid attachment group.

The capture agent of the instant invention can be attached to the quantum dot via the hydrophilic attachment group and forms a conjugate. The capture agent can be attached, such as by any stable physical or chemical association, to the hydrophilic attachment group of the water-soluble luminescent quantum dot directly or indirectly by any suitable means, through one or more covalent bonds, via an optional linker that does not impair the function of the capture agent or the quantum dot. For example, if the attachment group is mercaptoacetic acid and a nucleic acid biomolecule is being attached to the attachment group, the linker preferably is a primary amine, a thiol, streptavidin, neutravidin, biotin, or a like molecule. If the attachment group is mercaptoacetic acid and a protein biomolecule or a fragment thereof is being attached to the attachment group, the linker preferably is strepavidin, neutravidin, biotin, or a like molecule.

By using the quantum dot-capture agent conjugate, a PET-containing sample, when contacted with a conjugate as described above, will promote the emission of luminescence when the capture agent of the conjugate specifically binds to the PET peptide. This is particularly useful when the capture agent is a nucleic acid aptamer or an antibody. When the aptamer is used, an alternative embodiment may be employed, in which a fluorescent quencher may be positioned adjacent to the quantum dot via a self-pairing stem-loop structure when the aptamer is not bound to a PET-containing sequence. When the aptamer binds to the PET, the stem-loop structure is opened, thus releasing the quenching effect and generates luminescence.

In another related embodiment, arrays of nanosensors comprising nanowires or nanotubes as described in US2002/0117659A1 may be used for detection and/or quantitation of PET-capture agent interaction. Briefly, a "nanowire" is an elongated

nanoscale semiconductor, which can have a cross-sectional dimension of as thin as 1 nanometer. Similarly, a "nanotube" is a nanowire that has a hollowed-out core, and includes those nanotubes know to those of ordinary skill in the art. A "wire" refers to any material having a conductivity at least that of a semiconductor or metal. These nanowires / nanotubes may be used in a system constructed and arranged to determine an analyte (e.g., PET peptide) in a sample to which the nanowire(s) is exposed. The surface of the nanowire is functionalized by coating with a capture agent. Binding of an analyte to the functionalized nanowire causes a detectable change in electrical conductivity of the nanowire or optical properties. Thus, presence of the analyte can be determined by determining a change in a characteristic in the nanowire, typically an electrical characteristic or an optical characteristic. A variety of biomolecular entities can be used for coating, including, but not limited to, amino acids, proteins, sugars, DNA, antibodies, antigens, and enzymes, etc. For more details such as construction of nanowires, functionalization with various biomolecules (such as the capture agents of the instant invention), and detection in nanowire devices, see US2002/0117659A1 (incorporated by reference). Since multiple nanowires can be used in parelle, each with a different capture agent as the functionalized group, this technology is ideally suited for large scale arrayed detection of PET-containing peptides in biological samples without the need to label the PET peptides. This nanowire detection technology has been successfully used to detect pH change (H^+ binding), biotin-streptavidin binding, antibody-antigen binding, metal (Ca^{2+}) binding with picomolar sensitivity and in real time (Cui *et al.*, *Science* **293**: 1289-1292).

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), uses a laser pulse to desorb proteins from the surface followed by mass spectrometry to identify the molecular weights of the proteins (Gilligan *et al.*, Mass spectrometry after capture and small-volume elution of analyte from a surface plasmon resonance biosensor. *Anal. Chem.* **74** (2002), pp. 2041–2047). Because this method only measures the mass of proteins at the interface, and because the desorption protocol is sufficiently mild that it does not result in fragmentation, MALDI can provide straightforward useful information such as confirming the identity of the bound PET peptide, or any enzymatic modification of

a PET peptide. For this matter, MALDI can be used to identify proteins that are bound to immobilized capture agents. An important technique for identifying bound proteins relies on treating the array (and the proteins that are selectively bound to the array) with proteases and then analyzing the resulting peptides to obtain sequence data.

IV. Samples and Their Preparation

The capture agents or an array of capture agents typically are contacted with a sample, *e.g.*, a biological fluid, a water sample, or a food sample, which has been fragmented to generate a collection of peptides, under conditions suitable for binding a PET corresponding to a protein of interest.

Samples to be assayed using the capture agents of the present invention may be drawn from various physiological, environmental or artificial sources. In particular, physiological samples such as body fluids or tissue samples of a patient or an organism may be used as assay samples. Such fluids include, but are not limited to, saliva, mucous, sweat, whole blood, serum, urine, amniotic fluid, genital fluids, fecal material, marrow, plasma, spinal fluid, pericardial fluids, gastric fluids, abdominal fluids, peritoneal fluids, pleural fluids and extraction from other body parts, and secretion from other glands. Alternatively, biological samples drawn from cells taken from the patient or grown in culture may be employed. Such samples include supernatants, whole cell lysates, or cell fractions obtained by lysis and fractionation of cellular material. Extracts of cells and fractions thereof, including those directly from a biological entity and those grown in an artificial environment, can also be used. In addition, a biological sample can be obtained and/or deribed from, for example, blood, plasma, serum, gastrointestinal secretions, homogenates of tissues or tumors, synovial fluid, feces, saliva, sputum, cyst fluid, amniotic fluid, cerebrospinal fluid, peritoneal fluid, lung lavage fluid, semen, lymphatic fluid, tears, or prostatitc fluid.

A general scheme of sample preparation prior to its use in the methods of the instant invention is described in Figure 6 (slide 45 of D2). Briefly, a sample can be pretreated by extraction and/or dilution to minimize the interference from certain

substances present in the sample. The sample can then be either chemically reduced, denatured, alkylated, or subjected to thermo-denaturation. Regardless of the denaturation step, the denatured sample is then digested by a protease, such as trypsin, before it is used in subsequent assays. A desalting step may also be added just after protease digestion if chemical denaturation if used. This process is generally simple, robust and reproducible, and is generally applicable to main sample types including serum, cell lysates and tissues.

The sample may be pre treated to remove extraneous materials, stabilized, buffered, preserved, filtered, or otherwise conditioned as desired or necessary. Proteins in the sample typically are fragmented, either as part of the methods of the invention or in advance of performing these methods. Fragmentation can be performed using any art-recognized desired method, such as by using chemical cleavage (*e.g.*, cyanogen bromide); enzymatic means (*e.g.*, using a protease such as trypsin, chymotrypsin, pepsin, papain, carboxypeptidase, calpain, subtilisin, gluc-C, endo lys-C and proteinase K, or a collection or sub-collection thereof); or physical means (*e.g.*, fragmentation by physical shearing or fragmentation by sonication). As used herein, the terms "fragmentation" "cleavage," "proteolytic cleavage," "proteolysis" "restriction" and the like are used interchangeably and refer to scission of a chemical bond, typically a peptide bond, within proteins to produce a collection of peptides (*i.e.*, protein fragments).

The purpose of the fragmentation is to generate peptides comprising PET which are soluble and available for binding with a capture agent. In essence, the sample preparation is designed to assure to the extent possible that all PET present on or within relevant proteins that may be present in the sample are available for reaction with the capture agents. This strategy can avoid many of the problems encountered with previous attempts to design protein chips caused by protein-protein complexation, post translational modifications and the like.

In one embodiment, the sample of interest is treated using a pre-determined protocol which: (A) inhibits masking of the target protein caused by target protein-protein non covalent or covalent complexation or aggregation, target protein degradation or denaturing, target protein post-translational modification, or

environmentally induced alteration in target protein tertiary structure, and (B) fragments the target protein to, thereby, produce at least one peptide epitope (*i.e.*, a PET) whose concentration is directly proportional to the true concentration of the target protein in the sample. The sample treatment protocol is designed and empirically tested to result reproducibly in the generation of a PET that is available for reaction with a given capture agent. The treatment can involve protein separations; protein fractionations; solvent modifications such as polarity changes, osmolarity changes, dilutions, or pH changes; heating; freezing; precipitating; extractions; reactions with a reagent such as an endo-, exo- or site specific protease; non proteolytic digestion; oxidations; reductions; neutralization of some biological activity, and other steps known to one of skill in the art.

For example, the sample may be treated with an alkylating agent and a reducing agent in order to prevent the formation of dimers or other aggregates through disulfide/dithiol exchange. The sample of PET-containing peptides may also be treated to remove secondary modifications, including but are not limited to, phosphorylation, methylation, glycosylation, acetylation, prenylation, using, for example, respective modification-specific enzymes such as phosphatases, etc.

In one embodiment, proteins of a sample will be denatured, reduced and/or alkylated, but will not be proteolytically cleaved. Proteins can be denatured by thermal denaturation or organic solvents, then subjected to direct detection or optionally, further proteolytic cleavage.

The use of thermal denaturation (50-90 °C for about 20 minutes) of proteins prior to enzyme digestion in solution is preferred over chemical denaturation (such as 6-8 M guanidine HCl or urea) because it does not require purification / concentration, which might be preferred or required prior to subsequent analysis. Park and Russell reported that enzymatic digestions of proteins that are resistant to proteolysis are significantly enhanced by thermal denaturation (*Anal. Chem.*, 72 (11): 2667 -2670, 2000). Native proteins that are sensitive to proteolysis show similar or just slightly lower digestion yields following thermal denaturation. Proteins that are resistant to digestion become more susceptible to digestion, independent of protein size, following thermal denaturation. For example, amino

acid sequence coverage from digest fragments increases from 15 to 86% in myoglobin and from 0 to 43% in ovalbumin. This leads to more rapid and reliable protein identification by the instant invention, especially to protease resistant proteins.

Although some proteins aggregate upon thermal denaturation, the protein aggregates are easily digested by trypsin and generate sufficient numbers of digest fragments for protein identification. In fact, protein aggregation may be the reason thermal denaturation facilitates digestion in most cases. Protein aggregates are believed to be the oligomerization products of the denatured form of protein (Copeland, R. A. *Methods for Protein Analysis*; Chapman & Hall: New York, NY, 1994). In general, hydrophobic parts of the protein are located inside and relatively less hydrophobic parts of the protein are exposed to the aqueous environment. During the thermal denaturation, intact proteins are gradually unfolded into a denatured conformation and sufficient energy is provided to prevent a fold back to its native conformation. The probability for interactions with other denatured proteins is increased, thus allowing hydrophobic interactions between exposed hydrophobic parts of the proteins. In addition, protein aggregates of the denatured protein can have a more protease-labile structure than nondenatured proteins because more cleavage sites are exposed to the environment. Protein aggregates are easily digested, so that protein aggregates are not observed at the end of 3 h of trypsin digestion (Park and Russell, *Anal. Chem.*, 72 (11): 2667 -2670, 2000). Moreover, trypsin digestion of protein aggregates generates more specific cleavage products.

Ordinary proteases such as trypsin may be used after denaturation. The process may be repeated by one or more rounds after the first round of denaturation and digestion. Alternatively, this thermal denaturation process can be further assisted by using thermophilic trypsin-like enzymes, so that denaturation and digestion can be done simultaneously. For example, Nongporn Towatana et al. (*J of Bioscience and Bioengineering* 87(5): 581-587, 1999) reported the purification to apparent homogeneity of an alkaline protease from culture supernatants of *Bacillus sp.* PS719, a novel alkaliphilic, thermophilic bacterium isolated from a thermal spring soil sample. The protease exhibited maximum activity towards azocasein at

pH 9.0 and at 75°C. The enzyme was stable in the pH range 8.0 to 10.0 and up to 80°C in the absence of Ca^{2+}. This enzyme appears to be a trypsin-like serine protease, since phenylmethylsulfonyl fluoride (PMSF) and 3,4-dichloroisocoumarin (DCI) in addition to N-α-p-tosyl-L-lysine chloromethyl ketone (TLCK) completely inhibited the activity. Among the various oligopeptidyl-p-nitroanilides tested, the protease showed a preference for cleavage at arginine residues on the carboxylic side of the scissile bond of the substrate, liberating p-nitroaniline from N-carbobenzoxy (CBZ)-L-arginine-p-nitroanilide with the K_m and V_{max} values of 0.6 mM and 1.0 μmol min^{-1}mg $protein^{-1}$, respectively.

Alternatively, existing proteases may be chemically modified to achieve enhanced thermostability for use in this type of application. Mozhaev et al. (*Eur J Biochem.* 173(1):147-54, 1988) experimentally verified the idea presented earlier that the contact of nonpolar clusters located on the surface of protein molecules with water destabilizes proteins. It was demonstrated that protein stabilization could be achieved by artificial hydrophilization of the surface area of protein globules by chemical modification. Two experimental systems were studied for the verification of the hydrophilization approach. In one experiment, the surface tyrosine residues of trypsin were transformed to aminotyrosines using a two-step modification procedure: nitration by tetranitromethane followed by reduction with sodium dithionite. The modified enzyme was much more stable against irreversible thermo-inactivation: the stabilizing effect increased with the number of aminotyrosine residues in trypsin and the modified enzyme could become even 100 times more stable than the native one. In another experiment, alpha-chymotrypsin was covalently modified by treatment with anhydrides or chloroanhydrides of aromatic carboxylic acids. As a result, different numbers of additional carboxylic groups (up to five depending on the structure of the modifying reagent) were introduced into each Lys residue modified. Acylation of all available amino groups of alpha-chymotrypsin by cyclic anhydrides of pyromellitic and mellitic acids resulted in a substantial hydrophilization of the protein as estimated by partitioning in an aqueous Ficoll-400/Dextran-70 biphasic system. These modified enzyme preparations were extremely stable against irreversible thermal inactivation at elevated temperatures (65-98°C); their thermostability was practically equal to the stability of proteolytic

enzymes from extremely thermophilic bacteria, the most stable proteinases known to date. Similar approaches may be used to any other chosen proteases for the subject method.

In other embodiments, samples can be pre-treated with reducing agents such as b-mercaptoethanol or DTT to reduce the disulfide bonds to facilitate digestion.

Fractionation may be performed using any single or multidimentional chromatography, such as reversed phase chromatography (RPC), ion exchange chromatography, hydrophobic interaction chromatography, size exclusion chromatography, or affinity fractionation such as immunoaffinity and immobilized metal affinity chromatography. Preferably, the fractionation involves surface-mediated selection strategies. Electrophoresis, either slab gel or capillary electrophoresis, can also be used to fractionate the peptides in the sample. Examples of slab gel electrophoretic methods include sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and native gel electrophoresis. Capillary electrophoresis methods that can be used for fractionation include capillary gel electrophoresis (CGE), capillary zone electrophoresis (CZE) and capillary electrochromatography (CEC), capillary isoelectric focusing, immobilized metal affinity chromatography and affinity electrophoresis.

Protein precipitation may be performed using techniques well known in the art. For example, precipitation may be achieved using known precipitants, such as potassium thiocyanate, trichloroacetic acid and ammonium sulphate.

Subsequent to fragmentation, the sample may be contacted with the capture agents of the present invention, *e.g.*, capture agents immobilized on a planar support or on a bead, as described herein. Alternatively, the fragmented sample (containing a collection of peptides) may be fractionated based on, for example, size, post-translational modifications (*e.g.*, glycosylation or phosphorylation) or antigenic properties, and then contacted with the capture agents of the present invention, *e.g.*, capture agents immobilized on a planar support or on a bead.

Figure 7 provides an illustrative example of serum sample pre-treatment using either the thermo-denaturation or the chemical denaturation. Briefly, for thermo-denaturation, 100 μL of human serum (about 75 mg/mL total protein) is first

diluted 10-fold to about 7.5 mg/mL. The diluted sample is then heated to 90°C for 5 minutes to denature the proteins, followed by 30 minutes of trypsin digestion at 55°C. The trypsin is inactivated at 80°C after the digestion.

For chemical denaturation, about 1.8 mL of human serum proteins diluted to about 4 mg/mL is denatured in a final concentration of 50mM HEPES buffer (pH 8.0), 8M urea and 10mM DTT. Iodoacetamide is then added to 25mM final concentration. The denatured sample is then further diluted to about 1 mg/mL for protease digestion. The digested sample will pass through a desalting column before being used in subsequent assays.

Figure 8 shows the result of thermo-denaturation and chemical denaturation of serum proteins, cell lysates (MOLT4 and Hela cells). It is evident that denaturation was successful for the majority, if not all of the proteins in both the thermo- and chemical-denaturation lanes, and both methods achieved comparable results in terms of protein denaturation and fragmentation.

The above example is for illustrative purpose only and is by no means limiting. Minor alterations of the protocol depending on specific uses can be easily achieved for optimal results in individual assays.

V. Selection of PET

One advantages of the PET of the instant invention is that PET can be determined in sillico and generated in vitro (such as by peptide synthesis) without cloning or purifying the protein it belongs. PET is also advantageous over the full-length tryptic fragments (or for that matter, any other fragments that predictably results from any other treatments) since full-length tryptic fragments tend to contain one or more PETs themselves, though the tryptic fragment itself may be unique simply because of its length (the longer a stretch of peptide, the more likely it will be unique). A direct implication is that, by using relatively short and unique PETs rather than the full-length (tryptic) peptide fragments, the method of the instant invention has greatly reduced, if not completely eliminated, the risk of having

multiple antibodies with unique specificities against the same peptide fragment – a source of antibody cross-reactivity. An additional advantage may be added due to the PET selection process, such as the nearest-neighbor analysis and ranking prioritization(see below), which further eliminates the chance of cross-reactivity. All these features make the PET-based methods particularly suitable for genome-wise analysis using multiplexing techniques.

The PET of the instant invention can be selected in various ways. In the simplest embodiment, the PET for a given organism or biological sample can be generated or identified by a brute force search of the relevant database, using all theoretically possible PET with a given length. This process is preferably carried out computationaly using, for example, any of the sequence search tools available in the art or variations thereof. For example, to identify PET of 5 amino acids in length (a total of 3.2 million possible PET candidates, see table 2.2.2 below), each of the 3.2 million candidates may be used as a query sequence to search against the human proteom as described below. Any candidate that has more than one hit (found in two or more proteins) is immediately eliminated before further searching is done. At the end of the search, a list of human proteins that have one or more PETs can be obtained (see Example 1 below). The same or similar procedure can be used for any pre-determined organism or database.

For example, PETs for each human protein can be identified using the following procedure. A Perl program is developed to calculate the occurrence of all possible peptides, given by 20^N, of defined length N (amino acids) in human proteins. For example, the total tag space is 160,000 (20^4) for tetramer peptides, 3.2 M (20^5) for pentamer peptides, and 64 M (20^6) for hexamer peptides, so on. Predicted human protein sequences are analyzed for the presence or absence of all possible peptides of N amino acids. PET are the peptide sequences that occur only once in the human proteome. Thus the presence of a specific PET is an intrinsic property of the protein sequence and is operational independent. According to this approach, a definitive set of PETs can be defined and used regardless of the sample processing procedure (operational independence).

In one embodiment, to speed up the searching process, computer algorithms

may be developed or modified to eliminate unnecessary searches before the actual search begins.

Using the example above, two highly related (say differ only in a few amino acid positions) human proteins may be aligned, and a large number of candidate PET can be eliminated based on the sequence of the identical regions. For example, if there is a stretch of identical sequence of 20 amino acids, then sixteen 5-amino acid PETs can be eliminated without searching, by virtue of their simultaneous appearance in two non-identical human proteins. This elimination process can be continued using as many highly related protein pairs or families as possible, such as the evolutionary conserved proteins such as histones, globins, etc.

In another embodiment, the identified PET for a given protein may be rank-ordered based on certain criteria, so that higher ranking PETs are preferred to be used in generating specific capture agents.

For example, certain PET may naturally exist on protein surface, thus making good candidates for being a soluble peptide when digested by a protease. On the other hand, certain PET may exist in an internal or core region of a protein, and may not be readily soluble even after digestion. Such solubility property may be evaluated by available softwares. The solvent accessibility method described in Boger, J., Emini, E.A. & Schmidt, A., Surface probability profile-An heuristic approach to the selection of synthetic peptide antigens, Reports on the Sixth International Congress in Immunology (Toronto) 1986 p.250 also may be used to identify PETs that are located on the surface of the protein of interest. The package MOLMOL (Koradi, R. *et al.* (1996) *J. Mol. Graph.* 14:51–55) and Eisenhaber's ASC method (Eisenhaber and Argos (1993) *J. Comput. Chem.* 14:1272–1280; Eisenhaber *et al.* (1995) *J. Comput. Chem.* 16:273–284) may also be used. Surface PETs generally have higher ranking than internal PETs. In one embodiment, the logP or logD values that can be calculated for a PET, or proteolytic fragment containing a PET, can be calculated and used to rank order the PET's based on likely solubility under conditions that a protein sample is to be contacted with a capture agent.

Regardless of the manner the PETs are generated, an ideal PET preferably is

8 amino acids in length, and the parental tryptic peptide should be smaller than 20 amino acid long. This is because antibodies typically recognize peptide epitopes of 4 – 8 amino acids, thus peptides of 12-20 amino acids are conventionally used for antibody production.

Since trypsin is a preferred digestion enzyme in certain embodiments, a PET in these embodiments should not contain K or R in the middle of the sequence so that the PET will not be cleaved by trypsin during sample preparation. In a more general sense, the selected PET should not contain or overlap a digestion site such that the PET is expected to be destroyed after digestion, unless an assay specifically prefer that a PET be destroyed after digestion.

In addition, an ideal PET preferably does not have hydrophobic parental tryptic peptide, is highly antigenic, and has the smallest numbers (preferably none) of closest related peptides (nearest neighbor peptides or NNP) defined by nearest neighbor analysis.

Any PET may also be associated with an annotation, which may contain useful information such as: whether the PET may be destroyed by a certain protease (such as trypsin), whether it is likely to appear on a digested peptide with a relatively rigid or flexible structure, etc. These characteristics may help to rank order the PETs for use if generating specific capture agents, especially when there are a large number of PETs associated with a given protein. Since PET may change depending on particular use in a given organism, ranking order may change depending on specific usages. A PET may be low ranking due to its probability of being destroyed by a certain protease may rank higher in a different fragmentation scheme using a different protease.

In another embodiment, the computational algorithm for selecting optimal PET from a protein for antibody generation takes antibody-peptide interaction data into consideration. A process such as Nearest-Neighbor Analysis (NNA), can be used to select most unique PET for each protein. Each PET in a protein is given a relative score, or PET Uniqueness Index, that is based on the number of nearest neighbors it has. The higher the PET Uniqueness Index, the more unique the PET is. The PET Uniqueness Index can be calculated using an Amino Acid Replacement

Matrix such as the one in Table VIII of Getzoff, ED, Tainer JA and Lerner RA. *The chemistry and meachnism of antibody binding to protein antigens*. 1988. *Advances. Immunol.* **43**: 1-97. In this matrix, the replaceability of each amino acid by the remaining 19 amino acids was calculated based on experimental data on antibody cross-reactivity to a large number of peptides of single mutations (replacing each amino acid in a peptide sequence by the remaining 19 amino acids). For example, each octamer PET from a protein is compared to 8.7 million octamers present in human proteome and a PET Uniqueness Index is calculated. This process not only selects the most unique PET for particular protein, it also identifies Nearest Neighbor Peptides for this PET. This becomes important for defining cross-reactivity of PET-specific antibodies since Nearest Neighbor Peptides are the ones most likely will cross-react with particular antibody.

Besides PET Uniqueness Index, the following parameters for each PET may also be calculated and help to rank the PETs:

a) PET Solubility Index: which involves calculating LogP and LogD of the PET.

b) PET Hydrophobicity & water accessibility: only hydrophilic peptides and peptides with good water accessibility will be selected.

c) PET Length: since longer peptides tend to have conformations in solution, we use PET peptides with defined length of 8 amino acids. PET-specific antibodies will have better defined specificity due to limited number of epitopes in a shorter peptide sequences. This is very important for multiplexing assays using these antibodies. In one embodiment, only antibodies generated by this way will be used for multiplexing assays.

d) Evolutionary Conservation Index: each human PET will be compared with other species to see whether a PET sequence is conserved cross species. Ideally, PET with minimal conservation, for example, between mouse and human sequences will be selected. This will maximize the possibility to generate good immunoresponse and monoclonal antibodies in mouse.

VI. Applications of the Invention

A. Investigative and Diagnostic Applications

The capture agents of the present invention provide a powerful tool in probing living systems and in diagnostic applications (*e.g.*, clinical, environmental and industrial, and food safety diagnostic applications). For clinical diagnostic applications, the capture agents are designed such that they bind to one or more PET corresponding to one or more diagnostic targets (*e.g.*, a disease related protein, collection of proteins, or pattern of proteins). Specific individual disease related proteins include, for example, prostate-specific antigen (PSA), prostatic acid phosphatase (PAP) or prostate specific membrane antigen (PSMA) (for diagnosing prostate cancer); Cyclin E for diagnosing breast cancer; Annexin, *e.g.*, Annexin V (for diagnosing cell death in, for example, cancer, ischemia, or transplant rejection); or β-amyloid plaques (for diagnosing Alzheimer's Disease).

Thus, PETs and the capture agents of the present invention may be used as a source of surrogate markers. For example, they can be used as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of protein expression.

As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the causation of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using a PET corresponding to a protein associated with a cardiovascular disease as a surrogate marker, and an analysis of HIV infection may be made using a PET corresponding to an HIV protein as a surrogate marker, well in advance of the

undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35:258-264; and James (1994) *AIDS Treatment News Archive* 209.

Perhaps the most significant use of the invention is that it enables practice of a powerful new protein expression analysis technique: analyses of samples for the presence of specific *combinations of proteins* and specific *levels of expression of combinations of proteins.* This is valuable in molecular biology investigations generally, and particularly in development of novel assays. Thus, this invention permits one to identify proteins, groups of proteins, and protein expression patterns present in a sample which are characteristic of some disease, physiologic state, or species identity. Such multiparametric assay protocols may be particularly informative if the proteins being detected are from disconnected or remotely connected pathways. For example, the invention might be used to compare protein expression patterns in tissue, urine, or blood from normal patients and cancer patients, and to discover that in the presence of a particular type of cancer a first group of proteins are expressed at a higher level than normal and another group are expressed at a lower level. As another example, the protein chips might be used to survey protein expression levels in various strains of bacteria, to discover patterns of expression which characterize different strains, and to determine which strains are susceptible to which antibiotic. Furthermore, the invention enables production of specialty assay devices comprising arrays or other arrangements of capture agents for detecting specific patterns of specific proteins. Thus, to continue the example, in accordance with the practice of the invention, one can produce a chip which can be exposed to a cell lysate preparation from a patient or a body fluid to reveal the presence or absence or pattern of expression informative that the patient is cancer free, or is suffering from a particular cancer type. Alternatively, one might produce a protein chip that would be exposed to a sample and read to indicate the species of bacteria in an infection and the antibiotic that will destroy it.

A junction PET is a peptide which spans the region of a protein corresponding to a splice site of the RNA which encodes it. Capture agents designed to bind to a junction PET may be included in such analyses to detect splice variants as well as gene fusions generated by chromosomal rearrangements, *e.g.*, cancer-

associated chromosomal rearrangements. Detection of such rearrangements may lead to a diagnosis of a disease, *e.g.*, cancer. It is now becoming apparent that splice variants are common and that mechanisms for controlling RNA splicing have evolved as a control mechanism for various physiological processes. The invention permits detection of expression of proteins encoded by such species, and correlation of the presence of such proteins with disease or abnormality. Examples of cancer-associated chromosomal rearrangements include: translocation t(16;21)(p11;q22) between genes FUS-ERG associated with myeloid leukemia and non-lymphocytic, acute leukemia (see Ichikawa H. *et al.* (1994) *Cancer Res.* 54(11):2865-8); translocation t(21;22)(q22;q12) between genes ERG-EWS associated with Ewing's sarcoma and neuroepithelioma (see Kaneko Y. *et al.* (1997) *Genes Chromosomes Cancer* 18(3):228-31); translocation t(14;18)(q32;q21) involving the bcl2 gene and associated with follicular lymphoma; and translocations juxtaposing the coding regions of the PAX3 gene on chromosome 2 and the FKHR gene on chromosome 13 associated with alveolar rhabdomyosarcoma (see Barr F.G. *et al.* (1996) *Hum. Mol. Genet.* 5:15-21).

For applications in environmental and industrial diagnostics the capture agents are designed such that they bind to one or more PET corresponding to a biowarfare agent (*e.g.*, anthrax, small pox, cholera toxin) and/or one or more PET corresponding to other environmental toxins (Staphylococcus aureus a-toxin, Shiga toxin, cytotoxic necrotizing factor type 1, Escherichia coli heat- stable toxin, and botulinum and tetanus neurotoxins) or allergens. The capture agents may also be designed to bind to one or more PET corresponding to an infectious agent such as a bacterium, a prion, a parasite, or a PET corresponding to a virus (*e.g.*, human immunodeficiency virus–1 (HIV-1), HIV-2, simian immunodeficiency virus (SIV), hepatitis C virus (HCV), hepatitis B virus (HBV), Influenza, Foot and Mouth Disease virus, and Ebola virus).

The following part illustrates the general idea of diagnostic use of the instant invention in one specific setting – serum biomarker assays.

The proteins found in human plasma perform many important functions in the body. Over or under expression of these proteins can thus cause disease directly,

or reveal its presence. Studies have shown that complex serum proteomic patterns might reflect the underlying pathological state of an organ such as the ovary (Petricoin et al., Lancet 359: 572-577, 2002). Therefore, the easy accessibility of serum samples, and the fact that serum comprehensively samples the human phenotype - the state of the body at a particular point in time – make serum an attractive option for a broad array of applications, including clinical and diagnostics applications (early detection and diagnosis of disease, monitor disease progression, monitor therapy etc.), discovery applications (such as novel biomarker discovery), and drug development (drug efficacy and toxicity, and personalized medicine). In fact, over $1 billion annually is spent on immunoassays to measure proteins in plasma as indicators of disease (Plasma Proteome Institute (PPI), Washington, D.C.).

Despite decades of research, only a handful of proteins (about 20) among the 500 or so detected proteins in plasma are measured routinely for diagnostic purposes. These include: cardiac proteins (troponins, myoglobin, creatine kinase) as indicators of heart attack; insulin, for management of diabetes; liver enzymes (alanine or aspartate transaminases) as indicators of drug toxicity; and coagulation factors for management of clotting disorders. About 150 proteins in plasma are measured by some laboratory for diagnosis of less common diseases.

IN addition, proteins in plasma differ in concentration by at least one billion-fold. For example, serum albumin has a normal concentration range of 35-50 mg/mL (35-50 x 10^9 pg/mL) and is measured clinically as an indication of severe liver disease or malnutrition, while interleukin 6 (IL-6) has a normal range of just 0-5 pg/mL, and is measured as a sensitive indicator of inflammation or infection.

Thus, there is a need for reference levels of all serum proteins, and reliable assays for measuring serum protein levels under any conditions. However, standardization of immunoassays for heterogeneous antigens is nearly impossible about 10 years ago (Ekins, Scand J Clin Lab Invest. 205: 33–46, 1991). One of the major obstacle is the apparent need of having identical standard and analyte. This is the case with only a few small peptides. With larger peptides and proteins, the problems tend to become more complicated because biological samples often

contain proforms, splice variants, fragments, and complexes of the analyte (Stenman, Clinical Chemistry 47: 815-820, 2001). One such problem is illustrated by measuring serum TGF-beta levels.

The TGF-beta superfamily proteins are a collection of structurally related multi-function proteins that have a diverse array of biological functions including wound healing, development, oncogenesis, and atherosclerosis. There are at least three known mammalian TGF-beta proteins (beta1, beta2 and beta3), which are thought to have similar functions, at least in vitro. Each of the three isoforms are produced as pre-pro-proteins, which rapidly dimerizes. After the loss of the signal sequences, sugar moieties are added to the proproteins regions known as the Latency Associated Peptide, or LAP. In addition, there is proteolytic cleavage between the LAPs and the mature dimers (the functional portion), but the cleaved LAPs still associate with the mature dimer, forming a complex known as the small latent complex. Either prior to secretion, or in the extracellular milieu, the small latent complex can bind to a large number of other proteins forming a large number of higher molecular weight latent complexes. The best characterized of these proteins are the latent TGF-beta binding protein family LTBP1-4 and fibrillin-1 and -2 (see Figure 9). Once in the extracellular environment, the TGF-beta complex may bind even more proteins to form other complexes. Known soluble TGF-beta binding proteins include: decorin, alpha-fetoprotein (AFP), betaglycan extracellular domain, ß-amyloid precursor, and fetuin. Given the various isoforms, complexes, processing stages, etc., it is very difficult to accurately measure serum TGF-beta protein levels, and a range of 100-fold differences in serum level of TBG-beta1 are reported by different groups (see Grainger et al., *Cytokine & Growth Factor Reviews* 11: 133-145, 2000).

The other problem arises from the false positive / negative effects of anti-animal antibodies on immunoassays. Specifically, in a sandwich-type assay for a specific antigen in a serum sample, instead of capturing the desired antigen, the immobilized capture antibody may bind to anti-animal antibodies in the serum sample, which in turn can be bound by the labeled secondary antibody and gives rise to false positive result. On the other hand, too much anti-animal antibodies may block the interaction between the capture antibody and the desired antigen, and the

interaction between the labeled secondary antibody and the desired antigen, leading to false negative result. This is a serious problem demonstrated in a recent study by Rotmensch and Cole (Lancet 355: 712-715, 2000), which shows that in all 12 cases where women were diagnosed of having postgestational choriocarcinoma on the basis of persistently positive human chorionic gonadotropin (hCG) test results in the absence of pregnancy, a false diagnosis had been made, and most of the women had been subjected to needless surgery or chemotherapy. Such diagnostic problems associated with anti-animal antibodies have also been reported elsewhere (Hennig et al., The influence of naturally occurring heterophilic anti-immunoglobulin antibodies on direct measurement of serum proteins using sandwich ELISAs. *Journal of Immunological Methods* 235: 71–80, 2000; Covinsky et al., *An IgMl Antibody to Escherichia coli Produces False-Positive Results in Multiple Immunometric Assays. Clinical Chemistry* 46: 1157-1161, 2000).

All these problems can be efficiently solved by the methods of the instant invention. By digesting serum samples and converting all forms of the target protein to a uniform PET-containing peptide, the methods of the instant invention greatly reduce the complexity of the sample. Anti-animal antibodies, proteins complexes, various isoforms are no longer expected to be a significant factor in the digested serum sample, thus facilitating more reliable, reproducible, and accurate results from assay to assay.

The method of the instant invention is by no means limited to one particular serum protein such as TGF-beta. It has broad applications in a wide range of serum proteins, including peptide hormones, candidate disease biomarkers (such as PSA, CA125, MMPs, etc.), serum disease and non-disease biomarkers, and acute phase response proteins. For example, measuring the following types of serum biomarkers will have broad applications in clinical and diagnostic uses: 1) disease state markers (such as markers for inflammation, infection, etc.), and 2) non-disease state markers, including markers indicating drug and hormone effects (e.g., alcohol, androgens, anti-epileptics, estrogen, pregnancy, hormone replacement therapy, etc.). Exemplary serum proteins that can be measured include: ApoA-I, Andogens, AAT, AAG, A2M, Alb, Apo-B, AT III, C3, Cp, C4, CRP, SAA, Hp, AGP, Fb, AP, FIB, FER, PAL, PSM, Tf, IgA, IgG, IgM, IgE, FN, B2M, and RBP.

One preferred assay method for these serum proteins is the sandwich assay using a PET-specific capture agent and at least one labeled secondary capture agent(s) for detection of binding. These assays may be performed in an array format according to the teaching of the instant application, in that different capture agents (such as PET-specific antibodies) can be arrayed on a single (or a few) microarrays for use in simultaneous detection / quantitation of a large number of serum biomarkers.

Foundation for Blood Research (FBR, Scarborough, ME) has developed a 152-page guide on serum protein utility and interpretation for day to day use by practitioners and laboratorians. This guide contains a distillation of the world's literature on the subject, is fully indexed, and is presented by a given disease state (Section I), as well as by individual proteins (Section II). This book is generally useful for interpretation of test results, as well as providing guidance regarding which test is (or is not) appropriate to order and why (or why not). Section II, which covers general information on serum proteins, is also helpful regarding background information about each protein. The entire content of which is incorporated herein by reference.

B. High-Throughput Screening

Compositions containing the capture agents of the invention, *e.g.*, microarrays, beads or chips enable the high-throughput screening of very large numbers of compounds to identify those compounds capable of interacting with a particular capture agent, or to detect molecules which compete for binding with the PETs. Microarrays are useful for screening large libraries of natural or synthetic compounds to identify competitors of natural or non-natural ligands for the capture agent, which may be of diagnostic, prognostic, therapeutic or scientific interest.

The use of microarray technology with the capture agents of the present invention enables comprehensive profiling of large numbers of proteins from normal and diseased-state serum, cells, and tissues.

For example, once the microarray has been formed, it may be used for high-

throughput drug discovery (*e.g.*, screening libraries of compounds for their ability to bind to or modulate the activity of a target protein); for high-throughput target identification (*e.g.*, correlating a protein with a disease process); for high-throughput target validation (*e.g.*, manipulating a protein by, for example, mutagenesis and monitoring the effects of the manipulation on the protein or on other proteins); or in basic research (*e.g.*, to study patterns of protein expression at, for example, key developmental or cell cycle time points or to study patterns of protein expression in response to various stimuli).

In one embodiment, the invention provides a method for identifying a test compound, *e.g.*, a small molecule, that modulates the activity of a ligand of interest. According to this embodiment, a capture agent is exposed to a ligand and a test compound. The presence or the absence of binding between the capture agent and the ligand is then detected to determine the modulatory effect of the test compound on the ligand. In a preferred embodiment, a microarray of capture agents, that bind to ligands acting in the same cellular pathway, are used to profile the regulatory effect of a test compound on all these proteins in a parallel fashion.

C. Pharmacoproteomics

The capture agents or arrays comprising the capture agents of the present invention may also be used to study the relationship between a subject's protein expression profile and that subject's response to a foreign compound or drug. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, use of the capture agents in the foregoing manner may aid a physician or clinician in determining whether to administer a pharmacologically active drug to a subject, as well as in tailoring the dosage and/or therapeutic regimen of treatment with the drug.

D. Protein Profiling

As indicated above, capture agents of the present invention enable the characterization of any biological state via protein profiling. The term "protein profile," as used herein, includes the pattern of protein expression obtained for a given tissue or cell under a given set of conditions. Such conditions may include, but are not limited to, cellular growth, apoptosis, proliferation, differentiation, transformation, tumorigenesis, metastasis, and carcinogen exposure.

The capture agents of the present invention may also be used to compare the protein expression patterns of two cells or different populations of cells. Methods of comparing the protein expression of two cells or populations of cells are particularly useful for the understanding of biological processes. For example, using these methods, the protein expression patterns of identical cells or closely related cells exposed to different conditions can be compared. Most typically, the protein content of one cell or population of cells is compared to the protein content of a control cell or population of cells. As indicated above, one of the cells or populations of cells may be neoplastic and the other cell is not. In another embodiment, one of the two cells or populations of cells being assayed may be infected with a pathogen. Alternatively, one of the two cells or populations of cells has been exposed to a chemical, environmental, or thermal stress and the other cell or population of cells serves as a control. In a further embodiment, one of the cells or populations of cells may be exposed to a drug or a potential drug and its protein expression pattern compared to a control cell.

Such methods of assaying differential protein expression are useful in the identification and validation of new potential drug targets as well as for drug screening. For instance, the capture agents and the methods of the invention may be used to identify a protein which is overexpressed in tumor cells, but not in normal cells. This protein may be a target for drug intervention. Inhibitors to the action of the overexpressed protein can then be developed. Alternatively, antisense strategies to inhibit the overexpression may be developed. In another instance, the protein expression pattern of a cell, or population of cells, which has been exposed to a drug or potential drug can be compared to that of a cell, or population of cells, which has not been exposed to the drug. This comparison will provide insight as to whether the drug has had the desired effect on a target protein (drug efficacy) and whether other

proteins of the cell, or population of cells, have also been affected (drug specificity).

E. Protein Sequencing, Purification and Characterization

The capture agents of the present invention may also be used in protein sequencing. Briefly, capture agents are raised that interact with a known combination of unique recognition sequences. Subsequently, a protein of interest is fragmented using the methods described herein to generate a collection of peptides and then the sample is allowed to interact with the capture agents. Based on the interaction pattern between the collection of peptides and the capture agents, the amino acid sequence of the collection of peptides may be deciphered. In a preferred embodiment, the capture agents are immobilized on an array in pre-determined positions that allow for easy determination of peptide-capture agent interactions. These sequencing methods would further allow the identification of amino acid polymorphisms, *e.g.*, single amino acid polymorphisms, or mutations in a protein of interest.

In another embodiment, the capture agents of the present invention may also be used in protein purification. In this embodiment, the PET acts as a ligand/affinity tag and allows for affinity purification of a protein. A capture agent raised against a PET exposed on a surface of a protein may be coupled to a column of interest using art known techniques. The choice of a column will depend on the amino acid sequence of the capture agent and which end will be linked to the matrix. For example, if the amino-terminal end of the capture agent is to be linked to the matrix, matrices such as the Affigel (by Biorad) may be used. If a linkage via a cysteine residue is desired, an Epoxy-Sepharose-6B column (by Pharmacia) may be used. A sample containing the protein of interest may then be run through the column and the protein of interest may be eluted using art known techniques as described in, for example, J. Nilsson *et al.* (1997) "Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins," *Protein Expression and Purification*, 11:11-16, the contents of which are incorporated by reference. This embodiment of the invention also allows for the characterization of protein-protein interactions under native conditions, without the need to introduce artificial affinity

tags in the protein(s) to be studied.

In yet another embodiment, the capture agents of the present invention may be used in protein characterization. Capture agents can be generated that differentiate between alternative forms of the same gene product, e.g., between proteins having different post-translational modifications (*e.g.*, phosphorylated versus non-phosphorylated versions of the same protein or glycosylated versus non-glycosylated versions of the same protein) or between alternatively spliced gene products.

The utility of the invention is not limited to diagnosis. The system and methods described herein may also be useful for screening, making prognosis of disease outcomes, and providing treatment modality suggestion based on the profiling of the pathologic cells, prognosis of the outcome of a normal lesion and susceptibility of lesions to malignant transformation.

F. Detection of Post-translational Modifications

The subject computer generated PETs can also be analyzed according to the likely presence or absence of post-translational modifications. More than 100 different such modifications of amino acid residues are known, examples include but are not limited to acetylation, amidation, deamidation, prenylation (such as farnesylation or geranylation), formylation, glycosylation, hydroxylation, methylation, myristoylation, phosphorylation, ubiquitination, ribosylation and sulphation. Sequence analysis softwares which are capable of determining putative post-translational modification in a given amino acid sequence include the NetPhos server which produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins (available through http://www.cbs.dtu.dk/services/Net-Phos/), GPI Modification Site Prediction (available through http://mendel.imp.univie.ac.at/gpi) and the ExPASy proteomics server for total protein analysis (available through www.expasy.ch/tools/)

In certain embodiments, preferred PET moieties are those lacking any post-translational modification sites, since post-translationally modified amino acid

sequences may complicate sample preparation and/or interaction with a capture agent. Notwithstanding the above, capture agents that can discriminate between post-translationally forms of a PET, which may indicate a biological activity of the polypeptide-of-interest, can be generated and used in the present invention. A very common example is the phosphorylation of OH group of the amino acid side chain of a serine, a threonine, or a tyrosine group in a polypeptide. Depending on the polypeptide, this modification can increase or decrease its functional activity. In one embodiment, the subject invention provides an array of capture agents that are variegated so as to provide discriminatory binding and identification of various post-translationally modified forms of one or more proteins. In a preferred alternative embodiment, the subject invention provides an array of capture agents that are variegated so as to provide specific binding to one or more PET uniquely associated with a modification of interest, which modification itself can be readily detected and/or quantitated by additional agents, such as a labeled secondary antibody specifically recognizing the modification (e.g., a phospho-tyrosine antibody).

In a general sense, the invention provides a general means to detect / quantitate protein modifications. "Modification" here refers generally to any kind of non-wildtype changes in amino acid sequence, including post-translational modification, alternative splicing, polymorphysm, insertion, deletion, point mutation, etc. To detect / quantitate a specific modification within a potential target protein present in a sample, the sequence of the target protein is first analyzed to identify potential modification sites (such as phosphorylation sites for a specific kinase). Next, a potential fragment of the target protein containing such modification site is identified. The fragment is specific for a selected method of treatment, such as tryptic digestion or digestion by another protease or reliable chemical fragmentation. PET within (and unique) to the modification site-containing fragment can then be identified using the method of the instant invention. Fragmentation using a combination of two or more methods is also contemplated. Absolute predictability of the fragment size is desired, but not necessary, as long as the fragment always contains the desired PET and the modification site.

Antibody or other capture agents specific for the identified PET is obtained. The capture agent is then used in a sandwich ELISA format to detect captured

fragments containing the modification (see Figure 22). The site of the PET is proximal to the post-translational modification site(s). Thus a binding to the PET by a capture agent will not interfere with the binding of a detection agent specific for the modified residue.

A few specific embodiments of this aspect of the invention are described in more detail below (see Figure 23). For illustrative purpose only, the capture agents described below in various embodiments of the invention are antibodies specific for PETs. However, it should be understood that any capture agents described above can be used in each of the following embodiments.

(i) Phosphorylation

The reversible addition of phosphate groups to proteins is important for the transmission of signals within eukaryotic cells and, as a result, protein phosphorylation and dephosphorylation regulate many diverse cellular processes. To detect the presence and/or quantitate the amount of a phosphorylated peptide in a sample, anti-phospho-amino acid antibodies can be used to detect the presence of phosphopeptides.

There are numerous commercially available phospho-tyrosine specific antibodies that can be adapted to be used in the instant invention. Merely to illustrate, phosphotyrosine antibody (ab2287) [13F9] of Abcam Ltd (Cambridge, UK) is a mouse IgG1 isotype monoclonal antibody reacts specifically with phosphotyrosine and shows minimal reactivity by ELISA and competitive ELISA with phosphoserine or phosphothreonine. The antibody reacts with free phosphotyrosine, phosphotyrosine conjugated to carriers such as thyroglobulin or BSA, and detects the presence of phosphotyrosine in proteins of both unstimulated and stimulated cell lysates.

Similarly, RESEARCH DIAGNOSTICS INC (Flanders, NJ) provides a few similar anti-phosphotyrosine antibodies. Among them, RDI-PHOSTYRabmb is a mouse mIgG2b isotype monoclonal antibody reacts strongly and specifically with phosphotyrosine-containing proteins and can be blocked specifically with

phosphotyrosine. No reaction with either phosphothreonine or phosphoserine is detected. This antibody appears to have broad cross-species reactivity, and is reactive with various tyrosine-phosphorylated proteins of human, chick, frog, rat, mouse and dog origin.

RESEARCH DIAGNOSTICS INC also provides phosphoserine-specific antibodies, such as RDI-PHOSSERabr, which is an affinity-purified rabbit antibody made against phosphoserine containing proteins. The antibody reacts specifically with serine phosphorylated proteins and shows no significant cross reactivity to other phosphothreonine or phosphotyrosine by western blot analysis. This antibody is suitable for ELISA according to the manufacture's suggestion. The company also provides a mouse IgG1 monoclonal anti-phosphoserine antibody RDI-PHOSSEabm, which reacts specifically with phosphorylated serine, both as free amino acid or conjugated to carriers as BSA or KLH. No cross reactivity is observed with non-phosphorylated serine, phosphothreonine, phosphotyrosine, AmpMP or ATP.

RDI-PHOSTHRabr is an affinity isolated rabbit anti-phosphothreonine antibody (anti-pT) provided by RESEARCH DIAGNOSTICS INC. Both antigen-capture and antibody-capture ELISA indicated that the anti-phosphothreonine antibodies can recognize threonine-phosphorylated protein, phosphothreonine and lysine-phosphothreonine-glycine random polymer, respectively. Direct, competitive antigen-capture ELISA demonstrated that the antibodies are specifically inhibited by free phosphothreonine, phosvitin but not by free phosphoserine, phosphotyrosine, threonine and ATP. The company also provides a mouse IgG2b monoclonal anti-phosphothreonine antibody RDI-PHOSTHabm, which reacts specifically with phosphorylated threonine, both as free amino acid or conjugated to carriers as BSA or KLH. No cross reactivity is observed with non-phosphorylated threonine, phophoserine, phosphotyrosine, AmpMP or ATP.

Molecular Probe (Eugene, OR) has developed a small molecule fluorophore phosphosensor, referred to as Pro-Q Diamond dye, which is capable of ultrasensitive global detection and quantitation of phosphorylated amino acid residues in peptides and proteins displayed on microarrays. The utility of the fluorescent Pro-Q Diamond phosphosensor dye technology is demonstrated using phosphoproteins and

phosphopeptides as well as with protein kinase reactions performed in miniaturized microarray assay format (Martin, *et al.*, *Proteomics* 3: 1244–1255, 2003). Instead of applying a phosphoamino acid-selective antibody labeled with a fluorescent or enzymatic tag for detection, a small, fluorescent probe is employed as a universal sensor of phosphorylation status. The detection limit for phosphoproteins on a variety of different commercially available protein array substrates was found to be 312–625 fg, depending upon the number of phosphate residues. Characterization of the enzymatic phosphorylation of immobilized peptide targets with Pro-Q Diamond dye readily permits differentiation between specific and non-specific peptide labeling at picogram to subpicogram levels of detection sensitivity. Martin et al. (*supra*) also describe in detail the suitable protocols, instruments for using the Pro-Q stain, especially for peptides on microarrays, the entire contents of which are incorporated herein by reference.

One of the advantageous of the method over other methods, such as identification of modified amino acids in proteins by mass spectrometry, is that the instant invention provides a much simpler technique that does not rely on expensive instruments, and thus can be easily adapted to be used in small or large laboratories, in industry or academic settings alike.

In one embodiment, the instant invention can be used to identify potential substrates of a specific kinase or kinase subfamily. As the number of known protein kinases has increased at an ever-accelerating pace, it has become more challenging to determine which protein kinases interact with which substrates in the cell.

The determination of consensus phosphorylation site motifs by amino acid sequence alignment of known substrates has proven useful in this pursuit. These motifs can be helpful for predicting phosphorylation sites for specific protein kinases within a potential protein substrate. The table below summarizes merely some of the known data about specificity motifs for various well-studied protein kinases, along with examples of known phosphorylation sites in specific proteins (for a more extensive list, see Pearson, R. B., and Kemp, B. E. (1991). In T. Hunter and B. M. Sefton (Eds.), Methods in Enzymology Vol. 200, pp. 62–81. San Diego: Academic Press, incorporated by reference). Phosphoacceptor residue is indicated in bold,

amino acids which can function interchangeably at a particular residue are separated by a slash (/), and residues which do not appear to contribute strongly to recognition are indicated by an "X." Some protein kinases such as CKI and GSK-3 contain phosphoamino acid residues in their recognition motifs, and have been termed "hierarchical" protein kinases (see Roach, *J. Biol. Chem.* 266, 14139–14142, 1991 for review). They often require prior phosphorylation by another kinase at a residue in the vicinity of their own phosphorylation site. S(p) represents such preexisting phosphoserine residues.

Protein Kinase	**Recognition Motifs**[a]	**Phosphorylation Sites**[b]	**Protein Substrate (reference)**
cAMP-dependent Protein Kinase (PKA, cAPK)	R-X-S/T[c] R-R/K-X-S/T	Y_7LRRASLAQLT F_1RRLSIST A_{29}GARRKASGPP	pyruvate kinase (2) phosphorylase kinase, a chain (2) histone H1, bovine (2)
Casein Kinase I (CKI, CK-1)	S(*P*)-X-X-S/T	R_4TLS(*P*)VSSLPGL D_{43}IGS(*p*)ES(*p*)TEDQ	glycogen synthase, rabbit muscle (4) a_{s1}-casein (4)
Casein Kinase II (CKII, CK-2)	S/T-X-X-E	A_{72}DSESEDEED L_{37}ESEEEGVPST E_{26}DNSEDEISNL	PKA regulatory subunit, R_{II} (2) $p34^{cdc2}$, human (5) acetyl-CoA carboxylase (2)
Glycogen Synthase Kinase 3 (GSK-3)	S-X-X-X-S(*p*)	S_{641}VPPSPSLS(*p*) S_{641}VPPS(*p*)PSLS(*p*)	glycogen synthase, human (site 3b) (6,2) glycogen synthase, human (site 3a) (6,2)
Cdc2 Protein Kinase	S/T-P-X-R/K[c]	P_{13}AKTPVK H_{122}STPPKKKRK	histone H1, calf thymus (2) large T antigen (2)
Calmodulin-dependent Protein Kinase II (CaMK II)	R-X-X-S/T R-X-X-S/T-V	N_2YLRRRLSDSN K_{191}MARVFSVLR	synapsin (site 1) (2) calcineurin (2)
Mitogen-activated Protein Kinase (Extracellular Signal-regulated	P-X-S/T-P[d] X-X-S/T-P	P_{244}LSP P_{92}SSP V_{420}LSP	c-Jun (7) cyclin B (7) Elk-1 (7)

Kinase) (MAPK, Erk)			
cGMP-dependent Protein Kinase (cGPK)	R/K-X-S/T R/K- X -X-S/T	G_{26}KKRKRSRKES F_1RRLSIST	histone H2B (2) phosphorylase kinase (a chain) (2)
Phosphorylase Kinase (PhK)	K/R-X-X-S-V/I	D_6QEKRKQISVRG P_1LSRTLSVSS	phosphorylase (2) glycogen synthase (site 2) (2)
Protein Kinase C (PKC)	S/T-X-K/R K/R- X -X-S/T K/R-X-S/T	H_{594}EGTHSTKR P_1LSRTLSVSS Q_4KRPSQRSKYL	fibrinogen (2) glycogen synthase (site 2) (2) myelin basic protein (2)
Abl Tyrosine Kinase	I/V/L-Y-X-X-P/F[e]		
Epidermal Growth Factor Receptor Kinase (EGF-RK)	E/D-Y-X E/D-Y-I/L/V	R_{1168}ENAEYLRVAP A_{767}EPDYGALYE	autophosphorylation (2) phospholipase C-g(2)

Single-letter Amino Acid Code:
A = alanine, **C** = cysteine, **D** = aspartic acid, **E** = glutamic acid, **F** = phenylalanine, **G** = glycine, **H** = histidine, **I** = isoleucine, **K** = lysine, **L** = leucine, **M** = methionine, **N** = asparagine, **P** = proline, **Q** = glutamine, **R** = arginine, **S** = serine, **T** = threonine, **W** = tryptophan, **V** = valine, **Y** = tyrosine, **X** = any amino acid

[a] Recognition motifs are taken from Pearson and Kemp (*supra*) except where noted. Consult Pearson and Kemp for a comprehensive list of phosphorylation site sequences and specificity motifs.

[b] Subscripted numbers refer to the position of the first residue within the given polypeptide chain.

[c] From (1).

[d] From (7).

[e] From (8). See refs (8) and (9) for discussion of substrate recognition by Abl.

References used in the table above:

1. Kennelly, P. J., and Krebs, E. G. (1991) J. Biol. Chem. 266, 15555–15558.

2. Pearson, R. B., and Kemp, B. E. (1991). In T. Hunter and B. M. Sefton

(Eds.), Methods in Enzymology Vol. 200, (pp. 62–81). San Diego: Academic Press.

3. Roach, P. J. (1991) J. Biol. Chem. 266, 14139–14142.
4. Flotow, H. et al. (1990) J. Biol. Chem. 265, 14264–14269.
5. Russo, G. L. et al. (1992) J. Biol. Chem. 267, 20317–20325.
6. Fiol, C. J. et al. (1990) J. Biol. Chem. 265, 6061–6065.
7. Davis, R. J. (1993) J. Biol. Chem. 268, 14553–14556.
8. Songyang, Z. et al. (1995) Nature 373, 536–539.
9. Geahlen, R. L. and Harrison, M. L. (1990). In B. E. Kemp (Ed.), Peptides and Protein Phosphorylation, (pp. 239–253). Boca Raton: CRC Press.

However, since the determinants of protein kinase specificity involve complex 3-dimensional interactions, these motifs, short amino-acid sequences describing the primary structure around the phosphoacceptor residue, are a significant oversimplification of the issue. They do not take into account possible secondary and tertiary structural elements, or determinants from other polypeptide chains or from distant locations within the same chain. Furthermore, not all of the residues described in a particular specificity motif may carry the same weight in determining recognition and phosphorylation by the kinase. In addition, the potential recognition sequence maybe buried deep inside a tertiary structure of within a protein complex under physiological conditions and thus may never be accessible in vivo. As a consequence, they should be used with some caution. The instant invention provides a fast and convenient way to determine, on a proteome-wide basis, the identity of all potential kinase substrates that actually do become phopshorylated by the kinase of interest in vivo (or in vitro).

Specifically, consensus recognition sequences of a kinase (or a kinase subfamily sharing substrate specificity) can be identified based on, for example, Pearson and Kemp or other kinase substrate motif database. For example, AKT (or PKB) kinase has a consensus phosphorylation site sequence of RXRXXS/T. All proteins in an organism (e.g., human) that contains this potential recognition sequence can be readily identified through routine sequence searches. Using the

method of the instant invention, peptide fragments of these potential substrates, after a pre-determined treatment (such as trypsin digestion), which contain both the recognition motif and at least one PET can then be generated. Antibodies (or other capture agents) against each of these identified PETs can be raised and printed on an array to generate a so-called "kinase chip," in this case, an AKT chip. Using this chip, any sample to be studied can be treated as described above and then be incubated with the chip so that all potential recognition site-containing fragments are captured. The presence or absence of phosphorylation on any given "spot" – a specific potential substrate - can be detected / quantitated by, for example, labeled secondary antibodies (see Figure 10). Thus, the identity of all AKT substrates in this organism under this condition may be identified in one experiment. The array can be reused for other samples by eluting the bound peptides on the array. Different arrays can be used in combination, preferably in the same experiment, to determine the substrates for multiple kinases.

The reversible phosphorylation of tyrosine residues is an important mechanism for modulating biological processes such as cellular signaling, differentiation, and growth, and if deregulated, can result in various types of cancer. Therefore, an understanding of these dynamic cellular processes at the molecular level requires the ability to assess changes in the sites of tyrosine phosphorylation across numerous proteins simultaneously as well as over time. Thus in another embodiment, the instant invention provides a method to identify the various signal transduction pathways activated after a specific treatment to a sample, such as before and after a specific growth factor or cytokine treatment to a sample cell. The same method can also be used to compare the status of signal transduction pathways in a diseased sample from a patient and a normal sample from the same patient.

Know ledges about the various signal transduction pathways existing in various organisms are accumulating at an astonishing pace. Science magazine's STKE (Signal Transduction Knowledge Environment) maintains a comprehensive and expanding list of known signal transduction pathways, their important components, relationship between the components (inhibit, stimulation, etc.), and

cross-talk between key members of the different pathways. The "Connections Map" provides a dynamic graphical interface into a cellular signaling database, which currently covers at least the following broad pathways: immune pathways (IL-4, IL-13, Token-like receptor); seven-transmembrane receptor pathways (Adrenergic, PAC1 receptor, Dictyostelium discoideum cAMP Chemotaxis, Wnt/Ca^{2+}/cyclic GMP, G Protein-Independent 7 Transmembrane Receptor); Circadian Rhythm pathway (murine and Drosophila); Insulin pathway; FAS pathway; TNF pathway; G-Protein Coupled Receptor pathways; Integrin pathways; Mitogen-Activated Protein Kinase Pathways (MAPK, JNK, p38); Estrogen Receptor Pathway; Phosphoinositide 3-Kinase Pathway; Transforming Growth Factor-β (TGF-β) Pathway; B Cell Antigen Receptor Pathway; Jak-STAT Pathway; STAT3 Pathway; T Cell Signal Transduction Pathway; Type 1 Interferon (α/β) Pathway; Jasmonate Biochemical Pathway; and Jasmonate Signaling Pathway. Many other well-known signal transduction pathways not yet included are described in detail in other scientific literatures which can be readily identified in PubMed or other common search tools. Activation of most, if not all of these signal transduction pathways are generally characterized by changes in phosphorylation levels of one or more members of each pathway.

Thus in a general sense, the status of any given number of signaling pathways in a sample can be determined by taking a "snap shot" of the phosphorylation status of one or more key members of these selected pathways. For example, the Mitogen-activated protein (MAP)1 kinase pathways are evolutionarily conserved in eukaryotic cells. The pathways are essential for physiological processes, such as embryonic development and immune response, and regulate cell survival, apoptosis, proliferation, differentiation, and migration. In mammals, three major classes of MAP kinases (MAPKs) have been identified, which differ in their substrate specificity and regulation. These subgroups compose the extracellular signal-regulated kinases (ERKs), the c-Jun N-terminal kinases (JNKs), and the p38/RK/CSBP kinases. ERKs are activated by a range of stimuli including growth factors, cell adhesion, tumor-promoting phorbol esters, and oncogenes, whereas JNK and p38 are preferentially activated by proinflammatory cytokines, and a variety of environmental stresses such as UV and osmotic stress. For this reason, the latter are

classified as stress-activated protein kinases. Activation of the MAPKs is achieved by dual phosphorylation on threonine and tyrosine residues within a Thr-Xaa-Tyr motif located in the kinase subdomain VIII. This phosphorylation is mediated by a dual specificity protein kinase, MAPK kinase (MAPKK), and MAPKK is in turn activated by phosphorylation mediated by a serine/threonine protein kinase, MAPKK kinase. In addition to these activating kinases, several types of protein phosphatases have been also shown to control MAPK pathways by dephosphorylating the MAPKs or their upstream kinases. These protein phosphatases include tyrosine-specific phosphatases, serine/threonine-specific phosphatases, and dual specificity phosphatases (DSPs). Therefore, the activities of MAPKs can be regulated by upstream activating kinases and protein phosphatases, and the activation status can be determined by the phosphorylation status of, for example, ERK1/2, JNK, and p38.

Specifically, fragments of ERK1/2, JNK, and p38 containing the signature phosphorylation sites and PETs can be identified using the methods of the instant invention. Capture agents specifically recognizing such phosphorylation site-associated PETs can then be raised and immobilized on an array / chip. A sample (treated or untreated, thus containing high or low levels of phosphorylation of these pathway markers) can be digested and incubated with the chip, so as to determine the presence / absence of activation, and degree, time course, duration of activation, etc.

In the same principal, many other related or perceived unrelated pathways may be manufactured on the same chip, since each pathway may be represented by from just one, to possibly all of the known pathway components. This type of chip may provide a comprehensive view of the various pathways that may be activated after a drug treatment. Pathway specific chips may also be used in conjunction to further determine the status of individual components within a pathway of interest.

Because of the important functions of the kinases in virtually all kinds of signal transduction pathways, it is not surprising to see that many drugs directly or indirectly affects phosphorylation status of carious kinase substrates. Thus this type of array may also be used in drug target identification. Briefly, samples treated by

different drug candidates may be incubated with the same kind of array to generate a series of activation profiles of certain chosen targets. These profiles may be compared, preferably automatically, to determine which drug candidate has the same or similar activation profile as that of the lead molecule.

This type of experiment will also yield useful information concerning the selectivity of candidate drugs, since it can be easily determined whether a candidate drug or drug analog actually have differential effects on various pathways, and if so, whether the difference is significant.

The same type of experiments can also be adapted to screen for drug candidates that lacks undesired side effects or toxicity.

One aspect of this type of application relates to the selection of specific protease(s) for fragmentation. The following table presents data resulting from analysis of protease sensitivity of potential phosphorylation sites in the human "kinome" (all kinases). This table may aid the selection of proteases among the several most frequently used proteases.

Enzymes	**Total Peptide Fragments**	**Peptide Fragments with S/T/Y**	
		=<10 aa	>10 aa
Chymotrypsin	34,094	10930 (43%)	14985 (57%)
S.A. V-8 E specific Enzyme	34,233	6753 (32%)	14917 (68%)
Post-Proline Cleaving Enzyme	29,715	7077 (37%)	12224 (63%)
Trypsin	54,260	15, 217 (53%)	13311 (47%)

(ii) Glycosylation

A wide variety of eukaryotic membrane-bound and secreted proteins are glycosylated, that is they contain covalently-bound carbohydrate, and therefore are

termed glycoproteins. In addition, certain intracellular eukaryotic proteins are also glycoproteins. Glycosylation of polypeptides in eukaryotes occurs principally in three ways (Parekh et al., *Trends Biotechnol.* 7: 117, 1989). Glycosylation through a glycosidic bond to an asparagine side-chain is known as N-glycosylation. Such asparagine residues only occur in the amino acid triplet sequence of Asn-Xaa-Ser/Thr, where Xaa can be any amino acid. The carbohydrate portion of a glycoprotein is also known as a glycan. O-glycans are linked to serine or threonine side-chains, through O-glycosidic bonds. In human, 284,535 octamer tags contains this NX(S/T) sequence, and 228,256 octamer PETs contains the NX(S/T) sequence. The latter is about 2.6% of the total octamer peptide tags in human. The N- and O-linked glycosylation are two of the most complex post-translational modifications. The polypeptide may also be linked to a phosphatidylinositol lipid anchor through a carbohydrate "bridge", the whole assembly being known as the glycosyl-phosphatidylinositol (GPI) anchor.

In recent years, the functional significance of the carbohydrate moieties has been increasingly appreciated (Rademacher et al., *Ann. Rev. Biochem.* 57: 785, 1988). Carbohydrates covalently attached to polypeptide chains can confer many functions to the glycoprotein, for example resistance to proteolytic degradation, the transduction of information between cells, and intercellular adhesion through ligand-receptor interactions (Gesundheit et al., *J. Biol. Chem.* 262: 5197, 1987; Ashwell & Harford, *Ann. Rev. Biochem.* 51: 531, 1982; Podskalny et al., *J. Biol. Chem.* 261: 14076, 1986; Dennis et al., *Science* 236: 582, 1987). As glycoforms are the product of a series of biochemical modifications, perturbations within a cell can have profound effects on their structure. With the increase in understanding of carbohydrate functions, the need for rapid, reliable and sensitive methods for carbohydrate detection and analysis has grown considerably.

Lectins are proteins that interact specifically and reversibly with certain sugar residues. Their specificity enables binding to polysaccharides and glycoproteins (even agglutination of erythrocytes and tumor cells). The binding reaction between a lectin and a specific sugar residue is analogous to the interaction between an antibody and an antigen. Substances bound to lectin may be resolved with a competitive binding substance or an ionic strength gradient. In addition,

among other procedures, lectins can be labeled with biotin or digoxigenin, and subsequently detected by avidin-conjugated peroxidase or anti-digoxigenin antibodies coupled with alkaline phosphatase, respectively (Carlsson SR: *Isolation and characterization of glycoproteins*. In: *Glycobiology. A Practical Approach.* Fukuda M and Kobata A (eds). Oxford University Press, Oxford, pp1-25, 1993, incorporated herein by reference).

For example, Concanavalin A (Con A) binds molecules that contain α-D-mannose, α-D-glucose and sterically related residues with available C-3, C-4, or C-5 hydroxyl groups. Like Con A, lentil lectin binds α-D-mannose, α-D-glucose, and sterically related residues, but lentil lectin distinguishes less sharply between glucosyl and mannosyl residues and binds simple sugars with lower affinity. Agarose wheat germ lectin specifically binds to N-acetyl-β-glucosaminyl residues. Wheat germ lectin specifically binds to N-acetyl-β-D-glucosaminyl residues. Psathyrella velutina lectin (PVL) preferentially interacts with the N-acetylglucosamine beta 1-->2Man group. All these lectins can be used to detect the presence of various kinds of glycosylated peptides fragments after these PET-associated glycosylated peptide fragments are captured from the sample by capture agents.

The GlycoTrack Kit from Glyko, Inc. (a Prozyme company, San Leandro, CA) detect glycosylation by using a specific carbohydrate oxidation reaction prior to binding of a high amplification color generating reagent. Briefly, a sample, either in solution or already immobilized to a support, is oxidized with periodate. This generates aldehyde groups that can react spontaneously with certain hydrazides at room temperature in aqueous conditions. Use of biotin-hydrazide following periodate oxidation leads to the incorporation of biotin into the carbohydrate (9). The biotinylated compound is detected by reaction with a streptavidin-alkaline phosphatase conjugate. Subsequently visualization is achieved using a substrate that reacts with the alkaline phosphatase bound to glycoproteins on the membrane, forming a colored precipitate.

Molecular Probes (Eugene, OR) offer a proprietary Pro-Q Emerald 300 fluorescent glycoprotein stain for detection of glycoproteins. The new Pro-Q

Emerald 300 fluorescent glycoprotein stain reacts with periodate-oxidized carbohydrate groups, creating a bright green-fluorescent signal on glycoproteins. Depending upon the nature and the degree of glycosylation, this stain may be 50-fold more sensitive than the standard periodic acid–Schiff base method using acidic fuchsin dye. According to the manufacture, detection using the Pro-Q Emerald 300 glycoprotein stain is much easier than detection of glycoproteins using biotin hydrazide with streptavidin–horseradish peroxidase and ECL detection (Amersham Pharmacia Biotech). The stain can detect 50ng of a typical glycosylated protein. Since the captured glycosylated PET-containing peptide fragments are much smaller than a typocal peptide, as little as low nanogram to high picograms of captured peptides can be detected using this dye.

Thus to detect the presence and quantitation of glycosylation in a sample, all proteins or a subpopulation thereof which contains the potential glycosylation site NXS/T may be identified, and peptide fragments resulting from a specific pre-determined treatment may be analyzed to identify associated PETs. Capture agents against these PETs can then be raised. In a method analogous to the phosphorylation detection as described above, glycosylation can be detected / quantitated using the various detection methods

(iii) Other Post-translational modifications

Capture agents, such as antibodies specific for other post-translationally modified residues are also readily availble.

There are at least 46 anti-ubiquitin commercial antibodies available from 14 different vendors. For example, Cell Signaling Technology (Beverly, MA) offers mouse anti-Ubiquitin monoclonal antibody, clone P4D1 (IgG1 isotype, Cat. No. 3936), which is specific for all species of ubiquitin, polyubiquitin, and ubiquitinated peptides.

Anti-acetylated amino acid antibodies have also been commerciallized. See anti-acetylated-histon H3 and H4 antibodies (Catalog # 06-599 and Catalog # 06-598) from Upstate Biotechnology (Lake Placid, NY). In fact, Alpha Diagnostic

International, Inc. (San Antonio, TX) offers custom synthesis of anti-acetylated amino acid antibodies.

Arginine methylation, a protein modification discovered almost 30 years ago, has recently experienced a renewed interest as several new arginine methyltransferases have been identified and numerous proteins were found to be regulated by methylation on arginine residues. Mowen and David published detailed protocols on Science's STKE (www.stke.org/cgi/content/full/OC_sigtrans;2001/93/pl1) that provide guidelines for the straightforward identification of arginine-methylated proteins, made possible by the availability of novel, commercially available reagents. Specifically, two anti-methylated arginine antibodies are described: mouse monoclonal antibody to methylarginine, clone 7E6 (IgG1) (Abcam, Cambr idge, UK) (Data sheet: www.abcam.com/public/ab_detail.cfm?intAbID=412, which reacts with mono- and asymmetric dimethylated arginine residues; and mouse monoclonal antibody to methylarginine, clone 21C7 (IgM) (Abcam) (Data sheet: www.abcam.com/public/ab_detail.cfm?intAbID=413), which reacts with asymmetric dimethylated arginine residues. Detailed protocols for in vitro and in vivo analysis of arginine methylation are provided. See Mowen et al., *Cell* 104: 731-741, 2001.

Even if there is no reported antibodies at present for certain specific modifications, it is well within the capability of a skilled artisan to raise antibodies against that specific type of modified residues. There is no compelling reason to believe that such antibodies cannot be obtained, especially in view of the prior success in raising antibodies against reletively small groups such as phosphorylated amio acids. The anti-post-translational modification antibody should be checked against the same antigen that is un-modified to verify that the reactivity is depending upon the presence of the post-translational modification.

G. Immunohistochemistry (IHC)

Immunohistochemical analysis of tumor tissues / biopsy has traditionally played an important role in diagnosis, monitoring, and prognosis analysis of cancer. IHC is typically performed on disease tissue sections using antibodies (monoclonal or polyclonal) to specific disease markers. However, two major problems have hampered this useful procedure, such that it is frequently difficult to get reproducible, quantitative data. One problem is associated with the poor quality of antibodies used in the assay. Many antibodies lack specificity to a target biomarker, and tend to cross-react with other proteins not associated with disease status, resulting in high background. The other complication is that antibody may have difficulties accessing unknown epitopes after tissue/cell fixation.

For example, Press et al. (*Cancer Res.* 54(10): 2771-7, 1994) compared immunohistochemical staining results obtained with 7 polyclonal and 21 monoclonal antibodies in sections from paraffin-embedded blocks of breast cancer samples. It was found that the ability of these antibodies to detect the HER2/neu antigen overexpression was extremely variable, providing an important explanation for the variable overexpression rate reported in the literature.

The other problem is associated with sample processing before IHC. Generally, the efficiency of antigen retrieval is unpredictable in the concurrent protocol. It is also reported that heating coupled with enzyme digestion tends to give better results. But since epitopes for antibodies are not known, heating/digestion may cause different degree of problems for antibody recognition.

Therefore, PET-derived antibodies represent a unique solution as standardized reagents for IHC. In certain preferred embodiments, PETs present on the surface of the target protein will be chosen for easy accessibility by the PET-specific antibodies. The chemistry of cell fixation may also be taken into account to select optimum amino acid sequences of PETs. For example, if certain residues are known to form cross-links after fixation, these residues will be selected against in PET selection. Similarly, epitopes that overlap with enzyme recognition sites will not be chosen. These measures will help to achieve consistent, reproducible results and high rate of success in IHC experiments.

VII. Other Aspects of the Invention

In another aspect, the invention provides compositions comprising a plurality of isolated unique recognition sequences, wherein the unique recognition sequences are derived from at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% 95% or 100% of an organism's proteome. In one embodiment, each of the unique recognition sequences is derived from a different protein.

The present invention further provides methods for identifying and/or detecting a specific organism based on the organism's Proteome Epitope Tag. The methods include contacting a sample containing an organism of interest (*e.g.*, a sample that has been fragmented using the methods described herein to generate a collection of peptides) with a collection of unique recognition sequences that characterize, and/or that are unique to, the proteome of the organism. In one embodiment, the collection of unique recognition sequences that comprise the Proteome Epitope Tag are immobilized on an array. These methods can be used to, for example, distinguish a specific bacterium or virus from a pool of other bacteria or viruses.

The unique recognition sequences of the present invention may also be used in a protein detection assay in which the unique recognition sequences are coupled to a plurality of capture agents that are attached to a support. The support is contacted with a sample of interest and, in the situation where the sample contains a protein that is recognized by one of the capture agents, the unique recognition sequence will be displaced from being bound to the capture agent. The unique recognition sequences may be labeled, *e.g.*, fluorescently labeled, such that loss of signal from the support would indicate that the unique recognition sequence was displaced and that the sample contains a protein is recognized by one or more of the capture agents.

The PETs of the present invention may also be used in therapeutic applications, *e.g.*, to prevent or treat a disease in a subject. Specifically, the PETs may be used as vaccines to elicit a desired immune response in a subject, such as an immune response against a tumor cell, an infectious agent or a parasitic agent. In this embodiment of the invention, a PET is selected that is unique to or is over-

represented in, for example, a tissue of interest, an infectious agent of interest or a parasitic agent of interest. A PET is administered to a subject using art known techniques, such as those described in, for example, U.S. Patent No. 5,925,362 and international publication Nos. WO 91/11465 and WO 95/24924, the contents of each of which are incorporated herein by reference. Briefly, the PET may be administered to a subject in a formulation designed to enhance the immune response. Suitable formulations include, but are not limited to, liposomes with or without additional adjuvants and/or cloning DNA encoding the PET into a viral or bacterial vector. The formulations, *e.g.*, liposomal formulations, incorporating the PET may also include immune system adjuvants, including one or more of lipopolysaccharide (LPS), lipid A, muramyl dipeptide (MDP), glucan or certain cytokines, including interleukins, interferons, and colony stimulating factors, such as IL1, IL2, gamma interferon, and GM-CSF.

EXAMPLES

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures are hereby incorporated by reference.

EXAMPLE 1: IDENTIFICATION OF UNIQUE RECOGNITION EQUENCES WITHIN THE HUMAN PROTEOME

As any one of the total 20 amino acids could be at one specific position of a peptide, the total possible combination for a tetramer (a peptide containing 4 amino acid residues) is 20^4; the total possible combination for a pentamer (a peptide containing 5 amino acid residues) is 20^5 and the total possible combination for a hexamer (a peptide containing 6 amino acid residues) is 20^6. In order to identify unique recognition sequences within the human proteome, each possible tetramer, pentamer or hexamer was searched against the human proteome (total number: 29,076; Source of human proteome: EBI Ensembl project release v 4.28.1 on Mar 12, 2002, http://www.ensembl.org/Homo_sapiens/).

The results of this analysis, set forth below, indicate that using a pentamer as a unique recognition sequence, 80.6% (23,446 sequences) of the human proteome have their own unique recognition sequence(s). Using a hexamer as a unique recognition sequence, 89.7% of the human proteome have their own unique recognition sequence(s). In contrast, when a tetramer is used as a unique recognition sequence, only 2.4% of the human proteome have their own unique recognition sequence(s).

Results and Data

2.1. Tetramer analysis:

2.1.1. Sequence space:

Total number of human protein sequences	29,076	100%
*Number of sequences with 1 or more unique tetramer tag	**684**	2.4%
Number of sequences with 0 unique tetramer tag	28,392	97.6%

*For these 684 sequences, average Tag/sequence: 1.1.

2.1.2. Tag space:

Total number of tetramers	20^4=160,000	100%
Tetramers found in 0 sequence	393	0.2%
#Tetramers found in 1 sequence only	**745**	0.5%
Tetramers found in more than 1 sequences	158,862	99.3%

#: These are signature tetra-peptides

2.2. Pentamer analysis:

2.2.1. Sequence space:

Total number of human protein sequences	29,076	100%
*Number of sequences with 1 or more unique pentamer tag	**23,446**	80.6%
Number of sequences with 0 unique pentamer tag	5,630	19.4%

*For these 23,446 sequences, Average Tag/sequence: 23.9

2.2.2. Tag space:

Total number of pentamers	20^5=3,200,000	100%
Pentamers found in 0 sequence	955,007	29.8%
#Pentamers found in 1 sequence only	**560,309**	17.5%
Pentamers found in more than 1 sequences	1,684,684	52.6%

#: These are signature penta-peptides

2.3. Hexamer analysis:

2.3.1. Sequence space:

Total number of human protein sequences	29,076	100%
*Number of sequences with 1 or more unique hexamer tag	**26,069**	89.7%
Number of sequences with 0 unique hexamer tag	3,007	10.3%

*For these 26069 sequences, Average Tag/sequence: 177

2.3.2. Tag space:

Total number of hexamers	20^6=64,000,000	100%
hexamers found in 0 sequence	57,040,296	89.1%
# hexamers found in 1 sequence only	**4,609,172**	7.2%
hexamers found in more than 1 sequences	2,350,532	3.7%

#: These are signature hexa-peptides.

Similar analysis in the human proteome was done for PET sequences of 7-10 amino acids in length, and the results are combinedly summarized in the table below:

PET Length (Amino Acids)	**Tagged Sequences** (Number)	**Tagged Sequences** (% of total - 29076)	**Average PET** (Number/ Tagged Protein)
4	684	2.35%	3
5	23,446	80.64%	24
6	26,069	89.66%	177
7	26,184	90.05%	254
8	26,216	90.16%	268
9	26,238	90.24%	272
10	26,250	90.28%	275

EXAMPLE 2: IDENTIFICATION OF UNIQUE RECOGNITION SEQUENCES (OR PETS) WITHIN ALL BACTERIAL PROTEOMES

In order to identify pentamer PETs that can be used to, for example, distinguish a specific bacterium from a pool of all other bacteria, each possible pentamer was searched against the NCBI database (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/eub_g.html, updated as of April 10, 2002). The results from this analysis are set forth below.

Results and Data:

Number of unique pentamers	Database ID (NCBI RefSeq ID)	Species Name
6	NC_000922	Chlamydophila pneumoniae CWL029
37	NC_002745	Staphylococcus aureus N315 chromosome
40	NC_001733	Methanococcus jannaschii small extra-chromosomal element
58	NC_002491	Chlamydophila pneumoniae J138
84	NC_002179	Chlamydophila pneumoniae AR39
135	NC_000909	Methanococcus jannaschii
206	NC_003305	Agrobacterium tumefaciens str. C58 (U. Washington) linear chromosome
298	NC_002758	Staphylococcus aureus Mu50 chromosome
356	NC_002655	Escherichia coli O157:H7 EDL933
386	NC_003063	Agrobacterium tumefaciens str. C58 (Cereon) linear chromosome
479	NC_000962	Mycobacterium tuberculosis
481	NC_002737	Streptococcus pyogenes
495	NC_003304	Agrobacterium tumefaciens str. C58 (U. Washington) circular chromosome
551	NC_003098	Streptococcus pneumonia R6
567	NC_003485	Streptococcus pyogenes MGAS8232
577	NC_002695	Escherichia coli O157
592	NC_003028	Streptococcus pneumonia, TIGR4
702	NC_003062	Agrobacterium tumefaciens str. C58 (Cereon) circular chromosome
729	NC_001263	Deinococcus radiodurans chromosome 1
918	NC_003116	Neisseria meningitidis Z2491

924	NC_000908	Mycoplasma genitalium
960	NC_002755	Mycobacterium tuberculosis CDC1551
977	NC_003112	Neisseria meningitidis MC58
979	NC_000921	Helicobacter pylori J99
1015	NC_000915	Helicobacter pylori 26695
1189	NC_000963	Rickettsia prowazekii
1284	NC_001318	Borrelia burgdorferi chromosome
1331	NC_002771	Mycoplasma pulmonis
1426	NC_000912	Mycoplasma pneumoniae
1431	NC_002528	Buchnera sp. APS
1463	NC_000868	Pyrococcus abyssi
1468	NC_000117	Chlamydia trachomatis
1468	NC_002162	Ureaplasma urealyticum
1478	NC_003212	Listeria innocua
1553	NC_003210	Listeria monocytogenes
1577	NC_000961	Pyrococcus horikoshii
1630	NC_002620	Chlamydia muridarum
1636	NC_003103	Rickettsia conorii Malish 7
1769	NC_003198	Salmonella typhi
1794	NC_000913	Escherichia coli K12
1894	NC_002689	Thermoplasma volcanium
1996	NC_003413	Pyrococcus furiosis
2081	NC_002578	Thermoplasma acidophilum
2106	NC_003197	Salmonella typhimurium LT2
2137	NC_003317	Brucella melitensis chromosome I
2402	NC_002677	Mycobacterium leprae
2735	NC_000918	Aquifex aeolicus
2803	NC_002505	Vibrio cholerae chromosome 1

2900	NC_000907	Haemophilus influenzae
3000	NC_003318	Brucella melitensis chromosome II
3120	NC_000854	Aeropyrum pernix
3229	NC_002662	Lactococcus lactis
3287	NC_002607	Halobacterium sp. NRC-1
3298	NC_003454	Fusobacterium nucleatum
3497	NC_001732	Methanococcus jannaschii large extra-chromosomal element
3548	NC_002163	Campylobacter jejuni
3551	NC_000853	Thermotoga maritima
3688	NC_003106	Sulfolobus tokodaii
3775	NC_002754	Sulfolobus solfataricus
3842	NC_000919	Treponema pallidum
3921	NC_003296	Ralstonia solanacearum GMI1000
3940	NC_000916	Methanobacterium thermoautotrophicum
4165	NC_001264	Deinococcus radiodurans chromosome 2
4271	NC_003047	Sinorhizobium meliloti 1021 chromosome
4338	NC_002663	Pasteurella multocida
4658	NC_003364	Pyrobaculum aerophilum
5101	NC_000917	Archaeoglobus fulgidus
5787	NC_003366	Clostridium perfringens
5815	NC_003450	Corynebacterium glutamicum
6520	NC_002696	Caulobacter crescentus
6866	NC_002506	Vibrio cholerae chromosome 2
6891	NC_003295	Ralstonia solanacearum chromosome
7078	NC_002488	Xylella fastidiosa chromosome
8283	NC_003143	Yersinia pestis chromosome
8320	NC_000911	Synechocystis PCC6803

8374	NC_002570	Bacillus halodurans
8660	NC_000964	Bacillus subtilis
8994	NC_003030	Clostridium acetobutylicum ATCC824
11725	NC_003552	Methanosarcina acetivorans
12120	NC_002516	Pseudomonas aeruginosa
12469	NC_002678	Mesorhizobium loti
14022	NC_003272	Nostoc sp. PCC 7120

EXAMPLE 3: IDENTIFICATION OF SPECIFIC PETS

Figure 11 outlines a general approach to identify all PETs of a given length in an organism with sequenced genome or a sample with known proteome. Briefly, all protein sequences within a sequenced genome can be readily identified using routine bioinformatic tools. These protein sequences are parsed into short overlapping peptides of 4-10 amino acids in length, depending on the desired length of PET. For example, a protein of X amino acids gives (X-N+1) overlapping peptides of N amino acids in length. Theoretically, all possible peptide tags for a given length of, for example, N amino acids, can be represented as 20^N (preferably, N = 4-10). This is the so-called peptide tag database for this particular length (N) of peptide fragments. By comparing each and every sequence of the parsed short overlapping peptides with the peptide tag database, all PET (with one and only one occurrence in the peptide tag database) can be identified, while all non-PET (with more than one occurrence in the peptide tag database) can be eliminated.

As indicated above, each possible tetramer, pentamer or hexamer was searched against the human proteome (total number: 29,076; Source of human proteome: EBI Ensembl project release 4.28.1 on Mar 12, 2002, http://www.ensembl.org/Homo_sapiens/) to identify unique recognition sequences (PETs).

Based on the foregoing searches, specific PETs were identified for the majority of the human proteome. *Figure 1* depicts the pentamer unique recognition

sequences that were identified within the sequence of the Interleukin-8 receptor A. Figure 2 depicts the pentamer unique recognition sequences that were identified within the Histamine H1 receptor that are not destroyed by trypsin digestion. Further Examples of pentamer unique recognition sequences that were identified within the human proteome are set forth below.

Sequence ID*	Number of pentamer PETs	Pentamer PETs
ENSP00000000233	9	AMPVS CATQG CFTVW ICFTV MPNAM PNAMP SRTWY TWYVQ WYVQA (SEQ ID NOs:1-9)
ENSP00000000412	30	CDFVC CGKEQ CWRTG DNFNP DNHCG FRVCR FYSCW GMEQF HLAFW IFNGS IMLIY IYIFR KGMEQ KTCDL MFPFY MISCN NETHI NWIML PFYSC QDCFY QFPHL RESWQ SNWIM VMISC YDNHC YIYIF YKGGD YLFEM YRGVG YSCWR (SEQ ID NOs:10-39)
ENSP00000000442	2	ASNEC PASNE (SEQ ID NOs:40-41)
ENSP00000000449	9	AQPWA ASTWR CLCLV FVICA LYCCP PRANR VNVLC YAQLW YCCPV (SEQ ID NOs:42-50)
ENSP00000001008	20	AIQRM AKPNE AMCHL AWDIA CQQRI ELKYE EMPMI FVHYT HSIVY HYTGW LYANM MIGDR QKSNT SWEMN SWLEY TEMPM WEMNS YAKPN YESSF YPNNK (SEQ ID NOs:51-70)
ENSP00000001146	32	ATRDK CPCEG DKSCK DTHDT EWPRS FEVYQ FQIPK FSGYR GCPCE GHLFE HDTAP IFSHE KEMTM KLQCT KSCKL KYGNV LKHPT MGEHH MTMQE MYSIR NVFDP QLWQL RGIQA RYLDC STEWP THDTA TRTFP VMYSI VRTCL VSTEW WQLRW WSVMY (SEQ ID NOs:71-102)
ENSP00000001178	8	ACKCF CKCFW FWLWY KCFWL LWYPH QKRRC WLWYP WYPHF (SEQ ID NOs:103-110)
ENSP00000001380	26	AMEQT APCTI AYMER CTIMK DGLCN EQTWR FRSYG GMAYM GYHMP HIPNY KGRIP KLDMG MAYME MEQTW MNKRE PGMNK QGYHM TMSPK TWRLD VEQGY VNDGL WDQTR WRLDP YEAME YHMPC YNPCQ (SEQ ID NOs:111-136)
ENSP00000001567	137	ATYYK CATYY CDNPY CEVVK CIKTD CINSR CKSPD CKSSN CNELP CQENY CSESF CYERE CYHFG CYMGK DFTWF DGWSA DIPIC DQTYP DREYH EEMHC EFDHN EFNCS EHGWA EINYR EKIPC EMHCS ESNTG ESTCG ESYAH EYHFG EYYCN FENAI FQYKC FTWFK GEWVA

		GNVFE GWTND HGRKF HGTIN HGWAQ HPGYA HPPSC HTVCI IHGVW IKHRT IMVCR INGRW IPCSQ IPVFM IVCGY IYKCR IYKEN KCNMG KGEWV KIPCS KPCDY KWSHP LPICY MENGW MGKWS MGYEY MIGHR NCSMA NDFTW NEGYQ NETTC NGWSD NMGYE NQNHG NSVQC NVFEY NYRDG NYREC PCDYP PEVNC PICYE PPQCE PPYYY PQCVA PYIPN QCYHF QIQLC QYKVG RDTSC REYHF RIKHR RKGEW RPCGH RVRYQ RWQSI SCDNP SDQTY SFTMI SITCG SRWTG STGWI SVEFN SWSDQ TAKCT TCIHG TCINS TCMEN TCYMG TMIGH TNDIP TSTGW TWFKL TYKCF VAIDK VCGYN VEFNC VFEYG VIMVC VNCSM VTYKC WDHIH WFKLN WIHTV WQSIP WSDQT WTNDI YCNPR YHENM YHFGQ YKCFE YKCNM YKCRP YKIEG YMGKW YNGWS YNQNH YPDIK YQCRN YQYGE YSERG YWDHI YYKMD (SEQ ID NOs:137-274)
ENSP00000001585	25	CVSKG EIIII GINYE GMKHA GWDLK HGMKH HHPKF IEKCV IIMDA INYEI KGYVF MEMIV MIVRA NYTIG QMEMI SHHPK TGSFR TRYKG VYGWD YGESK YGWDL YIHGM YNERE YTIGE YVFQM (SEQ ID NOs:275-299)
ENSP00000002125	7	GRYQR KNMGI MGERF PIKQH QRNAR RYQRN YDMLM (SEQ ID NOs:299-306)
ENSP00000002165	63	AHSAT AKFFN CKWGW CMTID DKLSW DQAKF DVWYT EYSWN FDQAK FEWFH FNANQ FWWYW FYTCS HKWEN HPKAI HQMPC HTWRS IHQMP IPKYV IYETH KFFNA KWENC KWGWA KWPTS LMNIG LPHKW MPCKW MRPQE NANQW NCMTI NYPPS NYQPE PCKWG PDQYW PHKWE QMGSW QYWNS RNRTD SCGGN SKHHE TCSDR THTWR TIHQM TNDRW TPDVW TRFDP TVVTN VRGTV VVTND WENCM WFDQA WFWWY WGSEY WGWAL WNWNA WRSQN WWYWQ YEDFG YETHT YNPGH YSWNW YVEFM YYSLF (SEQ ID NOs:307-369)
ENSP00000002494	74	AMNDA ANHGE AQWRN CVKLP CVQYK DAHKR DCVQY DIEQR DMAER DPDKW DTANH EVSFM EYVID FEQYE FFEQY FGDCV FMNET HEIYR HERFL HFDQT HKQWK HKRAF HTAMN HWIQQ KHFDQ KMLNQ KQMTS KQYAQ KRAFH KWERF LNGRW LPHWI MFATM MKFMN MKMEF MLNQS MPQEG MYVKA NLPHW NTDAH NVLKH PHWIQ PVMDA QADEM QENCK QHTAM QNYVS QWKDL QYAQA RVPVM SFYDS SHERF TCDEM TDAHK TKLMP TVVRY TYQIL VMDAQ VMKFM VPVMD VRYLF VSFMN WDRYG WERFE WIIKY WIQQH WISTN WKDYT WKKHV YAQAD YEVTY YGRRE YTDCV YVKAD (SEQ ID NOs:369-443)
ENSP00000002594	7	CFKEN DGGFD FDLGD KLCFK KPMPN MPNPN PNPNH (SEQ ID NOs:444-450)
ENSP00000002596	36	DRCLH EEHYS EHYSH ENEVH EYFHE FFDWE FHEPN FSWPH FYNHM GRDRC GVAPN HEYFH HFFDW HIVDG HKPYP HMQKH HMQNW HPQVD HVHMQ KGRAH KHKPY KTPAY MQNWL NHMQK QKHKP QNWLR RVYSM SMNPS SWPHQ TFDWH TQVFY WEEHY YCLRD YHVHM YNHMQ YPSIE (SEQ ID NOs:451-486)
ENSP00000002829	60	ADIRM AWPSF CLVNK CQAYG CTYVN DHDRM DPSFI DRMYV GHCCL GIETH GYWRH HCCLV HDINR HDRMY HQYCQ HRCQA IETHF IFYLE IHQYC IIHWA INFMR

		IQPWN KMPYP KWLFQ LIIHW LIQPW MCTYV MPYPR MRSHP NNFKH NPIRQ NSRWL NTTDY NYQWM PIRQC PRNRR PVKTM PWNRT QDYIF QGYWR QTAMR RCQAY RMVFN SKDYV SNANK TGAWP VGVTH VINFM VKWLF WDGQA WPSFP WRHVP YAGVY YCQGY YNPMC YNSRW YPLQR YQAVY YQWMP YWRHV (SEQ ID NOs:487-546)

* The Sequence IDs used are the ones provided in http://www.ensembl.org/Homo_sapiens/

Figure 12 lists the results of searching the whole human proteome (a total of 29,076 proteins, which correspond to about 12 million 4-10 overlapping peptides) for PETs, and the number of PETs identified for each N between 4-10.

Figure 13 shows the result of percentage of human proteins that have at least one PET(s). It is shown that for a PET of 4 amino acids in length, only 684 (or about 2.35% of the total human proteins) proteins have at least one 4-mer PETs. However, if PETs of at least 6 amino acids are used, at least about 90% of all proteins have at least one PET. In addition, it is somewhat surprising that there is a significant increase in average number of PETs per protein from 5-mer PETs to 6-mer (or more) PETs (see lower panel of Figure 13), and that average quickly reaches a platue when 7- or 8-mer PETs are used. These data indicates that PETs of at least 6 amino acids, preferably 7-9 amino acids, most preferably 8 amino acids have the optimal length of PETs for most applications. It is easier to identify a useful PET of that length, partly because of the large average number of PETs per protein when a PET of that length is sought.

Figure 14 provides further data resulting from tryptic digest of the human proteome. Specifically, the top panel lists the average number of PETs per tagged protein (protein with at least one PETs), with or without trypsin digestion. Trypsin digestion reduces the average number of PETs per tagged protein by roughly 1/3 to 1/2. The bottom right panel shows the distribution of tryptic fragments in the human proteome, listed according to peptide length. On average, a typical tryptic fragment is about 8.5 amino acids in length. The bottom left panel shows the distribution of number of tryptic fragments generated from human proteins. On average, a human protein has about 49 tryptic fragments.

Example 6 below provides a detailed example of identifying SARS virus-

specific 8-mer PETs. These PETs are potentially useful as SARS-specific antigens for immunization (vaccine production) in human or other mammals.

EXAMPLE 4: DETECTION AND QUANTITATION IN A COMPLEX MIXTURE OF A SINGLE PEPTIDE SEQUENCE WITH TWO NON-OVERLAPPING PET SEQUENCES USING SANDWICH ELISA ASSAY

A fluorescence sandwich immunoassay for specific capture and quantitation of a targeted peptide in a complex peptide mixture is illustrated herein.

In the example shown here, a peptide consisting of three commonly used affinity epitope sequences (the HA tag, the FLAG tag and the MYC tag) is mixed with a large excess of unrelated peptides from digested human protein samples (Figure 15). The FLAG epitope in the middle of the target peptide is first captured here by the FLAG antibody, then the labeled antibody (either HA mAb or MYC mAb) is used to detect the second epitope. The final signal is detected by fluorescence readout from the secondary antibody. Figure 15 shows that picomolar concentrations of HA-FLAG-MYC peptide was detected in the presence of a billion molar excess of digested unrelated proteins. The detection limit of this method is typically about 10 pM or less.

The sandwich assay was used to detect a tagged-human PSA protein, both as full length protein secreted in conditioned media of cell cultures, and as tryptic peptides generated by digesting the same conditioned media. The result of this analysis is shown in Figure 16. The PSA protein sandwich assay (left side of the figure) indicated that the PSA protein concentration is about 7.4 nM in conditioned media. SDS-PAGE analysis indicated that the tryptic digestion of all proteins in the sample was complete, manifested by the absence of any visible bands on the gel after digestion since most tryptic fragments are expected to be less than 1 kDa. The right side of the figure indicated that nearly the same concentration (8 nM) of the last fragment – the tag-containing portion of the recombinant PSA protein was present in the digested sample. The higher concentration could be attributed to the elimination of interfering substances in the sample, such as other proteins that bind

the full-length PSA protein and mask its interaction with the antibody. Although this type of interference is not so severe in this example since the relatively simple conditioned media was used, it is expected to be much more prevalent in real biological samples, where large interference is expected from unknown proteins in a non-digested and complicated bodily fluid such as serum.

The same sandwich assay may be used for detecting modified amino acids, such as phosphorylated proteins using anti-tyrosine, anti-serine, or anti-threonine antibodies. For example, Figure 17 shows that the phopshoprotein SHIP-2 contains a 28-amino acid tryptic fragment, which is phosphorylated on one tyrosine residue N-terminal to an 8-mer PET (YVLEGVPH) and on one serine residue C-terminal to the PET. Thus in the sandwich assay, the trypsin digested SHIP-2 protein can first be pulled-down using the PET-specific antibody, and the presence of phosphorylated tyrosine or serine may be detected / quantitated using the phospho-specific antibodies, such as those described elsewhere in the instant specification. Three of the nearest neighbors of the selected PET are also shown in the figure.

Similarly, the phosphoprotein ABL also contains an 8-mer PET on its tryptic fragment containing the phosphorylation site. The phosphorylated peptide is readily detectable by a phospho-tyrosine-specific antibody.

In fact, as a general approach, the sandwich assay may be used to detect N proteins with N+1 PET-specific antibodies: one PET is common to all N peptides to be detected, while each specific peptide also contains a unique PET. All N peptides can be pulled-down by a capture agent specific to the common PET, and the presence and quantity of each specific peptide can be individually assessed using antibodies specific to the unique PETs (see Figure 18).

To illustrate, most kinases are somehow related by sharing similar catalytic structures and/or catalytic mechanisms. Thus, it is interesting that only 88 5-mer PETs are needed to represent all known 518 human kinases, and 122 6-mer PETs are needed for the same purpose. Figure 18 also shows that the top 20 most common 6-mer PETs cover more than 70% of all known kinases. Since closely related kinases tend to share common features, the subject sandwich assay is suitable for simultaneous detection of family of kinases. Figure 19 provides such an example,

wherein one 5-mer PET is shared among tryptic fragments of 22 related kinases, each of which also has unique 7-mer or 8-mer PETs.

The same approach may be used for other protein families, including GPCRs, proteases, phosphotases, receptors, or specific enzymes. The Human Plasma Membrane Receptome is disclosed at http://receptome.stanford.edu/HPMR.

EXAMPLE 5: PEPTIDE COMPETITION ASSAY

In certain embodiments of the invention, a peptide competition assay may be used to determine the binding specificity of a capture agent towards its target PET, as compared to several nearest neighbor sequences of the PET.

For a typical peptide competition assay, the following illustrative protocol may be used: 1 μg/100 μl/well of each target peptide is coated in Maxisorb Plates with coating buffer (carbonate buffer, pH 9.6) overnight at 4°C, or 1 hour at room temperature. The plates are washed with 300 μl of PBST (1 x PBS / 0.05% tween 20) for 4 times. Then 300 μl of blocking buffer (2% BSA / PBST) is added and the plates are incubated for 1 hour at room temperature. Following blocking, the plates are washed with 300 μl of PBST for 4 times.

Synthesized competition peptides are dissolved in water to a final concentration of 2 mM solution. Serial dilution of competition peptides (for example, from 100 pM to 100 μM) in digested human serum are prepared. These competition peptides at particular concentrations are then mixed with equal amounts of primary antibodies against the target peptide. These mixtures are then added to plate wells with immobilized target peptides respectively. Binding is allowed to proceed for 2 hours at room temperature. The plates are washed with 300 μl of PBST for 4 times. Then labeled secondary antibody against the primary antibody, such as 100 μl of 5,000 x diluted anti-rabbit-IgG-HRP, is added and incubated for 1 more hour at room temperature. The plates are washed with 300 μl of PBST for 6 times. For detection of the HRP label activity, add 100 μl of TMB substrate (for HRP) and incubate for 15 minutes at room temperature. Add 100 μl of stop buffer (2N HCL) and read the plates at OD_{450}. A peptide competition curve is plotted using

the ABS at OD_{450} versus the competitor peptide concentrations.

EXAMPLE 6: IDENTIFICATION OF SARS-SPECIFIC PETS

Sequence Retrieval

A total of 2028 Coronavirus peptide sequences were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov:80/genomes/SARS/SARS.html). These sequences represent at least 10 different species of Coronavirus. Among them, 1098 non-redundant peptide sequences were identified. Each sequence that appeared identically within (was subsumed in) a larger sequence was removed, leaving the larger sequence as the representative. The resulting sequences were then broken up into overlapping regions of eight amino acids (8-mers), with a sequence difference of 1 amino acid between successive 8-mers. These 8-mers were then queried against a database consisting of all 8-mers similarly generated and present in the proteome of the species in question (or any other set of protein sequences deemed necessary). 8-mers found to be present only once (the sequence identified only itself) were considered unique. The remainder of the sequences were initially classified as non-unique with the understanding that with more in-depth analysis, they might actually be as useful as those sequences initially determined to be unique. For example, an 8-mer may be present in another isoform of its parent sequence, so it would still be useful in uniquely detecting that parental sequence and that isoform from all other unrelated proteins.

A total of ~650,000 8-mer peptide sequences were generated, ~50,000 of which were determined to be PETs. Among these, 605 were SARS-specific and 602 were PETs relative to human.

PET Prioritization:

Once PETs have been identified, the best candidates for a particular application must be chosen from the pool of all PETs.

Generally, PETs are ranked based upon calculations used to predict their hydrophobicity, antigenicity, and solubility, with hydrophilic, antigenic, and soluble PETs given the highest priority. The PETs are then further ranked by determining

each PET's closest nearest neighbors (similar looking 8-mers with at least one sequence difference(s)) in the proteome(s) in question. A matrix calculation is performed using a BLOSUM, PAM, or a similar proprietary matrix to determine sequence similarity and distance. PETs with the most distant nearest neighbors are given the priority.

The parental peptide sequence is then proteolytically cleaved *in silico* and the resulting fragments sorted by user-defined size / hydrophobicity / antigenicity / solubility criteria. The presence of PETs in each fragment is assessed, and fragments containing no PETs are discarded. The remaining fragments are analyzed in terms of PET placement within them depending upon the requirements of the type of assay to be performed. For example, a sandwich assay prefers two non-overlapping PETs in a single fragment. The ideal final choice would be the most antigenic PETs with only distantly-related nearest neighbors in an acceptable proteolytic fragment that fit the requirements of the assay to be performed.

Figure 20 shows two SARS-specific PETs and their nearest neighbors in both the human proteome and the related Coronaviruses.

All SARS-specific PETs identified using this method is listed below in Table SARS.

Table SARS List of SARS virus-specific PETs

>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	ISLCSCIC
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	SLCSCICT
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	LCSCICTV
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	CSCICTVV
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	SCICTVVQ
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	CICTVVQR
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	ICTVVQRC
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	CTVVQRCA
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	HVLEDPCK
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	VLEDPCKV
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	LEDPCKVQ
>gi|30795153|gb|AAP41045.1| Orf10 [SARS coronavirus Tor2]	EDPCKVQH
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1]	MNELTLID
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1]	NELTLIDF
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1]	ELTLIDFY
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1]	LTLIDFYL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1]	TLIDFYLC
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1]	LIDFYLCF

>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] IDFYLCFL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] DFYLCFLA
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] FYLCFLAF
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] YLCFLAFL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LCFLAFLL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] CFLAFLLF
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] FLAFLLFL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LAFLLFLV
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] AFLLFLVL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] FLLFLVLI
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LLFLVLIM
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LFLVLIML
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] FLVLIMLI
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LVLIMLII
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] VLIMLIIF
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LIMLIIFW
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] IMLIIFWF
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] MLIIFWFS
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LIIFWFSL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] IIFWFSLE
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] IFWFSLEI
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] FWFSLEIQ
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] WFSLEIQD
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] FSLEIQDL
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] SLEIQDLE
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LEIQDLEE
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] EIQDLEEP
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] IQDLEEPC
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] QDLEEPCT
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] DLEEPCTK
>gi|32187352|gb|AAP72981.1| Orf7b [SARS coronavirus HSR 1] LEEPCTKV
>gi|32187350|gb|AAP72979.1| Orf6 [SARS coronavirus HSR 1] DEEPMELB
>gi|32187350|gb|AAP72979.1| Orf6 [SARS coronavirus HSR 1] EEPMELBY
>gi|32187350|gb|AAP72979.1| Orf6 [SARS coronavirus HSR 1] EPMELBYP
>gi|30023959|gb|AAP13572.1| unknown [SARS coronavirus CUHK-W1] DEEPMELD
>gi|30023959|gb|AAP13572.1| unknown [SARS coronavirus CUHK-W1] EEPMELDY
>gi|30023959|gb|AAP13572.1| unknown [SARS coronavirus CUHK-W1] EPMELDYP
>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] SELDDEEL
>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] ELDDEELM
>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] LDDEELME
>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] DDEELMEL

>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] DEELMELD
>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] EELMELDY
>gi|30275674|gb|AAP30035.1| putative uncharacterized protein 3 [SARS coronavirus BJ01] ELMELDYP
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] MLPPCYNF
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LPPCYNFL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] PPCYNFLK
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] PCYNFLKE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] CYNFLKEQ
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] YNFLKEQH
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] NFLKEQHC
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] FLKEQHCQ
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LKEQHCQK
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] KEQHCQKA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EQHCQKAS
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] QHCQKAST
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] HCQKASTQ
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] CQKASTQR
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] QKASTQRE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] KASTQREA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] ASTQREAE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] STQREAEA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] TQREAEAA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] QREAEAAV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] REAEAAVK

>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EAEAAVKP
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AEAAVKPL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EAAVKPLL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AAVKPLLA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AVKPLLAP
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VKPLLAPH
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] KPLLAPHH
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] PLLAPHHV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LLAPHHVV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LAPHHVVA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] APHHVVAV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] PHHVVAVI
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] HHVVAVIQ
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] HVVAVIQE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VVAVIQEI
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VAVIQEIQ
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AVIQEIQL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VIQEIQLL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] IQEIQLLA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] QEIQLLAA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EIQLLAAV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] IQLLAAVG
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] QLLAAVGE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LLAAVGEI

>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LAAVGEIL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AAVGEILL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AVGEILLL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VGEILLLE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] GEILLLEW
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EILLLEWL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] ILLLEWLA
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LLLEWLAE
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LLEWLAEV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LEWLAEVV
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EWLAEVVK
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] WLAEVVKL
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] LAEVVKLP
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] AEVVKLPS
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] EVVKLPSR
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VVKLPSRY
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] VKLPSRYC
>gi|31747859|gb|AAP69660.1| uncharacterized protein 9c [SARS coronavirus ZJ-HZ01] KLPSRYCC
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] VLLFLAFM
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] LLFLAFMV
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] LFLAFMVF
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] FLAFMVFL
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] LAFMVFLL
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] AFMVFLLV

>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] FMVFLLVT
>gi|31416298|gb|AAP51230.1| envelope protein E [SARS coronavirus GZ01] MVFLLVTL
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] VLLFLAFV
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] LLFLAFVV
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] LFLAFVVF
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] FLAFVVFL
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] LAFVVFLL
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] AFVVFLLV
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] FVVFLLVT
>gi|29836499|ref|NP_828854.1| small envelope protein; protein sM; protein E [SARS coronavirus] VVFLLVTL
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] MCLKILVR
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] CLKILVRY
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] LKILVRYN
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] KILVRYNT
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] ILVRYNTR
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] LVRYNTRG
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] VRYNTRGN
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] TAAFRDVL
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] AAFRDVLV
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] AFRDVLVV
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] FRDVLVVL
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] RDVLVVLN
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] DVLVVLNK
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] VLVVLNKR
>gi|32187354|gb|AAP72983.1| Orf8b [SARS coronavirus HSR 1] LVVLNKRT
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01] MDPNQTNV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01] DPNQTNVV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01] PNQTNVVP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01] NQTNVVPP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01] QTNVVPPA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01] TNVVPPAL

>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
NVVPPALH
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
VVPPALHL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
VPPALHLV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PPALHLVD
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PALHLVDP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ALHLVDPQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LHLVDPQI
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
HLVDPQIQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LVDPQIQL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
VDPQIQLT
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
DPQIQLTI
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PQIQLTIT
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QIQLTITR
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
IQLTITRM
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QLTITRME
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LTITRMED
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
TITRMEDA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ITRMEDAM
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
TRMEDAMG
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
RMEDAMGQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
MEDAMGQG
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
EDAMGQGQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
DAMGQGQN
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
AMGQGQNS

>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
MGQGQNSA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
GQGQNSAD
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QGQNSADP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
GQNSADPK
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QNSADPKV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
NSADPKVY
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
SADPKVYP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ADPKVYPI
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
DPKVYPII
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PKVYPIIL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
KVYPIILR
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
VYPIILRL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
YPIILRLG
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PIILRLGS
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
IILRLGSQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ILRLGSQL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LRLGSQLS
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
RLGSQLSL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LGSQLSLS
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
GSQLSLSM
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
SQLSLSMA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QLSLSMAR
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LSLSMARR
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
SLSMARRN

>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LSMARRNL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
SMARRNLD
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
MARRNLDS
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ARRNLDSL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
RRNLDSLE
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
RNLDSLEA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
NLDSLEAR
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LDSLEARA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
DSLEARAF
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
SLEARAFQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LEARAFQS
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
EARAFQST
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ARAFQSTP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
RAFQSTPI
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
AFQSTPIV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
FQSTPIVV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QSTPIVVQ
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
STPIVVQM
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
TPIVVQMT
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PIVVQMTK
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
IVVQMTKL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
VVQMTKLA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
VQMTKLAT
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
QMTKLATT

>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
MTKLATTE
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
TKLATTEE
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
KLATTEEL
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LATTEELP
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ATTEELPD
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
TTEELPDE
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
TEELPDEF
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
EELPDEFV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
ELPDEFVV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
LPDEFVVV
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
PDEFVVVT
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
DEFVVVTA
>gi|31416303|gb|AAP51235.1| BGI-PUP7 [SARS coronavirus GZ01]
EFVVVTAK
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
ISLCSCIR
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
SLCSCIRT
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
LCSCIRTV
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
CSCIRTVV
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
SCIRTVVQ
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
CIRTVVQR
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
IRTVVQRC
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
RTVVQRCA
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
HVLEDPCP
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
VLEDPCPT
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
LEDPCPTG

>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
EDPCPTGY
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
DPCPTGYQ
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
PCPTGYQP
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
CPTGYQPE
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
PTGYQPEW
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
TGYQPEWN
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
GYQPEWNI
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
YQPEWNIR
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
QPEWNIRY
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
PEWNIRYN
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
EWNIRYNT
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
WNIRYNTR
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
NIRYNTRG
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
IRYNTRGN
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
TAAFRDVF
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
AAFRDVFV
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
AFRDVFVV
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
FRDVFVVL
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
RDVFVVLN
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
DVFVVLNK
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
VFVVLNKR
>gi|31416304|gb|AAP51236.1| BGI-PUP(GZ29-nt-Ins) [SARS coronavirus GZ01]
FVVLNKRT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
MKIILFLT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
KIILFLTL

>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1] IILFLTLI
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ILFLTLIV
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LFLTLIVF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FLTLIVFT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LTLIVFTS
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TLIVFTSC
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LIVFTSCE
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
IVFTSCEL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VFTSCELY
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FTSCELYH
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TSCELYHY
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
SCELYHYQ
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
CELYHYQE
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ELYHYQEC
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LYHYQECV
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
YHYQECVR
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
HYQECVRG
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
YQECVRGT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
QECVRGTT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ECVRGTTV
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
CVRGTTVL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VRGTTVLL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
RGTTVLLK
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
GTTVLLKE

>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TTVLLKEP
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TVLLKEPC
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VLLKEPCP
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LLKEPCPS
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LKEPCPSG
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
KEPCPSGT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
EPCPSGTY
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
PCPSGTYE
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
CPSGTYEG
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
PSGTYEGN
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
SGTYEGNS
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
GTYEGNSP
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TYEGNSPF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
YEGNSPFH
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
EGNSPFHP
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
GNSPFHPL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
NSPFHPLA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
SPFHPLAD
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
PFHPLADN
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FHPLADNK
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
HPLADNKF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
PLADNKFA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LADNKFAL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ADNKFALT

>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
DNKFALTC
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
NKFALTCT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
KFALTCTS
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FALTCTST
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ALTCTSTH
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LTCTSTHF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TCTSTHFA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
CTSTHFAF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TSTHFAFA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
STHFAFAC
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
THFAFACA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
HFAFACAD
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FAFACADG
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
AFACADGT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FACADGTR
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ACADGTRH
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
CADGTRHT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ADGTRHTY
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
DGTRHTYQ
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
GTRHTYQL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TRHTYQLR
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
RHTYQLRA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
HTYQLRAR
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
TYQLRARS

>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
YQLRARSV
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
QLRARSVS
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LRARSVSP
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
RARSVSPK
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ARSVSPKL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
RSVSPKLF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
SVSPKLFI
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VSPKLFIR
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
SPKLFIRQ
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
PKLFIRQE
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
KLFIRQEE
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LFIRQEEV
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FIRQEEVQ
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
IRQEEVQQ
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
RQEEVQQE
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
QEEVQQEL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
EEVQQELY
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
EVQQELYS
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VQQELYSP
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
QQELYSPL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
QELYSPLF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ELYSPLFL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LYSPLFLI
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
YSPLFLIV

>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
SPLFLIVA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
PLFLIVAA
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LFLIVAAL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FLIVAALV
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LIVAALVF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
IVAALVFL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VAALVFLI
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
AALVFLIL
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ALVFLILC
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LVFLILCF
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
VFLILCFT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FLILCFTI
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LILCFTIK
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
ILCFTIKR
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
LCFTIKRK
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
CFTIKRKT
>gi|31581511|gb|AAP33703.1| Orf7a [SARS coronavirus Frankfurt 1]
FTIKRKTE
>gi|30026017|gb|AAP04587.1| RNA-directed RNA polymerase [SARS coronavirus Taiwan] ILSDDGVX
>gi|30026017|gb|AAP04587.1| RNA-directed RNA polymerase [SARS coronavirus Taiwan] LSDDGVXV
>gi|30026017|gb|AAP04587.1| RNA-directed RNA polymerase [SARS coronavirus Taiwan] SDDGVXVL
>gi|30026017|gb|AAP04587.1| RNA-directed RNA polymerase [SARS coronavirus Taiwan] DDGVXVLN
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] LLIQQWIP
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] LIQQWIPF
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] IQQWIPFM

>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] QQWIPFMM
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] QWIPFMMS
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] WIPFMMSR
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] IPFMMSRR
>gi|30275671|gb|AAP30032.1| putative uncharacterized protein 2 [SARS coronavirus BJ01] PFMMSRRR
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] QIQLSLLQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] IQLSLLQV
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] QLSLLQVT
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] LSLLQVTA
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] SLLQVTAF
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] LLQVTAFQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] LQVTAFQH
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] QVTAFQHQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] STALQELQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] TALQELQI
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] ALQELQIQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] LQELQIQQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] QELQIQQW
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] ELQIQQWI
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] LQIQQWIQ
>gi|31416297|gb|AAP51229.1| BGI-PUP2 [SARS coronavirus GZ01] QIQQWIQF
>gi|30795147|gb|AAP41039.1| Orf4 [SARS coronavirus Tor2] LLIQQWIQ
>gi|30795147|gb|AAP41039.1| Orf4 [SARS coronavirus Tor2] LIQQWIQF
>gi|30314342|gb|AAP06763.1| RNA-directed RNA polymerase [SARS coronavirus Hong Kong/03/2003] QDAVASKI
>gi|30314342|gb|AAP06763.1| RNA-directed RNA polymerase [SARS coronavirus Hong Kong/03/2003] DAVASKIL

>gi|30314342|gb|AAP06763.1| RNA-directed RNA polymerase [SARS coronavirus Hong Kong/03/2003] YVDTENNL
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] LACFVLAV
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] ACFVLAVV
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] CFVLAVVY
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] FVLAVVYR
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] VLAVVYRI
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] LAVVYRIN
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] AVVYRINW
>gi|31581509|gb|AAP33701.1| membrane protein M [SARS coronavirus Frankfurt 1] VVYRINWV
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] HLRMAGHP
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] LRMAGHPL
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] RMAGHPLG
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] MAGHPLGR
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] AGHPLGRC
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] GHPLGRCD
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] HPLGRCDI
>gi|30027623|gb|AAP13444.1| M protein [SARS coronavirus Urbani] PLGRCDIK
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] LCWKCKSQ
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] CWKCKSQN
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] WKCKSQNP
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] KCKSQNPL
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] CKSQNPLL
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] KSQNPLLY
>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] SQNPLLYD

>gi|30275670|gb|AAP30031.1| putative uncharacterized protein 1 [SARS coronavirus BJ01] QNPLLYDA
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] TDTIVVTA
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] DTIVVTAG
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] TIVVTAGD
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] IVVTAGDG
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] VVTAGDGI
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] VTAGDGIS
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] TAGDGIST
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] AGDGISTP
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] IGGYSEDW
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] GGYSEDWH
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] GYSEDWHS
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] YSEDWHSG
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] SEDWHSGV
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] EDWHSGVK
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] DWHSGVKD
>gi|31416296|gb|AAP51228.1| BGI-PUP1 [SARS coronavirus GZ01] WHSGVKDY
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] FMRFFTLR
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] MRFFTLRS
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] RFFTLRSI
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] FFTLRSIT
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] FTLRSITA
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] TLRSITAQ
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] LRSITAQP
>gi|30795146|gb|AAP41038.1| Orf3 [SARS coronavirus Tor2] RSITAQPV
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] RSSSRSRC
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] SSSRSRCN
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] SSRSRCNS

>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] SRSRCNSR
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] RSRCNSRN
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] SRCNSRNS
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] RCNSRNST
>gi|30421455|gb|AAP30714.1| putative nucleocapsid protein [SARS coronavirus CUHK-Su10] CNSRNSTP
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] PQGLPNNI
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] QGLPNNIA
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] GLPNNIAS
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] LPNNIASW
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] PNNIASWF
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] NNIASWFT
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] NIASWFTA
>gi|31540949|gb|AAP49024.1| nucleocapsid protein [SARS coronavirus] IASWFTAL
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] HTSPDVDF
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] TSPDVDFG
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] SPDVDFGD
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] PDVDFGDI
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] DVDFGDIS
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] VDFGDISG
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] DFGDISGI
>gi|31581505|gb|AAP33697.1| spike protein S [SARS coronavirus Frankfurt 1] FGDISGIN
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01] RAILTAFL
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01] AILTAFLP
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01] ILTAFLPA

>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
LTAFLPAQ
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
TAFLPAQD
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
AFLPAQDT
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
FLPAQDTW
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
LPAQDTWG
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
NFRVVPSR
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
FRVVPSRD
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
RVVPSRDV
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
VVPSRDVV
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
VPSRDVVR
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
PSRDVVRF
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
SRDVVRFP
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
RDVVRFPN
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
VYAWERKR
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
YAWERKRI
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
AWERKRIS
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
WERKRISN
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
ERKRISNC
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
RKRISNCV
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
KRISNCVA
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
RISNCVAD
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
YRVVVLSY
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
RVVVLSYE
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
VVVLSYEL

>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
VVLSYELL
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
VLSYELLN
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
LSYELLNA
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
SYELLNAP
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
YELLNAPA
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
YKTPTLKD
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
KTPTLKDF
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
TPTLKDFG
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
PTLKDFGG
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
TLKDFGGF
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
LKDFGGFN
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
KDFGGFNF
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
DFGGFNFS
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
ILPDPLKS
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
LPDPLKST
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
PDPLKSTK
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
DPLKSTKR
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
PLKSTKRS
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
LKSTKRSF
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
KSTKRSFI
>gi|31416295|gb|AAP51227.1| spike glycoprotein S [SARS coronavirus GZ01]
STKRSFIE
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2]
ILDISPCA
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2]
LDISPCAF
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2]
DISPCAFG

>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2] ISPCAFGG
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2] SPCAFGGV
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2] PCAFGGVS
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2] CAFGGVSV
>gi|30795145|gb|AAP41037.1| spike glycoprotein [SARS coronavirus Tor2] AFGGVSVI
>gi|30023954|gb|AAP13567.1| putative E2 glycoprotein precursor [SARS coronavirus CUHK-W1] AFSPAQDT
>gi|30023954|gb|AAP13567.1| putative E2 glycoprotein precursor [SARS coronavirus CUHK-W1] FSPAQDTW
>gi|30023954|gb|AAP13567.1| putative E2 glycoprotein precursor [SARS coronavirus CUHK-W1] SPAQDTWG
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] DALCEKAS
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] ALCEKASK
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] LCEKASKY
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] CEKASKYL
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] EKASKYLP
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] KASKYLPI
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] ASKYLPID
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] SKYLPIDK
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] SVIDLLLN
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] LLLNDFVE
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] LLNDFVEI
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] LNDFVEII
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] NDFVEIIK
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] LVDSDLNE
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] VDSDLNEF
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] DSDLNEFV
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] SDLNEFVS
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] DLNEFVSD
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] LNEFVSDA
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] NEFVSDAD
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] EFVSDADS
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] ANYIFWRK
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] NYIFWRKT
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] YIFWRKTN
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] IFWRKTNP
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] FWRKTNPI
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] WRKTNPIQ
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] RKTNPIQL
>gi|31416293|gb|AAP51225.1| orf1ab [SARS coronavirus GZ01] KTNPIQLS
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2] SADASTFF

>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
ADASTFFK
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
DASTFFKR
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
ASTFFKRV
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
STFFKRVC
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
TFFKRVCG
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
FFKRVCGV
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
FKRVCGVS
>gi|30795144|gb|AAP41036.1| replicase 1AB [SARS coronavirus Tor2]
KRVCGVSA
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
ELFYSYAI
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
LFYSYAIH
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
FYSYAIHH
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
YSYAIHHD
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
SYAIHHDK
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
YAIHHDKF
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
AIHHDKFT
>gi|31581504|gb|AAP33696.1| polyprotein 1ab [SARS coronavirus Frankfurt 1]
IHHDKFTD

EXAMPLE 7: PET-SPECIFIC ANTIBODIES ARE HIGHLY SPECIFIC AND HAVE HIGH AFFINITY FOR THEIR PET ANTIGENS

There are numerous PET-specific antibodies that were shown to be highly specific and have high affinity for their respective antigens. The following table lists a few exemplary antibodies showing high affinity (low nanomolar to high picomolar range) for their respective antigens.

Peptide Sequence	Length (aa)	Affinity (K_D in nM)	Reference

GATPEDLNQKLAGN	14	1.4	Cell 91:799,1997
CRGTGSYNRSSFESSSG	17	2.8	JIM 249:253, 2001
NYRAYATEPHAKKKS	15	0.5	EJB 267: 1819, 2000
RYDIEAKVTK	10	3.5	JI 169: 6992, 2002
DRVYIHPF	8	0.5	JIM 254: 147, 2001
PQSDPSVEPPLS	12	16 (a scFv)	NG 21: 163, 2003
YDVPDYAS (HA tag)	8	2	engeneOS
MDYKAFDN (FLAG tag)	8	2.3	engeneOS
HHHHH (HIS tag)	5	25	Novagen

Further more, the table below shows three additional PET-specific antibodies with similar nanomolar-range affinity for the respective antigens:

PET Sequence	Ab name	Affinity (K_D in nM)	Parental Protein
EPAELTDA	P1	5	PSA
YEVQGEVF	C1	31	CRP
GYSIFSYA	C2	200	CRP

These PETs are selected based on the criteria set forth in the instant specification, including nearest neighbor analysis. Listed below are several nearest neighbors of two of the PETs above.

PET			AA Differences
	LS**EPAELTDA**VK		
-	NNP1	D**EP**V**ELTS**A**P**TGHTFS	2
-	NNP2	AG**E**A**AEL**Q**DA**EVESSAK	2
-	NNP3	LQ**EPAEL**V**ES**DGVPK	3
-	NNP4	A**Q****PAEL**V**DS**SGW	3

	- NNP5	GLDPTQLTDALTQR	3
PET		YEVQGEVFTK	AA Differences
	- NNP1	HVEVNGEVFQK	2
	- NNP2	SYEVLGEEFDR	2
	- NNP3	QYAVSGEIFVVDR	3
	- NNP4	VYEEQGEIILK	3
	- NNP5	LYEVRGETYLK	3

PET-specific antibodies are not only high affinity antibodies, but also highly specific antibodies showing little, if any cross-reactivity with other closely related peptide sequences.

For example, Figure 24 shows peptide competition results using the peptide competition assay described in Example 5. The left panel shows that antibody P1, which is specific for the PSA-derived 8-mer PET sequence EPAELTDA, can be effectively competed away by the antigen PET (EPAELTDA), with a half-maximum effective peptide concentration of around 40 nM. However, two of its nearest-neighbor 8-mer PETs found in the human proteome with only two- or three-amino-acid differences, EPVELTSA and DPTQLTDA, are completely ineffective even at 1000 μM (25,000-fold higher concentration). Similarly, the right panel shows that antibody C1, which is specific for the CRP-derived 8-mer PET sequence YEVQGEVF, can be effectively competed away by the antigen PET sequence YEVQGEVF, with a half-maximum effective peptide concentration of around 1 μM. However, two of its nearest-neighbor 8-mer PETs found in the human proteome with only two-amino-acid differences, VEVNGEVF and YEVLGEEF, are completely ineffective even at 1000 μM (at least 1,000-fold higher concentration).

EXAMPLE 8: ANTIBODY CROSS-REACTIVITY: KALLIKREIN Ab's

The kallikreins are a subfamily of the serine protease enzyme family (Bhoola et al., *Pharmacol Rev* 44: 1-80, 1992; Clements J. *The molecular biology of the*

kallikreins and their roles in inflammation. Farmer S. eds. *The kinin system* 1997: 71-97 Academic Press New York). The human kallikrein gene family was, until recently, thought to include only three members: KLK1, which encodes for pancreatic/renal kallikrein (hK1); KLK2, which encodes for human glandular kallikrein 2 (hK2); and KLK3, which encodes for prostate-specific antigen (PSA; hK3) (Riegman et al., *Genomics* 14: 6-11, 1992). The best known of the three classic human kallikreins is PSA, an important biomarker for prostate cancer diagnosis and monitoring. Recently, new serine proteases with high degrees of homology to the three classic kallikreins were cloned. These newly identified serine proteases have now been included in the expanded human kallikrein gene family. The entire human kallikrein gene locus on chromosome 19q13.4 now includes 15 genes, designated KLK1–KLK15; their respective proteins are known as hK1–hK15 (Diamandis et al., *Clin Chem* 46: 1855-1858, 2000).

KLK13, previously known as KLK-L4, is one of the newly identified kallikrein genes. The protein has 47% and 45% sequence identity with PSA and hK2, respectively (Yousef et al., *J Biol Chem* 275: 11891-11898, 2000). At the mRNA level, KLK13 expression is highest in the mammary gland, prostate, testis, and salivary glands (Yousef, supra). Although the function of KLK13 is still unknown, KLK13, like all other members of the human kallikrein family, is predicted to encode a secreted serine protease that is likely present in biological fluids. Given the prominent role of PSA as a cancer biomarker and the recent demonstration that other members of this gene family are also potential cancer biomarkers (Diamandis et al., *Clin Biochem* 33: 369-375, 2000; Luo et al., *Clin Chem* 47: 237-246, 2001; Diamandis et al., *Clin Biochem* 33: 579-583, 2000; Luo et al., *Clin Chim Acta* 7: 806-811, 2001; Diamandis et al., *Cancer Res* 62: 293-300, 2002), hK13 may also have utility as a disease biomarker. In order to develop a suitable method for measuring hK13 protein in biological fluids and tissues with high sensitivity and specificity, and to further investigate the diagnostic and other clinical applications of this protein, Kapadia et al. (*Clinical Chemistry* 49: 77-86, 2003) cloned and expressed the full-length recombinant human KLK13 in a yeast expression system, and raised KLK13-specific monoclonal and polyclonal antibodies. A sandwich-type assay revealed that the KLK13 antibody is quite

specific - recombinant hK1, hK2, hK3, hK4, hK5, hK6, hK7, hK8, hK9, hK10, hK11, hK12, hK14, and hK15 proteins did not produce measurable readings, even at concentrations 1000-fold higher than that of hK13.

However, it should be noted that this type of antibody specificity defined by cross-reactivity to other related proteins, without any epitope information, can frequently be misleading, and thus the data presented in Kapadia et al. should be interpreted with caution. For one thing, unrelated proteins may have higher sequence homology or conformation similarity than family proteins. It may be pure luck that any hK13 antibody does not cross-react with other highly related family members. However, there is no guarantee that the specific epitope recognized by the hK13 antibody does not appear in other proteins, such as an un-identified kallikrein family member, or an alternative splicing form of hK13. Therefore, antibody specificity is better defined by reactivity to peptides most homologous to a selected PET (nearest neighbor peptides). Antibody cross-reactivity is now readily measurable using peptide competitive assays at a wide dynamic range.

On the other hand, in certain situations, detection for the whole protein family or a specific subset of the family are needed. For example, it has already been demonstrated that multiple kallikreins are overexpressed in ovarian carcinoma (reviewed in Yousef and Diamandis, *Minerva Endocrinol* 27: 157-166, 2002). There is experimental evidence that these kallikreins may form a cascade enzymatic pathway similar to the pathways of coagulation and fibrinolysis. Therefore, one single antibody specific for the subset of ovarian carcinoma-associated kallikreins is of particular interest in clinical setting. Lastly, the concentrations of competitors used is limited in Kapadia's assay.

These problems can be readily tackled with the approach of the instant invention. For example, the table below lists a common PET for hK1-hK11 (except hK6 and 7, which have their common PETs), as well as PETs specific for each hK proteins listed. In addition, both the family-specific PET and the protein-specific PET are within the same tryptic fragment.

hK1	HSQPWQVAVYSHGWAHCGGVLVHR

hK2 IVGGWECEQHSQPWQAALYHFSTFQCGGILVHK

hK3 GSQPWQVSLFNGLSFHCAGVLVDR

hK4 NSQPWQVGLFEGTSLR

hK5 HECQPHSQPWQAALFQGQQLLCGGVLVGR

hK8 EDCSPHSQPWQAALVMENELFCSGVLVHR

hK9 VLNTNGTSGFLPGGYTCFPHSQPWQAALLVQGR

hK10 LLEGDECAPHSQPWQVALYER

hK11 PNSQPWQAGLFHLTR

hK6 CVTAGTSCLISGWGSTSSPQLR

Hk7 VMDLPTQEPALGTTCYASGWGSIEPEEFLTPK

By using these family- and individual-specific PET antibodies (or other suitable capture reagents), the same tryptic digestion can be used for a sandwich-type assay that captures all interested tryptic peptides (using the family-specific PET antibodies), followed by selective detection / quantitation of specific family members (using for example, differentially labeled individual-specific antibodies, preferably in a single experiment.

In addition, the same approach may be used to detect the presence of alternative splicing isoforms of any protein. For example, there are three alternative splicing forms of hK15 (* represents trypsin digestion sites):

hK15-V1

R*LNPQVR*PAVLPTR*CPHPGEACVVSGWGLVSHEPGTAGSPR*SQG

hK15-V2 R*LNPQ----------

hK15-V3 R*LNPQGDSGGPLVCGGILQGIVSWGDVPCDNTTK*PGVYTK

Thus, SGWGLVSH is a PET for detecting V1, with the three nearest neighbor peptides being AGWGIVNH, SGWGITNH, and SGWGMVTE. Similarly, WGDVPCDN is a PET for detecting V1, with the three nearest neighbor peptides

being WKDVPCED, WNDAPCDS, and WNDAPCDK.

EXAMPLE 9: DETECTING SERUM PROTEIN LEVELS

Due to the fundamental problems in measuring an antigen which exists in more than one form and/or present in different complexes, it may be difficult to reach a consensus on the level of total a serum protein (such as TGF-b1 protein) in normal human plasma. The instant invention provides a method that efficiently solves these problems.

Figure 21 shows a design for the PET-based assay for standardized serum TGF-beta measurement. The C-terminal monomer for the mature TGF-beta is represented in the top panel as a red bar. The sequences below indicates the PETs specific for each of the 4 TGF-beta isoforms and their respective nearest neighbors. The PET-based assay can be used to specifically detect one of the TGF-beta isoforms, as well as the total amount of all TGF-beta isoforms present in a serum sample.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Md. (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994);

Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, Calif. (1990); Marshak et al., "Strategies for Protein Purification and Characterization—A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

2004292226 05 Jun 2008

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:-

1. A method for detecting the presence or absence of a post-translational modification at a location on a target protein within a sample, comprising:

a. computationally analysing an amino acid sequence of said target protein to identify one or more potential sites for said post-translational modification;

b. computationally identifying an amino acid sequence of at least one fragment of said target protein, said fragment predictably resulting from a treatment of said target protein within said sample, and said fragment comprising at least one of said potential post-translational modification sites and, separate therefrom, a PET (proteome epitope tag) unique to said fragment within said sample;

c. generating a capture agent that specifically binds said PET separate from said post-translational modification on said fragment, and immobilizing said capture agent to a support;

d. subjecting said sample to the treatment to produce said fragment, rendering said fragment soluble in solution, and contacting said fragment with said capture agent to bind said fragment, at said PET, to said capture agent;

e. detecting, on said fragment bound to said capture agent, the presence or absence of said post-translational modification by using a secondary capture agent specific for said post-translational modification separate from said PET on said fragment, wherein said secondary capture agent is labelled by a detectable moiety.

2. The method of claim 1, wherein said post-translational modification is acetylation, amidation, deamidation, prenylation, formylation, glycosylation, hydroxylation, methylation, myristoylation, phosphorylation, ubiquitination, ribosylation or sulphation.

3. The method of claim 2, wherein said post-translation modification is phosphorylation on tyrosine, serine or threonine.

4. The method of any one of claims 1 to 3, wherein said step of computationally identifying an amino acid sequence includes a Nearest-Neighbor amino acid Analysis that identifies said PET based on criteria that also include one or more of pI, charge, steric, solubility, hydrophobicity, polarity and solvent exposed area.

5. The method of claim 4, further comprising determining the specificity of said

2004292226 05 Jun 2008

capture agent generated in (c) against one or more nearest neighbor(s), if any, of said PET.

6. The method of claim 5, wherein peptide competition assay is used in determining the specificity of said capture agent generated in (c) against said nearest neighbor(s) of said PET.

7. The method of any one of claims 1 to 6, wherein said step of computationally identifying an amino acid sequence includes a solubility analysis that identifies a said PET that is predicted to have at least a threshold solubility under a designated solution condition.

8. The method of any one of claims 1 to 7, wherein the length of said amino acid sequence of at least one fragment of said target protein is selected from 15-20 amino acids, 20-25 amino acids, 25-30 amino acids, or 30-40 amino acids.

9. The method of any one of claims 1 to 8, wherein said capture agent is a full-length antibody, or a functional antibody fragment selected from: an Fab fragment, an $F(ab')_2$ fragment, an Fd fragment, an Fv fragment, a dAb fragment, an isolated complementarity determining region (CDR), a single chain antibody (scFv), or derivative thereof.

10. The method of any one of claims 1 to 8, wherein said capture agent is selected from nucleotides; nucleic acids; PNA (peptide nucleic acids); proteins; peptides; carbohydrates; artificial polymers; or small organic molecules.

11. The method of any one of claims 1 to 10, wherein said treatment is denaturation and/or fragmentation of said sample by a protease, a chemical agent, physical shearing, or sonication.

12. The method of claim 11, wherein said denaturation is thermo-denaturation or chemical denaturation.

13. The method of claim 12, wherein said thermo-denaturation is followed by or concurrent with proteolysis using thermo-stable proteases.

14. The method of claim 11, wherein said fragmentation is carried out by a protease selected from trypsin, chymotrypsin, pepsin, papain, carboxypeptidase, calpain, subtilisin, gluc-C, endo lys-C, or proteinase K.

15. The method of any one of claims 1 to 14, wherein said sample is a body fluid selected from: saliva, mucous, sweat, whole blood, serum, urine, amniotic fluid, genital fluid, fecal material, marrow, plasma, spinal fluid, pericardial fluid, gastric fluid, abdominal fluid, peritoneal fluid, pleural fluid, synovial fluid, cyst fluid, cerebrospinal

2004292226 05 Jun 2008

fluid, lung lavage fluid, lymphatic fluid, tears, prostatitc fluid, extraction from other body parts, or secretion from other glands; or from supernatant, whole cell lysate, or cell fraction obtained by lysis and fractionation of cellular material, extract or fraction of cells obtained directly from a biological entity or cells grown in an artificial environment.

16. The method of claim 15, wherein said sample is obtained from human, mouse, rat, frog, fish, fly, nematode, fission or budding yeast, or plant.

17. The method of claim 15 or claim 16, wherein said sample comprises membrane bound proteins.

18. The method of any one of claims 1 to 17, wherein said treatment is carried out under conditions to preserve said post-translational modification.

19. The method of any one of claims 1 to 18, wherein said capture agent is optimized for selectivity for said PET under denaturing conditions.

20. The method of any one of claims 1 to 19, wherein said secondary capture agent is labeled by a detectable moiety selected from: an enzyme, a fluorescent label, a stainable dye, a chemilumninescent compound, a colloidal particle, a radioactive isotope, a near-infrared dye, a DNA dendrimer, a water-soluble quantum dot, a latex bead, a selenium particle, or a europium nanoparticle.

21. The method of claim 20, wherein said post-translational modification is phosphorylation, and said secondary capture agent is a labeled secondary antibody specific for phosphorylated tyrosine, phosphorylated serine, or phosphorylated threonine.

22. The method of claim 21, wherein said post-translational modification is phosphorylation, and said secondary capture agent is a labeled secondary antibody specific for phosphorylated tyrosine.

23. The method of claim 21 or claim 22, wherein said secondary antibody is labeled by an enzyme or a fluorescent group.

24. The method of claim 23, wherein said secondary antibody is labeled by a fluorescent group.

25. The method of any one of claims 1 to 24, wherein said sample contains billion molar excess of unrelated proteins or fragments thereof relative to said fragment.

26. The method of any one of claims 1 to 25, further comprising qauntitating the amount of said fragment bound to said capture agent.

27. The method of any one of claims 1 to 26, wherein step (c) is conducted by

2004292226 05 Jun 2008

immunizing an animal with an antigen comprising said PET sequence.

28. The method of claim 27, wherein the N- or C-terminus, or both, of said PET sequence are blocked to eliminate free N- or C-terminus, or both.

29. The method of claim 28, wherein the N- or C-terminus of said PET sequence are blocked by fusing the PET sequence to a heterologous carrier polypeptide, or blocked by a small chemical group.

30. The method of claim 3, wherein said post-translational modification is phosphorylation on tyrosine.

31. The method of claim 9, wherein said capture agent is a full-length antibody.

32. The method of claim 20, wherein said secondary capture agent is labeled by a fluorescent label.

33. A method for detecting the presence or absence of a post-translational modification at a location on a target protein within a sample, substantially as herein described with reference to any one of the Examples.

Interleukin-8 receptor A: SeqID: ENSP00000295683

Sequence and pentamer URS distribution

- MSNITDPQMWDFDDLNFTGMPPADEDYSPCMLETETLNKYVVIIAYALVFLLSLLGNSLVMLVILYSR

  VGRSVTDVYLLNLALADLLFALTLPIWAASKVNGWIFGTFLCKVVSLLKEVNFYSGILLLACISVDRY

  LAIVHATRTLTQKRHLVKFVCLGCWGLSMNLSLPFFLFRQAYHPNNSSPVCYEVLGNDTAKWRMVLRI

  LPHTFGFIVPLFVMLFCYGFTLRTLFKAHMGQKHRAMRVIFAVVLIFLLCWLPYNLVLLADTLMRTQV

  IQESCERRNNIGRALDATEILGFLHSCLNPIIYAFIGQNFRHGFLKILAMHGLVSKEFLARHRVTSYT

  SSSVNVSSNL

- ENSP00000295683: list of pentamer URS peptides (total number: 35) AHMGQ AKWRM AYHPN CERRN CLGCW CYGFT DLNFT DPQMW DTAKW FCYGF FKAHM GQNFR HTFGF IWAAS KWRMV KYVVI MGQKH MPPAD NFTGM NGWIF PQMWD PYNLV QKLRA QMWDF QNFRH TAKWR TDPQM VCLGC VGYEV VMLFC VNGWI WDFDD WIFGT WRMVL YSPCM
- ____: URS Penta-peptide

Figure 1

Histamine H1 receptor: SeqID: ENSP00000273023

Sequence and pentamer URS distribution

MSLPNSSCLLEDK|MCEGNK|TTMASPQLMPLVVVLSTICLVTVGLNLLVLYAVR|SER|K|LHTVGN

LYIVSLSVADLIVGAVVMPMNILYLLMSK|WSLGR|PLCLFWLSMDYVASTASIFSVFLLCIDR|YR|

SVQQPLR|YLK|YR|TK|TR|ASATILGAWFLSFLWVIPILGWNHFMQQTSVR|R|EDK|CETDFYDV

TWFK|VMTAIINFYLPTLLMLWFYAK|IYK|AVR|QHCQHR|ELINR|SLPSFSEIK|LR|PENPK|G

DAK|K|PGK|ESPWEVLK|R|K|PK|DAGGGSVLK|SPSQTPK|EMK|SPVVFSQEDDR|EVDK|LYC

FPLDIVHMQAAAEGSSR|DYVAVNR|SHCQLK|TDEQGLNTHGASEISEDQMLGDSQSFSR|TDSDTT

TETAPGK|GK|LR|SGSNTGLDYIK|FTWK|R|LR|SHSR|QYVSGLHMNR|ER|K|AAK|QLGFIMA

AFILCWIPYFIFFMVIAFCK|NCCNEHLHMFTIWLGYINSTLNPLIYPLCNENFK|K|TFK|R|ILHI

R|S

- URS Penta-peptide: AIINF CEGNK CNENF CWIPY DFYDV DQMLG FILCI FLWVI FMQQT FWLSM GWNHF HLHMF HMFTI HTVGN IFFMV ILCWI LLMSK LWFYA MCEGN MDYVA MFTIW MLWFY NCCNE NHFMQ PLCLF SPWEV VIAFC WEYAK WNHFM WVIPI YCFPL
- |: trypsin cutting site, _____: URS Penta-peptide

Figure 2

Figure 3

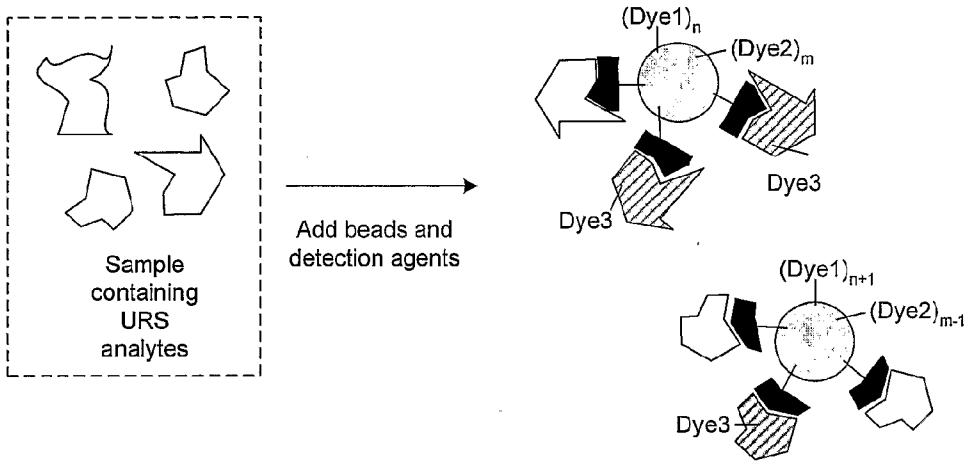

Sample containing URS analytes
Add beads and detection agents
(Dye1)n
(Dye2)m
Dye3
Dye3
(Dye1)n+1
(Dye2)m-1
Dye3

Figure 4

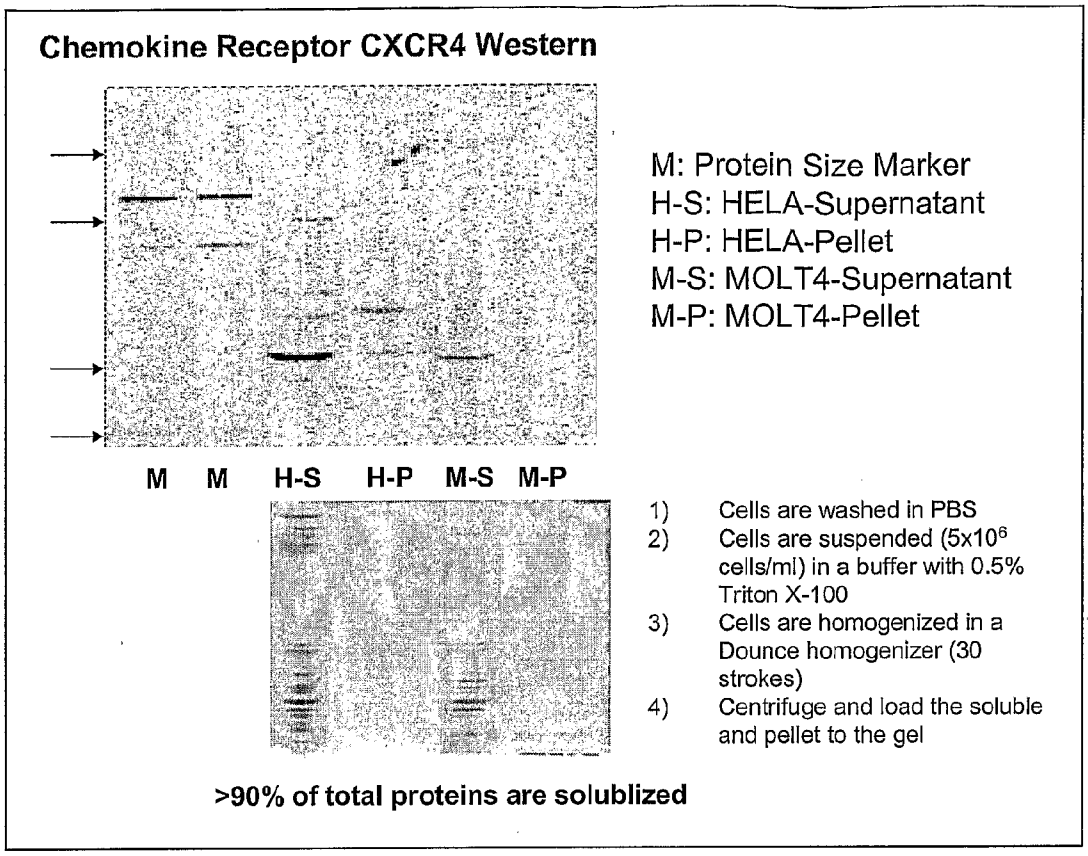

Chemokine Receptor CXCR4 Western
M: Protein Size Marker
H-S: HELA-Supernatant
H-P: HELA-Pellet
M-S: MOLT4-Supernatant
M-P: MOLT4-Pellet
M M H-S H-P M-S M-P
1) Cells are washed in PBS
2) Cells are suspended (5x10^6 cells/ml) in a buffer with 0.5% Triton X-100
3) Cells are homogenized in a Dounce homogenizer (30 strokes)
4) Centrifuge and load the soluble and pellet to the gel
>90% of total proteins are solublized

Figure 5

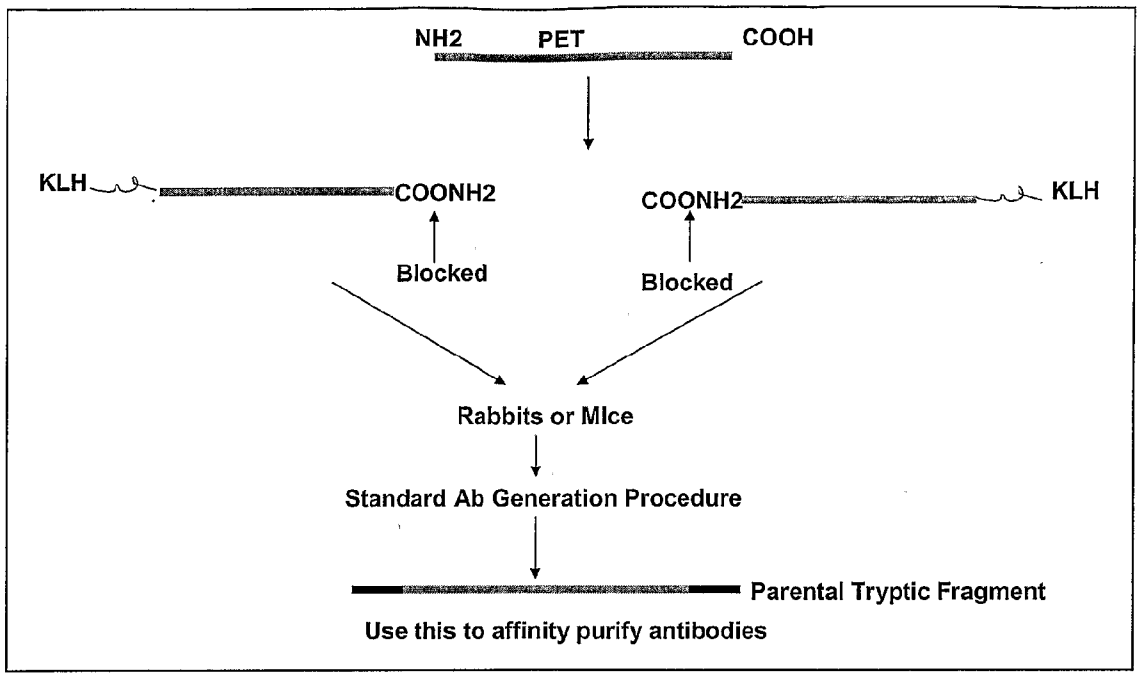

NH2
PET
COOH
KLH
COONH2
COONH2
KLH
Blocked
Blocked
Rabbits or Mice
Standard Ab Generation Procedure
Parental Tryptic Fragment
Use this to affinity purify antibodies

Figure 6

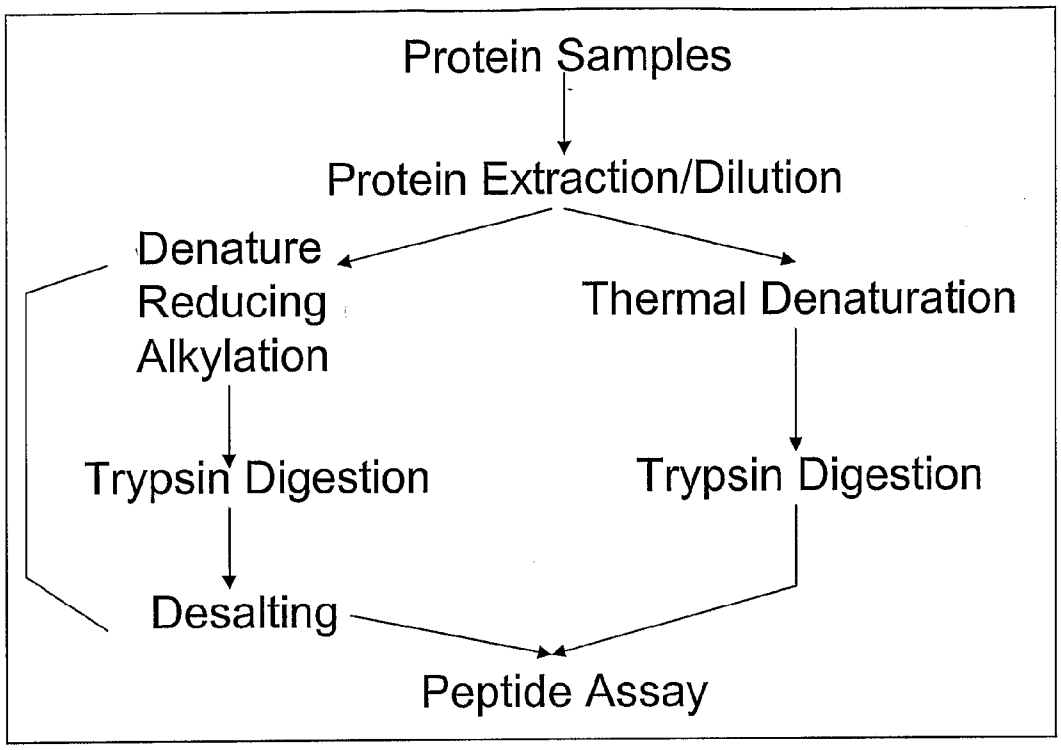

Protein Samples
Protein Extraction/Dilution
Denature
Reducing
Alkylation
Thermal Denaturation
Trypsin Digestion
Trypsin Digestion
Desalting
Peptide Assay

Figure 7

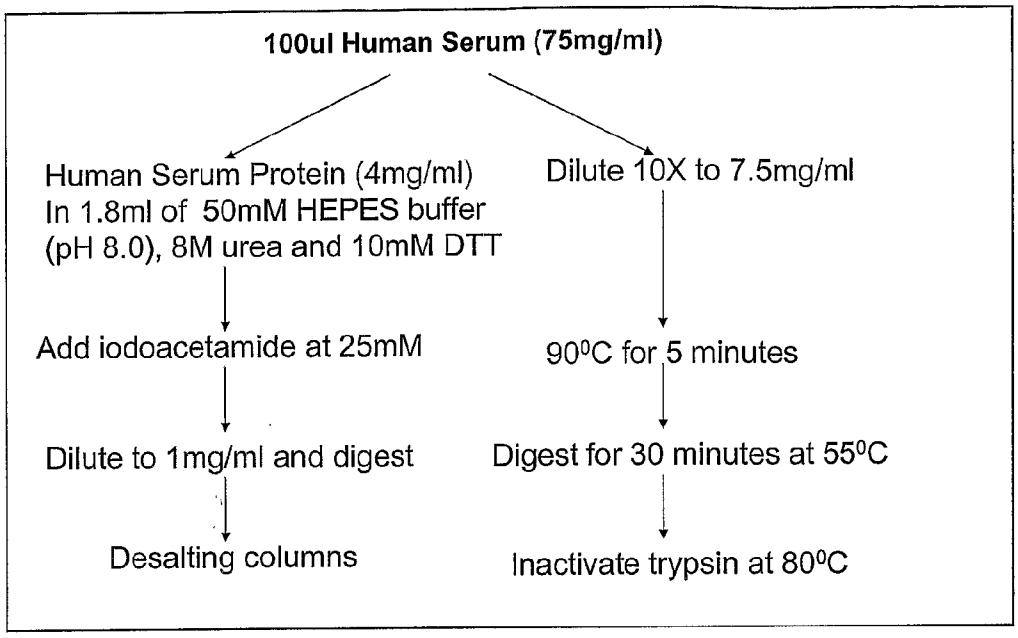

100ul Human Serum (75mg/ml)
Human Serum Protein (4mg/ml)
In 1.8ml of 50mM HEPES buffer
(pH 8.0), 8M urea and 10mM DTT
Add iodoacetamide at 25mM
Dilute to 1mg/ml and digest
Desalting columns
Dilute 10X to 7.5mg/ml
90°C for 5 minutes
Digest for 30 minutes at 55°C
Inactivate trypsin at 80°C

Figure 8

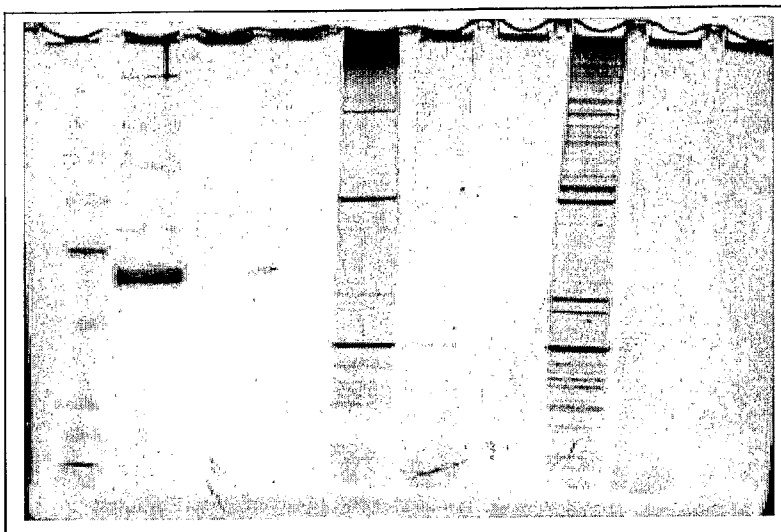

1 2 3 4 5 6 7 8 9 10
M S S-T S-TC M M-T M-C H H-T H-C
M: Marker
S: Serum
S-T: Serum-Thermal
S-C: Serum-Chemical
M: MOLT4-Thermal
M: MOLT4-Chemical
H: HELA
H-T: HELA-Thermal
H-C: HELA-Chemical

Figure 9

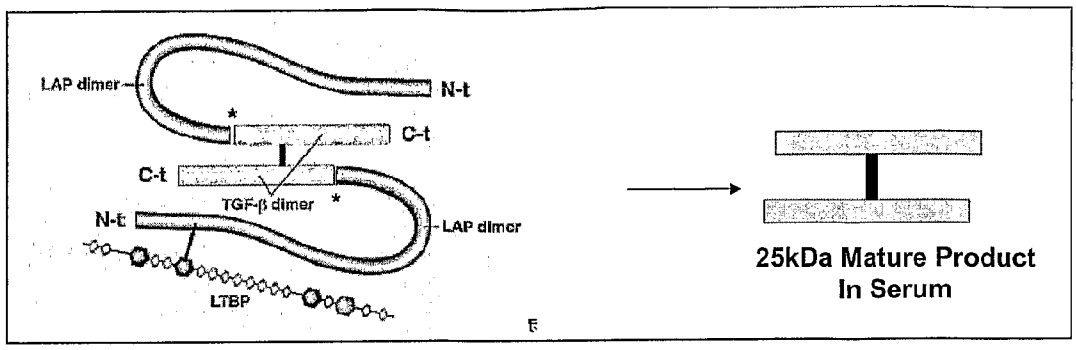

LAP dimer
N-t
C-t
C-t
TGF-β dimer
N-t
LAP dimer
LTBP
25kDa Mature Product
In Serum

Figure 10

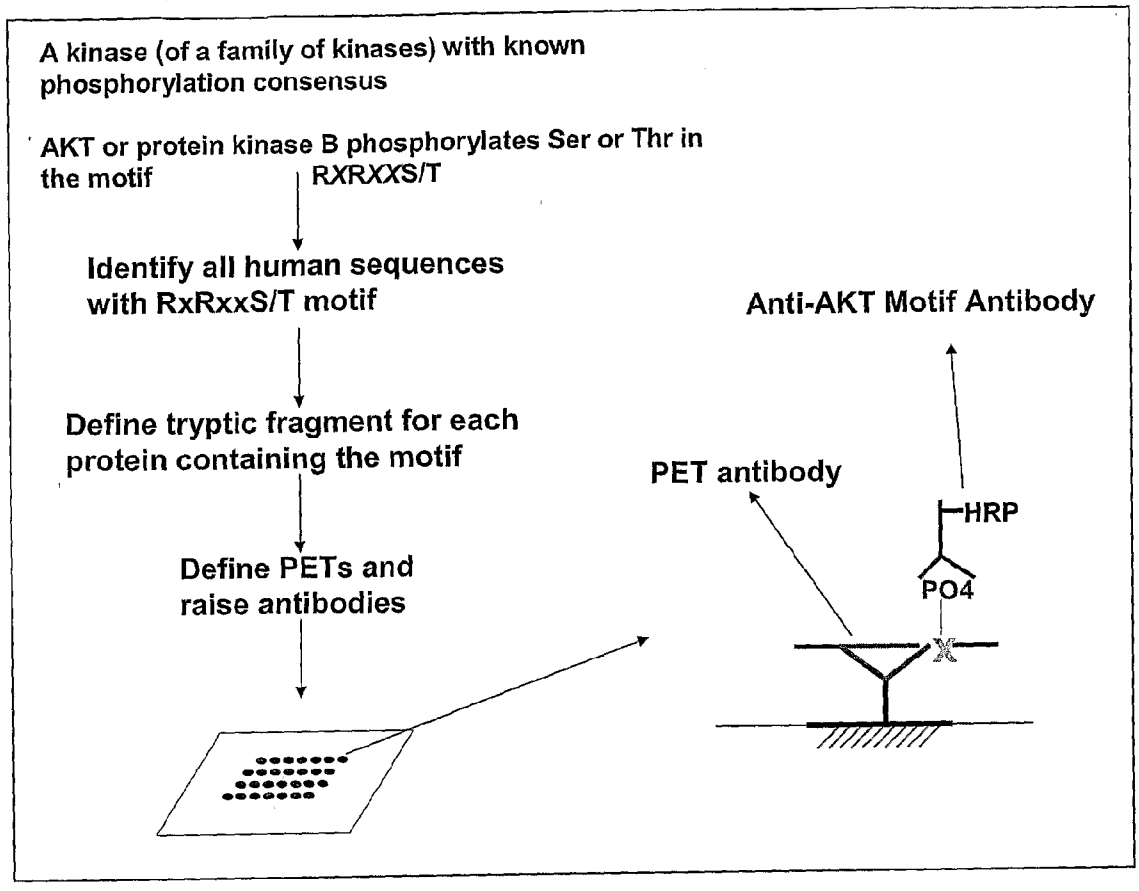

A kinase (of a family of kinases) with known phosphorylation consensus
AKT or protein kinase B phosphorylates Ser or Thr in the motif RXRXXS/T
Identify all human sequences with RxRxxS/T motif
Define tryptic fragment for each protein containing the motif
Define PETs and raise antibodies
Anti-AKT Motif Antibody
PET antibody
HRP
PO4
X

Figure 11

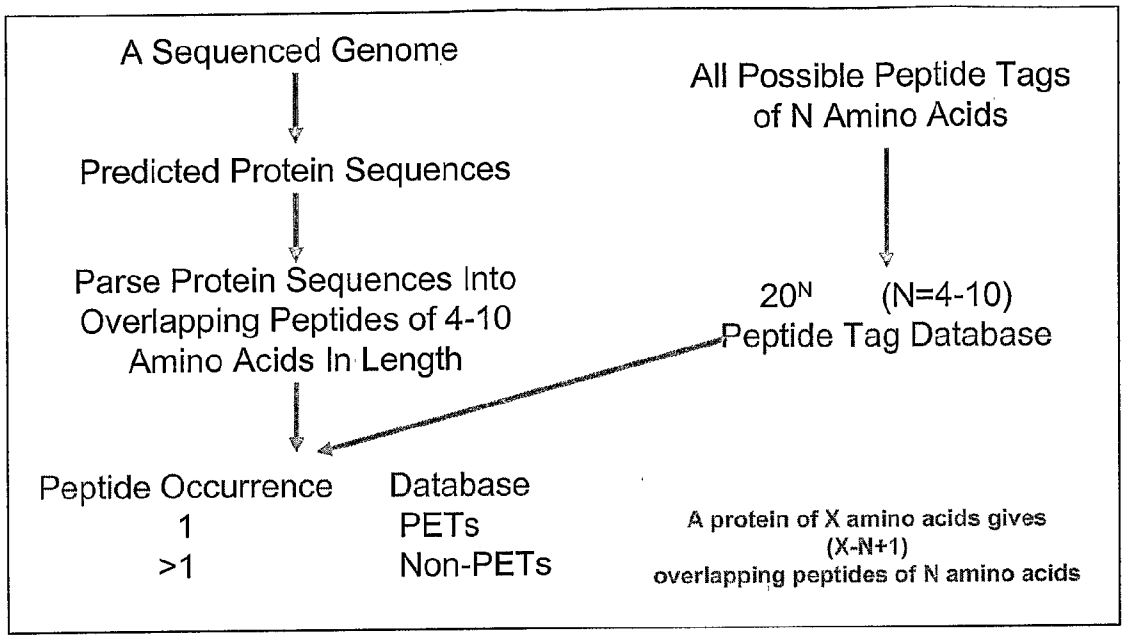

A Sequenced Genome
Predicted Protein Sequences
Parse Protein Sequences Into Overlapping Peptides of 4-10 Amino Acids In Length
Peptide Occurrence
1
>1
Database
PETs
Non-PETs
All Possible Peptide Tags of N Amino Acids
20N (N=4-10)
Peptide Tag Database
A protein of X amino acids gives (X-N+1) overlapping peptides of N amino acids

Figure 12

Tag Length (amino acids)	All Possible Tags	Total PETs	Total Non-PETs (non-redundant)
4	160,000	745	158,862
5	$3.2x10^6$	560,309	1,684,684
6	$6.4x10^7$	4,609,172	2,350,532
7	$1.28x10^9$	6,652,224	1,848,908
8	$2.56x10^{10}$	7,018,340	1,744,029
9	$5.12x10^{11}$	7,138,933	1,714,971
10	$1.02x10^{13}$	7,216,090	1,695,512

29,076 human protein sequences analyzed
~12M overlapping 4-10mr peptides

% of Human Peptide Tags in Total Tag Space

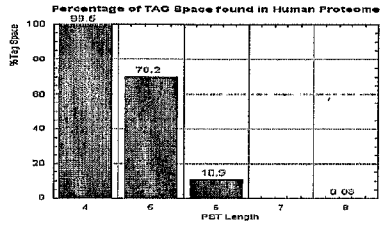


Figure 13

Length of PET (amino acids)	Proteins with PETs	Proteome Coverage
4	684	2.35%
5	23,446	80.64%
6	26,069	89.66%
7	26,184	90.05%
8	26,216	90.16%
9	26,238	90.24%
10	26,250	90.28%

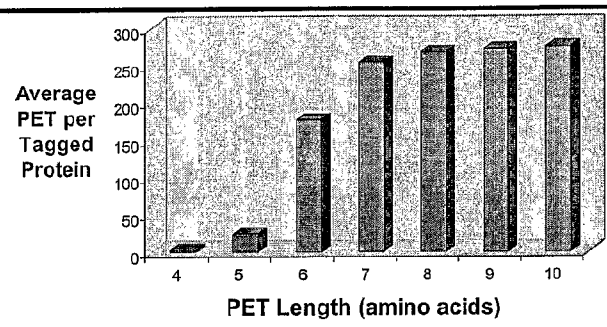


Figure 14

PET Length (Amino Acids)	Average PETs/Tagged Protein No Cleavage	Trypsin Cleavage
5	24	16
6	177	98
8	268	129

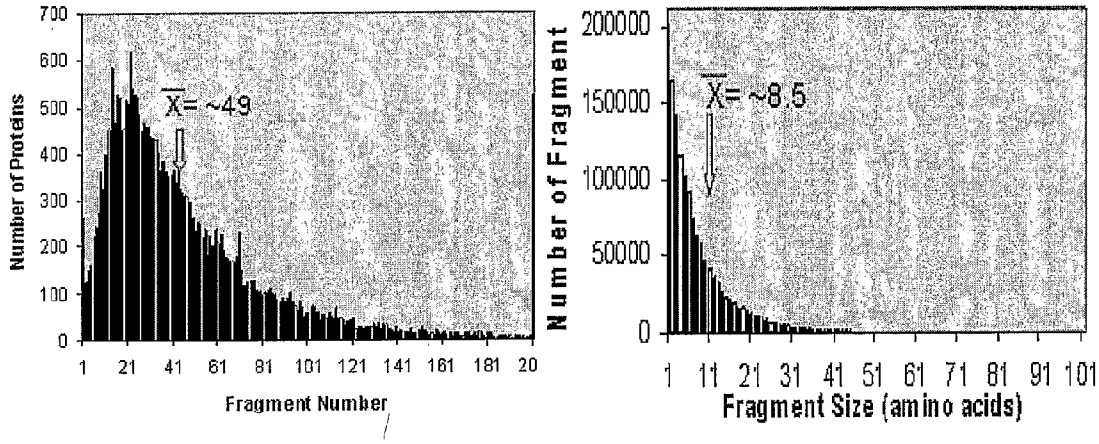


Figure 15

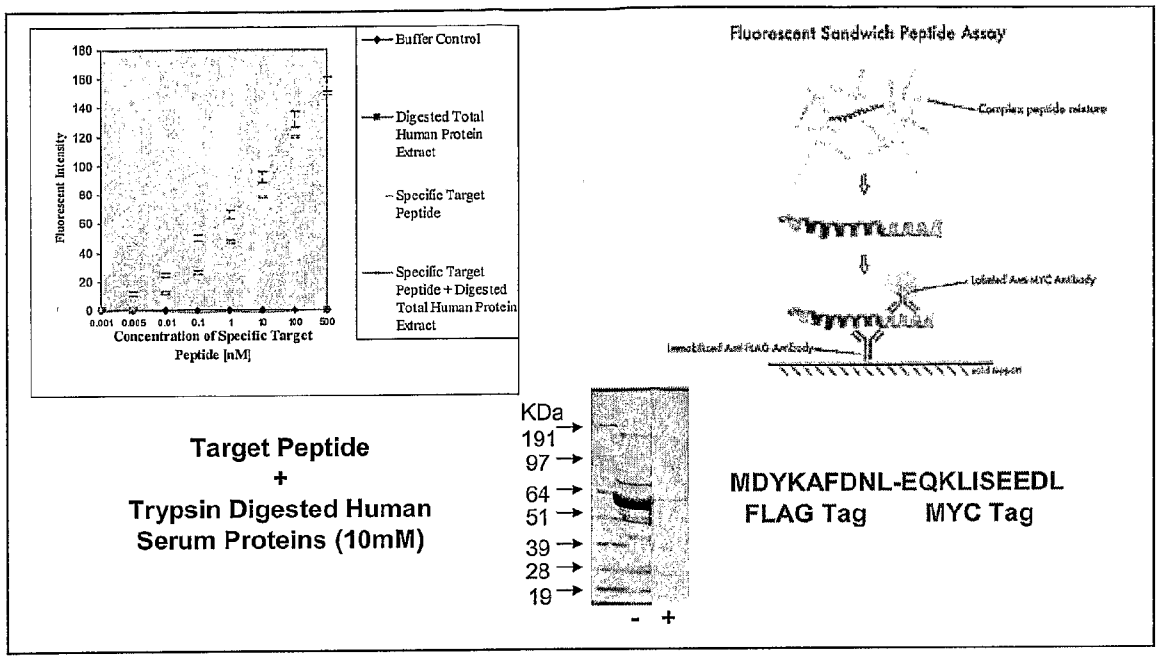

Buffer Control
Digested Total Human Protein Extract
Specific Target Peptide
Specific Target Peptide + Digested Total Human Protein Extract
Fluorescent Intensity
Concentration of Specific Target Peptide [nM]
Fluorescent Sandwich Peptide Assay
Complex peptide mixture
Target Peptide
+
Trypsin Digested Human Serum Proteins (10mM)
KDa
191
97
64
51
39
28
19
- +
MDYKAFDNL-EQKLISEEDL
FLAG Tag MYC Tag

Figure 16

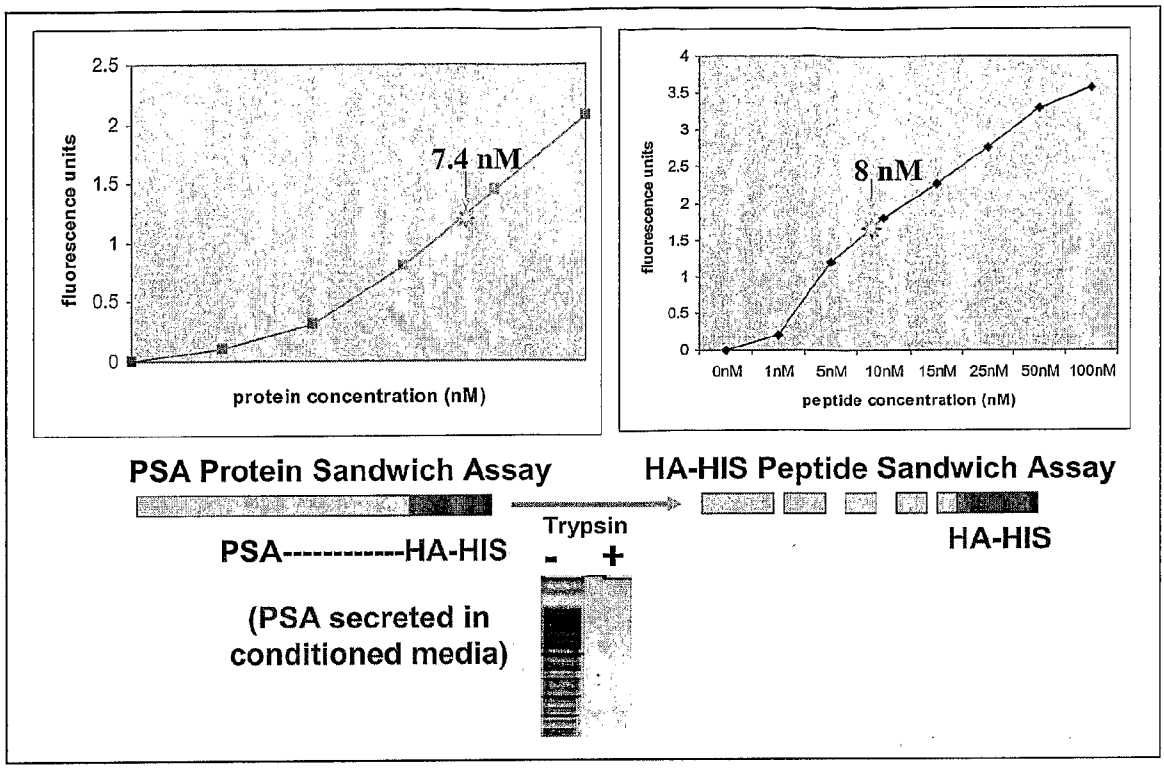

fluorescence units
protein concentration (nM)
7.4 nM
fluorescence units
peptide concentration (nM)
8 nM
0nM 1nM 5nM 10nM 15nM 25nM 50nM 100nM
PSA Protein Sandwich Assay
HA-HIS Peptide Sandwich Assay
Trypsin
PSA----------HA-HIS
HA-HIS
(PSA secreted in conditioned media)
- +

Figure 17

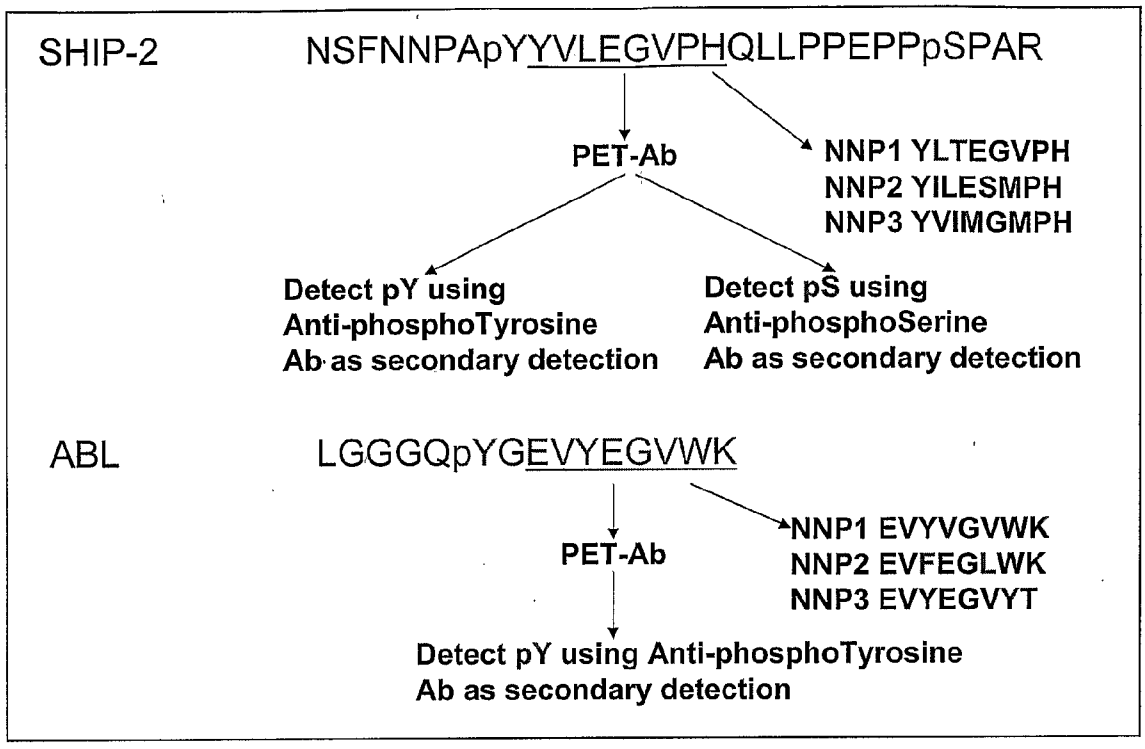
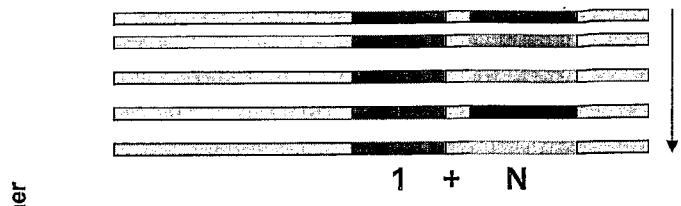
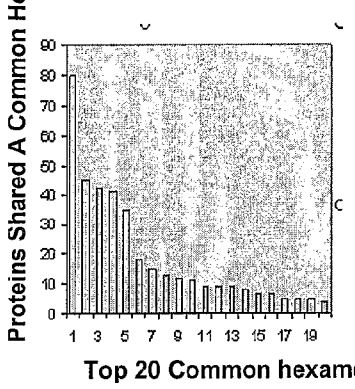

SHIP-2
NSFNNPApYYVLEGVPHQLLPPEPPpSPAR
PET-Ab
NNP1 YLTEGVPH
NNP2 YILESMPH
NNP3 YVIMGMPH
Detect pY using Anti-phosphoTyrosine Ab as secondary detection
Detect pS using Anti-phosphoSerine Ab as secondary detection
ABL
LGGGQpYGEVYEGVWK
PET-Ab
NNP1 EVYVGVWK
NNP2 EVFEGLWK
NNP3 EVYEGVYT
Detect pY using Anti-phosphoTyrosine Ab as secondary detection

Figure 18





Size of the Tag	Number To Represent All Kinases
5 aa	88
6 aa	122

Top 20 common hexamers are shared by 72% of 520 kinases

Figure 19

Protein	Parental Tryptic Peptide	Also Detect
BRAF	WSGSHQFEQLSGSILWMAPEVIR*	
DLK	MSFAGTVAWMAPEVIR	
GCK	SFIGTPYWMAPEVAAVER	
HH498	WMAPEVFTQCTR	
HPK1	LSFIGTPYWMAPEVAAVALK	
LOK	DSFIGTPYWMAPEVVMCETMK	KSH1,2,HPK1, SLK
LZK	MSFAGTVAWMAPEVIR	
MAP3K	SMHGTPYWMAPEVINESGYGR	
MST1	NTVIGTPFWMAPEVIQEIGYNCVR	MST2
MST4	NTFVGTPFWMAPEVIQQSAYDSK	
MYO3A	NTSVGTPFWMAPEVIACEQQLDTTYDAR	MY03B
MYO3B	NTSVGTPFWMAPEVIACEQQYDSSYDAR	MY03A
ZC1/HGK	NTFIGTPYWMAPEVIACDENPDATYDYR	ZC2,ZC3
OSR1	TFVGTPCWMAPEVMEQVR	
PAK1	STMVGTPYWMAPEVVTR	PAK2,3
PAK5	SLVGTPYWMAPEVISR	PAK6
RAF1	WSGSQQVEQPTGSVLWMAPEVIR	
STLK3	TFVGTPCWMAPEVMEQVR	
TAO1	ASMASPANSFVGTPYWMAPEVILAMDEGQYDGK	TAO3
TAO2	ASIMAPANSFVGTPYWMAPEVILAMDEGQYDGK	
TESK1	EPLAVVGSPYWMAPEVLR	
ZAK	TTHMSLVGTFPWMAPEVIQSLR	

BLUE = PET
RED = Commone Epitope

Figure 20

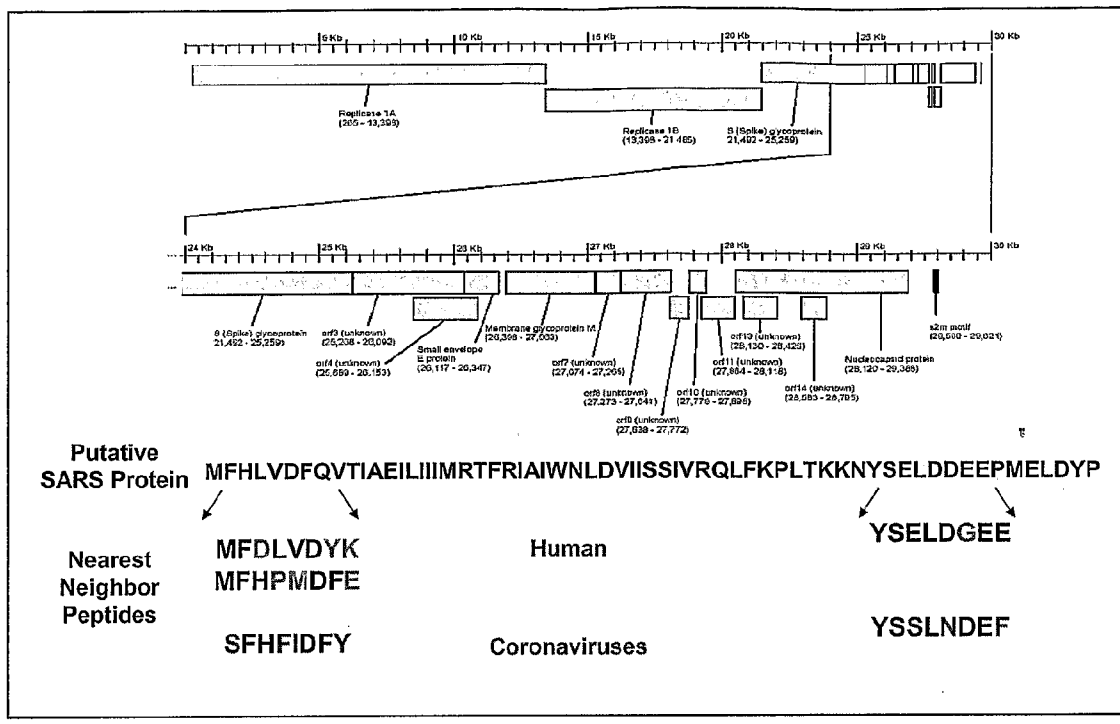
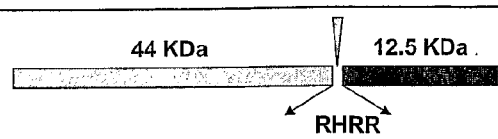

Putative SARS Protein
MFHLVDFQVTIAEILIIIMRTFRIAIWNLDVIISSIVRQLFKPLTKKNYSELDDEEPMELDYP
Nearest Neighbor Peptides
MFDLVDYK
MFHPMDFE
SFHFIDFY
Human
Coronaviruses
YSELDGEE
YSSLNDEF

Figure 21



TGFß-1	**RHRR***AL.. K*VLALY**NQHNPGAS**AAPCCVPQAL**EPLPIVYY**VGR*K
	NNP1 NHHSPG**G**S EPLT**I**LYY
	NNP2 **Q**QHNP**A**A**N** **D**PLP**V**RYY
	NNP3 N**K**H**G**PG**V**S EPLP**S**QYY
TGFß-2	**RKKR***AL...R*VLSLY**NTINPEAS**ASPCCV**SQDLEPLT**ILYYIGK*T
	NNP1 NT**L**NPEAS **P**QDLEPLT
	NNP2 N**K**L**D**PEAS S**E**DLEPL**A**
	NNP3 NT**A**NPE**R**S SQDL**D**P**M**A
TGFß-3	**RKKR***AL...R***SADTTHST**VLGLY**NTLNPEAS**ASPCCVPQDLEPLTILYYVGR*T
	NNP1 SA**HS**THST NTINPEAS
	NNP2 S**S**DTTH**AS** N**K**L**D**PEAS
	NNP3 **AAEA**THST NT**A**NPE**R**S
TGFß-4	MK*WAK***NWVLEPPG**FLAYECVGTCQ**QPPEALAF**NWPFLGPR*Q
	NNP1 NW**AVD**PPG QPPEA**FG**F
	NNP2 HWV**VSP**PG **K**PPEALA**M**
	NNP3 NWV**RL**PPG QPPEA**KK**F

Figure 22

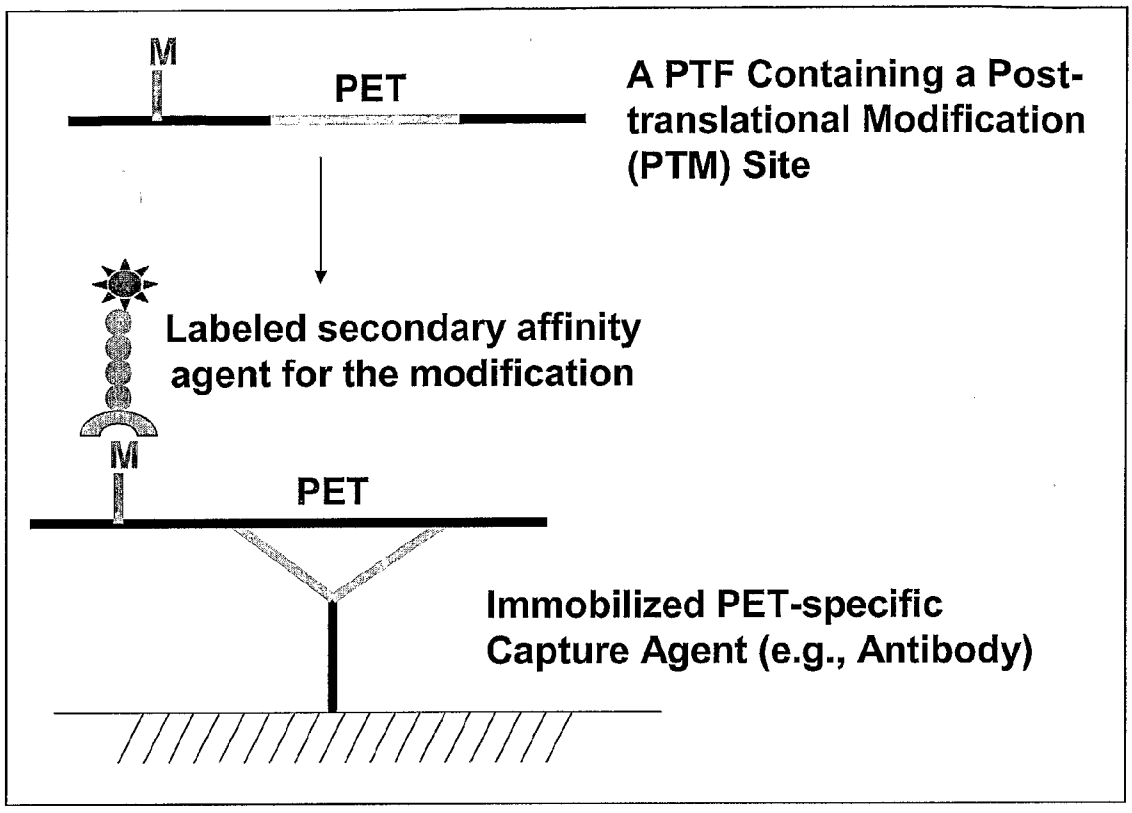

M
PET
A PTF Containing a Post-translational Modification (PTM) Site
Labeled secondary affinity agent for the modification
M
PET
Immobilized PET-specific Capture Agent (e.g., Antibody)

Figure 23

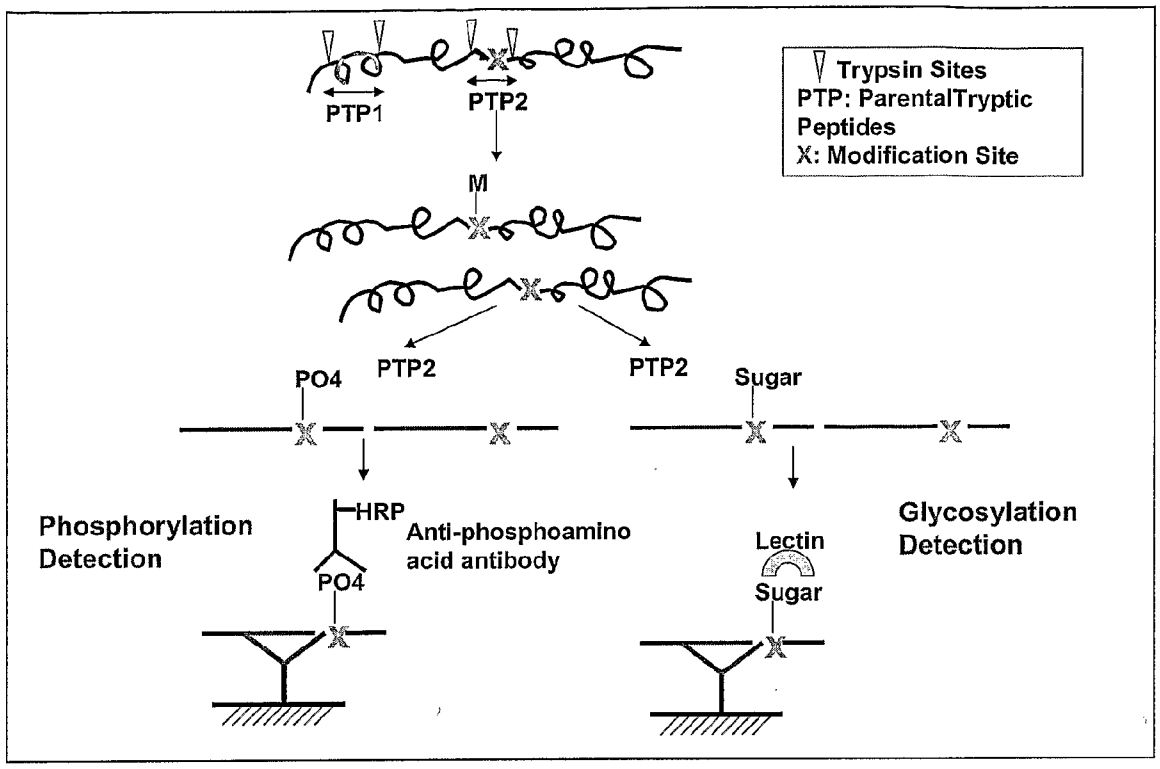
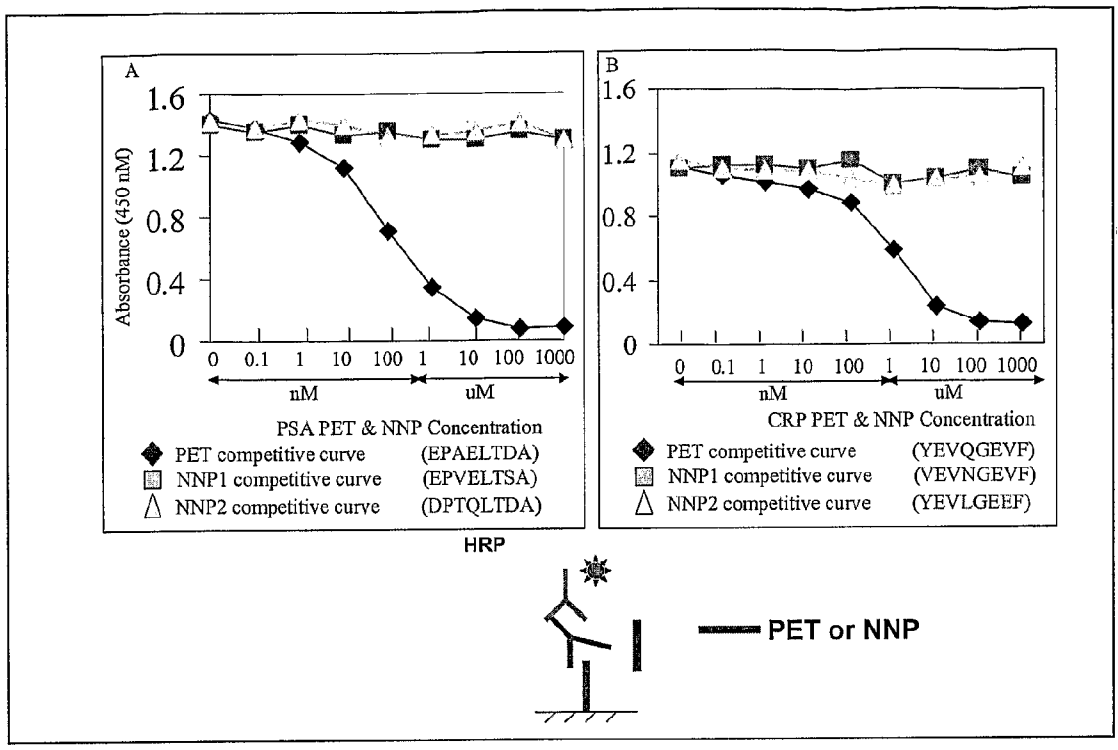

Trypsin Sites
PTP: ParentalTryptic Peptides
X: Modification Site
PTP1
PTP2
M
PO4
PTP2
PTP2
Sugar
Phosphorylation Detection
HRP
Anti-phosphoamino acid antibody
PO4
Lectin
Sugar
Glycosylation Detection

Figure 24


A
1.6
1.2
0.8
0.4
0
Absorbance (450 nM)
0 0.1 1 10 100 1 10 100 1000
nM uM
PSA PET & NNP Concentration
PET competitive curve (EPAELTDA)
NNP1 competitive curve (EPVELTSA)
NNP2 competitive curve (DPTQLTDA)
B
1.6
1.2
0.8
0.4
0
0 0.1 1 10 100 1 10 100 1000
nM uM
CRP PET & NNP Concentration
PET competitive curve (YEVQGEVF)
NNP1 competitive curve (VEVNGEVF)
NNP2 competitive curve (YEVLGEEF)
HRP
PET or NNP