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

(19) World Intellectual Property Organization
International Bureau

PCT

(43) International Publication Date
26 August 2010 (26.08.2010)

(10) International Publication Number
WO 2010/094300 A1

(51) International Patent Classification:
G06F 19/00 (2006.01)

(21) International Application Number:
PCT/EP2009/001230

(22) International Filing Date:
20 February 2009 (20.02.2009)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **NMI NATURWISSENSCHAFTLICHES UND MEDIZINISCHES INSTITUT AN DER UNIVERSITÄT TÜBINGEN** [DE/DE]; Markwiesenstrasse 55, 72770 Reutlingen (DE).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **JOOS, Thomas** [DE/DE]; Eichhaldenstrasse 18, 72074 Tübingen (DE). **TEMPLIN, Markus** [DE/DE]; Käthe-Kollwitz-Str. 1, 72074 Tübingen (DE). **STOLL, Dieter** [DE/DE]; Wilhelm-Röntgen-Strasse 24, 72116 Mössingen (DE). **POETZ, Oliver** [DE/DE]; Karl-Adam-Strasse 13, 72076 Tübingen (DE). **ZELL, Andreas** [DE/DE]; Griessstr. 28/1, 72827 Wannweil (DE). **PLANATSCHER, Hannes** [DE/DE]; Neckar-Alb-Str. 5, 72127 Kusterdingen (DE). **SUPPER, Jochen** [DE/DE]; Loichingerstrasse 3, 81245 München (DE).

(74) **Agent**: **OTTEN, Hajo**; Witte, Weller & Partner, Postfach 10 54 62, 70047 Stuttgart (DE).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, 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 (Art. 21(3))*

(54) **Title**: A METHOD FOR DETERMINING *IN SILICO*- A SET OF SELECTED TARGET EPITOPES

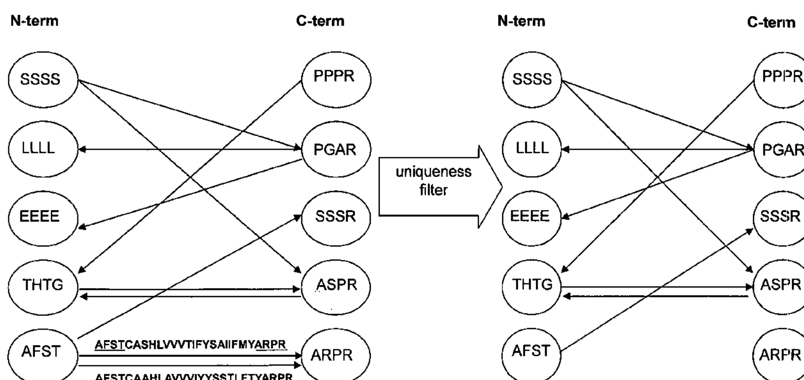

Fig. 1

(57) **Abstract**: A method for determining *in silico* a set of selected target epitopes suited for the detection of at least one protein of a set of proteins of interest within a pool of sample proteins, comprises the steps of: providing a first list containing the sequences of said sample proteins; providing a second list containing the sequences of said proteins of interest; applying an *in silico* fragmentation to the sample proteins in said first list to generate a third list containing sequences of possible target peptides having possible target epitopes at their free N- and C-terminal ends, which possible target epitopes each consisting of three to five N- or C-terminal amino acids; generating a fourth list of sequences of valid target peptides by removing from said third list such possible target peptides that are less suited for said protein of interest identification; selecting from said possible target epitopes on said valid target peptides in said fourth list a set of selected target epitopes, said set of selected target epitopes containing at least one target epitope for each protein of interest in said second list, while at the same time containing an as small as possible number of target epitopes.

WO 2010/094300 A1

A METHOD FOR DETERMINING *IN SILICO*- A SET OF SELECTED TARGET EPITOPES

Field of the Invention

The present invention relates to a method for determining in silico a set of selected target epitopes suited for the identification of at least one protein of a set of proteins of interest within a pool of sample proteins.

Further the present invention relates to a method for detecting proteins of interest in a biological sample, to a set of selected target epitopes being suited for identifying proteins, to binding molecules that bind to target epitopes contained in such a set, and to the use of such binding molecules.

Related Prior Art

Protein expression profiling is becoming a key concept of clinical analytics and diagnosis. In the past decades, many examples have shown that human diseases are closely associated to changes in the activity of particular sets of genes and consecutively to changes in protein levels.

Therefore, screening assays, that allow identifying the presence or absence and respectively the level of different proteins contained in a proteome, especially of sets of biomarkers, which are related to a certain disease, is a powerful means to specifically diagnose certain conditions.

Thus, with the analysis of the proteome, the condition of a cell, a tissue, an organ or an organism can be described, in particular by the quantitative profile of its proteome. For example, it is possible, that as a consequence of a disease, the expression of certain proteins is suppressed and the expression of other proteins is increased, or that certain proteins are not expressed at all, or that certain protein variants, isoforms or post-translational modifications are changed in level as compared to a normal or reference condition. Thus, the protein profile is suitable as a direct indicator for the respective condition of cells or tissues, organs or the whole organism at a certain point in time and therefore is suited as an indicator for the health status of the organism.

Moreover, changes in the protein profile allow following up the influence of pharmacological compounds as well as their side effects.

Contrary to the analysis of mRNA profiles, which are often used in this respect to detect changes in transcription, the investigation of protein profiles has the advantage, that the information about changes in protein profiles gives direct information to the molecular mechanisms involved in a certain disease condition, because the vast majority of cellular processes is directly based on the function of proteins, based

on changes in posttranslational protein modifications, protein localization or protein-protein interactions.

For data acquisition recording the qualitative and quantitative aspects of a protein profile, techniques have to be applied, which allow the fast and parallel identification and quantification of the vast majority of proteins even in highly complex samples. Thereby, it should be possible to detect and quantify these proteins in a large dynamic range with concentration differences in the order of magnitudes of 9 to 12. Moreover, in contrast to DNA or RNA, there are no methods to amplify proteins.

Furthermore, there is as yet no method with which all proteins, i.e. very acidic, very basic, very large, very small, hydrophobic and hydrophilic proteins, can be detected at the same sensitivity within the same sample.

Mainly two different approaches are currently used for most analyses of complex proteomes. In principle, either the proteins are separated to reduce sample complexity enabling protein identification based on established methods like mass spectrometric fingerprinting, or proteins are degraded to many thousands of proteolytic peptides, which are then separated and identified based on tandem mass spectrometry and spectrum comparison to protein databases.

In this connection, 2D electrophoresis enables electrophoretic separation of up to 10.000 different protein species in a gel matrix in two dimensions followed by subsequent defined proteolysis of the separated proteins in spots cut from the gel, followed by the identification of the respective protein species via its specific peptide masses determined by mass spectrometry.

Further applied is one-dimensional or multidimensional chromatographic separation of peptides from a defined proteolytic degradation of all proteins of a proteome with subsequent identification of the peptides by tandem mass spectrometry and bioin-

formatic assignment of the peptide fragments to the original proteins of the proteome via protein or genome databases.

These methods are supplemented by antibody-based methods in which proteins are detected qualitatively and quantitatively through the specific binding to corresponding antibodies. Examples thereof are Western blots, ELISA or antibody microarrays.

In this connection, it is of particular importance to generate certain criteria, according to which complex protein mixtures can be analyzed, whereby the aim is to reduce the effort of such an assay and to accomplish either a preferably full coverage of the analyzed proteome or alternatively to accomplish the full coverage of a subpopulation which is defined in regard to certain criteria.

Mass spectrometry (MS)-based strategies became the key technologies for identifying proteins in proteomic research within recent years. Sample preparation and efficient fractionation of target analytes are the major bottleneck in all MS-based protein analysis technologies so far. The complexity of biological samples needs to be greatly reduced and the protein of interest needs to be enriched significantly before low abundance proteins can be identified with MS. Sample preparation strategies that reduce the complexity of protein samples or of peptide mixes from e.g. tryptic digests using immunoaffinity-based methods lead to a substantial increase in throughput and sensitivity. However, the lack of thousands of appropriate peptide-specific capture reagents limits the application of such immunoaffinity-based approaches on a proteomic scale or for unbiased biomarker discovery. Currently, these methods were mainly used for efficient and specific high abundant protein depletion or for low abundant diagnostic peptide enrichment followed by mass spectrometric quantification of single peptides from a proteome.

Therefore, techniques are necessary, which enable the analysis of complex protein mixtures covering all or almost all proteins in the sample by first separating the sample into different subpopulations of proteins and later analyzing the individual

subpopulations. These methods, of course, allow also the analysis of individual subpopulations according to general physical or general functional aspects of the proteins.

Several methods are known in the prior art, that allow such a preseparation of complex protein samples

WO 2007/112927 of this assignee discloses a method wherein two different sets of binding molecules targeted against small terminal epitopes of proteolytic peptides are used in parallel or in succession to specifically identify each protein in a proteome. Each set of binding molecules has to comprise up to 8000 binding molecules.

WO 2004/081575 discloses an array of binding molecules, which array allows a subdivision of a complex protein sample into subpopulations, whereupon a subsequent analysis of the subpopulations by mass spectrometry is performed. For said preselection, binding molecules, in particular antibodies are utilized that specifically bind to short epitopes.

The binding molecules are selected by rational selection of epitopes aiming to avoid an overlap and to increase the coverage between the entirety of the binding molecules used and the proteins recognized. In one example provided, all possible C-terminal tetra-peptides are considered as the starting set for epitope selection. These 160.000 sequences are thereafter reduced by deleting certain sequences to find a final set of epitopes, whereby the sequences are selected as to be likely to generate good epitopes, and are said to be also decided based on their frequency in naturally occurring proteins.

This is a cumbersome and heuristic method that furthermore cannot lead to a promising at least not to an optimal set of epitopes. Furthermore, it cannot be guaranteed that at least some of the chosen epitopes also bind to proteins other than the proteins of interest, this leading to false positive results, and that some epitopes are

excluded that are needed for detection of certain proteins, this leading to false negative results.

Thus, although the determination of a set of target epitopes is the starting point of the method disclosed, there are no rules disclosed of how to select the set of epitopes, the less of how to select an optimal set of epitopes that allows a quick and reliable identification of all proteins or a defined set of proteins from a large pool of proteins.

This document nevertheless describes methods for producing binding molecules that specifically bind to the set of epitopes, and methods for determining proteins of interest by using such set of binding molecules.

Therefore, the disclosure of this document is incorporated into the disclosure of this application by reference.

WO 02/060377 A2 discloses a classification procedure wherein the proteins within a complex protein sample are subdivided into subpopulations according to the binding to certain binding molecules which are referred to as "slinkers" that may be antibodies or binding partners for larger protein classes. By repeated binding procedures to different slinkers or slinker combinations individual proteins shall be isolated from a protein sample.

The method disclosed in WO 02/060377 A2 is generally unsuited for quantitative analysis because the definition of the "slinkers" used for subfractionating includes binding molecules that cover a wide range of binding affinities. Further, this document does not disclose how to optimize sets of binding molecules.

The known methods described insofar all show one general disadvantage in that the known methods do not provide a straight forward route for selecting binding molecules in a way that ensures a maximum coverage of proteins of interest while at the same time ensuring a minimum number of binding molecules required. Therefore,

an analysis of a complex protein sample, for example the proteome of an eukaryotic cell, would for a full or almost full coverage of proteins to be specifically analysed require a great effort in terms of time and material.

Summary of the Invention

In view of the above, it is an object of the present invention to provide a novel method for selecting binding molecules in a way that ensures a maximum coverage while at the same time ensuring a minimum number of binding molecules required, and that is capable of adjusting the selection process to a variety of detection methods.

According to the invention, this object is achieved by a method for determining *in silico* a set of selected target epitopes suited for the detection of at least one protein of a set of proteins of interest within a pool of sample proteins, the method comprising the steps of:

a) providing a first list containing the sequences of said sample proteins;

b) providing a second list containing the sequences of said proteins of interest;

c) applying an *in silico* fragmentation to the sample proteins in said first list to generate a third list containing sequences of possible target peptides having possible target epitopes at their free N- and C-terminal ends, which possible target epitopes each consist of three to five N- or C-terminal amino acids;

d) generating a fourth list of sequences of valid target peptides by removing from said third list such possible target peptides that are less suited for said protein of interest identification;

e) selecting from said possible target epitopes on said valid target peptides in said fourth list a set of selected target epitopes, said set of selected target epitopes containing at least one target epitope for each protein of interest in said second list, while at the same time containing an as small as possible number of target epitopes.

With this method, a set of target epitopes can be determined that in turn can be used to generate a respective set of binding molecules which can be used to detect the proteins of interest in a given biological sample.

The object underlying the invention is thus solved in its entirety.

This invention is based on the surprising finding of the inventors that using the novel method allows to strongly reduce the number of binding molecules, which have to be generated in order to accomplish a full coverage of detection for a given group of proteins in a complex protein sample.

Thereby, it is advantageous that with the novel method for determining *in silico* a set of epitopes, the detection of one or more or preferably all proteins within a complex protein sample or a group of proteins serving as biomarkers can be accomplished in a relatively moderately priced and easy way.

The disadvantages of prior art methods mentioned above are overcome by the method of claim 1.

It should be appreciated that the second and first list may be identical lists or that the second list may be derived from the first list or otherwise compiled.

The "pool of sample proteins" may contain all peptides of a proteome, e.g. the human proteome, whereby the second list may cover all or several biomarkers.

The expression "less suited" in step d) shall mean that the excluded peptides are unsuited for or impeding the method applied for peptide identification and/or quantitation, e.g. are not suited for MS detection of the proteins of interest, possibly due to limited resolution of available MS equipment, or have unfavorable properties for envisaged measurements of any kind,. e.g. bad solubility, chemical instability, and the like.

The expression "as small as possible" in step e) shall mean "minimal number of epitopes" only in the optimal case, whereby "less than the number of proteins of interest" is already an advantage over the prior art.

In this connection "proteome" is understood as the entirety of proteins present in a cell, a tissue, an organ or an organism including all expressed proteins, all their respective isoforms, polymorphisms and post-translational modifications. Proteome analyses may thereby address their respective concentrations at a certain point in time and under certain external conditions.

"List of proteins of interest" thereby means a certain list of proteins to be analyzed within a given sample. Such list can contain putative or confirmed biomarkers for e.g. cardiovascular disease and cancer to be identified or quantified within a given proteome.

Several resources exist, from which such lists of proteins can be retrieved. Thereby for example the UniProt database (The Universal Protein Resource (UniProt) *Nucleic Acids Res.* Bd. 35, S. D193-D197) aims to cover the proteomes of a variety of species in its entirety. From the HUPO list protein data from the human plasma proteome can be retrieved; States et al. *Nature Biotechnology*, 24(3), March 2006: 333-338.

"Biomarkers" means proteins that alone or in combination with other such proteins serve as markers for normal and abnormal physiology. Data about putative or confirmed biomarkers for cardiovascular disease and cancer are listed in Anderson, J

Physiol 563.1 (2005) pp 23-60, listing 177 candidate biomarker proteins; in Polanski and Anderson, *Biomarker Insights*, 2006:2 1-48, listing 240 candidate biomarker proteins, and from the Human Biomarker Test Panel "HumanMAP" , Version 1.6, from Rules Based Medicine, Austin, Texas, listing 188 biomarkers. These publically available data-sources can be used as a data-base for in silico approaches to the definition of binding molecules and sets of binding molecules. After removal of redundant and high abundant proteins having concentrations in human plasma greater 0.3 mg/ml, the size of the Biomarker dataset is 300 proteins.

The term "epitope" herein refers to a chemical structure on a molecule that interacts with a binding molecule. More specifically, "epitope" means an amino acid sequence comprising 3 to 6 amino acids that is specific to a peptide/protein or a group of peptides/proteins. Preferably, the epitope is located within a protein sequence at the terminus of one or more peptides generated from a protein by proteolytic digest.

"Binding molecule" herein means any molecule or any substance which is able to bind to a peptide/protein or to which a peptide/protein can bind. It is known to the skilled person, that binding molecules may be for example antibodies, antibody fragments, aptamers, and recombinant binding molecules.

In connection with epitopes, "set" is referred to as a group of epitopes comprising at least one C-terminal epitope and/or at least one N-terminal epitope. In connection with binding molecules, the term "set" refers to at least one binding molecule capable of binding at least one of the epitopes from an epitope set.

In this connection "peptide" is defined as a linear chain of amino acids that may either represent a full-length protein, a truncation product, or a proteolytic fragment of a protein.

"Detection" and "detecting" in connection with a protein or peptide refers to binding, isolating, enriching, qualitatively and/or quantitatively identifying of one or more proteins or peptides within a complex sample.

The term "support" refers to a solid surface whereon binding molecules can be immobilized. Thereby, the surface may be formed from or coated with any material suitable for immobilizing binding molecules. Thereby, the surface may be either flat or comprising a number of cavities.

The term "array" herein refers to a set of binding molecules immobilized on a support or on addressable beads so that each binding molecule is present at a known or addressable location. For example, binding molecules may be arranged on the surface of a support in spatially defined manner, so that each binding molecule is located at a different and identifiable location.

The term "identifying" herein means the assignment of a specific known or hypothesized amino acid sequence to a peptide and/or protein.

The term "quantifying" refers to measuring the total amount or concentration of a peptide in comparison to the total amount or concentration of another peptide or other compound or in relation to an otherwise specified parameter.

In the context of the present application, "filter step" means a selection step performed on a list of peptides by removing from said list peptides fulfilling certain criteria or not fulfilling certain criteria.

In this connection, a filter is a piece of software or hardware that performs such filter steps *in silico*.

It should further be appreciated that step e) can be further split into two or more substeps without departing from the scope of captioned invention.

Although this is unlikely to happen, proteins of interest no longer represented in list 4 by at least one valid target peptide can be identified within the pool of sample proteins by standard assay techniques, e.g. by using a specific binding molecule not binding to any of the possible target peptides

The set of selected target epitopes in step d) can be used to detect any protein from the pool as long as list 1 equals list 2, whereby the number of binding molecules can be reduced when list 2 is smaller than list 1.

As the whole pool of proteins is the starting point for epitope selection, only such epitopes are considered that are indeed represented on the proteins from the pool thereof, thereby eliminating already at the starting point of the method such epitopes from further consideration that can by no means contribute to the identification of proteins of interest, this enhances the method.

Further, the method leads to a set of selected epitopes, wherein it is known which peptides carry the respective epitopes, and to which proteins these peptides belong, thereby eliminating false negative and false positive results.

In this connection, it should be noted that WO 02/086081 discloses a method for identifying proteins of interest that are present as single isolates or mixtures of proteins, including cleaving the proteins of interest with a proteolytic agent to produce peptide fragments, contacting the peptides to an array having affixed thereto at discrete locations a set of binding molecules, detecting binding between the peptides and the binding molecules, and comparing the detected binding to a reference set.

Epitopes to be bound by the binding molecules are selected by applying a proteolytic digestion step to a protein pool for determining epitopes, and performing a randomized greedy algorithm to identify a set of epitopes that can distinguish the protein pool proteins.

The goal is to optimize such set of binding molecules, or epitopes associated with such binding molecules, such that many or all of the binding molecules recognize epitopes common to peptide fragments produced from cleavage of a plurality of proteins within a given mixture. A concurrent goal is a minimum set that produces a unique binding pattern for the peptides of protein within the mixture.

This is done by *in silico* cleavage of each protein within the pool, by applying the specificity of a proteolytic agent, into peptide fragments, and then determining from these peptide fragments the epitopes of all proteins. Subsequently, each protein within the pool is compared with all other pool proteins for grouping the proteins into different groups thereof according to the occurrence of epitopes within the sequence of the proteins, thereby scoring the epitopes according to their occurrence. The most abundant epitope is selected for the set of epitopes, and this algorithm is repeated with the remaining epitopes, and so on, until all proteins are represented in the selected set of epitopes.

If a particular binding molecule does not show the desired specificity for an epitope in the computationally optimized set thereof, an additional binding molecule may be added to the set.

The peptide fragments may be labeled after cleavage to allow e.g. fluorescent based detection of peptides bound on the array of binding molecules. If the binding molecules are designed such as to recognize C- or N-terminal epitopes, and the peptide fragments may not be labeled, Surface Plasmon Resonance shall be used to detect the binding pattern on the array of binding molecules.

This document, therefore, describes how to produce binding molecules of different kind, and in which way different proteolytic agents may be employed. Further, it discloses a way of how to use a greedy algorithm, what may be employed for step e) of claim 1.

Therefore, the disclosure of this document is incorporated into the disclosure of this application by reference.

However, the computations necessary for performing the known method are cumbersome and do not lead to an optimal set of selected target epitopes that allows e.g. MS or sandwich immunoassay detection of the proteins of interest. It should be noted that the starting consideration of this document is to avoid the expensive and technologically demanding MS instrumentation.

To the contrary, with the new method, the number of epitopes in the selected set thereof is reduced to such an amount that MS detection becomes suited for the detection of proteins of interest in a large pool of sample proteins.

The new method is a suitable method for circumventing the limitations of the prior art. In contrast to classical peptide immunisation strategies where peptides with a length of 10 to 30 amino acids are used as specific antigenic agents, the antigen, i.e. the epitope, was reduced to three or five amino acids and contains either a free N- or a free C-terminus. Antibodies generated against these epitope therefore bind groups of peptides that share a common short motif at the N- or C-terminal end generated during the e.g. tryptic digestion of the biological sample. In the course of this application, such antibodies are also referred to as "TXP antibodies".

In contrast to classical peptide-specific antibodies, the new binding molecules, e.g. the TXP antibodies, are capable of binding up to several hundred peptides derived from e.g. a fragmented proteome of any species. They can be used for the immunofractionation of peptide classes. The peptide fraction can be analyzed and deconvoluted by MS and MS/MS, which results in binding molecules that can be used for several dozens of targets and targets from different species.

Under the assumption that an ideal tryptic digest was obtained, the bioinformatic analysis shows that approximately 2000 TXP antibodies are enough to capture signature peptides that cover all human proteins in the Uniprot database

(downloaded on 11 June 2007, http://www.uniprot.org/). Compared to the classical 'one antibody – one analyte' whole-genome approach, the application of TXP antibodies results in a ten-fold reduction of the number of capture antibodies required. Since the TXP epitopes are very short epitopes, amino acid differences occurring within the protein sequences of different species may become negligible.

Furthermore, due to the knowledge of the genome, any differences in captured signature peptides can be calculated. in silico. Thus the TXP antibodies are a universal source for the analysis of signature peptides in human, animal, and plant proteomes using an immunoaffinity MS approach.

Furthermore these TXP antibodies can be used as a universal toolbox for the fast generation of sandwich immunoassays based on binding of a C-terminal and an N-terminal TXP-antibody generating a specific peptide recognition by 6 to 10 amino acids (dependent on epitope length of TXP-antibodies). Based on the TXP-antibody toolbox TXP-sandwich assays can be established without the need for antibody generation. This results in an enormous advantage in assay development, as concerns time and costs.

As will be shown below, this TXP concept has been evaluated using ß-catenin as a model target analyte. Antibodies could be successfully generated against short TXP motifs and the immunoaffinity fractionation enabled the direct analysis of ß-catenin expression in complex biological samples as well as the expression of other analytes that shared the same TXP epitopes.

According to another object of the invention, step d) is performed by applying at least one of the following target peptide removing filters to the possible target peptides in said third list:

- Met Filter for removing target peptides with a methionine in at least one of its terminal sequences,

- Unknown AA Filter for removing target peptides with unknown amino acids,
- High Abundant Target Epitope Filter for removing target peptides comprising target epitopes with high abundance,
- Weight Filter for removing target peptides with similar weight and with the same target epitope, which cannot be separated by mass spectrometry,
- Length Filter for removing target peptides with a length not adapted to the resolution or the method of a subsequent detection technology, e.g. shorter than 15, 12 or 5 amino acids and/or longer than 30 amino acids,
- High Abundant Protein Filter for removing target peptides belonging to highly abundant proteins,
- Proteotypic Peptide Filter for removing target peptides not predicted to be proteotypic.

The "Met Filter" removes peptides having a methionine in at least one of its terminal sequences, since chemical modifications of methionine may hamper the recognition of the target epitope by a binding molecule, especially by an antibody.

The "Unknown AA Filter" removes peptides, selected based on database entries, containing unknown or uncertain amino acids. Thereby, the occurrence of false-negative results can be reduced irrespective of the detection method applied.

The "High Abundant Target Epitope filter" removes peptides containing epitopes present in preferably more than 700 peptides. Therefore, especially when applying detection methods with a strictly limited resolution, an exceeding of the resolvable

peptide number can be avoided. This holds true especially for applications, in which peptides are detected by one-step Affinity-MS or by Fluorescence.

The "Weight Filter" removes peptides that have a similar weight (e.g. $\Delta m = \pm 4Da$, selected as isotope window) to other peptides bound by the same binding molecule. This way, peptides, which are indistinguishable or hardly distinguishable by MS are excluded from the analysis.

The "Length Filter" removes either peptides too small or too long for the read-out technology of the subsequent detection method. Small peptides e.g. often do not bind to HPLC-media and therefore are lost during HPLC-MS analysis. Large peptides might have signals outside the selected detection window of the mass spectrometer. The selected values can be adapted to the detection method applied. In case the subsequent detection is performed using fluorescent techniques such as Sandwich-Immunoassay, peptides that are shorter than 15 amino acids and might thus cause false-positive results.

The "High Abundant Protein Filter" removes peptides that bind the same binding molecule as peptides belonging to highly abundant proteins such as Actin or Tubulin. Therefore, false-negative results caused by sequestering out certain peptides from the binding molecules by other highly abundant peptides are avoided.

The "Proteotypic Peptide Filter" identifies peptides that are repeatedly and consistently identified for any given protein present in the mixture. All peptides which are not predicted to be proteotypic are removed.

The above list of filters does not claim to be complete and exhaustive, depending on the background list of proteins, further filters may be used and/or certain of the above filters may not be necessary or may be adapted to the respective analysis method used for peptide identification or quantification.

Said target peptide removing filters are usually applied in a specific order, in which order the Met Filter, the Unknown AA Filter and the High Abundant Protein Filter are used at any point within the order. Preferably, the High Abundant Target Epitope Filter is applied within said order prior to the Weight Filter and prior to the Length Filter, and the Weight Filter is applied within said order prior to the Length Filter.

By this, the preselection of epitopes is adjusted to the experimental setup, the novel method ensuring for example, a full coverage in the identification and quantification of a given group of proteins within a complex protein sample, thereby at the same time reducing the number of binding molecules required in the process.

Additionally, this method offers the possibility to specifically adjust the search criteria for fitting binding molecules to the requirements and specifics of a given method for said identification and quantification of proteins within a sample.

It is further preferred, if in connection with step d) possible target peptides not belonging to a protein of interest are removed.

This step has the advantage that the set of selected target epitopes will be further minimized.

Preferably, prior to step d) possible target peptides comprising a combination of N- and C-terminal possible target epitopes which is not unique amongst the possible target peptides are removed.

This step enhances the selectivity of the set of selected target epitopes.

According to a further object, the *in silico* fragmentation in step b) is performed according to the sequence characteristics of a sequence specific proteolytic digest.

It is further preferred, if during step e) an optimization procedure is carried out for determining a globally optimal set of selected target epitopes, e.g. by using an ILP-solver, or by determining a locally optimal set of selected target epitopes, e.g. by applying a standard greedy approach.

Such optimization procedures are common in many technical fields, they provide solutions to optimization problems which are NP-hard.

Such procedures are carried out by electronic data processing. Therefore, the present invention also concerns a computer program for carrying out the new method, a computer programmed for carrying out the new method, and a data carrier containing information for carrying out this new program.

According to another object, steps c) and d) are repeated in order to generate a second set of selected target epitopes, whereby for each possible target peptide in said third list, the selected target epitopes in the one set thereof comprise an N- terminal epitope and in the other set thereof comprise a C-terminal epitope.

This enables a 2 step Affinity MS Approach or a Sandwich-Immunoassay Approach which results in a dramatic reduction of required antibodies compared to the standard one antibody one analyte method.

The present invention also relates to a set of selected target epitopes, determined according to the new method, as well as to a set of selected target epitopes suited for identifying proteins of interest, wherein the proteins of interest serve as Biomarkers, and the set of selected target epitopes contains at least 30%, preferably 50%, more preferably at least 70% of the aforementioned set of selected target epitopes.

Still further, the present invention relates to a method for preparing a set of binding molecule suited for the detection of peptides or proteins of interest from a pool of proteins, comprising the steps of:

a) determining a set of selected target epitopes according to the new method; and

b) generating for each target epitope in said set of selected target epitopes a binding molecule.

Even further, the present invention relates to a set of binding molecule generated according to this method, and to a set of binding molecules, containing at least one binding molecule for each target epitope within the set of selected target epitopes, said binding molecules being preferably selected from the group consisting of antibodies, antibody fragments, antibody variants, aptamers, polynucleotides, recombinant binding molecules.

Finally, the present invention relates to a method for detecting proteins of interest in a biological sample, comprising the steps:

a) providing said set of binding molecules,

b) incubating said biological sample with a proteolytic agent to provide a peptide solution, said proteolytic agent having the same cleavage rules as the fragmentation applied *in silico* for determining the set of target epitopes used for generating the set of binding molecules from step a),

c) incubating said peptide solution with said set of binding molecule, and

d) determining the peptides bound to said set of binding molecules;

wherein the determining step d) preferably comprising the step of

e) incubating said peptide solution with a second set of binding molecule, for each peptide there being provided in one set of binding molecules a binding

molecule for a N-terminal epitope, and in the other set of binding molecules a binding molecule for a C-terminal epitope;

and wherein further preferably the target epitope is selected from one of the set of selected target epitopes mentioned above.

In this connection, the protein of interest may be a biomarker protein.

In the context of a 2 step Affinity MS Approach, the peptides are first harvested with the first set of binding molecules and thereafter, in a second step, eluted and incubated with the second set of binding molecules.

In the Sandwich-Immunoassay Approach, a first set of peptides is captured with the first terminus specific binding molecule, followed by detection of the specific analyte peptide in said first set of captured peptides by binding of the second terminus specific antibody. In the sandwich immunoassay approach detection of the specific analyte peptide relies on binding of TXP antibodies to both termini at the same time.

Both approaches require a specific optimization strategy for epitope selection; see section 2.2 and cases 2 and 3 in section 3 of the detailed description below.

In the method for detecting proteins of interest, step d) is preferably being performed using at least one of the following methods: mass spectroscopy, immunoassay, chromatography, electrophoresis, electrochemistry, surface plasmon resonance, oscillating crystal.

Further, step c) and/or step e) is preferably performed by using binding molecules immobilized to a carrier, the carrier being selected from the group consisting of material for affinity chromatography, beads, microspheres, capillary surfaces, micro channel structures, sensory surfaces.

Last, the present invention relates to the use of the above selected binding molecules for identifying proteins serving as biomarkers.

Further advantages and features will become clear from the following description and from the attached figures.

It will be understood that the features which are mentioned above and those, which are still to be explained below can be used not only in the combination which is in each case specified but also in other combinations or on their own without departing from the scope of the present invention.

Brief description of the drawings

Fig. 1 and Fig. 2 show graphic representations of optimization procedures for 2 step Affinity MS Approach and Sandwich Immunoassay Approach.

Preferred Embodiments

1.1 Basics of the TXP Proteomics Concept

The use of peptide-specific antibodies that recognize peptide fragments generated by an enzymatic digest of a proteome is an effective possibility to enrich target peptides from highly complex samples. Combined with subsequent MS analysis it is a valuable tool for targeted proteomics.

The Triple X Proteomics approach of this embodiment is based on immunoaffinity enrichment of tryptic peptide groups subsequently identified in an MS read out. Instead of an immunoprecipitation of a single peptide the method is suitable for fractionating complex samples into classes of signature peptides that share the same terminal epitope, i.e. TXP epitope. The enriched peptide classes can be deconvoluted through subsequent MS read-out.

If the peptide signal turns out to be unique, then the peptides can be mapped back to their original proteins. Different signature peptides can be used to identify a protein of interest in a digested sample. This gives rise to several protein targeting possibilities, wherein each TXP antibody is multispecific. TXP antibodies can be generated from the epitopes of different signature peptides of the protein of interest enriching several other signature peptides at the same time. This is a unique possibility for reducing and optimising the number of antibodies required for an immunoaffinity approach on a proteome-wide scale.

A bioinformatics process was established to define the TXP epitopes and to determine the optimal set of TXP-type antibodies. All human proteins in the Uniprot database (downloaded on 11 June 2007, http://www.uniprot.org/) were digested with trypsin - *in silico*. All TXP- epitopes enriching tryptic fragments that would impede the measurements had to be removed (Figure 3). This was achieved through multiple filtering steps. Based on incomplete sequencing information, the first filter removed the peptides that contained unknown amino acids. The second filter removed all TXP epitopes that were shared by more than 700 peptides. Larger peptide quantities may not be deconvoluted with an MS-based read-out.

Subsequently, all peptides that shared the same TXP epitope and revealed a weight difference of less than 4 Da were removed by a weight filter in order to enable the detection of unique MS signals. A peptide length filter was set at a length of between 6 and 30 amino acids. The TXP epitopes of signature peptides derived from high abundant proteins were not taken into account since they would have prevented the binding of signature peptides derived from low abundant proteins with the same terminal sequences.

The filter sets can be reduced further in order to only cover peptides derived from the proteins of interest. The remaining peptides and antibodies constitute the basis for the optimization algorithm. The goal is to minimize the number of antibodies to a

degree that at least one peptide from every protein in the set is captured. The minimum set cover problem is a difficult (NP-hard) combinatorial optimization problem.

Currently, there is no suitable algorithm available. One possibility is to describe the covering problem as a Boolean linear program because excellent solvers such as CPLEX, GLPK, and lpsolve are available:

$$\min \sum_{i=1}^{n} x_i \quad (1)$$

$$s.t. \sum_{i=1}^{n} x_i s_{ij} \geq 1 \quad \forall p_j \in P \quad (2)$$

$$x_i \in \{0,1\}$$

There are n possible antibodies to select from. x_i is a Boolean variable which is only true if antibody i is selected. The constant s_{ij} is true if protein j contains a peptide whose terminus matches antibody i. P denominates the set of proteins, p_j the j-th protein in a given set. The number of selected antibodies is the objective function (1) to minimize. There are as many type (2) constraints as there are proteins. The sum is bigger or equal than 1 if one or more antibodies have been selected that capture measurable peptides from protein j.

In general, this approach works well for small to medium protein sets. Heuristics needs to be applied for larger protein sets. In this case, one normally uses the greedy approach as it has the best possible approximation to runtime ratio. Initially, an antibody is selected that captures the highest number of uncovered proteins in every iteration. Then, the protein coverage is recalculated. This procedure is repeated until at least one antibody has been found for every protein present.

The bioinformatic analysis reveals that 1622 TXP antibodies are enough to capture signature peptides that cover 19072 human proteins deposited in the Uniprot database (downloaded on 11 June 2007, http://www.uniprot.org/). This would be a reduction by 12 compared to the possibility to generate one antibody per protein.

1.2 Generation of TXP Antibodies

The TXP approach uses antibodies that are directed against short epitopes in order to fractionate peptide classes according to their common terminal sequences, the TXP epitopes. This requires the detailed characterization of the antibody binding activities. The specificity of the TXP epitope antibody needs to be assessed prior to applying the antibody in a bioanalytical workflow. The assessments can be done with peptide arrays.

To characterize the TXP antibodies, peptide arrays were designed that contained positional scanning peptide libraries for the TXP epitopes. Four different peptide libraries were synthesized for each TXP epitope. Within each library, one individual amino acid position was randomized; any of the 20 amino acids was possible. The sequence specificity is lost at the randomised position and the need for a specific amino acid is indicated by the loss of signal on a peptide bead made from such a positional scan peptide library.

The signal intensity was reduced dramatically at amino acid positions that were important for the proper interaction between the antigens and the antibodies. No substantial signal loss was observed at positions that were of subordinate importance. Such an approach allows the detailed analysis of the importance of a particular amino acid of the eptiope for antigen-antibody interaction.

Results have been obtained for an antibody that was generated against the C-terminus of a ß-catenin signature peptide with the AMTR amino acid sequence. In case of the AMTR antibody, the amino acids that contributed to the binding of the antigen were A, T and R. When these amino acids were replaced with the peptide libraries, the binding of the AMTR-specific antibody decreased to 5 % of the original signal. This was expected because A, T, and R were present at this position with a frequency of 5 %. However, replacing methionine only had a marginal effect on the

antigen-antibody binding. Thus, the AXTR binding motif was shown to be specific for the AMTR antibody.

1.3 TXP Affinity Enrichment MS from HepF1 Cell Lysates

ß-catenin was chosen as a model molecule to prove the feasibility of the TXP approach as the protein plays a key role in the WNT pathway and tumorigenesis. Mutations have been found at distinct phosphosites in several types of tumours. Two TXP antibodies were chosen that were directed against the TXP epitopes that were present in the ß-catenin signature peptides (GNPEEEDVDTSQVLYEWEQ-GFSQSFTQEQVADIDGQYAMTR, TEPMAWNETADLGLDIGAQGEALGYR). The TXP antibodies were specific for the LGYR and AMTR termini and were used for the immunoaffinity fractionation of a tryptic HepF1 mouse hepatoma cell line digest.

After enrichment, the two ß-catenin signature peptides could be detected and verified within the LGYR and AMTR fractions using MALDI TOF/TOF MS. Mascot ion scores of 65 were obtained for the AMTR and 57 for the LGYR fragment. This shows that ß-catenin could be detected independently by two different TXP antibodies. As expected, several other tryptic signature peptides were identified besides the ß-catenin fragments.

The terminal sequences identified within the group of peptides enriched with the AMTR antibody correlated well with the binding motif AXTR, described in section 1.2. The antibody was not very specific for the third C-terminal amino acid, where the amino acids alanine, phenylalanine, threonine, serine, glutamine, and tyrosine were found instead of methionine, which was used in the immunogen. The antibody at the methionine position had no preference for side chains of a basic, acidic, large or hydrophobic nature. Nine different proteins were found in the immunofraction of the AMTR antibody.

14 signature peptides could be identified for peptide classes enriched with the LGYR antibody. Two peptides were present in pyroglutamic acid or glutamine form. Four out of the 14 peptides analyzed contained LGYR at the C-terminus, 10 had amino acids other than tyrosine and four of these 10 had amino acides other than leucine. The majority of the fractionated peptide termini had LGXR binding motives that fitted to the LGYR TXP antibody.

As for the AMTR TXP antibody no amino acid preference was observed for the position after the amino acid arginine. In the peptide fraction, the amino acids leucine, glutamine, glutamate and phenylalanine were found instead of the amino acid tyrosine. At this position in the immunogen large, uncharged amino acids seemed to favour the antibodies´ unspecificity for other amino acids. This has to be investigated in detail by the epitope analysis of other TXP antibodies.

No specific software tool, which would enable the identification of proteins based on the TXP epitopes and the measured mass, is currently available for the TXP approach presented in this embodiment. Therefore, all major MS peaks were analyzed in detail using the MS/MS mode. But maybe in future, the combination of the knowledge of the TXP epitope with the accurate peptide mass will allow an accurate identification of signature peptides without MS/ MS. Furthermore an integration of isotope-labelled internal standards will allow the absolute quantification of the protein.

In summary, the Triple X Proteomics approach broadens the applicability of immunoaffinity MS. A feasible number of antibodies would enable the analysis of entire proteomes of any given species. Compared to one analyte-one antibody approaches, TXP reduces the number of required antibodies tremendously, thereby bringing the coverage of proteomes into reach. Furthermore, the TXP concept enables the combination of unbiased and biased proteomic approaches. The biomarker workflow is simplified in so far that the tools of the discovery phase – the TXP antibodies - can be used in the validation phase. Once a potential biomarker is identified with TXP immunoaffinity MS, the same antibody can be applied to validate the biomarker candidate.

1.4 Conclusion

This example shows the feasibility of using terminus-specific antibodies for the fractionation of tryptically digested biological samples. Peptides that shared the same terminal sequence motif could be fractionated and subsequently identified by MALDI TOF/TOF. The presence of ß-catenin could be verified using two independent TXP approaches. At the same time, several other peptides could be identified and the protein of interest be verified.

Antibodies could be produced that were specific for short peptide sequences with a length of 3 amino acids at the free terminus. The properties of the generated antibodies enabled TXP protein profiling to be performed. The generated antibodies could be integrated in a simple immunoaffinity workflow. It was shown for 23 proteins that the peptide classes could be enriched and the corresponding proteins be clearly identified. Thus, the difficulties experienced when developing antibodies for protein analysis on a proteomic scale are minor compared to the classical immunoaffinity MS approach.

The introduction of biostatistical analysis resulted in an optimised set of antibodies for the analysis of the human proteome. The calculation showed that less than 2,000 antibodies were necessary to cover the current set of human proteins stored in UNIProt. The simple immunisation strategy enabled the generation of hundreds of antibodies and the broad coverage of the proteome.

The TXP concept is suitable for immunoaffinity MS approaches on a proteomic scale. The speed and specificity of a single immunoaffinity enrichment step, combined with accurate mass determination, makes the combination of unbiased and biased proteomic approaches feasible. Additionally, the short epitopes of the TXP antibodies extend the application spectrum of immunoaffinity MS strategies to humans, animals and plants, as sequence heterologies become negligible.

2. Examples for optimizations procedures

2.1 Example of the optimization step for minimization of the set of antibodies required to analyze a group of proteins by immunoaffinity precipitation coupled to mass spectrometry

A list of theoretically measurable peptides and suitable antibodies has been calculated by the filter process described above. As an example, the list of proteins of interest is a set of 5 G-protein coupled receptor families, comprising acetylcholine, alpha-adrenergic, beta-adrenergic, histamine and serotonin receptors.

Receptor Type	Uniprot IDs of family members
acetylcholine	P11229, P08172, P20309, P08173
alpha-adrenergic	P35348, P35368, P25100, P08913, P18089, P18825
beta-adrenergic	P08588, P07550, P13945
histamine	P35367, P25021, Q9Y5N1, Q9H3N8
serotonin	P08908, P28222, P28221, P28566, P30939, P28223, P41595, P28335, Q13639, P47898, P50406, P34969

Table 1: List of proteins of interest sorted by protein families

The complete list of potentially signature peptides contains 123 theoretically selectable C and N-terminal epitopes of a length of 4 to 5 amino acids. The optimization task is to define the minimal number of epitopes, which cover all signature peptides of interest. If the number of epitopes is significantly smaller than the number of proteins of interest, this approach fits the goal of reducing the number of required antibodies to identify large sets of target proteins.

	AB1	AB2	AB3	AB4	AB5	AB6	AB7	AB8	AB9
Protein 1	0	0	1	0	1	1	0	0	1
Protein 2	0	0	0	1	0	0	0	1	0
Protein 3	0	0	1	0	1	0	1	0	0
Protein 4	1	0	0	0	0	0	1	0	0
Protein 5	0	1	0	0	1	0	0	0	1
Protein 6	1	0	0	0	0	0	0	0	1

Table 2: **Optimal Set-Cover, the smallest subset of columns (antibodies) with at least one 1 in every row protein. AB4, 7 and 9 (highlighted) represent the optimal solution for the set of proteins (Protein 1-6) here.**

This problem is known as the set-covering problem. The set-covering problem is NP-hard, which means that it is computationally hard to solve. The problem can be formulated a classical linear programming problem with boolean variables. Boolean variables can take values of 0 (zero) or 1 (one). Variables are designated with the name "x".

If the variable x_i is set to 1, the antibody number i in the list is selected, if the value is 0, the antibody is omitted. This problem can be mathematically formulated as the minimization of the sum of the values of all variables.

$$\text{minimize} \sum_{i=1}^{n} x_i$$

The smaller the sum, the fewer antibodies are selected. This problem is subject to the condition that all proteins in the list of interest have to be "covered". A protein is

covered if one or more selected antibodies capture at least one peptide of the tryptic digest of the protein.

A solution that violates this constraint, e.g. by not selecting any antibody, is invalid. These constraints can be formulated as follows for the protein P11229. The set of epitopes that would capture an identifiable peptide from the tryptic digest of P11229 can be derived from the output of the filter pipeline (table 3). These are: |DCVP, |DCVPE, ALCNK|, LCNK|, |EEEE, |EEEEE, EVVIK|, VVIK|, |ELAAL, |ELAA, |ETPGK, |TPGK, |GGGS, |GGGSS, SSSER|, |GPWQV, |GPWQ, ... , LAPGK|, APGK|, QAPTK|, APTK|. All terminal sequences have a corresponding variable x_i with an index i, e.g. x_{23} for ALCNK, x_{67} for VVIK, etc. A constant designated y_i is introduced which has the value 1 if the terminal sequence corresponding to x_i is contained in the set of epitopes for the protein of interest. y_{23} would have the value 1, because ALCNK is contained in the set.

Sequences	Valid termini
DCVPETLWELGYWLCYVNSTINPMCYALCNK	N4,N5,C4,C5
EEEEEDEGSMESLTSSEGEEPGSEVVIK	N4,N5,C4,C5
ELAALQGSETPGK	N4,N5,C4,C5
GGGSSSSSER	N4,N5,C5
GPWQVAFIGITTGLLSLATVTGNLLVLISFK	N4,N5,C4,C5
SQPGAEGSPETPPGR	N4,N5,C4,C5
TLSAILLAFILTWTPYNIMVLVSTFCK	N4,N5,C4,C5
TVLAGQCYIQFLSQPIITFGTAMAAFYLPVTVMCTLYWR	N4,N5,C4,C5
YFSVTRPLSYR	N4,N5,C4,C5
AALMIGLAWLVSFVLWAPAILFWQYLVGER	C4,C5
MNTSAPPAVSPNITVLAPGK	C4,C5
MPMVDPEAQAPTK	C4,C5

Table 3: **Valid targets in the tryptic digest of the muscarinic acetylcholine receptor M1 (P11229)**

Suppose x_{15} is the variable corresponding to the antibody specific for the C-terminal sequence EIER, which does not capture any peptide from P11229. This would lead to a 0 value for y_{15}.

A linear constraint of the form

$$\sum_{i=1}^{n} y_i x_i > 0$$

is introduced. The sum will be bigger than zero if, and only if at least one antibody is selected, that matches the terminal sequence of at least one identifiable peptide contained in P11229.

Such a constraint is constructed for each of the 29 proteins of interest. The size of a problem is therefore determined by the number of antibodies to select from and the number of proteins to cover. For all five GPCR protein families this leads to a boolean linear program with 123 variables and 29 constraints.

This program is then solved with a suitable solver (e.g. cplex), by exporting it in a standard format for linear programs (lp, gmpl). The used solvers are very efficient, state-of-the-art implementations of well-known algorithms like the simplex algorithm for linear programs, and branch-cut-techniques for integer programs. A solver returns the values for x variables, which are equivalent to the optimal subset of antibodies: antibody i is selected if the variable x_i has the value 1.

The optimal set for the 5 GPCR families contains 13 epitopes: |TDSDT, |DDEIT, LYWR|, IALDR|, |TLGII, ISFDR|, ISLDR|, VSLEK|, VYAR|, ISIDR|, PLFWR|, HSFR|, LGWK|.

This set is obviously considerably smaller than the set of 29 proteins of interest.

2.2 Short example for the optimization procedure 2 Step Affinity-MS and Sandwich-Immunoassay

After the preprocessing step (*in silico* digest) basic filters are applied to remove peptides with unknown positions or methionine in the terminal sequence from the peptide pool.

Then a data structure called „epitope combination graph" is built, which contains all existing epitope combinations. For each epitope a "node" is defined in the data structure. If a peptide exhibits a certain epitope combination, the nodes corresponding to the epitopes are connected with a "peptide edge". These edges are directed, meaning that they have start and end epitope. The direction of an edge defines which epitope is used for capture (start) and detection (end). Only peptides (length > 15 AA) which can reliably be bound by two antibodies are added to the graph. If the data are prepared for sequential Immunoaffinity MS the length filter can be set to shorter peptides. Only very short peptides (length <5 AA) are filtered out as they are hardly detectable in the MALDI Matrix background, often do not bind to HPLC-columns and give only little specific information in their fragment mass spectra for peptide identification based on database searches (Mascot, Phenyx etc.).

Parallel peptide edges occur when a terminal combination is not unique. If the optimization is done for the Sandwich-Immunoassay case, these edges are removed (uniqueness filter).

In Figure 1, the left side shows the "epitope combination graph" before the application of the uniqueness filter. There are parallel edges connecting the epitopes AFST and ARPR. The edges where inserted because two peptides (AFSTCASHLVVVTIFYSAIIFMYARPR, AFSTCAAHLAVVVIYYSSTLFTYARPR) show the same terminal epitopes . The right side shows the "epitope combination graph" after the application of the uniqueness filter. Parallel edges have been removed.

If the optimization is done for 2 step Immunoaffinity MS, parallel edges are only removed if the weights of the peptides corresponding to the edges differ less than the resolution of the used mass spectrometry device/protocol/technology can discriminate (e.g. 4 Dalton).

Before the optimization step, peptide edges are removed, which can not reliably bind two antibodies (< 15 AA). After this step, referred to as interference filter, the epitope combination graph is ready for the optimization step.

In the optimization step a score is assigned to every epitope. The score is a weighted sum of

1) the number of all peptides captured by a capture reagent for the capture (start) epitope from proteins, which were not yet captured or identified before (weighing factor 1); and

2) the number of identified peptides captured by a selected capture (start) epitope resulting in a pair of capture reagents to both terminal epitopes, from proteins, which were not yet captured or identified.before (weighing factor 2)

After the scoring, the node with the highest score is selected.. The weighed sum of this node is not used for selection anymore. If a protein has been identified, all peptide edges from other epitopes of this protein are removed. Then all residual scores are updated and the whole process is started again with the reduced set of epitopes.

Figures 2a to 2d illustrate the in succession the first four iterations of the optimization process. In Fig. 2a, PGAR's score is the highest and therefore this epitope is selected. The scores of the connected nodes (LLLL, EEEE, AFST) are updated (1 identified peptide) as shown in Fig. 2b.

In Fig. 2b, AFST scores highest because it leads to the identification of peptide AFST----PGAR captured by PGAR in Fig. 2c. The peptide corresponding to the peptide edge connecting PPPR and THTG belongs to the same protein as the identified peptide. This edge is therefore removed and the weighed sums are updated for SSSR as part of the identified AFST--- SSSR peptide.

Among the remaing epitopes in Fig. 2c, there are several nodes with the same weighed highscore of 2 (N-term: SSSS, LLLL, EEEE, C-term: SSSR).

SSSS is selected by chance resulting in the modified scheme Fig. 2d, where the weighed sum of the corresponding epitope ASPR is updated as the peptide SSSS---ASPR is identified.

3. Methods for protein detection using the TXP approach

The method for *in silico* determination of a set of selected target epitopes can be used in combination with a multitude of different analytical methods. Thereby, the particular advantage of the new method is that in the preprocessing step, the set of epitopes determined can be tightly adjusted to the particular needs and constraints of the respective method.

The novel method thereby comprises a predefinition phase. In this phase, it is determined, from which list the analytes are to be selected. Examples for Protein lists are the Uniprot and HUPO databases as well as the Biomarker dataset, as described hereinabove.

Further, the novel method comprises a preprocessing phase, in which an *in silico* proteolytic digest of the Proteins comprised in the list defined in the predefinition phase is performed. The *in silico* digest thereby can be performed according to the

specific features of a Tryptic digest, a LysC-digest or other known methods for sequence-specific proteolytic digests.

After the preprocessing phase the remaining peptides are subjected to a filtering phase, in which peptides, which exhibit unfavorable features for the chosen analytic method, are eliminated from the epitope list.

Three examples for methods, which can be used in combination with the novel method, are described below. Further is described, which adjustments are made in the novel method in order to generate an epitope set suitable for the specific method.

Case 1: **Affinity MS**

Protein fragments – sharing the same terminus sequence - are enriched by using 3 to 5 amino acid specific terminal antibodies after a proteolytic digest. This step results in an enrichment of a group of peptides with the same N- or C-terminal sequence. The detection of the proteins is performed in a following step with mass spectrometry.

In this method, a first set of binding molecules is immobilized on a support or on addressable beads. A complex protein sample is subjected to protein fragmentation by for example tryptic digest. Subsequently, the resulting peptide mixture is brought into contact with the support or the beads. Subsequently, unbound peptides are removed in a washing step. The bound peptide fractions are then released from the binding molecules and the peptide mixtures generated this way are subjected to MS-analysis.

One prerequisite of a successful MS-analysis is that the number of peptides present in the final solution does not exceed the resolution of the MS-method chosen. Therefore, binding molecules have to be selected, that bind only a comparably limited number of peptides from the initial peptide mixture.

A preprocessing phase is performed, in which the following filter criteria are applied in the following order:

a. Unknown AA Filter

b. Met Filter

c. Highly Abundant Target Epitope Filter

d. Weight Filter

e. Length Filter

f. High Abundant Protein Filter

g. Filter removing proteins not belonging to proteins of interest

Thereby, the peptide Length Filter in accordance with the specifications of the MS procedures will sort out peptides with a length shorter than 5 amino acids and/or longer than 30 amino acids.

The optimization problem in case 1 can be formalized as a set cover problem. The goal is to cover a set of proteins with a minimal set of antibodies. Small instances, up to several hundred proteins, can be approached by solving an integer linear program with standard solver software. It is unfeasible to formulate and solve integer programs for large instances. The inventors have applied heuristic algorithms to calculate epitope sets for coverage of the whole proteome - several thousand proteins. It has been shown in theory that a standard greedy approach gives generally satisfactory results. The heuristic algorithm chooses the antibody that captures the most yet uncovered proteins in every iteration. This is repeated until every protein is covered by at least one antibody.

Versus the standard one antibody one antibody approach, the terminus specific antibody approach reduces the required number of antibodies e.g. to cover the proteome, by the factor 12, see table 4 below.

database / list	Proteins covered	Used protease	antibodies required		improvement
			one antibody/ one analyte	TXP approach	
Uniprot	19072	Trypsin	19072	1622	12
Hupo Plasma Poteome	9018	Trypsin	9018	972	9
Biomarker Dataset	294	Trypsin	294	113	3
Uniprot	18520	LysC	18520	2093	9
Hupo Plasma Poteome	8818	LysC	8818	1191	7
Biomarker Dataset	287	LysC	287	129	2

Table 4: Antibodies required for Affinity MS Approach

Case 2: 2 step Affinity MS

Protein fragments – sharing the same terminus sequence - are enriched by using 3 to 5 amino acid specific terminal antibodies after a proteolytic digest. This step results in an enrichment of a group of peptides with the same N- or C-terminal sequence. With a second enrichment step using additional 3 to 5 amino acid specific terminal antibodies the complexity of this peptide population is further reduced, in general down to one species. The detection of the proteins is performed in a following step with mass spectrometry.

In this method, a first set of binding molecules is immobilized on a support, e.g. beads. A complex protein sample is subjected to protein fragmentation by for example tryptic digest. Subsequently, the resulting peptide mixture is brought into contact with the support or the beads. Subsequently, unbound peptides are removed in a washing step. The bound peptides are then released from the first binding molecules and the resulting peptide mixture is brought into contact with a second set of binding molecules. After unbound peptides have been removed, the bound peptides are released from the second binding molecules and subjected to MS-analysis.

Like for the 1 step Affinity MS of section 2.1, it is a prerequisite for a successful 2 step Affinity MS analysis, that the number of peptides present in the final solution does not exceed the resolution of the MS-method chosen. Therefore, pairs of binding molecules have to be selected, that bind at least one, preferably a multiple of one, at most however only a comparably limited number of peptides from the initial peptide mixture.

Two independent filter processes are performed for a first and a second set of binding molecules, respectively.

For the first set of binding molecules, a process is performed, in which the following filter criteria are applied in the following order:

h. Unknown AA Filter

i. Met Filter

j. High Abundant Target Epitope Filter

k. Length Filter

l. High Abundant Protein Filter

Thereby the peptide length filter in accordance with the specifications of the MS procedures will sort out peptides with a length shorter than 5 amino acids and/or longer than 30 amino acids.

For the second set of binding molecules a process is applied to the peptide pool from the preprocessing phase, in which the following filter criteria are applied in the following order:

m. Unknown AA Filter

n. Met Filter

o. Length Filter

The filter not contained in the filter process for the second set of binding molecules are excluded because peptides meeting the respective criteria will be removed from the theoretical peptide set during the first binding procedure.

The bioinformatics optimization of this assay can be separated into two steps. First, the generation of the data structure in which all combinations of first and second binding molecules which capture multiple one are stored.

At this step a weight filter is applied, removing all combinations which capture isobaric peptides.

p. Weight Filter (2 Termini)

Based on this structure, a weighted score for each binding molecule is calculated, this score returns the number of peptides it captures for which the corresponding first or second binding molecule already exists plus the number of peptides it captured for which the corresponding first or second binding molecule does not exist yet. This score is the base for the iterative optimization scheme. In this scheme the set of binding molecules is incremented by the binding molecule which obtains the highest score.

After each iteration, the data structure is updated to account for inclusion of a new binding molecule into the assay. This update procedure removes all peptide information of proteins which are already covered though a peptide. This procedure is repeated until for all proteins at least one peptide is captured by a pair of one first and one second binding molecule.

database / list	used Protease	Proteins covered	required TXP antibodies	required antibodies for one antibody one analyte approach	reduction
Hupo Plasma Proteome	Lys C	8563	2444	17126	7.0
Hupo Plasma Proteome	Trypsin	8907	2041	17814	8.7
Uniprot + Hupo Plasma Proteome	Lys C	17784	3805	35568	9.3

Uniprot + Hupo Plasma Proteome	Trypsin	18683	2937	37366	12.7
Biomarker List	Lys C	279	362	558	1.5
Biomarker List	Trypsin	288	370	576	1.6

Table 5: 2 Step Affinity MS Approach

The usage of TXP antibodies in the 2 step Affinity MS Approach results in a dramatic reduction of required antibodies compared to the standard one antibody one analyte method. The reduction is 9.3 in the case of a digest with LysC and the aim to cover the whole proteome, as can be taken from table 5 above.

This approach and/or these results become more interesting if other species are taken into account. Unless the TXP antibodies bind to short sequences they can be used in different species (animals, plants, microbials, etc.). This transspecies application is a big advantage as the number of antibodies which have to be prepared for the study of the proteomes of different species is dramatically lower than today, where usually antibodies are applied which are specific for one or only some related species.

Case 3: Sandwich Immunoassay

Protein fragments – sharing the same terminal sequence - are enriched by using 3 to 5 amino acid specific terminal antibodies after a proteolytic digest. In a detection step using a second 3-5 aminoacid specific terminal antibody, the peptide is identified. In this case 3, two antibodies bind simultaneously – not like in case 2 sequentially - to the peptide. This process is further called Sandwich assay

In this method, a first set of binding molecules, the capture set, is immobilized on a support or on e.g. addressable beads. A complex protein sample is subjected to protein fragmentation by for example tryptic digest. Subsequently, the resulting peptide mixture is brought into contact with the support or the beads. Subsequently, unbound peptides are removed in a washing step. Then second binding molecules, the detection set, are applied on the support or the beads, respectively. After washing, specific peptides or proteins can be detected. Thereby, it is preferred, if the binding molecules bind to the termini of the detected peptides. More specifically, it is preferred, if the first binding molecules bind to the N-terminus of the peptides while the second binding molecules bind to the C-terminus of the peptides. Of course, this order can be reversed.

The prerequisite for the successful Sandwich-Immunoassay is the uniqueness of the combination of the first and the second binding molecules.

In this connection, the advantage of the novel method is that a multitude of analytes can be analyzed successively or in parallel requiring only a relatively small number of binding molecules.

A process is performed, in which the following filter criteria are applied in the following order to the peptide pool from the preprocessing phase:

t. High Abundant Protein Filter in first step (for capture step) (stop list)

u. Uniqueness Filter (only unique Target Epitope combinations are considered)

v. Length Filter for detection: 15

w. Length Filter for interference: 12

Thereby, the high protein abundance filter has to be applied only to the epitope selection process for the first set of binding molecules. This is the case, because in the prefractionation step of the Sandwich-Immunoassay, epitopes, which match the criteria of these filters, will be selected out and can therefore not disturb the following analysis steps.

A prerequisite for a Sandwich-Immunoassay is that two antibodies bind to their target peptide. This is possible to a degree for peptides with a length larger than 12 amino acids, but reliable only for peptides with a length larger than 15 amino acids. The binding does not work for peptides with a length smaller than 12 amino acids. Because of this, only peptides with a length longer than 15 amino acids are considered as targets.

The optimization procedure of this bioinformatics workflow can be separated into two steps. First, the generation of the data structure in which all epitope combinations which capture multiple one are stored.

Target epitope-combinations which are not unique, and thus can not be assigned to a specific peptide and protein, have to be removed from the data structure.The length considerations explained above have implications on the selection of peptides with unique terminal epitope combinations.

Uniqueness is defined by the hypothesis that there is no other peptide in the solution with the same terminal sequences, which is likely to bind both antibodies. As this is considered possible from a length of 12 amino acids, peptides from 12 to 15 amino acids have to be considered in the uniqueness filter, even if they are not considered as targets. These peptides can "interfere" with the detection process. Peptides with a length below 12 amino acids do not disturb the detection process, because they are too short to bind both antibodies, and hence are ignored in the uniqueness filter. This means e.g. that an epitope combination of a target (>15 amino acids) would be considered unique, even if a peptide shorter than 12 amino acids has

the same combination. On the other hand, a peptide from 12 to 15 amino acids in length would not be used for optimization, even if the epitope combination is unique, because the binding of both antibodies is not considered to be reliable.

To obtain information about individual antibodies the structure contains additional information for each binding molecule. Thus, for each antibody the number of capture-detector pairs it takes part in is stored. Furthermore, it is stored if for a capture-detector pair one of the binding molecules is already contained in the assay.

Based on this structure a weighted score for each binding molecule is calculated, this score returns the number of peptides it captures for which the corresponding capture or detector binding molecule already exists plus the number of peptides it captures for which the corresponding capture or detector binding molecule does not exist yet. This score is the base for the iterative optimization scheme. In this scheme the set of binding molecules is incremented by the antibody which obtains the highest score. This scoring function is corrected for the capturing of two or more peptides from the same protein.

After each iteration, the data structure is updated to account for inclusion of a new binding molecule into the assay. This update procedure removes all peptide information of proteins which are already covered through a peptide. This procedure is repeated until for all proteins at least one peptide is captured by such a capture-detector pair.

The usage of TXP antibodies in the Sandwich Immunoassay approach results in a dramatic reduction of required antibodies compared to the standard one antibody one analyte method. The reduction is 8 in the case of a digest with Trypsin with the aim to cover the whole proteome, as can be seen in table 6 below.

database / list	used Protease	Proteins covered	required TXP antibodies	required antibodies for one antibody one analyte approach	reduction
Hupo Plasma Proteome	Lys C	8779	3155	17558	5,6
Hupo Plasma Proteome	Trypsin	8676	3164	17352	5,5
Uniprot + Hupo Plasma Proteome	Lys C	8355	4766	36710	7,7
Uniprot + Hupo Plasma Proteome	Trypsin	18229	4556	36458	8,0
Biomarker Dataset	Lys C	282	380	564	1,5
Biomarker Dataset	Trypsin	282	397	564	1,4

Table 6: Sandwich Immunoassay Approach

The approach or the results becomes even more interesting if other species are taken into account. Unless the TXP antibodies bind to short sequences they can be used in different t species – usually antibodies are specific for one or some species.

Furthermore based on an existing set of TXP antibodies sandwich immunoassays can be established without the need for generation of capture molecules. This results in a tremendous saving of assay development time and costs for capture molecule generation.

Claims

1. A method for determining *in silico* a set of selected target epitopes suited for the detection of at least one protein of a set of proteins of interest within a pool of sample proteins, the method comprising the steps of:

   a) providing a first list containing the sequences of said sample proteins;

   b) providing a second list containing the sequences of said proteins of interest;

   c) applying an *in silico* fragmentation to the sample proteins in said first list to generate a third list containing sequences of possible target peptides having possible target epitopes at their free N- and C-terminal ends, which possible target epitopes each consisting of three to five N- or C-terminal amino acids;

   d) generating a fourth list of sequences of valid target peptides by removing from said third list such possible target peptides that are less suited for said protein of interest identification;

   e) selecting from said possible target epitopes on said valid target peptides in said fourth list a set of selected target epitopes, said set of selected target epitopes containing at least one target epitope for each protein of interest in said second list, while at the same time containing an as small as possible number of target epitopes.

2. The method according to claim 1, wherein step d) is performed by applying at least one of the following target peptide removing filters to the possible target peptides in said third list:

- Met Filter for removing target peptides with a methionine in at least one of its terminal sequences,
- Unknown AA Filter for removeing target peptides with unknown amino acids,
- High Abundant Target Epitope Filter for removing target peptides comprising target epitopes with high abundance,
- Weight Filter for removing target peptides with similar weight and with the same target epitope, which cannot be separated by mass spectrometry,
- Length Filter for removing target peptides with a length not adapted to the resolution or the method of a subsequent detection technology, e.g. shorter than 15, 12 or 5 amino acids and/or longer than 30 amino acids,
- High Abundant Protein Filter for removing target peptides belonging to high abundant proteins,
- Protetypic Peptide Filter for removing target peptides not predicted to be proteotypic.

3. The method according to claim 2, wherein said target peptide removing filters are applied in a specific order, in which order the Met Filter, the Unknown AA Filter and the High Abundant Protein Filter are used at any point within the order.

4. The method according to claim 3, wherein the High Abundant Target Epitope Filter is applied within said order prior to the Weight Filter and prior to the Length Filter.

5. The method according to claim 3 or claim 4, wherein the Weight Filter is applied within said order prior to the Length Filter.

6. The method according to anyone of claims 1 to 5, wherein in connection with step d) possible target peptides not belonging to a protein of interest are removed.

7. The method according to anyone of claims 1 to 6, wherein prior to step d) possible target peptides comprising a combination of N- and C-terminal possible target epitopes which is not unique amongst the possible target peptides are removed.

8. The method according to anyone of claims 1 to 7, wherein the in silico fragmentation in step b) is performed according to the sequence characteristics of a sequence specific proteolytic digest.

9. The method according to anyone of claims 1 to 8, wherein during step e) an optimization procedure is carried out for determining a globally optimal set of selected target epitopes, e.g. by using an ILP-solver or by determining a locally optimal set of selected target epitopes, e.g. by applying a standard greedy approach.

10. The method according to anyone of claims 1 to 9, wherein steps c) and d) are repeated in order to generate a second set of selected target epitopes, whereby for each possible target peptide in said third list, the selected target epitopes in the one set thereof comprise an N- terminal epitope and in the other set thereof comprise a C-terminal epitope.

11. A set of selected target epitopes, determined according to the method of anyone of claims 1 though 10.

12. A set of selected target epitopes suited for identifying proteins of interest, wherein the proteins of interest serve as Biomarkers, and the set of selected target epitopes contains at least 30%, preferably 50%, more preferably at least 70% of the set of claim 11

13. A method for preparing a set of binding molecule suited for the detection of peptides or proteins of interest from a pool of proteins, comprising the steps of:

    a) determining a set of selected target epitopes according to the method in anyone of claims 1 to 10; and

    b) generating for each target epitope in said set of selected target epitopes a binding molecule.

14. A set of binding molecules generated according to the method of claim 13.

15. A set of binding molecules, containing at least one binding molecule for each target epitope within the set of selected target epitopes of claim 11 or claim 12.

16. The set of binding molecules of claim 14 or claim 15, said binding molecules being selected from the group consisting of antibodies, antibody fragments, antibody variants, aptamers, polynucleotides, recombinant binding molecules.

17. A method for detecting proteins of interest in a biological sample, comprising the steps:

    a) providing a set of binding molecules according to any one of claims 14 through 16,

b) incubating said biological sample with a proteolytic agent to provide a peptide solution, said proteolytic agent having the same cleavage rules as the fragmentation applied *in silico* for determining the set of target epitopes used for generating the set of binding molecules from step a),

c) incubating said peptide solution with said set of binding molecule, and

d) determining the peptides bound to said set of binding molecules.

18. The method according to claim 17, wherein the determining step d) comprising the step of

e) incubating said peptide solution with a second set of binding molecules, for each peptide there being provided in one set of binding molecules a binding molecule for an N-terminal epitope, and in the other set of binding molecules a binding molecule for a C-terminal epitope.

19. The method according to claim 17 or 18, wherein the target epitope is selected from one of the set of selected target epitopes of claim 11 or 12.

20. The method according to anyone of claims 17 to 19, wherein the protein of interest is a Biomarker protein:

21. The method according to anyone of claims 17 to 20, wherein step d) is being performed using at least one of the following methods: mass spectroscopy, immunoassay, chromatography, electrophoresis, electrochemistry, surface plasmon resonance, oscillating crystal.

22. The method according to anyone of claims 17 to 21, wherein step c) and/or step e) is performed by using binding molecules immobilized to a carrier, the

carrier being selected from the group consisting of material for affinity chromatography, beads, microspheres, capillary surfaces, micro canal structures, sensory surfaces.

23. Use of the binding molecules of claim 15 or claim 16 for identifying proteins serving as Biomarkers.

24. A computer program for carrying out the method of anyone of claims 1 though 10.

25. A computer programmed for carrying out the method of anyone of claims 1 though 10.

26. A data carrier, containing information for carrying out the program of claim 24.

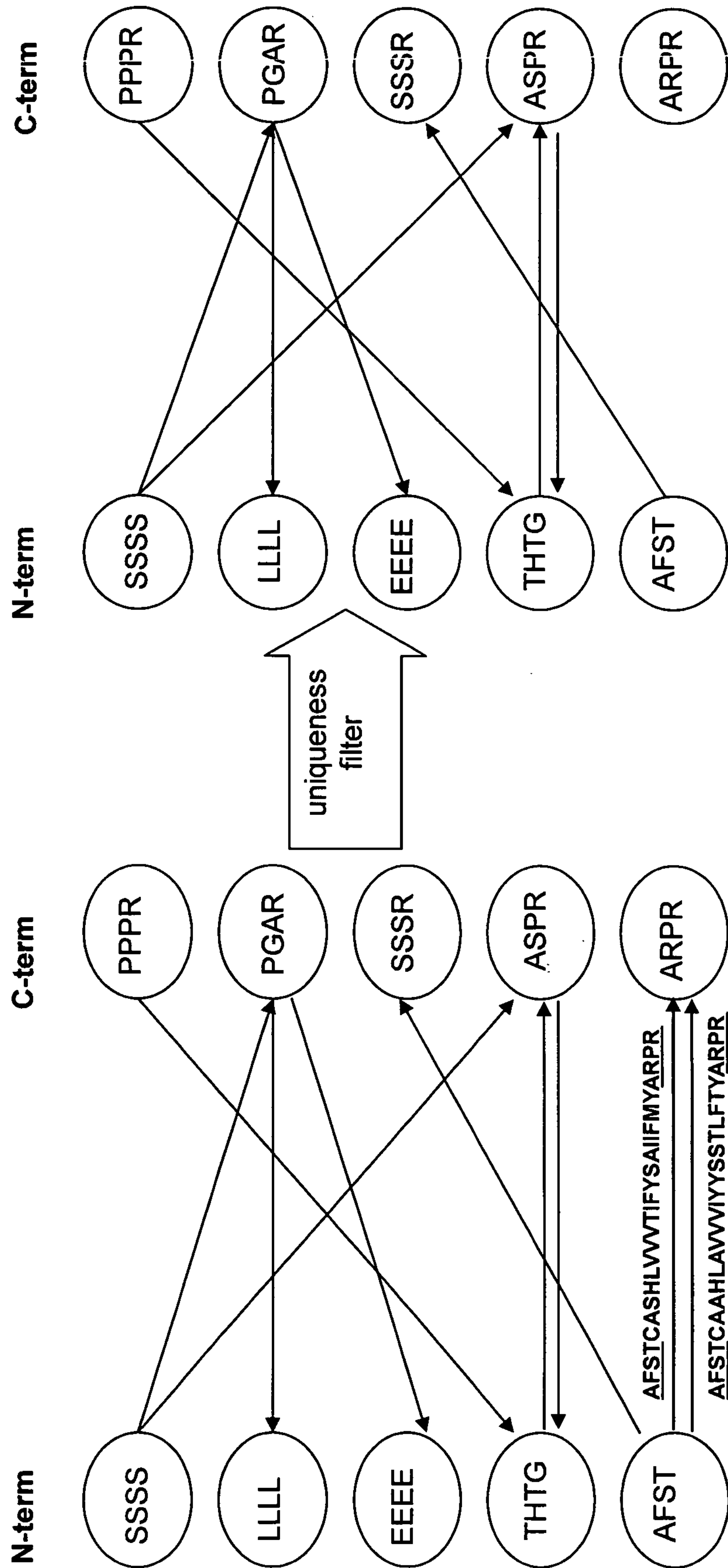

Fig. 1

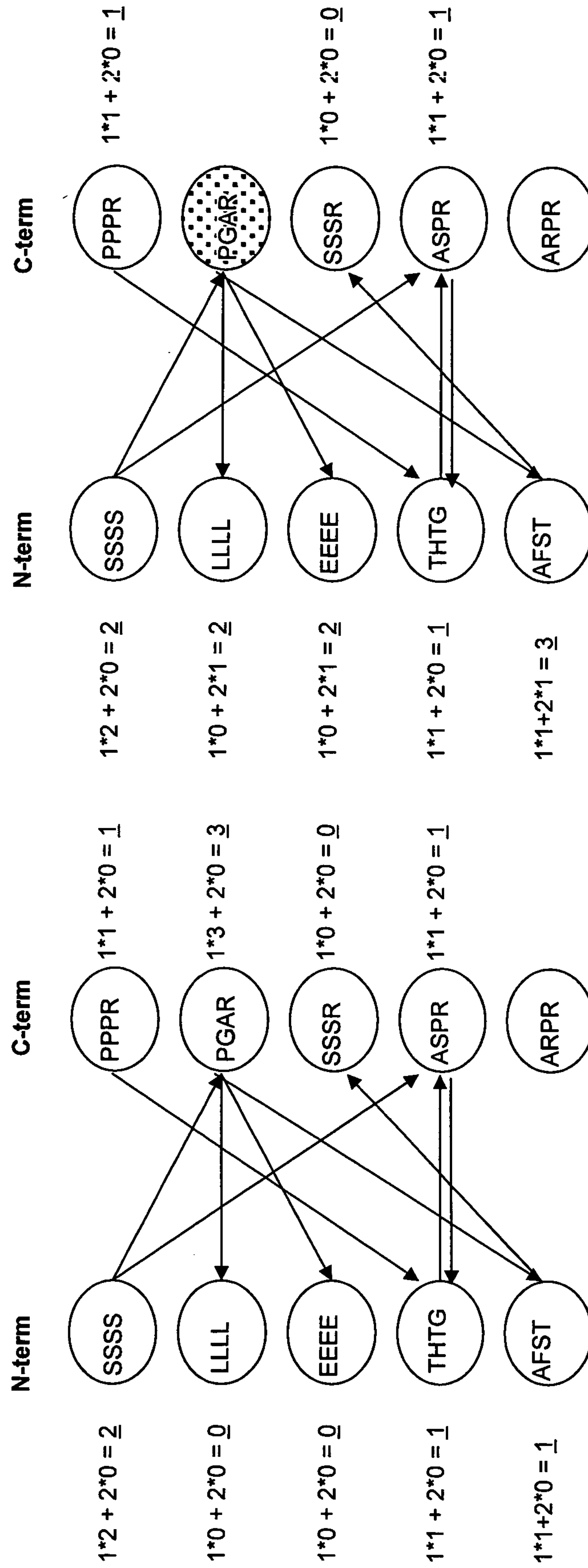

Fig. 2a

Fig. 2b

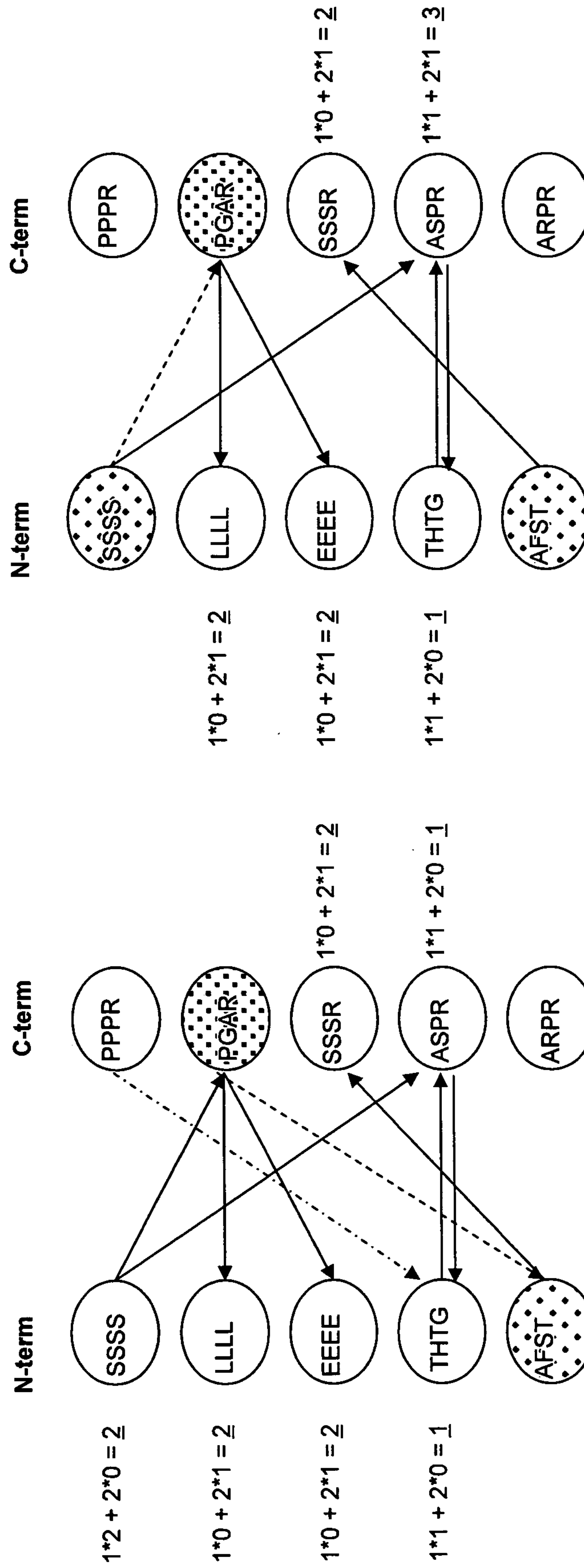

Fig. 2c

Fig. 2d

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/001230

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02/086081 A2 (UNIV CARNEGIE MELLON [US]; MINDEN JONATHAN [US]; RAVI RAMAMOORTHI [US]) 31 October 2002 (2002-10-31) cited in the application abstract; figures 2-6 paragraph [0041] - paragraph [0047] paragraph [0077] - paragraph [0085] paragraph [0098] - paragraph [0101]	1-10, 24-26
X	US 2005/255491 A1 (LEE FRANK D [US] ET AL) 17 November 2005 (2005-11-17) paragraph [0110] paragraph [0117] - paragraph [0119] paragraph [0472] - paragraph [0488] abstract examples 1-3	1-10, 24-26

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 July 2009	05/11/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Türkeli, Yasemin

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2004/081575 A1 (BIOINVENT INT AB [SE]; NILSSON FREDRIK [SE]) 23 September 2004 (2004-09-23) cited in the application abstract page 27, line 16 - page 28, line 6 page 22, line 19 - page 24, line 3	1-10, 24-26
A	SIEST G ET AL: "Functional genomics towards personalized healthcare" PERSONALIZED MEDICINE, FUTURE MEDICINE LTD., LONDON, GB, vol. 6, no. 1, January 2009 (2009-01), pages 19-32, XP008107739 ISSN: 1741-0541 page 21, right-hand column, paragraph 3 - page 22, left-hand column, paragraph 2	1-10, 24-26

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2009/001230

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers allsearchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search reportcovers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-10, 24-26

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-10, 24-26

   Method for determining in silico a set of selected target epitopes

   ---

2. claims: 11-12

   Sets of target epitopes

   ---

3. claims: 13-16, 23

   Sets of binding molecules and their use

   ---

4. claims: 17-22

   A method of detecting proteins of interest in a sample

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2009/001230

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 02086081	A2	31-10-2002	AU	2002338428	A1	05-11-2002
US 2005255491	A1	17-11-2005	NONE			
WO 2004081575	A1	23-09-2004	AT	412187	T	15-11-2008
			AU	2004219906	A1	23-09-2004
			CA	2518632	A1	23-09-2004
			DK	1604209	T3	23-02-2009
			EP	1604209	A1	14-12-2005
			ES	2319422	T3	07-05-2009
			JP	2006523829	T	19-10-2006