



(51) International Patent Classification:
G01N 27/62 (2006.01)

(21) International Application Number:
PCT/KR2008/002390

(22) International Filing Date:
28 April 2008 (28.04.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10-2008-0023411 13 March 2008 (13.03.2008) KR

(71) Applicant (for all designated States except US): **KOREA RESEARCH INSTITUTE OF BIOSCIENCE AND BIOTECHNOLOGY** [KR/KR]; 52, Eoeun-dong, Yuseong-gu, Daejeon 305-806 (KR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HUR, Cheol Goo** [KR/KR]; 133-604, Hanbit Apt., 99, Eoeun-dong, Yuseong-gu, Daejeon 305-755 (KR). **LEE, Seung Won** [KR/KR]; Room 1114, SK HUB Apt., Jijok-dong, Yuseong-gu, Daejeon 305-330 (KR). **CHOI, Jae Pil** [KR/KR]; Room 101, Jinseong House, 164-12, Seonhwa-dong, Jung-gu, Daejeon 301-825 (KR).

(74) Agent: **CHOI, Kyu Whan**; 416, Sigmavill, 936, Dunsan-dong, Seo-gu, Daejeon 302-120 (KR).

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

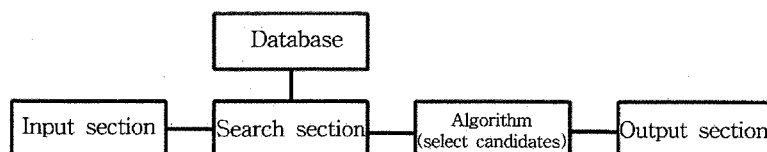
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: SYSTEM COMPRISING SCORING ALGORITHM AND METHOD FOR IDENTIFYING ALTERNATIVE SPLICING ISOFORMS USING PEPTIDE MASS FINGERPRINTING, AND RECORDING MEDIA HAVING PROGRAM THEREFOR

[Fig. 1]



(57) Abstract: The present invention relates to a system comprising a scoring algorithm and a method for identifying alternative splicing isoforms by using peptide mass fingerprinting, and recording media having a computer readable program to implement the method. By using a database including information of tissue-specific gene and an algorithm which can efficiently distinguish alternative splicing isoforms of a gene from another, it is possible according to the present invention that alternative splicing isoforms can be efficiently characterized and identified and tissue-specific isoforms can be determined.

Description

SYSTEM COMPRISING SCORING ALGORITHM AND METHOD FOR IDENTIFYING ALTERNATIVE SPLICING ISOFORMS USING PEPTIDE MASS FINGERPRINTING, AND RECORDING MEDIA HAVING PROGRAM THEREFOR

Technical Field

- [1] The present invention relates to a system comprising a scoring algorithm and a method for identifying alternative splicing isoforms by using peptide mass fingerprinting, and recording media having a computer readable program to implement the method.

Background Art

- [2] Inside cells hundreds or thousands of different proteins are constantly being expressed in various amounts. As such, it is very challenging to determine the kinds and the amounts of such proteins. Even if a protein of interest is isolated in pure state, molecular weight determination alone is not enough for a precise identification of protein since there are many different kinds of proteins having similar molecular weight present in a cell. Meanwhile, if a protein is digested with an enzyme which can hydrolyze the protein at a specific site and molecular weights of the resulting peptide fragments are determined, the volume of information that can be used for protein identification will dramatically increase and it would lead a way to characterize the original protein.
- [3] Peptide mass fingerprinting (herein after, sometimes referred to as "PMF") is a technique used to determine the mass distribution of peptide fragments that are obtained from protein hydrolysis. Similar to a line spectrum which serves as a fingerprint for an element, mass of peptide fragments is a fingerprint of a protein. In this case, it is unnecessary to measure the molecular weight of all peptides obtained from the protein hydrolysis. Instead, an accurate measurement of the molecular weight of several peptides is usually enough and it is because for two separate proteins a probability of obtaining a combination of peptides having exactly the same molecular weight is extremely low.
- [4] Alternative splicing is a mechanism of producing two or more various and separate proteins from a single gene in a complex organism. With such alternative splicing, a great variety of proteins can be produced from a limited number of genes at different time and location. It was also reported that alternative splicing is related to the occurrence of some diseases.

- [5] For an identification of an alternative splicing isoform, usually proteins and transcripts are experimentally determined.
- [6] As a method for determining a transcript, a microarray system is used. There is a method which designs and uses exon probes and exon-exon junction probes (Nagao, K. (2005) Human Molecular Genetics, 14, 3379-3388; Shoemaker, D. (2001) Nature, 409, 922-927). However, because not all of those determined at transcription level are translated into a protein, it is advantageous that the determination is carried out at protein level.
- [7] For a protein level identification, mass analysis is used. For a determination of splicing sites, there is a method based on protein sequencing with a combined mass spectrometry and a mass spectrometry (Tanner, S. (2007) Genome research, 17, 231-239). In addition, there is a method by which a genome is translated into six frames of proteins and then splicing site is identified (Giddings, M. (2003) Proc Natl Acad Sci USA, 100, 20-25). Although being suitable for locating a splicing site in a gene, these methods are not useful for obtaining an isoform. The reason is that, not all of the fragment mass of the peptides that are produced by proteolytic enzymes are confirmed during mass analysis. In addition, if an alternative splicing isoform is not taken into consideration, a wrong type of isoform can be identified from peptide mass fingerprinting, because there can be many isoforms present that have not yet been verified by an experiment.
- [8] Meanwhile, in order to determine an alternative splicing isoform by using PMF technique, a database wherein the presence of such alternative splicing is considered should be constructed. A gene having an event of alternative splicing in such database should have a common exon so that it may have a protein sequence that is partially same to each other. In addition, there can be more than one hundred isoforms which can be generated from a single gene. Due to such characteristics, a new algorithm that can distinguish isoforms different to each other is in need.
- [9] Korean Patent Registration No. 10-0789430 discloses a method for determining isotopic clusters and monoisotopic masses of polypeptides on mass spectra of complex polypeptide mixtures and computer-readable medium thereof. Korean Patent Registration No. 10-0757040 discloses a system and method for analysis of interfaces based on protein domain and recording medium therefor. In addition, PCT publication WO 01/57519 discloses a method for identifying and/or characterizing a (poly)peptide.

Disclosure of Invention

Technical Problem

- [10] The present intention, which is devised in view of said necessity, relates to a development of a system and a method for identifying alternative splicing isoforms that

have been known or not known from prior art, and an identification of alternative splicing isoforms using such system and method.

- [11] According to the present invention, by using an algorithm which can efficiently distinguish alternative splicing forms of a gene from another and a database which includes tissue-specific genetic information, the alternative splicing isoforms can be efficiently characterized and identified and tissue-specific isoforms can be confirmed.

Technical Solution

- [12] In order to solve the problems described above, the present invention provides a system and a method for identifying alternative splicing isoforms using peptide mass fingerprinting. In addition, the present invention provides recording media having a computer readable program that can be used to implement said method.

Advantageous Effects

- [13] According to the present invention, alternative splicing isoforms can be identified from the mass distribution of peptides that are generated by mass spectrometry. As a result, isoforms that have been previously unknown and isoforms that are related to a certain disease or specific for certain tissues can be identified.

Brief Description of the Drawings

- [14] Figure 1 is a schematic diagram representing the system for identifying alternative splicing isoforms according to the present invention.
- [15] Figure 2 is a diagram describing the algorithm of the present invention.
- [16] Figure 3 shows a screen for input section.
- [17] Figure 4 shows a screen for output section.

Mode for the Invention

- [18] In order to achieve said object of the invention, the present invention provides a system encompassing entire processes for identifying alternative splicing isoforms by using peptide mass fingerprinting, which comprises,
- [19] - an input section to which mass data measured for peptide fragments that are obtained from protein hydrolysis are input;
- [20] - a database comprising mass and isopotential value of a protein, site information of an amino acid which might be involved in protein modification and sequence information of a peptide that is produced after the hydrolysis of the protein;
- [21] - a search section by which the peptide mass data is searched against an established database;
- [22] - an algorithm for scoring a potential of candidate alternative splicing isoforms having a mass with match found in database; and
- [23] - an output section by which the candidate isoforms are aligned according to their score.

[24] According to a system of one embodiment of the present invention, transcripts of splicing isoforms are collected for a species having an alternative splicing, and then they can be translated into a protein sequence to construct the database.

[25] According to a system of one embodiment of the present invention, said algorithm can be an algorithm which determines priority of each candidate with a process that mass data measured from an experiment is searched against the established database and then the results are calculated for masses that are in match with respective isoforms by using the following formulae 1, 2 and 3.

[26]
$$r_{m_i} = f_{m_i} / T \quad \text{Formula 1}$$

$$E_{m_i} = \log_{10}(r_{m_i}) \quad \text{Formula 2}$$

$$P_s = \sum_{j=1}^k E_{Mj} \quad \text{Formula 3}$$

[27] In the above-described formulae 1, 2 and 3, symbols used are as follows. Specifically, when measured mass value is represented by $m = \{m_1, m_2, \dots, m_i\}$ and a mass value with match in the database is represented by $M = \{M_1, M_2, \dots, M_k\}$, m_i indicates mass value of m while M_k indicates mass value of M . M_k is a single value that is in match with one candidate isoform among m_i . When a search is carried out using a measured data and total number of peptides that are searched from the database is T and the number of peptides that are searched by using each of the mass values from the database is f_{m_i} , r_{m_i} can be obtained from the above formula 1. In this case, since the values can be too small to handle, r_{m_i} can be converted to E_{m_i} based on the above-described logarithmic formula 2. P_s value is a sum of E_{Mj} for the mass values in match with respective isoforms. In other words, if there are k mass values in match with a certain isoform, P_s is a sum of E_{Mj} that are generated from applying k number of matching mass values to formulae 1 and 2. Smaller the P_s value is, a more significant candidate it becomes. This means that, smaller the possibility of an accidental match with measured mass values, a more significant candidate it becomes. Because alternative isoforms originating from a single gene share a common peptide, they will have a similar score and thus can be concentrated.

[28] The present invention further provides a method for identifying alternative splicing isoforms by using peptide mass fingerprinting.

[29] More specifically, said method comprises following steps:

[30] - constructing a database for searching alternative splicing isoforms;

[31] - searching the mass data of the peptide fragments obtained from protein hydrolysis against said database;

[32] - scoring by algorithm which can score the potential of each candidate that have been

searched for the alternative splicing isoforms having a mass with match in the database; and,

[33] - identifying the isoforms by sorting them out by the score.

[34] In the present invention, transcripts from the species having an alternative splicing event are virtually translated into proteins and the resulting protein sequences are also virtually digested with a proteolytic enzyme as if they are digested with it in real. Consequently, generated peptides are stored in the database of the present invention.

[35] According to a method of one embodiment of the present invention, the database is constructed by processes including collecting transcripts for the species having an alternative splicing event and translating them into protein sequences, and the database may comprise mass and isopotential value of protein, site information of amino acid which might be involved in protein modification and sequence information of a peptide that is produced after the hydrolysis of the protein.

[36] According to a method of one embodiment of the present invention, the algorithm can be an algorithm which determines priority of each candidate with a process that mass data measured from an experiment is searched against the established database and then the results are calculated for masses that are in match with respective isoforms by using the following formulae 1, 2 and 3.

[37]
$$r_{m_i} = f_{m_i} / T \quad \text{Formula 1}$$

$$E_{m_i} = \log_{10}(r_{m_i}) \quad \text{Formula 2}$$

$$P_s = \sum_{j=1}^k E_{Mj} \quad \text{Formula 3}$$

[38] Definition of the symbols described in said formulae 1 to 3 is the same as described before.

[39] According to a method of one embodiment of the present invention, the implementation of said algorithm can be achieved by processes including, searching within error range of mass the mass data that has been supplied from the input section and the mass considering the error of mass against the database, sorting the results for each isoforms, scoring them by using said algorithm, and sorting them in an ascending order to decide the priority among candidates.

[40] The present invention still further provides recording media having a computer readable program to implement said method for identifying alternative splicing isoforms by using peptide mass fingerprinting.

[41] Herein after, preferred examples of the present invention will be explained in greater detail in view of the attached drawings.

[42] Figure 1 is a schematic diagram representing the system for identifying alternative

splicing isoforms according to the present invention.

[43] The system for identifying alternative splicing isoforms by using peptide mass fingerprinting according to the present invention includes an input section; a database; a search section; an algorithm; and an output section.

[44] The input section is where the mass data input of the peptide fragments obtained from protein hydrolysis occurs. Figure 3 shows a screen for an input section. Essential mass values, optional protein mass values and isopotential values are fed to the input section.

[45] The above-described database includes information regarding mass and isopotential value of protein sequence, site of amino acid that might be involved in protein modification, and peptide sequences that are produced after the enzyme treatment, etc. To construct the database, after collecting transcripts for the species having an alternative splicing, they are translated into protein sequences and redundant protein sequences are removed.

[46] The above-described search section is where a search against the database is carried out.

[47] The above-described algorithm scores potential of candidates that are found during search step in relation with the alternative splicing isoforms which have masses with match in the database constructed as above. More specifically, said algorithm can be an algorithm which determines priority of each candidate with a process that mass data measured from an experiment is searched against the established database and then the results are calculated for masses that are in match with respective isoforms by using the following formulae 1, 2 and 3.

[48]
$$r_{m_i} = f_{m_i} / T \quad \text{Formula 1}$$

$$E_{m_i} = \log_{10}(r_{m_i}) \quad \text{Formula 2}$$

$$P_s = \sum_{j=1}^k E_{Mj} \quad \text{Formula 3}$$

[49] Definition of the symbols described in said formulae 1 to 3 is the same as described before.

[50] The output section is where the candidate isoforms are sorted by their score. Figure 4 shows a screen for output section. Closer to the top of the output screen, it is more likely that it is a correct isoform.

[51] The present invention furthermore provides a method for identifying alternative splicing isoforms by using peptide mass fingerprinting. Said method for identifying alternative splicing isoforms by using peptide mass fingerprinting according to the present invention can provide a database which can be used for program search of

mass data that is generated from mass spectrometry. That is, after searching mass data generated from mass spectrometry against said database, the isoforms having a mass with match in the database are scored by using said algorithm, and the candidate isoforms are aligned in an ascending order with their scores.

- [52] In order to identify alternative splicing isoforms by using peptide mass fingerprinting, a database for searching mass data obtained from mass spectrometry, wherein an alternative splicing is taken into consideration, is required. In addition, an algorithm guaranteeing precise identification is necessary. Using a protein sequence comprising alternative splicing isoforms, a database which can be used for analyzing the results obtained from mass spectrometry is constructed based on the program for producing the search database according to the present invention. This database includes information regarding site of amino acid that might be involved in protein modification, isopotential value of each isoforms and mass of each isoforms, etc. The amino acid site at which a protein sequence can be modified indicates a site wherein the chemical structure of a certain amino acid can be modified after the translation of corresponding transcript.
- [53] Considering that an alternative splicing may occur, there can be several to hundred or more of isoforms can be present for a single gene. However, they tend to have a peptide sequence that is partially same to each other. In this regard, a new algorithm which can take an advantage of such characteristic is in need to efficiently distinguish the isoforms.
- [54] To identify the isoforms from the masses that are generated from an experiment, a search using a web-based software can be carried out. In this case, mass values that are essential for data input, optional protein mass values, and isopotential values are added (Figure 3). The added data is searched against the database described above, scored by said algorithm, and aligned in an ascending order with their scores (Figure 4). Closer to the top, it is more likely that it corresponds to a correct isoform. In other words, more significant candidate will have a lower score.
- [55] Detailed description of the algorithm of the present invention is given as a diagram of Figure 2. As it can be seen from Figure 2, mass value (201) obtained from a mass spectrometer is searched against database (202). A, B, C, D, L, M, and N in 202 are imaginary circles having a size proportional to the number of peptides that are in match after the search with said measured mass values. Bigger sized-circle indicates that there are more peptides in match. On the other hand, smaller the circle is, probability of accidental occurrence is lower. With the search of the measured mass values against the database, isoforms having a matching mass value is presented as scores as shown in 203. Each isoforms are aligned in an ascending order of score such as from C1, C2, ..., to C7. The isoforms originating from a single gene tend to share the same peptide

fragments, thus they end up having a similar score. As it is shown in 204, C1, C2, C3 isoforms originating from a single gene show an order of 1, 2 and 3 from the top. Then, an order of significance of each candidate is decided based on matched mass values of L, M, and N that are different to each other. In this case, because the size of L is smallest, C1 becomes the most significant candidate.

[56]
$$r_{m_i} = f_{m_i} / T \quad \text{Formula 1}$$

$$E_{m_i} = \log_{10}(r_{m_i}) \quad \text{Formula 2}$$

$$P_s = \sum_{j=1}^k E_{Mj} \quad \text{Formula 3}$$

[57] Definition of the symbols described in said formulae 1 to 3 is the same as described before.

[58] Mass data (201) measured by a mass spectrometer is searched against the mass value of the database (202) considering an error value. The number determined by search with all of the measured mass values corresponds to the size of dotted-line circle of Figure 2. Each mass value can be searched against the database. Circles A, B, C, D, L, M and N of (202) are represented by f_{m_i} and the formula 2 can be solved using the values obtained before. In addition, for the isoforms that have been searched using all of the measured mass values, E_{Mj} values of formula 2 which correspond to the mass values in match with each isoform are all added up to yield a specific score.

[59] The present invention still further provides recording media having a computer readable program to implement the method for identifying alternative splicing isoforms by using peptide mass fingerprinting.

[60] Recording media having a computer readable program can be any media that can be directly read and accessed through a computer. Examples of such recording media includes; magnetic recording media such as a floppy disk, a hard disk and a magnetic tape, etc., optical recording media such as CD-ROM, CD-R, CD, RW, DVD-ROM, DVD-RAM, and DVD-RW, etc., and electric recording media such as RAM and ROM, etc., and combinations thereof (e.g., magnetic/optical recording media such as MO, etc.), but are not limited thereto.

[61] Selection of an instrument that is used for recording and storing data into the above-described recording media and an instrument or a device that is used for reading the data information from recording media is determined by the kinds of recording media and an access method being used. Further, various kinds of data processor program, software, comparator, and format are used for recording in corresponding media the program that is required for the implementation of the method of the present invention. Information can be presented in a form of, for example, a binary file formatted with

commercially available software, text file or ASCII file.

- [62] A skilled person in the pertinent art would understand that the present invention can be implemented in another specific forms without changing the technical spirit or essential characteristics of the invention. Therefore, it must be understood that the examples described above are illustrative only and no way limit the present invention. The scope of the present invention is described by the claims that are given below, rather than by the detailed description of the invention. In addition, it should be understood that meaning and scope of the claims and variants or varied forms that can be deduced from the claims and their equivalents are all within the scope of the present invention.

Claims

- [1] A system encompassing entire processes for identifying alternative splicing isoforms using peptide mass fingerprinting, which comprises;
- an input section to which mass data measured for peptide fragments that are obtained from protein hydrolysis are input;
 - a database comprising mass and isopotential value of a protein, site information of an amino acid which might be involved in protein modification and sequence information of a peptide that is produced after the hydrolysis of the protein;
 - a search section wherein a search is carried out against the database;
 - an algorithm for scoring a potential of candidate alternative splicing isoforms having a mass with match found in database; and
 - an output section by which the candidate isoforms are aligned according to their score.
- [2] The system according to Claim 1, characterized in that said database is constructed by processes including collecting transcripts for a species having an alternative splicing event, translating them into protein sequences, and removing the overlapped protein sequences.
- [3] The system according to Claim 1, characterized in that said algorithm is an algorithm which determines priority of each candidate with a process that mass data measured from an experiment is searched against the database and then the results are calculated for masses that are in match with respective isoforms by using the following formulae 1, 2 and 3:

$$r_{m_i} = f_{m_i} / T \quad \text{Formula 1}$$

$$E_{m_i} = \log_{10}(r_{m_i}) \quad \text{Formula 2}$$

$$P_s = \sum_{j=1}^k E_{M_j} \quad \text{Formula 3}$$

(wherein, when measured mass value is represented by $m = \{m_1, m_2, \dots, m_i\}$ and a mass value with match in the database is represented by $M = \{M_1, M_2, \dots, M_k\}$, m_i indicates mass value of m , M_k indicates mass value of M and M_k is a single value that is in match with one candidate isoform among m_i , T represents total number of peptides that are searched from the database when the search is carried out using measured data, f_{m_i} represents the number of peptides that are searched with each of the mass values from the database, E_{m_i} is a logarithmic value of r_{m_i} , and P_s (score) is a sum of E_{M_j} for the mass values in match with respective isoforms).

- [4] A method for identifying alternative splicing isoforms by using peptide mass fingerprinting, which comprises;
- constructing a database for searching alternative splicing isoforms;
 - searching the mass data of the peptide fragments obtained from protein hydrolysis against said database;
 - scoring by algorithm which can score the potential of each candidate that has been searched for the alternative splicing isoforms having a mass with match in the database; and
 - identifying the isoforms by sorting them out by the score.

- [5] The method according to Claim 4, characterized in said algorithm is an algorithm which determines priority of each candidate with a process that mass data measured from an experiment is searched against the database and then the results are calculated for masses that are in match with respective isoforms by using the following formulae 1, 2 and 3:

$$r_{m_i} = f_{m_i} / T \quad \text{Formula 1}$$

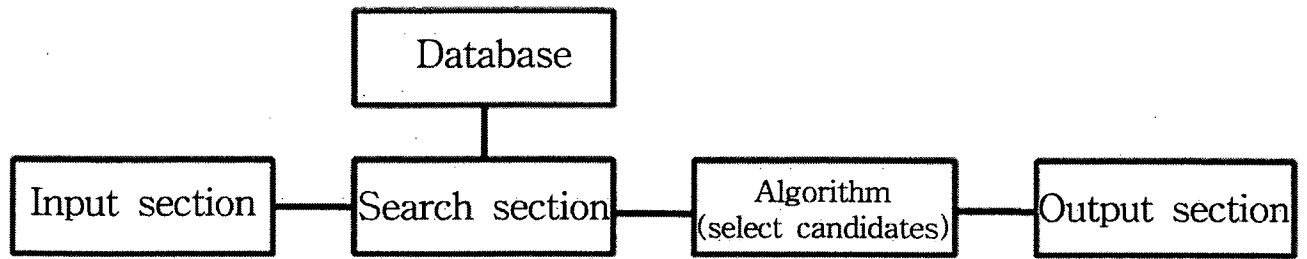
$$E_{m_i} = \log_{10}(r_{m_i}) \quad \text{Formula 2}$$

$$P_s = \sum_{j=1}^k E_{Mj} \quad \text{Formula 3}$$

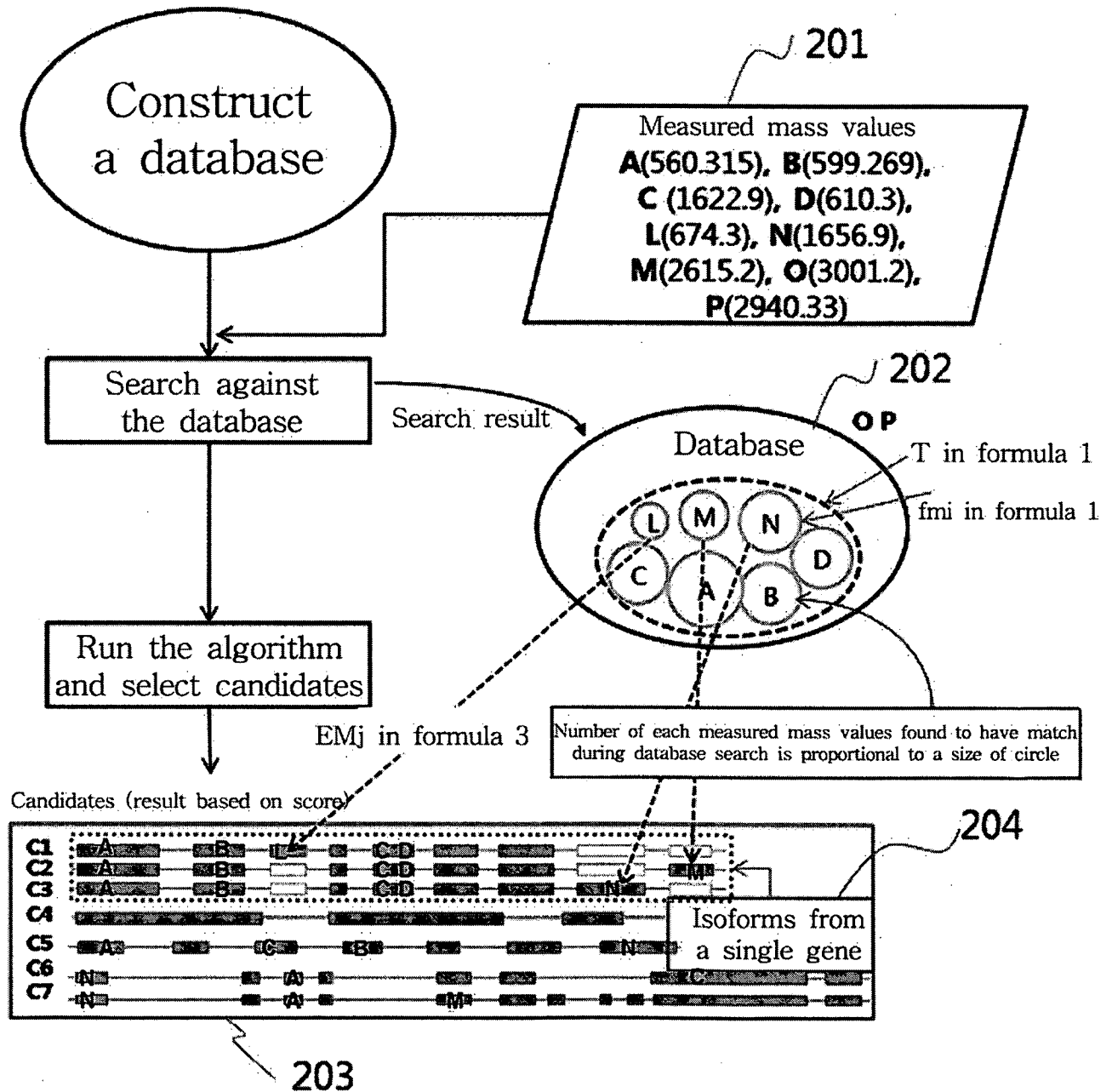
(wherein, definition of the symbols described in said formulae 1 to 3 is the same as described in Claim 3).

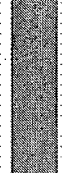
- [6] The method according to Claim 5, characterized in that the implementation of said algorithm is achieved by processes including, searching within error range of mass the mass data that is supplied from the input section and the mass considering the error of mass against the database, sorting the results for each isoforms, scoring them by using said algorithm, and sorting them in an ascending order to decide the priority among the candidates.
- [7] Recording media having a computer readable program to implement the method described in any one of Claims 4 to 6.

[Fig. 1]



[Fig. 2]





ASPMF

Alternative Splicing Isoform
using Peptide Mass Fingerprinting

[Home](#) | [Peptide Mass Fingerprinting](#) | [Instruction](#) | [Links](#)

Peptide Mass Fingerprinting

Species	☐ Arabidopsis ☒ Human ☐ Mouse		
Algorithm	☒ ASPMF ☐ MOWSE		
pI		within range	$\pm$
M.W.		kD within range	$\pm$ kD
Mass Fingerprint	1239.56 1447.68 1387.7 1552.84 2715.39		
Mass Tolerance	$\pm$ Da	Mass value	☒ [M] ☐ [M+H] ⁺
Type	☒ Monoisotopic ☐ Average		
Missed Cleavage	☐ 0 ☒ 1 ☐ 2		
Modification	No modification Acetyl (K) Acetyl (N-term) Amide (C-term) Biotinylated (K)		
Max Reported hit		Min. # peptides	3
E-mail			

[Fig. 4]

Protein ID	H10C12853.12	Linked DB	ECgene
Description	length=158 [Human]		
Molecular Weight	16294.2	pI Value	9.0824
Score	-4.0083160817		
Number of Matches	5		
Protein ID	H5C19486.1	Linked DB	ECgene
Description	length=44 [Human]		
Molecular Weight	4587.45	pI Value	7.8112
Score	-1.19067549412		
Number of Matches	1		
Protein ID	H8C7103.1	Linked DB	ECgene
Description	length=62 [Human]		
Molecular Weight	7236.86	pI Value	9.2792
Score	-1.19067549412		
Number of Matches	1		
Protein ID	H10C12853.15	Linked DB	ECgene
Description	length=77 [Human]		
Molecular Weight	8089.09	pI Value	4.2885
Score	-1.19067549412		
Number of Matches	1		
Protein ID	H10C12853.19	Linked DB	ECgene
Description	length=100 [Human]		
Molecular Weight	10544.4	pI Value	4.6946
Score	-1.19067549412		
Number of Matches	1		
Protein ID	H10C12853.18	Linked DB	ECgene
Description	length=133 [Human]		
Molecular Weight	14420.7	pI Value	6.6267
Score	-1.19067549412		
Number of Matches	1		

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2008/002390**A. CLASSIFICATION OF SUBJECT MATTER****G01N 27/62(2006.01)i**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 8 G01N 27/62, 33/00, 33/48, 33/50

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

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

D/B : eKIPASS

KEY WORD : Protein, Peptid, Mass, Alternative splicing isoform

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-0531207 B1 (CBS Soft Co.) 21 Nov. 2002 see the whole document	1-7
A	KR 10-2007-0017676 A (Korea Basic Science Institute) 13 Feb. 2007 see the whole document	1-7
A	KR 10-0757040 B1 (Object Interaction Tech. Inc.) 3 Sep. 2007 see the whole document	1-7
A	JP 2006-58111 A (Simazu Sesakusyo KK.) 2 Mar. 2006 see the whole document	1-7
A	US 2007-0224704 A1 (Neal Gordon) 27 Sep. 2007 see the whole document -----END OF DOCUMENT-----	1-7



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

21 AUGUST 2008 (21.08.2008)

Date of mailing of the international search report

21 AUGUST 2008 (21.08.2008)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-
gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

CHOI, SOK JIN

Telephone No. 82-42-481-8472



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/KR2008/002390Patent document
cited in search reportPublication
datePatent family
member(s)Publication
date

KR 10-0531207	B1	21.11.2002	NONE		
KR 10-2007-0017676	A	13.02.2007	WO 2007-018327	A1	15.02.2007
KR 10-0757040	B1	03.09.2007	KR 2007062025	A	15.06.2007
JP 2006-58111	A	02.03.2006	JP 2006-058111		02.03.2006
			JP 2006-058111	A2	02.03.2006
US 2007-0224704	A1	27.09.2007	US 2007-224628	A1	27.09.2007
			US 2007-224628	AA	27.09.2007
			US 2007-224704	AA	27.09.2007
			WO 2007-112012	A2	04.10.2007
			WO 2007-112012	A3	22.11.2007

-----END OF DOCUMENT-----