



US 20060154318A1

(19) **United States**

(12) **Patent Application Publication**
Anderson

(10) **Pub. No.: US 2006/0154318 A1**

(43) **Pub. Date: Jul. 13, 2006**

(54) **STABLE ISOTOPE LABELED POLYPEPTIDE
STANDARDS FOR PROTEIN
QUANTITATION**

(76) Inventor: **Norman L. Anderson**, Washington, DC
(US)

Correspondence Address:
Glenna Hendricks, Esq.
P.O. Box 2509
Fairfax, VA 22031-2509 (US)

(21) Appl. No.: **11/147,397**

(22) Filed: **Jun. 8, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/578,274, filed on Jun. 9, 2004. Provisional application No. 60/602,908, filed on Aug. 18, 2004.

Publication Classification

(51) **Int. Cl.**
G01N 33/537 (2006.01)
C12P 21/06 (2006.01)
C07K 14/47 (2006.01)
G01N 33/543 (2006.01)
(52) **U.S. Cl.** **435/7.92; 436/518; 435/68.1;**
530/350

(57) **ABSTRACT**

This invention relates to proteins having an amino acid sequence containing several amino acid subsequences found in nature and wherein at least two different subsequences act as monitor sequences, said subsequences being part of at least one natural protein which is a target protein, wherein the end of each of said two different subsequences have a cleavage site that will be cleaved by the same site-specific proteolytic treatment to release said subsequences.

Figure 1

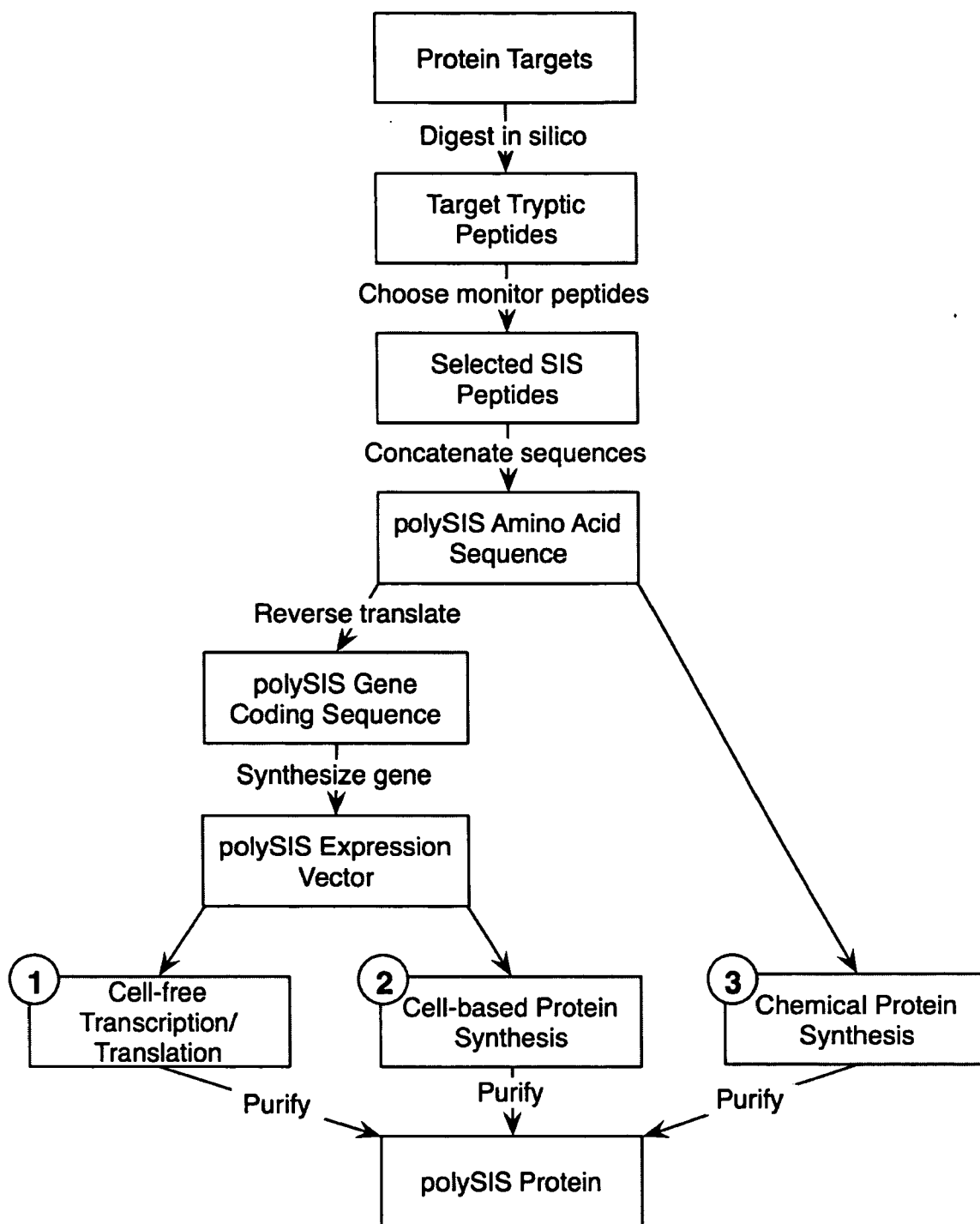


Figure 2

Protein	Peptide selected	Sequence	whole peptide				c-term residue		
			¹⁶ N	¹³ C	² H	¹⁸ O	¹⁶ N	¹³ C	¹⁶ N+ ¹³ C
IL-6	P05231_30_212_trypsin_10_54_65	EALAENLNLPK (Seq. ID No. 4)	16	57	89	19	2	6	8
Hemopexin	P02790_24_462_trypsin_8_69_79	NFPSPVDAAFR (Seq ID No. 5)	11	56	71	15	4	6	10
Alpha-1-Antichymotrypsin	P01011_24_423_trypsin_31_284_292	EIGELYLPK (Seq. ID No. 6)	10	50	75	14	2	6	8
TNF-alpha	P01375_1_233_trypsin_8_66_78	DLSLISPLAQAVR (Seq. ID No. 7)	17	61	96	18	4	6	10
Influenza haemagglutinin	HA-tag	YPYDVPDYA (Seq ID No. 8)							
Histidine tag	His-tag	HHHHHH (Seq. ID No. 9)							

Figure 3

Term	Explanation
asn * -2 +	Susceptible to deamidation, which would change the mass of a peptide containing it.
gln * -2 +	Susceptible to deamidation, which would change the mass of a peptide containing it.
met * -3 +	Susceptible to oxidation, which would change the mass of a peptide containing it.
trp * -3 +	Susceptible to oxidation, which would change the mass of a peptide containing it.
cys * -10 +	Susceptible to oxidation, which would change the mass of a peptide containing it.
pro * 5 +	Produces enhanced peptide structure and can improve immunogenicity
kp * 2 +	Results in an additional positive charge in a peptide
rp * 2 +	Results in an additional positive charge in a peptide
dp * 5 +	Introduces a site of efficient gas-phase fragmentation and improved MS/MS detection
chymo_sites * -3 +	Peptide more likely to be degraded by chymotryptic activity
occurrences_in_protein * 2 +	Multiply occurring peptide would give better detection
carbohyd * -50 +	Glycosylation of a peptide changes its mass and decreases MS detection
mod_res * -50 +	Any documented chemical modification of a residue in the peptide likely changes its mass
conflicts * -20 +	Any potential sequence errors in the peptide decrease its usefulness as an MS analyte
variants * -10 +	Any genetic variants in the peptide decrease its usefulness as an MS analyte by restricting its use to a subset of patients
sign(calc_mass-800)*100 +	Applies a negative value to mass less than 800
sign(1800-calc_mass)*100 ;	Applies a negative value to mass greater than 2,000
sign(detect - 0.0001) * 10 + detect * 20 +	Gives 10 points for detection in the Pounds experimental MS/MS data, up to 30 if the peptide is the most frequently detected for its protein (detect is the aforementioned index of detectability)

Figure 4

Protein	Peptide compound accession	fmol per ml	Peptide sequence	Occurrences _in_protein	Variants	Mod_res	Conflicts	Carbohyd
Apolipoprotein A-I	P02647_25_267_trypsin_32_196_208	4.8E+07	ATEHLSTLSEK (Seq. ID No. 10)	1	1	0	0	0
Alpha-1-acid glycoprotein 1	P02763_19_201_trypsin_12_121_135	3.2E+07	NWGLSVYADKPETTK (Seq. ID No. 11)	1	0	0	0	0
Haptoglobin beta chain	P00737_162_406_trypsin_1_1_9	2.3E+07	ILGGHLDK (Seq. ID No. 12)	1	0	0	0	0
Fibrinogen gamma chain	P02679_27_453_trypsin_17_141_151	1.5E+07	DTVQIHDITGK (Seq. ID No. 13)	1	0	0	0	0
Fibrinogen alpha chain	P02671_36_640_trypsin_56_433_441	1.5E+07	TVIGPDGHK (Seq. ID No. 14)	2	0	0	0	0
Fibrinogen beta chain	P02675_45_491_trypsin_31_257_269	1.5E+07	QFGGNVATNTDGK (Seq. ID No. 15)	1	0	0	0	0
Alpha-1-antichymotrypsin	P01011_24_423_trypsin_31_284_292	8.8E+06	EIGELYLPK (Seq. ID No. 16)	1	0	0	0	0
Complement C3	P01024_23_1663_trypsin_105_883_891	7.1E+06	TGLQEVEVK (Seq. ID No. 17)	4	0	0	0	0
Antithrombin-III	P01008_33_464_trypsin_48_360_370	4.1E+06	DDLIVSDAFHK (Seq. ID No. 18)	1	0	0	0	0
Ceruloplasmin	P00450_20_1065_trypsin_17_159_168	2.3E+06	IYHSHIDAPK (Seq. ID No. 19)	1	0	0	0	0
Prothrombin	P00734_44_622_trypsin_46_444_455	1.5E+06	ETAASLLQAGYK (Seq. ID No. 20)	2	0	0	1	0
Complement C4 gamma chain	P01028_1454_1744_trypsin_26_194_202	1.4E+06	ITQVLHFTK (Seq. ID No. 21)	1	0	0	1	0
Apolipoprotein B-100	P04114_28_4563_trypsin_416_3764_3772	1.4E+06	FPEVDLTK (Seq. ID No. 22)	1	0	0	0	0
Alpha-2-antiplasmin	P08697_40_491_trypsin_2_13_25	1.4E+06	LGNQEPGGQTALK (Seq. ID No. 23)	1	0	0	0	0
Plasminogen	P00747_20_810_trypsin_73_652_661	1.2E+06	LSSPAVITDK (Seq. ID No. 24)	2	0	0	0	0
Apolipoprotein E	P02649_19_317_trypsin_46_275_282	1.2E+06	QWAGLVEK (Seq. ID No. 25)	1	0	0	0	0
Coagulation factor XIIIa heavy chain	P00748_20_372_trypsin_1_1_8	4.6E+05	IPPWEAPK (Seq. ID No. 26)	1	0	0	0	0
Apolipoprotein(a)	P08519_20_4548_trypsin_249_4385_4398	2.8E+05	LFLEPTQADIALLK (Seq. ID No. 27)	1	0	0	0	0
Coagulation factor X	P00742_41_488_trypsin_56_436_448	2.0E+05	SHAPEVITSSPLK (Seq. ID No. 28)	3	0	0	0	0
Adiponectin	Q15848_19_244_trypsin_13_117_131	2.0E+05	IFYNQNNHYDGSTGK (Seq. ID No. 29)	1	0	0	0	0
Beta-2-glycoprotein I	P02749_20_345_trypsin_34_309_317	1.5E+05	EHSSLAFWK (Seq. ID No. 30)	1	1	0	0	0
Coagulation factor IX	P00740_47_461_trypsin_16_135_142	1.1E+05	VSVSQTSK (Seq. ID No. 31)	2	0	0	0	0
C-reactive protein	P02741_19_224_trypsin_4_14_23	1.0E+05	ESDTSYVSLK (Seq. ID No. 32)	1	0	0	0	0
Vitamin K-dependent protein C	P04070_43_461_trypsin_32_234_241	7.8E+04	WELDLDIK (Seq. ID No. 33)	2	1	0	0	0
Coagulation factor XIII A chain	P00488_38_731_trypsin_62_587_598	6.6E+04	STVLTPEIIK (Seq. ID No. 34)	1	0	0	0	0
Coagulation factor XIII B chain	P05160_21_661_trypsin_25_198_208	6.6E+04	LIENGYFHPVK (Seq. ID No. 35)	1	0	0	0	0
Cholesteryl ester transfer protein	P11597_18_493_trypsin_5_38_47	3.6E+04	ASYPDITGEK (Seq. ID No. 36)	2	0	0	0	0
Coagulation factor V	P12259_29_2224_trypsin_92_898_907	2.7E+04	DPPSDLLLLK (Seq. ID No. 37)	1	0	0	0	0
Angiotensinogen	P01019_34_485_trypsin_6_50_61	2.0E+04	ALQDQLVLVAAK (Seq. ID No. 38)	1	0	0	0	0
L-selectin	P14151_39_372_trypsin_6_33_40	1.8E+04	AEIEYLEK (Seq. ID No. 1)	1	0	0	0	0

Figure 5

1 M A K A T E H L S T L S E K N W G L S V
3 ATGGCAAAGCAACAGAACATTTAAGTACATTATCAGAAAAAATTGGGGATTATCAGTA

21 Y A D K P E T T K I L G G H L D A K D T
63 TATGCAGATAAACAGAACTACAAAAATTTTAGGAGGACATTTAGACGCAAAGACACA

41 V Q I H D I T G K T V I G P D G H K Q G
123 GTTCAAATTCATGATATCACAGGAAAAACAGTAATCGGACCAGACGGACATAAACAAGGT

61 F G N V A T N T D G K E I G E L Y L P K
183 TTCGGTAACGTAGCAACAAATACAGACGGAAAAGAAATTGGAGAATTATATTTACCAAAG

81 T G L Q E V E V K D D L Y V S D A F H K
243 ACAGGTTTACAAGAAGTAGAAGTAAAAGACGATTTATATGTATCAGACGCATTTTACATAAA

101 I Y H S H I D A P K E T A A S L L Q A G
303 ATTTACCATTACACATCGACGCACCAAAGAAACAGCAGCATCATTATTACAAGCAGGT

121 Y K I T Q V L H F T K F P E V D V L T K
363 TATAAAATCACACAAGTATTACATTTCCACAAAATTTCCAGAAGTAGACGTATTAACAAAA

141 L G N Q E P G G Q T A L K L S S P A V I
423 TTAGGAAATCAGGAACCAGGAGGTCAAACAGCATTAAAATTATCATCACCAGCAGTAATC

161 T D K Q W A G L V E K I P P W E A P K L
483 ACAGATAAACAATGGGCAGGATTAGTAGAAAAGATTCCACCTTGGGAAGCTCCTAAATTA

181 F L E P T Q A D I A L L K S H A P E V I
543 TTCTTAGAACCAACACAAGCAGATATTGCATTATTAATCTCACGCACCAGAAGTAATT

201 T S S P L K I F Y N Q Q N H Y D G S T G
603 ACCAGTTCACCATTAAAAATTTTTTACAACCAACAAAATCATTACGACGGATCAACAGGA

221 K E H S S L A F W K V S V S Q T S K E S
663 AAAGAACACTCATCATTAGCATTTTGGAAAGTATCAGTTAGTCAAACCTTCAAAGGAATCA

241 D T S Y V S L K W E L D L D I K S T V L
723 GACACAAGTTATGTATCATTAATGGGAATTAGACTTAGACATCAAATCAACAGTACTT

261 T I P E I I I K L I E N G Y F H P V K A
783 ACCATTCCAGAAATTATTATTAATTAATCGAAAACGGATATTTTACCCAGTTAAAGCA

281 S Y P D I T G E K D P P S D L L L L K A
843 TCATACCCAGATATTACAGGAGAAAAAGACCCACCAAGTGACTTATTATTATTAAGCA

301 L Q D Q L V L V A A K A E I E Y L E K Q
903 TTACAAGACCAATTAGTATTAGTAGCAGCAAAGCAGAAATTGAATATTTAGAAAAACAG

321 P G (Seq. ID No. 39)
963 CCCGGG (Seq. ID No. 40)

STABLE ISOTOPE LABELED POLYPEPTIDE STANDARDS FOR PROTEIN QUANTITATION

[0001] This application takes priority from U.S. Provisional Patent Application 60/578,274 filed Jun. 9, 2004, and U.S. Provisional Patent Application 60/602,908 filed Aug. 19, 2004.

FIELD OF THE INVENTION

[0002] This invention relates to quantitative assays for evaluation of proteins in complex samples such as human plasma, and specifically to the generation and use of labeled peptides as Stable Isotope Standards (SIS). It would be useful to be able to produce large numbers of different SIS peptides more cheaply than can be accomplished by chemical synthesis, to purify them more efficiently than can be accomplished by individual HPLC purification, and to quantify them by some means more efficiently than amino acid analysis of each peptide individually. Here I describe a strategy for making sets of SIS standards by protein expression. The invention can be used both for analysis of samples from a single individual source or, for purposes of evaluating the level of a particular protein in a population, can be used to analyze pooled samples from the target population.

[0003] There is a need for quantitative assays for proteins in various complex protein samples, e.g., in human plasma, serum and urine. Conventionally these assays have been implemented as immunoassays, making use of specific antibodies against target proteins as specificity and detection reagents. The current expansion of the diagnostic proteome suggests that the use of many protein measurements together as a panel provides superior diagnostic information compared to a single protein: here patterns of change can be associated with disease or treatment, instead of relying on single protein markers interpreted alone. This development presages the need to assay many more proteins than is currently feasible with existing immunoassays. New methods, particularly involving internal standardization with isotopically labeled peptides, allow mass spectrometry (MS) to provide large panels of such quantitative peptide and protein assays (as MS does in the measurement of low molecular weight drug metabolites currently). The efficient production, quantitative calibration and use of such standards remains an issue, however. The present invention addresses this problem by providing improvements in the manufacturing of multiple peptide standards, arranging such standards in fixed stoichiometries, and using them efficiently in assays of complex protein and peptide samples.

[0004] A general mass-spectrometry-based approach to protein quantitation involves digesting the proteins (e.g., with trypsin) into peptides that can be further fragmented (MS/MS) in a mass spectrometer to generate a sequence-based identification. The approach can be used with either electrospray (ESI) or MALDI ionization, and is typically applied after one or more dimensions of chromatographic fractionation to reduce the complexity of peptides introduced into the MS at any given instant. Optimized systems of multidimensional chromatography, ionization, mass spectrometry and data analysis (e.g., the multidimensional protein identification technology, or "MudPIT" approach of Yates, also referred to as shotgun proteomics) have been shown to be capable of detecting and identifying ~1,500 yeast proteins in one analysis (Washburn, Wolters and Yates,

Nat Biotechnol 19:242-7, 2001), while a single dimensional LC separation, combined with the extremely high resolution of a Fourier-transform ion cyclotron resonance (FTICR) MS identified more than 1,900 protein products of distinct open reading frames (i.e., predicted proteins) in a bacterium. In human urine, a sample much more like plasma than the microbial samples mentioned above, Patterson used a single LC separation ahead of ESI-MS/MS to detect 751 sequences derived from 124 different gene products. Recently, Adkins et al have used two chromatographic separations with MS to identify a total of 490 different proteins in human serum (Adkins, Varnum, Auberry, Moore, Angell, Smith, Springer and Pounds, Mol Cell Proteomics 1:947-55, 2002), and Anderson et al combined four datasets to generate a list of 1,175 non-redundant plasma components (Anderson, Polanski, Pieper, Gatlin, Tirumalai, Conrads, Veenstra, Adkins, Pounds, Fagan and Loble, Mol Cell Proteomics 2004). Such methods should have the ability to deal with the numerous post-translational modifications characteristic of many proteins in plasma, as demonstrated by the ability to characterize the very complex post-translational modifications occurring in aging human lens (MacCoss, McDonald, Saraf, Sadygov, Clark, Tasto, Gould, Wolters, Washburn, Weiss, Clark and Yates, Proc Natl Acad Sci USA 99:7900-5, 2002). Since 1995 a single peptide has been used as a surrogate for the presence of a parent protein (from which the peptide was derived by proteolytic digestion) in a complex protein mixture, based on, e.g., MALDI-PSD (Griffin, MacCoss, Eng, Blevins, Aaronson and Yates, Rapid Commun Mass Spectrom 9:1546-51, 1995) or ion trap (Yates, Eng, McCormack and Schieltz, Anal Chem 67:1426-36, 1995) MS/MS spectra. Regnier et al have pursued an equivalent "signature peptide" quantitation approach (Chakraborty and Regnier, J Chromatogr A 949:173-84, 2002, Zhang, Sioma, Wang and Regnier, Anal Chem 73:5142-9, 2001), also the subject of a published patent application (Regnier, F. E., X. Zhang, et al. US 2002/0037532), in which protein samples are digested to peptides by an enzyme, differentially labeled with isotopically different versions of a protein reactive agent, purified by means of a selective enrichment column, and combined for MS analysis using MALDI or ESI-MS.

[0005] The protein discovery methods described above focus on identifying peptides and proteins in complex samples, but they generally offer poor quantitative precision and reproducibility when used without internal standards. The well-known idiosyncrasies of peptide ionization arise in large part because the presence of one peptide can affect the ionization and, thus, signal intensity of another. These have been major impediments to accurate quantitation by mass spectrometry. This problem can be overcome, however, through the use of stable isotope-labeled internal standards. At least four suitable isotopes (^2H , ^{13}C , ^{15}N , ^{18}O) are commercially available in suitable highly enriched (>98 atom %) forms. In principle, abundance data as accurate as that obtained in MS measurement of drug metabolites with internal standards (coefficients of variation <5%) should ultimately be obtainable. In the early 1980's ^{18}O -labeled enkephalins were prepared and used to measure these peptides in tissues at ppb levels. In the 1990's GC/MS methods were developed to precisely quantitate stable isotope-labeled amino acids, and hence protein turnover, in human muscle and plasma proteins labeled in vivo. The extreme sensitivity and precision of these methods suggested that

stable isotope approaches could be applied in quantitative proteomics investigations, given suitable protein or peptide labeling schemes.

[0006] Over the past several years, a variety of such labeling strategies have been developed. The most straightforward approach (incorporation of label to a high substitution level during biosynthesis), has been successfully applied to microorganisms (Lahm and Langen, *Electrophoresis* 21:2105-14, 2000) and mammalian cells in culture, but is unlikely to be usable directly in humans for cost and ethical reasons. A related approach (which is applicable to human proteins) is the now-conventional chemical synthesis of monitor peptides containing heavy isotopes at specific positions. Post-synthetic methods have also been developed for labeling of peptides to distinguish those derived from an "internal control" sample from those derived from an experimental sample, with a labeled/unlabeled pair subsequently being mixed and analyzed together by MS. These methods include Aebersold's isotope-coded affinity tag (ICAT) approach, (Goodlett, Keller, Watts, Newitt, Yi, Purvine, Eng, von Haller, Aebersold and Kolker, *Rapid Commun Mass Spectrom* 15:1214-21, 2001) as well as deuterated acrylamide and iodoacetamide for labeling peptide sulphydrals, deuterated acetate to label primary amino groups, n-terminal-specific reagents, permethyl esterification of peptides carboxyl groups, and addition of twin ^{18}O labels to the c-terminus of tryptic peptides during cleavage.

[0007] An early quantitative MS-based assay for a peptide was published in 1989 by Jardine et al (Lisek, Bailey, Benson, Yaksh and Jardine, *Rapid Commun Mass Spectrom* 3:43-6, 1989). The reference discloses use of a single stable isotope labeled peptide (substance P sequence. Prepared by chemical peptide synthesis) spiked into neuronal tissue, followed (after extraction from the tissue) by binding to an immobilized anti-substance-P-specific antibody, to enrich the neuropeptide substance P, and finally quantitation by MS. Substance P abundance was calculated from the ratio of natural peptide ion current to the internal labeled standard peptide of the same sequence: i.e., demonstrating all elements of the single analyte peptide standard/antibody enrichment process. Jardine et al used a 10-fold molar excess of the labeled version of substance P to act as both internal standard and carrier, and measured masses by fast-atom bombardment (FAB) selected-ion monitoring (SIM) MS. Crowther published a similar approach in 1994 (Crowther, Adusumalli, Mukherjee, Jordan, Abuaf, Corkum, Goldstein and Tolan, *Anal Chem* 66:2356-61, 1994) to detect peptide drugs in plasma using deuterated synthetic internal standards. Rose used synthetic stable isotope labeled insulin to standardize an MS method for quantitation of insulin (a small protein or large peptide), in which the spiked sample was separated by reverse phase chromatography to fractionate the sample. Gygi used stable-isotope-labeled synthetic peptides to quantitate the level of phosphorylated vs non-phosphorylated peptides in the digest of a protein isolated on a 1-D gel (Stemmann, Zou, Gerber, Gygi and Kirschner, *Cell* 107:715-26, 2001, Gerber, Rush, Stemman, Kirschner and Gygi, *Proc Natl Acad Sci USA* 100:6940-5, 2003) and has described a method for peptide quantitation (WO03016861) that uses the approach of Jardine with the addition of greater mass spectrometer resolution (selected reaction monitoring [SRM] in which the desired peptide is isolated by a first mass analyzer, the peptide is fragmented in flight, and a specific fragment is

detected using a second mass analyzer). In each of these cases, the labeled peptide standards have been made by conventional solid-phase peptide synthesis.

[0008] The instant invention uses several of the cited methods of the prior art together with other technologies related to cell-free protein synthesis in an entirely novel combination. In the descriptions that follow, quantitation of proteins, peptides and other biomolecules is addressed in a general sense, and hence the invention disclosed is in no way limited to the analysis of plasma and other body fluids.

SUMMARY OF THE INVENTION

[0009] The present invention provides methods for the production, purification, characterization and use of stable-isotope-labeled peptide sequences which can be used together or separately as internal standards in the mass spectrometric quantitation of peptides and proteins. Briefly, one or more monitor peptide sequences are selected to represent each protein to be measured (the "analytes"). In the case of trypsin cleavage of the analyte-containing sample, candidate monitor peptides will be tryptic peptides (i.e., generally ending in K or R). A set of selected monitor peptide sequences representing multiple protein analytes is then concatenated to yield an extended amino acid sequence (a "polySIS" sequence) that can be reverse-translated to yield a DNA sequence, which can be prepared by chemical DNA synthesis and incorporated into an expression vector. Appropriate polySIS-containing vectors can be introduced into any of a variety of cell-based (e.g., *E. coli*) or cell-free (e.g., *E. coli* or rabbit reticulocyte) expression systems capable of linked transcription and translation, wherein the protein can be produced. Stable isotope labels can be incorporated into the polySIS protein product by providing as substrates to the expression system either a heavily isotope-substituted nutrient source (for a cell based system), or one or more heavily isotope-substituted amino acids (for an in vitro cell-free system). In either case isotopically-enriched ^{15}N or ^{13}C (preferably >99%) can be used as the input label to achieve a highly substituted product. The polySIS protein can be purified using specific tags incorporated into the expression vector sequence (e.g., poly-histidine at one or both ends or internally between SIS sequences) or based on physical properties such as solubility or size (i.e., on an SDS electrophoresis gel).

[0010] The intact polySIS protein can be quantitated once by amino acid analysis, yielding a molar concentration that applies to all the component SIS peptides subsequently liberated by proteolysis, thereby saving the cost and effort of individual amino acid analysis of each peptide separately. The polySIS protein can be added at known amounts to complex protein samples prior to proteolytic digestion, and digested with the sample proteins to produce a series of SIS peptides whose stoichiometry to one another is known, and whose absolute concentration is also known. Alternatively the polySIS can be pre-digested to yield a stoichiometric mixture of SIS peptides to be added to a sample before or after sample digestion. These SIS peptides are then used as standards for quantitation of sample protein derived peptides by mass spectrometry (e.g., as in the previously disclosed SISCAPA method disclosed in U.S. patent application Ser. No. 10/676,005 "High Sensitivity Quantitation of Peptides by Mass Spectrometry").

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] FIG. 1 shows a schematic diagram of the process for designing and producing polySIS proteins, beginning with a set of protein targets (analytes to be measured by MS).

[0012] FIG. 2 shows examples of four monitor peptides.

[0013] FIG. 3 shows a series of additive terms defining an index used to prioritize tryptic peptides in silico.

[0014] FIG. 4 shows monitor peptide sequences chosen to represent 30 proteins associated with cardiovascular disease and some of their relevant properties.

[0015] FIG. 5 shows DNA sequence of the assembled polySIS synthetic gene, and the corresponding amino acid sequence translated in the correct frame.

[0016] FIG. 6 shows the complete amino acid sequence of the expressed polySIS protein CVD_1, including n-terminal and c-terminal regions added by expression from the pIVEX2.4d vector.

[0017] FIG. 7 is a diagram showing the use of a polySIS protein.

DETAILED DESCRIPTION OF THE INVENTION

[0018] A principle object of the current invention is to provide a convenient means for producing stable-isotope-labeled peptide standards useful in quantitative analysis of a mixture of peptides (typically a proteolytic digest of a complex protein sample such as human serum or plasma). The object is to produce such standards by a method that 1) is less expensive overall than conventional individual synthesis approaches, 2) allows more efficient purification (many SIS at once instead of one at a time), 3) provides an efficient means of assaying the quantity of the standard in absolute terms, and 4) ensures proper stoichiometry of a series of different SIS standards.

[0019] The terms “analyte”, and “ligand” may be any of a variety of different molecules, or components, pieces, fragments or sections of different molecules that one desires to measure or quantitate in a sample.

[0020] The term “monitor fragment” may mean any piece of an analyte up to and including the whole analyte which can be produced by a reproducible fragmentation process (or without a fragmentation if the monitor fragment is the whole analyte) and whose abundance or concentration can be used as a surrogate for the abundance or concentration of the analyte.

[0021] The term “monitor peptide” means a peptide chosen as a monitor fragment of a protein or peptide, and is typically a peptide of length 8-24 amino acids resulting from proteolytic treatment of the analyte (or target) protein.

[0022] The terms “proteolytic treatment” or “enzyme” may refer any of a large number of different enzymes, including trypsin, chymotrypsin, lys-C, V8 protease and the like, as well as chemicals, such as cyanogen bromide. In this context, a proteolytic treatment acts to cleave peptide bonds in a protein or peptide in a sequence-specific manner, generating a collection of shorter peptides (a digest).

[0023] The term “denaturant” includes a range of chaotropic and other chemical agents that act to disrupt or loosen the 3-D structure of proteins without breaking covalent bonds, thereby rendering them more susceptible to proteolytic treatment. Examples include urea, guanidine hydrochloride, ammonium thiocyanate, as well as solvents such as acetonitrile, methanol and the like.

[0024] The term “reverse-phase matrix” and “C18” are meant to include any of a variety of hydrophobic surface phases (such as C18 or C8 aliphatic hydrocarbons) presented on the surface of a solid support and in contact with aqueous solvent.

[0025] The terms “internal standard”, “isotope-labeled monitor fragment”, or “isotope-labeled monitor peptide” may be any altered version of the respective monitor fragment or monitor peptide that is 1) recognized as equivalent to the monitor fragment or monitor peptide in any separation process employed before MS detection and 2) differs from it in a manner that can be distinguished by a mass spectrometer, either through direct measurement of molecular mass or through mass measurement of fragments (e.g., through MS/MS analysis), or by another equivalent means.

[0026] By a “SIS” or “stable isotope standard” I mean a peptide internal standard having a unique sequence derived from a protein of interest and including a label of some kind (e.g., a stable isotope) that allows its use as an internal standard for quantitation (see U.S. patent application Ser. No. 10/676,005 “High Sensitivity Quantitation of Peptides by Mass Spectrometry”).

[0027] By “polySIS” I mean a polypeptide or protein composed of multiple SIS peptide sequences, and which may or may not include stable isotope labels.

[0028] The term “multiple reaction monitoring”, abbreviated MRM, means a mass spectrometric assay based on two stages of mass selection. In MRM, the first mass analyzer within the MS (MS1, also called quadrupole 1 or Q1) is set to pass the parent molecule (the monitor peptide), rejecting components of other mass-to-charge ratios (m/z). The monitor peptide is then fragmented in a collision chamber and passed to a second mass analyzer (MS2, also called quadrupole 3 or Q3) set to pass a known specific fragment of the monitor peptide. This two-stage selection of parent and fragment ions (selected reaction monitoring: SRM, plural MRM) affords great specificity, with the result that the detected signal usually traces a peak in the chromatogram at the expected retention time corresponding to the selected analyte. Integrating this peak gives a measure of the quantity of the analyte.

[0029] The term “cell-free” expression system means a combination of molecules capable of producing protein from an input DNA sequence. Examples include, but are not limited to, cell-free extracts of bacteria (like *E. coli*) or eukaryotic cells (like rabbit reticulocytes) containing transcription and translation systems, together with appropriate accessory activities required to make mRNA and protein.

Embodiments

[0030] In each of the following embodiments, it is to be assumed that the preferred method of use can include other elements of the SISCAPA system described in US2003/031126.

[0031] 1) In a first embodiment, a polySIS protein is prepared according to the steps shown in **FIG. 1** (track 1). First a set of protein targets is selected whose amounts or concentrations are to be measured in one or more samples. These targets are “digested” in silico using an algorithm appropriate for the desired protease (e.g., for trypsin cut at K and R, except where followed by P) to yield a set of target tryptic peptides. From these candidate peptides, monitor peptides may be selected using information including the predicted physical properties of these peptides and available experimental data (e.g., which “fly” best in a mass spectrometer), selecting those optimal properties for detection, enrichment, etc. Multiple peptides can be selected from a single target protein in order to provide multiple independent measurements of the target, thus improving measurement statistics.

[0032] The monitor peptide sequences selected for use as stable isotope labeled internal standards (SIS), each including the cleavage site-defining K or R residue recognized by trypsin, are concatenated together in silico to yield a single polypeptide sequence. The number of peptides combined in this way can range from 2 to 100 or more, depending on the number of monitor peptides required to provide adequate measurements of the set of protein targets selected. In this embodiment, each monitor peptide sequence is included once in the concatenated polypeptide (although multiple copies of one monitor peptide can be used to achieve different, but integral, stoichiometries). The order of the monitor peptide sequence in the concatenated polypeptide is not of great significance, provided that the final proteolytic digestion is complete, as desired. Some adjustment of peptide order may be required if concatenation brings together sequences that inhibit complete cleavage at every intended cleavage site. Optionally, additional peptide sequences may be added to one or both ends of the concatenated monitor peptide sequence to provide “handles” for use in specific affinity purification of the concatenated protein product. For example, influenza hemagglutinin (HA) tag sequences can be added at one or both ends of the polySIS product to assist in purification of the polySIS protein. The tag sequences are separated from the n- and c-terminal monitor peptides by protease (e.g., trypsin) cleavage sites (“separator sequences”; e.g., the added K in **FIG. 2**) so that the tags are separated from the monitor peptides upon digestion. Multiple different purification tags may be used (e.g., HA and polyhistidine tags in **FIG. 2** case 2). Different monitor peptide sequences may be included in different copy numbers in order to achieve different (integral) stoichiometries upon digestion (**FIG. 2.**, case 4)

[0033] The complete polySIS sequence (comprising the monitor peptides, optional purification tags, and any required separator sequences) is reverse-translated into a DNA sequence using the appropriate genetic code, with codon usage optimized for translation in a suitable production organism such as *E coli* or a cell-free system based on *E coli* or rabbit reticulocytes, to yield a polySIS gene coding sequence.

[0034] Next, this DNA sequence is synthesized to produce a double-stranded polySIS DNA sequence (“polySIS gene”) using commercially available services and expertise (e.g., Blue Heron Biotechnology, GeneScript Corp., or SeqWright Inc.). In this process, the polySIS gene may be introduced into a temporary vector to facilitate generation of more

DNA, or introduced directly into an expression vector appropriate for expression in a coupled in vitro transcription/translation system. A 1kb DNA sequence (approximately 330 amino acids) is easily produced by current commercial technology, and can accommodate 30 SIS peptides of 11 amino acids. Codon usage is preferably optimized to suit the source of the translation system (e.g., *E coli*). In this embodiment, the polySIS expression vector (e.g., Roche Applied Science pIVEX2.4d vector) includes additional sequences required to initiate transcription (e.g., by a bacterial or phage DNA-dependent RNA polymerase), initiate translation on the resulting RNA (ribosome binding and translation initiation sites) and stop translation (a stop codon). This DNA construct can be made entirely by synthesis and ligation, without the need for cloning into a vector, or the extra sequences can be included in a vector optimized for in vitro transcription/translation.

[0035] In either case, the polySIS molecule is introduced into a suitable linked in vitro transcription/translation system (e.g., the commercially available systems based on *E coli* or rabbit reticulocyte lysates) and polySIS protein product is generated. The translation system used preferably requires an exogenous source of amino acids, and in this embodiment at least one amino acid is provided that contains a stable isotope at high enrichment. The different SIS sequences comprising the polySIS product contain varying amino acids, and thus the mass increments in the various peptides resulting from use of a collection of labeled amino acids can be quite variable. A useful simplification results, however, if labeled K and R are used exclusively, since each tryptic SIS peptide contains only one such residue (either K or R) per peptide (except for rare cases in which a KP or RP occurs within the peptide). Using this K/R labeling approach, each SIS peptide is 6 amu heavier than the natural version if K and R fully substituted with ^{13}C is used, or 2 and 4 amu respectively if K and R fully substituted with ^{15}N is used, or 8 and 10 if K and R fully substituted with both isotopes are used (**FIG. 3**). A difference of at least 6 amu is preferred so that the SIS and natural peptides are far enough apart to avoid any overlap of SIS with the normal isotopic distribution of the natural unlabeled form. The polySIS protein product formed in the linked transcription/translation system is purified for use as an internal standard as described in the first embodiment.

[0036] Standard techniques, including affinity capture by chelated nickel adsorbents (in the case of histidine tags) or immobilized anti-HA antibodies (in the case of HA tags). The polySIS protein is recovered in a state of high purity (preferably greater than 95%). Alternatively a physical separation such as SDS gel electrophoresis can be used, and the polySIS protein band excised. An aliquot of purified polySIS protein is hydrolyzed in HCl to liberate amino acids, and these are quantitated by amino acid analysis to establish the absolute amount of polySIS protein present. Alternatively the polySIS protein can be assayed by other means such as quantitation of a substituent such as biotin introduced at fixed stoichiometry during synthesis. Using this quantitative information, solutions or dried aliquots of polySIS containing accurately known amounts of material are prepared as standards.

[0037] A known amount of polySIS (i.e., a known volume of standardized solution) is then added to a measured volume of a sample of proteins in which the target proteins

are to be quantitated (in this case a sample of human blood plasma). This combined sample, including spiked polySIS standard, is then proteolytically digested by exposure to trypsin using any of a variety of well-known protocols. In one such protocol, plasma is denatured by addition of 9 volumes of 6 M guanidinium HCl/50 mM Tris-HCl/10 mM dithiothreitol and incubation for 2 hr at 60° C.; addition of 1 volume of 200 mM iodoacetamide followed by incubation for 30 min at 25° C.; addition of 1 volume of 200 mM dithiothreitol followed by incubation for 30 min at 25° C.; dilution to <1 M guanidinium HCl by addition of 50 mM NaHCO₃, addition of sequencing grade modified trypsin (e.g., from Promega, Madison, Wis.) at a 1:50 ratio (trypsin:plasma protein) and incubation overnight at 37° C. Digestion is allowed to proceed until substantially complete, liberating the monitor peptides from both target proteins and polySIS protein essentially to completion. Alternatively a mixture of SIS resulting from prior digestion of polySIS protein can be added to the sample before or after sample digestion. This sample digest now contains versions of monitor peptides containing natural isotopes (from peptides derived from the original sample) and stable isotopes (in the SIS peptides derived from the polySIS protein). In this embodiment, each SIS sequence is present only once in the polySIS product, and thus each is present at the same stoichiometry (i.e., the same number of moles per volume) as the initial polySIS standard added to the sample before digestion (after correction for any dilution or concentration occurring during or after the digestion protocol). Each sample-derived natural monitor peptide can then be quantitated by measuring its concentration relative to the stable isotope version (which has a known absolute concentration calculable from the amount spiked into the sample or sample digest), and this then allows calculation of the concentration of the associated target protein in the initial sample (as described in published U.S. patent application 20040072251, High sensitivity quantitation of peptides by mass spectrometry, Anderson, Norman. L). The relative concentrations of natural and stable isotope labeled monitor peptides are preferably measured by mass spectrometry as the relative ion currents recorded for the two peptides or their fragmentation products. The two versions perform essentially identically in any chromatographic or affinity based separation or enrichment process (provided N, C or O are used as labels), and thus co-elute, facilitating direct comparison of ion currents. In this embodiment, one polySIS protein replaces an entire collection of separate SIS peptides described in earlier disclosures, and eliminates the requirement to synthesize, purify, and standardize concentrations of the separate SIS peptide reagents. Quantitative MS measurements can be made using a variety of ionization sources (e.g., electrospray ionization [ESI] and matrix-assisted laser desorption ionization [MALDI]) and mass analyzers (e.g., time-of-flight [TOF], triple quadrupole [TQMS], Fourier transform ion cyclotron resonance [FTICR], and ion trap).

[0038] 2) In a second embodiment, the process of the first embodiment is altered so as to use a vector suitable for expression in a selected cell-based expression system (FIG. 1, track 2). This vector, containing the polySIS coding sequence in the correct frame and orientation is introduced into the cells of such an expression system (e.g., *E coli* cells), which transcribe the polySIS gene into mRNA and translate this mRNA into a polySIS protein with high

efficiency. In the case of *E coli*, additional sequences can be designed into the polySIS product to target it to the periplasmic space or to render it insoluble so as to form inclusion bodies. The *E coli* growth medium provided during the growth and product synthesis phase includes nutrients wherein at least one of the elements N, C, O or H is present in the form of an enriched ($\geq 98\%$ isotopic purity) stable isotope (¹⁵N, ¹³C, ¹⁸O or ²H respectively), thus ensuring that the polySIS product contains a high proportion of one or more stable isotopes. Under such conditions, SIS sequences such as the Hx and AAT peptides (FIG. 2 case 1 and FIG. 3) have masses greater than the natural versions by respectively 11 and 10 amu (if ¹⁵N is used) or 56 and 50 amu (if ¹³C is used). Once sufficient protein is produced, the cells are harvested, disrupted using conventional techniques, the protein contents recovered and the polySIS protein purified, making use of purification tags optionally included in the sequence.

[0039] 3) In a third embodiment (FIG. 1, track 3)), the polySIS amino acid sequence of concatenated monitor peptides is synthesized using well-known methods of chemical peptide synthesis. These are typically carried out on a solid phase resin (Merrifield, Methods Enzymol 289:3-13, 1997), and can include steps to ligate together multiple synthetic peptides to produce larger, 30-100 kD proteins (Dawson, Muir, Clark-Lewis and Kent, Science 266:776-9, 1994, Dawson and Kent, Annu Rev Biochem 69:923-60, 2000). As in the first embodiment, the preferred case makes use of stable isotope labeled K and R, since each tryptic SIS peptide contains only one such residue (either K or R) per peptide. Incorporation of labeled K or R is achieved through use of the corresponding labeled K or R synthons commercially available for solid phase peptide synthesis. Alternatively any amino acid containing stable isotope labels can be used.

[0040] 4) In a fourth embodiment, multiple polySIS products are made in order to facilitate standardized measurement of proteins having widely different abundances in the sample. Thus a first polySIS product can include monitor peptide sequences derived from proteins having expected concentrations around 1 mg/ml in human plasma (e.g., hemopexin and alpha-1-antichymotrypsin: (FIG. 2, case 3) while a second polySIS product is made containing monitor peptide sequences from low abundance (e.g., 10-1000 pg/ml) proteins such as IL-6 and TNF-alpha. Since the mass spectrometer detection systems used to measure the relative abundances of natural and SIS peptides have limited dynamic range (typically 100 to 1000), it is preferred to add an amount of each SIS peptide close to the expected amount of the equivalent natural monitor peptide. Thus the second polySIS described would optimally be added at a level approximately 1,000,000-fold less than the first polySIS above. In cases where the numbers of SIS peptides required in quantitative studies exceed the number that can conveniently be prepared as one polySIS protein, due to limitations on protein product size in many cell-free and solid phase chemical synthesis approaches, it is natural and efficient to group the desired SIS peptides into classes according to the expected concentration of the proteins from which they arise in the sample. If a set of monitor peptides were selected within a decade of concentration range (i.e., all members within a factor of 10 in expected concentration), then 6 polySIS products would be required to span a total dynamic range of 1,000,000 between the most and least

abundant target protein. Six such products would accommodate a total of 200 or more SIS sequences if each were limited to a synthesized gene length of 1 kb.

[0041] 5) In a fifth embodiment, unequal stoichiometries between individual SIS peptides are achieved by the incorporation of more than one copy of some SIS sequences in a polySIS product in which two copies of one SIS are concatenated with one copy of another SIS). In this case, exact ratios between the amounts of different SIS peptides are achieved by virtue of the necessarily integral numbers of copies present in the gene and the protein. Thus a polySIS product with 1 copy of a SIS sequence denoted A, 2 copies of B, 4 copies of C and 10 copies of D can provide peptide standards at concentrations that match the amounts of monitor peptides derived from proteins expected to be present at relative concentrations of 1:2:4:10 in the original sample. Many approaches will be apparent to those skilled in the art for inserting multiple copies of specific SIS sequences into a polySIS gene.

[0042] 6) In a sixth embodiment, two or more monitor peptide sequences are selected from the digest products of a single target analyte protein, and SIS sequences for each of these are incorporated into the polySIS product, but at different ratios. Thus SIS sequences A, B and C from a given target protein may be incorporated into the polySIS at multiplicities of 1 copy (A), 4 copies (B) and 16 copies (C). These three SIS peptides then provide an effective standard curve for measuring target protein concentration and establishing linearity over a range of at least 16-fold and generally more. The natural monitor peptides corresponding to SIS A, B and C will be present in equal amounts (in the typical case where one molecule of each is derived by digestion from one molecule of the target protein), and thus will be detected at consistent ratios versus the SIS standards: e.g., the ratios of natural monitor:SIS standard for A, B and C sequences will be x:1, x:4 and x:16. Use of multiple monitor peptides provides improved measurement precision through better statistics, and better accuracy through use of a multipoint calibration curve.

[0043] 7) In a seventh embodiment, calibrants for quantitative mass spectrometry are provided. Here two polySIS sequences are created each comprising the same series of peptides (which can be monitor peptides but can be other sequences as well). One polySIS sequence (here called X) may be comprised of a single copy of each component monitor sequence (i.e., sequences A,B,C,D present at 1,1,1,1 copies), and is produced without an incorporated stable isotope label. The other polySIS sequence may be comprised of the same monitor sequences but present in different copy numbers, e.g., A,B,C,D present in 1,2,4,8 copies respectively, and produced in an expression system so as to incorporate a stable isotope label. When equal numbers of molecules of the first and second polySIS are combined and digested to release SIS sequences, the peptide sequences A,B,C,D will each be present in unlabeled (from the first polySIS) and labeled (from the second polySIS) forms. These forms will be present in precise quantitative ratios of 1:1 (A), 1:2 (B), 1:4 (C) and 1:8 (D). These accurately defined ratios provide a precise means for calibrating the linearity of response of the mass spectrometer.

[0044] 8) In an eighth embodiment, DNA sequences for SIS peptides are inserted into "cassettes" allowing them to

be joined into expressible polySIS genes by standard molecular biology techniques. These include the techniques of recombinational cloning as well as PCR-based methods. This approach allows a series of SIS peptide sequences to be assembled into polySIS genes in different ways (i.e., different orders or at different multiplicities) by DNA fragment manipulation rather than by repeated synthesis of the entire polySIS gene.

[0045] 9) In a ninth embodiment, an easily assayed substituent is incorporated into the polySIS during synthesis and used for later quantitation of the polySIS protein. An example is the incorporation of a single biotin group into a specific lysine of the polySIS through use of the Roche "RTS AviTag Biotinylation Reagents for Enzymatic Monobiotinylation of Proteins". This site is added to the polySIS protein through use of the appropriate pVEX vector. The presence of the biotin group at 1 mole per mole of protein then allows absolute quantitation of the polySIS standard protein through use of a standard assay for the biotin tag (e.g., a competition assay using immobilized streptavidin as capture agent and a biotinylated acid phosphatase as the competing ligand able to generate a colorimetric signal). In addition, the biotin tag can be used for purification of the bulk polySIS protein by binding to a streptavidin column. The polySIS can be released from such a column by selective elution or by cleavage at a peptide sequence linking the SIS sequences to the biotinylated site using a specific protease (e.g., Factor Xa) with a specificity different from the protease used to liberate SIS (e.g., trypsin).

[0046] 10) In a tenth embodiment, entire domains of target proteins are combined into the polySIS instead of short peptides. In this approach, each domain contains at least one and preferably several peptides (e.g., tryptic peptides), and thus offers multiple opportunities to quantitate the target. More importantly, by including entire domains likely to fold in a manner more similar to the fold of part of the intact whole target protein, the polySIS better replicates the environment within which the proteolysis will occur for the native target protein—i.e., the cleavage of the peptides in the polySIS is likely to better parallel the efficiency in the target.

[0047] 11) In an eleventh embodiment, polySIS digestion products (SIS peptides), either labeled or unlabeled, are used as test materials for the optimization of MS/MS detection of the peptides. Since the relative abundances of various fragments produced in MS/MS is difficult to predict, and since one wants to maximize the production and detection sensitivity of a specific parent/fragment mass pair (particularly in triple quadrupole selected reaction monitoring as a quantitation technique), the availability of test samples of each selected target peptide provides a valuable test material for tuning MS parameters. By digesting the polySIS and infusing the resulting mix of the selected SIS peptides in a continuous infusion experiment, one can select one SIS (target) sequence at a time and systematically vary MS parameters (e.g., collision energy, mass selection windows, etc) to maximize detection of any of its fragments. One can also systematically select the best fragment for each SIS peptide in terms of detection sensitivity, signal-to-noise, and limit of quantitation. This optimization can improve the lower limit of quantitation (LLOQ) of an MS assay by a factor of 10 or more.

EXAMPLE

[0048] A series of 177 proteins and protein forms that are demonstrated or potential plasma markers of some aspect of cardiovascular disease was assembled (Anderson, *J Physiology* 563.1:23-60, 2005). Protein sequence information for the candidate markers was obtained using Swissprot accession numbers in two stages. First, when the protein was already listed in the non-redundant list of human plasma proteins described previously (Anderson, Polanski, Pieper, Gatlin, Tirumalai, Conrads, Veenstra, Adkins, Pounds, Fagan and Lobley, *Mol Cell Proteomics* 2004), the relevant accession in that non-redundant set was used. If the protein was not in this list, it was located, where possible, by query of the Swissprot web database using protein names, and added to the non-redundant list. In some cases the name used in the literature was not sufficiently specific to allow selection of a single gene product, and the candidate was not taken forward. Sequence and Swissprot annotation data was obtained in text format from the Swissprot server (<http://au.expasy.org/sprot/sprot-retrieve-list.html>) and placed in a relational database implemented using the PostgreSQL open-source database software running on an Apple Macintosh Powerbook G4 computer. Database functions were written in the PL/pgSQL language to parse the Swissprot information into fields containing the sequence, annotation related to the beginning and end of the mature protein (the CHAIN, SIGNAL, PEPTIDE and PROPEPTIDE descriptors), as well as the presence of sites where the sequence is modified in ways relevant to MS of peptides (the MOD_RES, CONFLICT, VARIANT, CARBOHYD descriptors). A separate sequence table was constructed using a PL/pgSQL function to extract that part of each sequence defined by a Swissprot CHAIN, PEPTIDE or PROPEPTIDE annotation and store it as a possible mature protein product. The "mature" products thus obtained were labeled as the Swissprot accession followed by the starting and ending amino acid positions separated by underscore characters (e.g., P08519_20_4548 for the CHAIN of Apolipoprotein(a)), and each was tagged with the name of that segment (e.g., haptoglobin alpha and beta chains, derived from a single translation product) in the Swissprot annotation (important where a single protein product is cleaved to yield multiple sequences with different names and functions).

[0049] Additional PL/pgSQL functions were used to "digest" each mature protein "in silico" to yield a list of its predicted tryptic peptides (29,155 total entries), which were stored in a separate table. Of these, 21,609 peptides occurred in only a single protein within the set of plasma proteins, and, because monitor peptides used for protein quantitation should uniquely represent a single protein analyte, only these peptides were carried forward for further analysis. The number of occurrences of each peptide in its parent protein was tabulated (in some cases more than one), in order to provide a conversion factor between moles of protein and moles of each peptide derived from it. The tryptic digestion algorithm cleaved a protein at each Arg or Lys residue, except those followed by Pro. The peptides generated were labeled by extending the mature product name with the "enzyme" used and the beginning and ending amino acid positions of the peptide within the mature sequence (e.g., P08519_20_4548_trypsin_110_2071_2080).

[0050] Computation of peptide parameters. Using a combination of PL/pgSQL functions and SQL steps, a series of parameters was calculated for each of the 21,609 peptides and stored in the database. Amino acid composition was obtained by counting the number of occurrences of each amino acid in a peptide, as was the number of occurrences of important dipeptides such as KP and RP (the only occurrences of K and R inside our predicted peptide sequences) and DP, a site within which peptide fragmentation is predicted to be especially efficient, yielding intense MS/MS signals. Peptide mass was computed in the same way as for the whole proteins, i.e., from the amino acid composition and the amino acid masses. Hoop-Woods hydrophilicity was computed by summing the standard coefficients for each residue weighted by the number of the corresponding amino acid residues (Hopp and Woods, *Proc Natl Acad Sci USA* 78:3824-8, 1981). A predicted retention time in reversed-phase (C18) chromatography was computed using the algorithm of Krokhin (Krokhin, Craig, Spicer, Ens, Standing, Beavis and Wilkins, *Mol Cell Proteomics* 3:908-19, 2004). Likely chymotryptic cleavages sites were counted. Several additional peptide attributes proved useful in the final selection process. An index of the likelihood of experimental detection was derived from a data set reported by Adkins (Adkins, Varnum, Auberry, Moore, Angell, Smith, Springer and Pounds, *Mol Cell Proteomics* 1:947-55, 2002): peptides detected in that MS/MS analysis of serum were given values equal to the number of separate "hits" for the peptide in the data set divided by the number of hits for the most frequently detected peptide from the same protein. Thus the index ranged from 1.0 for the most frequently detected peptide in a protein down to 0.1 or less for minor but still detected peptides. Predicted tryptic peptides that were not detected experimentally in the Pounds data set were given index values of 0.0. Normal plasma protein concentration values obtained from the literature were converted to a uniform scale (pg/ml). For multi-subunit proteins (e.g., fibrinogen composed of alpha, beta and gamma subunits) a factor was generated that reflected the fraction of the normal concentration attributable to that subunit. Finally a figure was derived for the molar concentration of these proteins, expressed as fmol/ml. The molar concentration of each peptide derived from such proteins is equal to the protein molar concentration times the number of occurrences of the peptide within the protein sequence. Since in some cases particular peptides occur many times (e.g., GTYSTTVTGR (Seq. ID No. 2) occurs 31 times in apolipoprotein (a)-P08519_20_4548), this correction is critical to obtaining accurate quantitative values. It also suggests that peptides of high multiplicity should yield improved detectability compared to singly represented peptides, all other factors being equal.

[0051] An overall index was generated by combining the various quantitative features described above according to a formula in which various favorable numerical criteria (e.g., content of proline) were multiplied by positive coefficients, while unfavorable criteria were multiplied by negative coefficients (FIG. 3). Peptides derived from each target protein were ranked by the overall index resulting from this formula and finally selected manually through consideration of several additional criteria in addition to the rank. Peptides that are preceded by a dipeptide of (K or R) were avoided where possible to avoid the likelihood of incomplete trypsin cleavage at KK, RR, KR and RK and thus lack of stoichiometric

release of the monitor peptide from the target protein. The proteins were ranked according to plasma concentration on a molar basis, beginning with albumin and decreasing towards the low abundance cytokines. The objective was to select monitor peptides for a series of protein targets, starting at the high abundance end of the distribution and extending downwards.

[0052] A practical polySIS gene length of 1,000 bases (selected due to commercial availability through synthesis) can code for 333 amino acids, which, given the average size of peptides selected here for MS/MS (8-14 amino acids), allows polySIS products comprising 28 to 30 SIS peptides. Two different sets of monitor peptides were selected for each of a set of 30 protein marker candidates (**FIG. 4**) selected from among the candidate markers of cardiovascular disease: one set of peptides ending in c-terminal Arg and one ending in Lys (the two amino acids at which trypsin cleaves). The mass increment due to full ^{13}C and ^{15}N labeling of the c-terminal amino acid is 8 amu for Lys and 10 amu for Arg, both sufficient to ensure adequate separation from the natural peptide isotopic distribution to give good quantitation by MS. In this example, Lys peptides were selected for further study for inclusion in polySIS protein CVD_1. In general it is possible to select good peptides having few recorded post-translational modifications (mod_res), genetic variants, sequence conflicts or glycosylation sites (carbohyd), the existence of which would alter the MS properties of the monitor peptide and disturb the equivalence of the labeled (polySIS) and unlabeled (sample-derived) versions in at least some samples.

[0053] It was noted that 5 of the final monitor peptides selected for the polySIS sequence occurred unmodified in the mouse cognate protein sequence, and thus could be useful in quantitative standardization of plasma measurements in that species. The other human sequences, which do not appear to occur in the mouse proteome, could be useful as negative quantitative controls (for which there should be no corresponding peptides in mouse plasma).

[0054] The selected Lys-ending monitor peptide sequences were concatenated into a linear sequence, in this case ordered from high to lower expected target abundance. The first peptide was preceded by an added Lys in order to release it from n-terminal vector-provided sequence. The CVD_1 amino acid sequence was backtranslated into a DNA sequence, optimizing codon usage for the *E. coli*-based cell-free system, avoiding NcoI and SmaI sites in the coding region in order to permit their use for cloning later, and introducing short 3' and 5' extensions providing appropriate restriction enzyme recognition sites. A synthetic CVD_1 gene (**FIG. 5**) was synthesized commercially (Blue Heron Technologies, Bothell, Wash.) and amplified by PCR using gene specific oligos with a 15 bp overhang specific to the pIVEX2.4d vector (Roche Applied Science, Indianapolis, Ind.). The template was digested with DpnI and the remaining PCR product purified. The amplified gene was mixed with pIVEX2.4d that had been linearized with NcoI and SmaI, and ligated into the vector with Clontech's In-Fusion Cloning enzyme (BD Biosciences Clontech, Mountain View, Calif.). This vector provides an n-terminal His6 purification tag and Factor Xa protease site in the expressed protein (sequence in **FIG. 6**). The predicted CVD_1 protein has a computed molecular mass of 38,525.76, a computed pI

of 6.08, and should yield 35 tryptic peptides (5 arising from the c- and n-terminal extensions plus the 30 monitor peptides).

[0055] It will be clear from this example that a wide variety of known and novel vectors could be used as vehicles for expression of the polySIS sequence, in both cell-based and cell-free expression systems, and to amplify it, and that a plethora of cloning strategies could be used to insert the polySIS sequence into a vector. It is also possible to expand the polySIS by PCR without use of a cloning vector.

[0056] For convenience, it is advantageous to arrange that the mass increment added to each of the labeled SIS peptides in comparison with its natural version is the same for all peptides. This can be achieved by arranging that one amino acid is labeled, and that this amino acid occurs only once per peptide. Since trypsin cleaves at most Lys and Arg residues, these are the obvious choices for labeling. Use of a single labeled amino acid also allows production of the polySIS protein, and the SIS peptides it comprises, most economically, since the cost of each different labeled amino acid is substantial. In the case of lysine, a version in which all 6 carbons are replaced with ^{13}C and both nitrogens with ^{15}N ($\text{U-}^{13}\text{C}_6\text{U-}^{15}\text{N}_2$; a total mass increment of 8 amu compared to the natural peptide) is available commercially at high (98-99%) substitution levels. As described above, a different set of monitor peptides could have been selected ending in Arg, for which an analogous commercial product is available with 10 amu mass increment.

[0057] The positioning of the label atoms at the extreme c-terminus of each peptide has the effect that all fragments that contain the c-terminus (i.e., the y-ions) will show the mass shift due to the label, whereas all the fragments that contain the n-terminus (and hence have lost one or more c-term residues: the b-series ions) will have the same masses as the corresponding fragments from the natural (sample-derived) target protein. These features (shifted y-ions, normal b-ions) provide a simplification in interpreting the fragmentation patterns of the SIS peptides. By selecting y-ions for use in relative quantitation of labeled (SIS) and sample-derived, unlabeled monitor peptides of the same sequence, the ions have identical properties except for a shift of 8 amu (for the Lys label used here). This mass increment appears as a +4 amu shift for +2 charge peptides ($z=2$), and +8/3 amu for +3 charged peptides ($z=3$).

[0058] An *E. coli*-based cell-free expression system (Roche Applied Science "RTS" coupled transcription/translation system) was used to produce the polySIS protein CVD_1. Use of a cell-free system avoids the interconversion between labeled and unlabeled amino acids that occurs in cell-based systems. Recent advances in the output of cell-free systems have made it possible to prepare milligram quantities of protein by this route: quantities sufficient to provide polySIS for many analyses given that 1 mg of the 38.5 kD polySIS is 26 nmol of product, or 29,000,000,000 amol (where 100 amol is a quantifiable amount of peptide in MS/MS). The RTS cell-free approach (commercially available kit) was used, with a mixture of 19 unlabeled amino acids and labeled lysine ($\text{U-}^{13}\text{C}_6\text{U-}^{15}\text{N}_2$ labeled: +8 amu).

[0059] Once all the reagents were mixed, the plasmid was added and the reaction proceeded for 18 hours at 30 C. and shaking at 750 rpm in a RTS ProteoMaster (Roche Applied Science). The CVD_1 polySIS protein proved to be

insoluble (despite having been constructed from relatively hydrophilic peptides) and was recovered as a major component of the pellet after centrifugation. Although the protein contains purification tags, no tag-based purification was used here.

[0060] When polySIS CVD_1 was digested with trypsin, the peptides modified with o-methylisourea and analyzed by MALDI-MS, ten of the expected peptides were detected at the expected masses, accounting for a majority of the observed peaks in the appropriate mass range. When the polySIS digest was analyzed by reversed-phase liquid chromatography and tandem mass spectrometry (using an Applied Biosystems 4000 Q-TRAP linear ion trap instrument), all 30 expected SIS peaks were observed at the expected masses (typically as doubly-charged ions). Using MS/MS data acquired on the SIS peptides, multiple reaction monitoring (M) assays were devised for each, providing three parameters: parent ion mass (Q1, typically doubly-charged), a high-mass specific y-ion fragment (Q3, typically singly charged and thus having a higher m/z than the parent), and collision energy appropriate for fragmentation in the collision cell of the 4000 Q-TRAP instrument. MRM assay parameters for the sample-derived unlabeled monitor peptides were obtained by subtracting the mass increments due to the stable isotopes labels from the Q1 and Q3 mass parameters of the labeled SIS peptides. The MS/MS data was also analyzed to assess single cleavage failures by scanning for the presence of molecules containing any two adjacent SIS peptides. Only one such failure was detected at high abundance (the peptide ILGGHLDAKTVIGPDGHK (Seq. ID No. 3), containing SIS peptides 28 and 29 in the polySIS protein).

[0061] For use for internal standardization in peptide quantitation, an amount of the polySIS protein (the "spiked" standard) is added to a sample of plasma or serum. In this case, the polySIS protein was digested before addition of the resulting SIS peptide mixture to a digest of normal human plasma from which 6 major proteins had been previously subtracted using the Agilent MARS column. Quantitative mass spectrometry was used to measure the ratios between

the ion currents of monitor peptides and same-sequence SIS standards using the 4000 Q-TRAP instrument in triple quadrupole mode. This ratio, when multiplied by the known concentration of the polySIS, provides the concentration of the monitor peptides, and thus of the target proteins in the sample at the time it was spiked.

[0062] Thus 1,300 amol of a tryptic digest of polySIS CVD_1 protein (containing 1,300 amol of each of the 30 SIS peptides) was added to the peptides derived from digestion of 0.0 ul of normal human plasma (from which 6 major proteins had been previously subtracted). The resulting peptide mixture was injected onto a 75 micron diameter C18 reversed phased LC column (LC Packings, a division of Dionex, Synnyvale Calif.), and eluted with a 40 minute gradient of 3-30% acetonitrile with 0.1% formic acid. A total of 137 MRM's were observed by time-slice multiplexing, and the peak areas of each obtained using Analyst software (Applied Biosystems). A set of 17 of the 30 SIS peptides were followed by specific MRM's, and of these 14 were detected at a signal-to-noise (S/N) ratio>10 (the usual criterion for quantitation in MS assays). The unlabeled, sample-derived same-sequence monitor peptides were detected at S/N>10 for 15 of the 17 SIS sequences, thus permitting calculation of the ratio of peak areas for SIS and monitor peptides for use in quantitation.

[0063] The peak areas for the L-selectin monitor peptide (AEIEYLEK (Seq ID No. 1)) and SIS standard were 17,620 and 79,930 respectively, yielding a ratio of 0.216. When multiplied by the 1,300 amol SIS loading, and considering that there is one copy of this peptide per molecule of intact L-selectin, this yields an L-selectin concentration of 280 amol per 0.01 ul, or 28 pmol/ml. Given a molecular weight for plasma L-selectin of ~35,000, this gives a measured concentration of 980 ng/ml. This may be compared with the published normal value of 670 ng/ml obtained by immunoassay. Given that the L-selectin monitor peptide was detected with a signal-to-noise ratio of 22, and that the lower limit of quantitation (LLOQ) is generally defined as a S/N of 10, L-selectin could have been quantitated using this MS assay at a level of ~450 ng/ml.

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 43

<210> SEQ ID NO 1
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide

<400> SEQUENCE: 1

Ala Glu Ile Glu Tyr Leu Glu Lys
 1                               5

<210> SEQ ID NO 2
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
    
```


-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 2

Gly Thr Tyr Ser Thr Thr Val Thr Gly Arg
1 5 10

<210> SEQ ID NO 3

<211> LENGTH: 18

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 3

Ile Leu Gly Gly His Leu Asp Ala Lys Thr Val Ile Gly Pro Asp Gly
1 5 10 15

His Lys

<210> SEQ ID NO 4

<211> LENGTH: 12

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 4

Glu Ala Leu Ala Glu Asn Asn Leu Asn Leu Pro Lys
1 5 10

<210> SEQ ID NO 5

<211> LENGTH: 11

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 5

Asn Phe Pro Ser Pro Val Asp Ala Ala Phe Arg
1 5 10

<210> SEQ ID NO 6

<211> LENGTH: 9

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 6

Glu Ile Gly Glu Leu Tyr Leu Pro Lys
1 5

<210> SEQ ID NO 7

<211> LENGTH: 13

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 7

-continued

Asp Leu Ser Leu Ile Ser Pro Leu Ala Gln Ala Val Arg
1 5 10

<210> SEQ ID NO 8
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 8

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

<210> SEQ ID NO 9
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 9

His His His His His His
1 5

<210> SEQ ID NO 10
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 10

Ala Thr Glu His Leu Ser Thr Leu Ser Glu Lys
1 5 10

<210> SEQ ID NO 11
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 11

Asn Trp Gly Leu Ser Val Tyr Ala Asp Lys Pro Glu Thr Thr Lys
1 5 10 15

<210> SEQ ID NO 12
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 12

Ile Leu Gly Gly His Leu Asp Ala Lys
1 5

<210> SEQ ID NO 13

-continued

<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 13

Asp Thr Val Gln Ile His Asp Ile Thr Gly Lys
1 5 10

<210> SEQ ID NO 14
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 14

Thr Val Ile Gly Pro Asp Gly His Lys
1 5

<210> SEQ ID NO 15
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 15

Gln Gly Phe Gly Asn Val Ala Thr Asn Thr Asp Gly Lys
1 5 10

<210> SEQ ID NO 16
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 16

Glu Ile Gly Glu Leu Tyr Leu Pro Lys
1 5

<210> SEQ ID NO 17
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 17

Thr Gly Leu Gln Glu Val Glu Val Lys
1 5

<210> SEQ ID NO 18
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

-continued

<400> SEQUENCE: 18

Asp Asp Leu Tyr Val Ser Asp Ala Phe His Lys
1 5 10

<210> SEQ ID NO 19

<211> LENGTH: 10

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 19

Ile Tyr His Ser His Ile Asp Ala Pro Lys
1 5 10

<210> SEQ ID NO 20

<211> LENGTH: 12

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 20

Glu Thr Ala Ala Ser Leu Leu Gln Ala Gly Tyr Lys
1 5 10

<210> SEQ ID NO 21

<211> LENGTH: 9

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 21

Ile Thr Gln Val Leu His Phe Thr Lys
1 5

<210> SEQ ID NO 22

<211> LENGTH: 9

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 22

Phe Pro Glu Val Asp Val Leu Thr Lys
1 5

<210> SEQ ID NO 23

<211> LENGTH: 13

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 23

Leu Gly Asn Gln Glu Pro Gly Gly Gln Thr Ala Leu Lys
1 5 10

-continued

<210> SEQ ID NO 24
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 24

Leu Ser Ser Pro Ala Val Ile Thr Asp Lys
1 5 10

<210> SEQ ID NO 25
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 25

Gln Trp Ala Gly Leu Val Glu Lys
1 5

<210> SEQ ID NO 26
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 26

Ile Pro Pro Trp Glu Ala Pro Lys
1 5

<210> SEQ ID NO 27
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 27

Leu Phe Leu Glu Pro Thr Gln Ala Asp Ile Ala Leu Leu Lys
1 5 10

<210> SEQ ID NO 28
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 28

Ser His Ala Pro Glu Val Ile Thr Ser Ser Pro Leu Lys
1 5 10

<210> SEQ ID NO 29
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 29

Ile Phe Tyr Asn Gln Gln Asn His Tyr Asp Gly Ser Thr Gly Lys
1 5 10 15

<210> SEQ ID NO 30

<211> LENGTH: 9

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 30

Glu His Ser Ser Leu Ala Phe Trp Lys
1 5

<210> SEQ ID NO 31

<211> LENGTH: 8

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 31

Val Ser Val Ser Gln Thr Ser Lys
1 5

<210> SEQ ID NO 32

<211> LENGTH: 10

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 32

Glu Ser Asp Thr Ser Tyr Val Ser Leu Lys
1 5 10

<210> SEQ ID NO 33

<211> LENGTH: 8

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 33

Trp Glu Leu Asp Leu Asp Ile Lys
1 5

<210> SEQ ID NO 34

<211> LENGTH: 12

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 34

Ser Thr Val Leu Thr Ile Pro Glu Ile Ile Ile Lys

-continued

1	5	10	
---	---	----	--

<210> SEQ ID NO 35
 <211> LENGTH: 11
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 35

Leu Ile Glu Asn Gly Tyr Phe His Pro Val Lys
 1 5 10

<210> SEQ ID NO 36
 <211> LENGTH: 10
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 36

Ala Ser Tyr Pro Asp Ile Thr Gly Glu Lys
 1 5 10

<210> SEQ ID NO 37
 <211> LENGTH: 10
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 37

Asp Pro Pro Ser Asp Leu Leu Leu Leu Lys
 1 5 10

<210> SEQ ID NO 38
 <211> LENGTH: 12
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 38

Ala Leu Gln Asp Gln Leu Val Leu Val Ala Ala Lys
 1 5 10

<210> SEQ ID NO 39
 <211> LENGTH: 322
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polySIS protein

<400> SEQUENCE: 39

Met Ala Lys Ala Thr Glu His Leu Ser Thr Leu Ser Glu Lys Asn Trp
 1 5 10 15

Gly Leu Ser Val Tyr Ala Asp Lys Pro Glu Thr Thr Lys Ile Leu Gly
 20 25 30

Gly His Leu Asp Ala Lys Asp Thr Val Gln Ile His Asp Ile Thr Gly

-continued

	35						40							45					
Lys	Thr	Val	Ile	Gly	Pro	Asp	Gly	His	Lys	Gln	Gly	Phe	Gly	Asn	Val				
	50					55					60								
Ala	Thr	Asn	Thr	Asp	Gly	Lys	Glu	Ile	Gly	Glu	Leu	Tyr	Leu	Pro	Lys				
	65				70					75					80				
Thr	Gly	Leu	Gln	Glu	Val	Glu	Val	Lys	Asp	Asp	Leu	Tyr	Val	Ser	Asp				
				85					90					95					
Ala	Phe	His	Lys	Ile	Tyr	His	Ser	His	Ile	Asp	Ala	Pro	Lys	Glu	Thr				
			100					105						110					
Ala	Ala	Ser	Leu	Leu	Gln	Ala	Gly	Tyr	Lys	Ile	Thr	Gln	Val	Leu	His				
		115					120					125							
Phe	Thr	Lys	Phe	Pro	Glu	Val	Asp	Val	Leu	Thr	Lys	Leu	Gly	Asn	Gln				
	130					135					140								
Glu	Pro	Gly	Gly	Gln	Thr	Ala	Leu	Lys	Leu	Ser	Ser	Pro	Ala	Val	Ile				
	145				150					155					160				
Thr	Asp	Lys	Gln	Trp	Ala	Gly	Leu	Val	Glu	Lys	Ile	Pro	Pro	Trp	Glu				
				165					170					175					
Ala	Pro	Lys	Leu	Phe	Leu	Glu	Pro	Thr	Gln	Ala	Asp	Ile	Ala	Leu	Leu				
			180					185					190						
Lys	Ser	His	Ala	Pro	Glu	Val	Ile	Thr	Ser	Ser	Pro	Leu	Lys	Ile	Phe				
		195					200					205							
Tyr	Asn	Gln	Gln	Asn	His	Tyr	Asp	Gly	Ser	Thr	Gly	Lys	Glu	His	Ser				
	210					215					220								
Ser	Leu	Ala	Phe	Trp	Lys	Val	Ser	Val	Ser	Gln	Thr	Ser	Lys	Glu	Ser				
	225				230					235				240					
Asp	Thr	Ser	Tyr	Val	Ser	Leu	Lys	Trp	Glu	Leu	Asp	Leu	Asp	Ile	Lys				
				245					250					255					
Ser	Thr	Val	Leu	Thr	Ile	Pro	Glu	Ile	Ile	Ile	Lys	Leu	Ile	Glu	Asn				
			260					265					270						
Gly	Tyr	Phe	His	Pro	Val	Lys	Ala	Ser	Tyr	Pro	Asp	Ile	Thr	Gly	Glu				
		275					280					285							
Lys	Asp	Pro	Pro	Ser	Asp	Leu	Leu	Leu	Leu	Lys	Ala	Leu	Gln	Asp	Gln				
	290					295					300								
Leu	Val	Leu	Val	Ala	Ala	Lys	Ala	Glu	Ile	Glu	Tyr	Leu	Glu	Lys	Gln				
	305				310					315					320				

Pro Gly

```

<210> SEQ ID NO 40
<211> LENGTH: 966
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polysIS gene
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(966)

<400> SEQUENCE: 40

atg gca aaa gca aca gaa cat tta agt aca tta tca gaa aaa aat tgg      48
Met Ala Lys Ala Thr Glu His Leu Ser Thr Leu Ser Glu Lys Asn Trp
  1           5           10           15

gga tta tca gta tat gca gat aaa cca gaa act aca aaa att tta gga      96
Gly Leu Ser Val Tyr Ala Asp Lys Pro Glu Thr Thr Lys Ile Leu Gly
  20           25           30
    
```


-continued

gga cat tta gac gca aaa gac aca gtt caa att cat gat atc aca gga	144
Gly His Leu Asp Ala Lys Asp Thr Val Gln Ile His Asp Ile Thr Gly	
35 40 45	
aaa aca gta atc gga cca gac gga cat aaa caa ggt ttc ggt aac gta	192
Lys Thr Val Ile Gly Pro Asp Gly His Lys Gln Gly Phe Gly Asn Val	
50 55 60	
gca aca aat aca gac gga aaa gaa att gga gaa tta tat tta cca aag	240
Ala Thr Asn Thr Asp Gly Lys Glu Ile Gly Glu Leu Tyr Leu Pro Lys	
65 70 75 80	
aca ggt tta caa gaa gta gaa gta aaa gac gat tta tat gta tca gac	288
Thr Gly Leu Gln Glu Val Glu Val Lys Asp Asp Leu Tyr Val Ser Asp	
85 90 95	
gca ttt cat aaa att tac cat tca cac atc gac gca cca aaa gaa aca	336
Ala Phe His Lys Ile Tyr His Ser His Ile Asp Ala Pro Lys Glu Thr	
100 105 110	
gca gca tca tta tta caa gca ggt tat aaa atc aca caa gta tta cat	384
Ala Ala Ser Leu Leu Gln Ala Gly Tyr Lys Ile Thr Gln Val Leu His	
115 120 125	
ttc aca aaa ttc cca gaa gta gac gta tta aca aaa tta gga aat cag	432
Phe Thr Lys Phe Pro Glu Val Asp Val Leu Thr Lys Leu Gly Asn Gln	
130 135 140	
gaa cca gga ggt caa aca gca tta aaa tta tca tca cca gca gta atc	480
Glu Pro Gly Gly Gln Thr Ala Leu Lys Leu Ser Ser Pro Ala Val Ile	
145 150 155 160	
aca gat aaa caa tgg gca gga tta gta gaa aag att cca cct tgg gaa	528
Thr Asp Lys Gln Trp Ala Gly Leu Val Glu Lys Ile Pro Pro Trp Glu	
165 170 175	
gct cct aaa tta ttc tta gaa cca aca caa gca gat att gca tta tta	576
Ala Pro Lys Leu Phe Leu Glu Pro Thr Gln Ala Asp Ile Ala Leu Leu	
180 185 190	
aaa tct cac gca cca gaa gta att acc agt tca cca tta aaa att ttt	624
Lys Ser His Ala Pro Glu Val Ile Thr Ser Ser Pro Leu Lys Ile Phe	
195 200 205	
tac aac caa caa aat cat tac gac gga tca aca gga aaa gaa cac tca	672
Tyr Asn Gln Gln Asn His Tyr Asp Gly Ser Thr Gly Lys Glu His Ser	
210 215 220	
tca tta gca ttt tgg aaa gta tca gtt agt caa act tca aag gaa tca	720
Ser Leu Ala Phe Trp Lys Val Ser Val Ser Gln Thr Ser Lys Glu Ser	
225 230 235 240	
gac aca agt tat gta tca tta aaa tgg gaa tta gac tta gac atc aaa	768
Asp Thr Ser Tyr Val Ser Leu Lys Trp Glu Leu Asp Leu Asp Ile Lys	
245 250 255	
tca aca gta ctt acc att cca gaa att att att aaa tta atc gaa aac	816
Ser Thr Val Leu Thr Ile Pro Glu Ile Ile Ile Lys Leu Ile Glu Asn	
260 265 270	
gga tat ttt cac cca gtt aaa gca tca tac cca gat att aca gga gaa	864
Gly Tyr Phe His Pro Val Lys Ala Ser Tyr Pro Asp Ile Thr Gly Glu	
275 280 285	
aaa gac cca cca agt gac tta tta tta tta aaa gca tta caa gac caa	912
Lys Asp Pro Pro Ser Asp Leu Leu Leu Leu Lys Ala Leu Gln Asp Gln	
290 295 300	
tta gta tta gta gca gca aaa gca gaa att gaa tat tta gaa aaa cag	960
Leu Val Leu Val Ala Ala Lys Ala Glu Ile Glu Tyr Leu Glu Lys Gln	
305 310 315 320	
ccc ggg	966
Pro Gly	

-continued

<210> SEQ ID NO 41
 <211> LENGTH: 350
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 polySIS protein sequence

<400> SEQUENCE: 41

Met Ser Gly Ser His His His His His Ser Ser Gly Ile Glu Gly
 1 5 10 15
 Arg Gly Arg Leu Ile Lys His Met Thr Met Ala Lys Ala Thr Glu His
 20 25 30
 Leu Ser Thr Leu Ser Glu Lys Asn Trp Gly Leu Ser Val Tyr Ala Asp
 35 40 45
 Lys Pro Glu Thr Thr Lys Ile Leu Gly Gly His Leu Asp Ala Lys Asp
 50 55 60
 Thr Val Gln Ile His Asp Ile Thr Gly Lys Thr Val Ile Gly Pro Asp
 65 70 75 80
 Gly His Lys Gln Gly Phe Gly Asn Val Ala Thr Asn Thr Asp Gly Lys
 85 90 95
 Glu Ile Gly Glu Leu Tyr Leu Pro Lys Thr Gly Leu Gln Glu Val Glu
 100 105 110
 Val Lys Asp Asp Leu Tyr Val Ser Asp Ala Phe His Lys Ile Tyr His
 115 120 125
 Ser His Ile Asp Ala Pro Lys Glu Thr Ala Ala Ser Leu Leu Gln Ala
 130 135 140
 Gly Tyr Lys Ile Thr Gln Val Leu His Phe Thr Lys Phe Pro Glu Val
 145 150 155 160
 Asp Val Leu Thr Lys Leu Gly Asn Gln Glu Pro Gly Gly Gln Thr Ala
 165 170 175
 Leu Lys Leu Ser Ser Pro Ala Val Ile Thr Asp Lys Gln Trp Ala Gly
 180 185 190
 Leu Val Glu Lys Ile Pro Pro Trp Glu Ala Pro Lys Leu Phe Leu Glu
 195 200 205
 Pro Thr Gln Ala Asp Ile Ala Leu Leu Lys Ser His Ala Pro Glu Val
 210 215 220
 Ile Thr Ser Ser Pro Leu Lys Ile Phe Tyr Asn Gln Gln Asn His Tyr
 225 230 235 240
 Asp Gly Ser Thr Gly Lys Glu His Ser Ser Leu Ala Phe Trp Lys Val
 245 250 255
 Ser Val Ser Gln Thr Ser Lys Glu Ser Asp Thr Ser Tyr Val Ser Leu
 260 265 270
 Lys Trp Glu Leu Asp Leu Asp Ile Lys Ser Thr Val Leu Thr Ile Pro
 275 280 285
 Glu Ile Ile Ile Lys Leu Ile Glu Asn Gly Tyr Phe His Pro Val Lys
 290 295 300
 Ala Ser Tyr Pro Asp Ile Thr Gly Glu Lys Asp Pro Pro Ser Asp Leu
 305 310 315 320
 Leu Leu Leu Lys Ala Leu Gln Asp Gln Leu Val Leu Val Ala Ala Lys
 325 330 335
 Ala Glu Ile Glu Tyr Leu Glu Lys Gln Pro Gly Gly Ile Arg
 340 345 350

-continued

<210> SEQ ID NO 42
 <211> LENGTH: 316
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 protein sequence

<400> SEQUENCE: 42

```

Ala Thr Glu His Leu Ser Thr Leu Ser Glu Lys Asn Trp Gly Leu Ser
  1           5           10           15
Val Tyr Ala Asp Lys Pro Glu Thr Thr Lys Ile Leu Gly Gly His Leu
          20           25           30
Asp Ala Lys Asp Thr Val Gln Ile His Asp Ile Thr Gly Lys Thr Val
          35           40           45
Ile Gly Pro Asp Gly His Lys Gln Gly Phe Gly Asn Val Ala Thr Asn
          50           55           60
Thr Asp Gly Lys Glu Ile Gly Glu Leu Tyr Leu Pro Lys Thr Gly Leu
          65           70           75           80
Gln Glu Val Glu Val Lys Asp Asp Leu Tyr Val Ser Asp Ala Phe His
          85           90           95
Lys Ile Tyr His Ser His Ile Asp Ala Pro Lys Glu Thr Ala Ala Ser
          100          105          110
Leu Leu Gln Ala Gly Tyr Lys Ile Thr Gln Val Leu His Phe Thr Lys
          115          120          125
Phe Pro Glu Val Asp Val Leu Thr Lys Leu Gly Asn Gln Glu Pro Gly
          130          135          140
Gly Gln Thr Ala Leu Lys Leu Ser Ser Pro Ala Val Ile Thr Asp Lys
          145          150          155          160
Gln Trp Ala Gly Leu Val Glu Lys Ile Pro Pro Trp Glu Ala Pro Lys
          165          170          175
Leu Phe Leu Glu Pro Thr Gln Ala Asp Ile Ala Leu Leu Lys Ser His
          180          185          190
Ala Pro Glu Val Ile Thr Ser Ser Pro Leu Lys Ile Phe Tyr Asn Gln
          195          200          205
Gln Asn His Tyr Asp Gly Ser Thr Gly Lys Glu His Ser Ser Leu Ala
          210          215          220
Phe Trp Lys Val Ser Val Ser Gln Thr Ser Lys Glu Ser Asp Thr Ser
          225          230          235          240
Tyr Val Ser Leu Lys Trp Glu Leu Asp Leu Asp Ile Lys Ser Thr Val
          245          250          255
Leu Thr Ile Pro Glu Ile Ile Ile Lys Leu Ile Glu Asn Gly Tyr Phe
          260          265          270
His Pro Val Lys Ala Ser Tyr Pro Asp Ile Thr Gly Glu Lys Asp Pro
          275          280          285
Pro Ser Asp Leu Leu Leu Leu Lys Ala Leu Gln Asp Gln Leu Val Leu
          290          295          300
Val Ala Ala Lys Ala Glu Ile Glu Tyr Leu Glu Lys
          305          310          315

```

<210> SEQ ID NO 43
 <211> LENGTH: 294
 <212> TYPE: PRT

-continued

```

<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 43
Trp Thr Tyr His Tyr Ser Glu Lys Pro Met Asn Trp Gln Arg Ala Arg
 1           5           10           15
Arg Phe Cys Arg Asp Asn Tyr Thr Asp Leu Val Ala Ile Gln Asn Lys
 20           25           30
Ala Glu Ile Glu Tyr Leu Glu Lys Thr Leu Pro Phe Ser Arg Ser Tyr
 35           40           45
Tyr Trp Ile Gly Ile Arg Lys Ile Gly Gly Ile Trp Thr Trp Val Gly
 50           55           60
Thr Asn Lys Ser Leu Thr Glu Glu Ala Glu Asn Trp Gly Asp Gly Glu
 65           70           75           80
Pro Asn Asn Lys Lys Asn Lys Glu Asp Cys Val Glu Ile Tyr Ile Lys
 85           90           95
Arg Asn Lys Asp Ala Gly Lys Trp Asn Asp Asp Ala Cys His Lys Leu
 100          105          110
Lys Ala Ala Leu Cys Tyr Thr Ala Ser Cys Gln Pro Trp Ser Cys Ser
 115          120          125
Gly His Gly Glu Cys Val Glu Ile Ile Asn Asn Tyr Thr Cys Asn Cys
 130          135          140
Asp Val Gly Tyr Tyr Gly Pro Gln Cys Gln Phe Val Ile Gln Cys Glu
 145          150          155          160
Pro Leu Glu Ala Pro Glu Leu Gly Thr Met Asp Cys Thr His Pro Leu
 165          170          175
Gly Asn Phe Ser Phe Ser Ser Gln Cys Ala Phe Ser Cys Ser Glu Gly
 180          185          190
Thr Asn Leu Thr Gly Ile Glu Glu Thr Thr Cys Gly Pro Phe Gly Asn
 195          200          205
Trp Ser Ser Pro Glu Pro Thr Cys Gln Val Ile Gln Cys Glu Pro Leu
 210          215          220
Ser Ala Pro Asp Leu Gly Ile Met Asn Cys Ser His Pro Leu Ala Ser
 225          230          235          240
Phe Ser Phe Thr Ser Ala Cys Thr Phe Ile Cys Ser Glu Gly Thr Glu
 245          250          255
Leu Ile Gly Lys Lys Lys Thr Ile Cys Glu Ser Ser Gly Ile Trp Ser
 260          265          270
Asn Pro Ser Pro Ile Cys Gln Lys Leu Asp Lys Ser Phe Ser Met Ile
 275          280          285
Lys Glu Gly Asp Tyr Asn
 290

```

What we claim is:

1. A polySIS protein having an amino acid sequence containing several amino acid subsequences found in nature and wherein at least two different subsequences act as monitor sequences, said subsequences being part of at least one natural protein which is a target protein, wherein the end of each of said two different subsequences have a cleavage site that will be cleaved by the same site-specific proteolytic treatment to release said subsequences.

2. The protein of claim 1 wherein at least one subsequence is present in more than one copy.

3. The protein of claim 1 wherein said polySIS protein is not naturally produced by any organism.

4. The protein of claim 1 wherein said proteolytic treatment includes exposure to at least one enzyme.

5. The protein of claim 1 which is be cleaved by one or more of trypsin, Lys-C, Arg-C chymotrypsin, proteinase K, Asp N or Glu-C.

6. The protein of claim 1 wherein said protein is cleaved by cyanogen bromide, formic acid or BNPS-skatole.

7. The protein of claim 1 which can be cleaved by a combination of at least one enzyme and at least one chemical reagent which is not an enzyme.

8. A method of producing a protein for use in quantitative analysis of other protein comprising the steps of

1) choosing two or more peptide subsequences from different proteins found in nature wherein at least one end of each of said subsequences represents a cleavage site, and

2) concatenating said two or more peptide sequences together into a polySIS protein.

9. The process of claim 8 where each of said two or more peptide sequences used in step 2 is obtained from a naturally occurring animal protein.

10. The process of claim 9 wherein said animal protein is a blood protein.

11. The protein of claim 1 containing, additionally, amino acid sequences which facilitate affinity purification of the protein.

12. The protein of claim 1 containing, additionally, a stoichiometrically determined amount of one or more detectable ligands which facilitate affinity capture or purification of said protein.

13. The protein of claim 12 containing subsequences from biotin, a sulfhydryl group, a sugar moiety or a nucleic acid.

14. A method of producing a polySIS protein of claim 1 by

(1) producing a DNA sequence capable of directing synthesis of said protein, then

(2) expressing the protein in an appropriate expression system.

15. The method of claim 14 wherein the expression system is a vector.

16. The method of claim 16 wherein at least part of said DNA sequence is generated by nucleic acid synthesis.

17. A DNA sequence which encodes a polySIS protein of claim 1.

18. The DNA sequence of claim 17 which contains at least two non-identical manipulatable cassettes wherein such cassettes are assembled into a protein-producing sequence, said sequences being linked together at the nucleic acid sequence level.

19. The method of claim 14 wherein said protein expression system is a cell-free system capable of linked transcription/translation.

20. The method of claim 14 wherein the expression system comprises living cells.

21. A method of producing a polySIS protein by chemical peptide synthesis.

22. The method of claim 21 wherein said synthesis is accomplished on a solid phase.

23. The method of claim 21 wherein said protein is prepared by ligation of two or more peptides produced by chemical peptide synthesis.

24. The protein of claim 1 further including at least one amino acid containing at least one stable isotope at a high state of isotopic enrichment.

25. The protein of claim 24 wherein said stable isotope-containing amino acid occurs at the c-terminus of said monitor peptide sequences.

26. The protein of claim 24 wherein said stable isotope-containing amino acid is lysine or arginine.

27. The protein of claim 24 wherein said stable isotope is one or more of ^{15}N , ^{13}C , ^{18}O or ^2H .

28. The protein of claim 1 wherein said target proteins are selected as a group expected to have concentrations in a sample of interest that differ from one another by no more than a factor of 100.

29. The protein of claim 1 wherein said at least one target protein represents a biomarker.

30. The protein of claim 29 wherein said biomarker is a biomarker for cardiovascular disease.

31. The protein of claim 29 wherein said biomarker is a biomarker for inflammation.

32. The protein of claim 29 wherein said biomarker represents a biomarker involved in hemostasis of thrombolysis.

33. The protein of claim 29 wherein said biomarker represents a biomarker associated with malignancy.

34. A composition comprising proteins containing at least two different polySIS proteins of claim 1 wherein at least one of said proteins is present at a concentration at least 10 fold greater than another of said proteins.

35. The composition of claim 34 wherein said protein in higher concentration contains at least one monitor peptide sequence derived from a target protein wherein said target protein is expected to be present in a sample at a 10-fold or higher concentration than monitor peptide sequences contained in at least one other said protein in a sample to be tested.

* * * * *