

SYSTEM AND METHOD FOR MASS SPECTROMETRY DATA SEARCH USING ADAPTED QUANTUM GROVER ALGORITHM

Field of the Invention

[0001] The present invention generally relates to mass spectrometry technologies for identification/verification of chemical molecules, and more particularly to a system and method for mass spectrometry data search using adapted quantum Grover algorithm.

Background of the Invention

[0002] In recent years, protein separation methods coupled with various mass spectrometry (MS) technologies have evolved as the dominant tools in the field of proteomics for identification of amino acid sequences of peptides and proteins. A mass spectrometer is an instrument that measures the mass-to-charge (m/z) ratio of individual molecules that have been converted into electrically charged gas-phase molecules, or ions. These ions are filtered in such a way as to produce an ordered separation of the ions as they pass through the mass spectrometer, ordered from low to high m/z ratios. The data is typically displayed as a plot of intensity vs. m/z ratio.

[0003] Two “soft ionization” methods, namely matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), have made it possible to analyze large biomolecules, such as peptides and proteins. One of the common methods is that a mixture of proteins is first separated by liquid chromatography (LC) followed by tandem mass spectrometry (MS/MS). In this procedure, a mixture of charged peptides is separated in the first MS according to their m/z ratios to create a list of the most intense peptide peaks. In the second MS analysis, the instrument is adjusted so that only a specific m/z species is directed into a collision cell to generate “daughter” ions derived from the “parent” species. Using the appropriate collision energy, fragments, each of which differs by the mass of a single amino acid, is generated. The daughter fragments are separated according to their m/z creating the MS/MS spectrum, and the sequence of the peptide can then be deduced from the resulting fragments. By comparison with predicted sequences in the databases, the identity of the peptide is revealed.

[0004] Mass spectrometry is experiencing a period of rapid growth based on applications in proteomic analyses. As a consequence of this rapid growth, the need to collect the large amounts of information produced and make it available to researchers worldwide has resulted in a number of MS databases. For example, the Illinois Bio-Grid Mass Spectrometry Database (IBG-MSD) is a public database curated and annotated of empirically derived mass spectra of peptides. The goal of the database is to address the need for a public database of mass spectrometry data and implement a useful web interface that will allow researchers to access the data and perform a variety of tasks based on their individual needs. The identification of proteins is becoming an easier and more accurate process due to the use of tandem mass spectrometry (MS/MS).

[0005] Although the generation of raw MS/MS spectra has become easier, the analysis and identification of that data is a difficult process to perform manually. For this reason, database searching is the most popular approach. Among the popular ones are PEAKS (See, Ma, B. et al., *Rapid Communications in Mass Spectrometry*. 17: 2337 (2003)), MASCOT (Web address: <http://www.matrixscience.com/home.html>), Phenyx (Web address: <http://phenyx-ms.com/>) and OMSSA (Geer, L.Y. *J Proteome Res.* 3: 958-64 (2004)). In the case of high throughput proteomics, it involves analysis of hundreds of thousands of peptide spectra derived from biological samples. These spectra can be identified by four general types of algorithms: de novo calling of the sequence directly from the spectrum (see, e.g., Johnson and Taylor. *Mol Biotechnol* 22:301-315 (2002)), the use of unambiguous peptide sequence tags derived from spectra that are used to search known sequences (see, e.g., Mann and Wilm. *Anal Chem.* 66:4390-4399 (1994)), cross-correlation methods that correlate experimental spectra with theoretical spectra (see, e.g., Eng, J.K., et al. *J Am Soc for Mass Spectrometry* 5: 976-989 (1994)), and probability-based matching that calculates a score based on the statistical significance of a match between an observed peptide fragment and those calculated from a sequence search library (see, e.g., Field, H.I., et al. *Proteomics* 2: 36-47 (2002)). However, as shown in FIG 1, using an experimental MS spectra to search a large database of theoretical MS spectra in the protein identification process employing tandem MS/MS technique is still tedious and time-consuming with current methods.

Summary of the Invention

[0006] One embodiment of the present invention provides a process executable within an electronic microprocessor. The process comprises procuring experimental mass spectrometry spectrum data of a chemical entity from experiments; procuring a theoretical mass spectrometry database containing all theoretical mass spectrometry spectrum data that cover the chemical entity according to pre-determined criteria; and searching the experimental mass spectrometry spectrum data using individual member of the theoretical mass spectrometry database for the match between one member of the theoretical mass spectrometry database with the experimental mass spectrometry data according to preset matching criteria; thereby when a match is found, the identity of the experimented chemical entity is accorded with the one of the matched theoretical chemical entity.

[0007] In another embodiment of the process, the step of procuring a theoretical mass spectrometry database comprises generating theoretical mass spectrometry spectra according to the m/z ratio of the experimented chemical identity.

[0008] In yet another embodiment of the process, the experimented chemical entity is a polypeptide. In a further embodiment of the process, the experimental mass spectrometry spectrum data is from tandem mass spectrometry analysis, and wherein the theoretical mass spectrometry database contains theoretical mass spectrometry spectrum data of polypeptides having substantially the same m/z ratio of the experimented polypeptide according to pre-determined criteria.

[0009] In still another embodiment of the process, the step of procuring experimental mass spectrometry spectrum data of a chemical entity from experiments comprises padding the data to make it $N=2^n$, and separating the experimental mass spectrometry data into two states so that the experimental mass spectrometry spectrum data can be searched by quantum Grover algorithm. In a further embodiment of the process, the step of padding comprises adding spectrum points that are small enough so that they will not be matched by any theoretical mass spectrometry points. In another further embodiment of the process, the two states are noise mass spectra and fragment mass spectra.

[0010] Another embodiment of the present invention provides a system comprising: an electronic microprocessor embedded therein a set of computer programs

that can be executed upon instructions; wherein the set of computer programs when executed perform the followings: procuring experimental mass spectrometry spectrum data of a chemical entity from experiments; procuring a theoretical mass spectrometry database containing all theoretical mass spectrometry spectrum data that cover the chemical entity according to pre-determined criteria; and searching the experimental mass spectrometry spectrum data using individual member of the theoretical mass spectrometry database for the match between one member of the theoretical mass spectrometry database with the experimental mass spectrometry data according to preset matching criteria; thereby when a match is found, the identity of the experimented chemical entity is accorded with the one of the matched theoretical chemical entity.

[0011] These and additional objectives and advantages of the invention will become apparent from the following detailed description of preferred embodiments thereof in connection with the accompanying drawings.

Brief Description of the Drawings

[0012] Preferred embodiments according to the present invention will now be described with reference to the Figures, in which like reference numerals denote like elements.

[0013] FIG 1 illustrates the conventional approach of using the experimental MS to search the theoretical MS database.

[0014] FIG 2 illustrates the new approach of using each member of the theoretical MS database to search the experimental MS spectra in accordance with one embodiment of the present invention.

[0015] FIG 3 shows a graphic presentation of the overlay of experimental MS spectra and theoretical MS spectra of the same peptide, **MTDQEAIQDLWQWR**.

[0016] FIG 4 shows the changes in the amplitude of the states at different iteration r.

[0017] FIG 5 is a graphic presentation of the probability of success with respect to the number of iteration.

[0018] FIG 6 is a panel of graphic presentations showing how the number of match j affects the probability of success in the case where N=64, where the “--*--*--” plot is the

fidelity of the input states compared to the final states, and the “-----” plot is the probability of success.

[0019] FIG 7 shows the plots of r Number of Iteration Versus Number of Matched m/z Ratio j.

Detailed Description of the Invention

[0020] The present invention may be understood more readily by reference to the following detailed description of certain embodiments of the invention.

[0021] Throughout this application, where publications are referenced, the disclosures of these publications are hereby incorporated by reference, in their entireties, into this application in order to more fully describe the state of art to which this invention pertains.

[0022] The present invention provides a system and method that identifies a chemical molecule using an experimental MS spectrum of the chemical molecule to search a theoretical MS spectrum database of all possible chemical molecules with substantially the same m/z ratio. This approach makes the search suitable for quantum Grover algorithm. While the following description will use peptides as the model of chemical molecules, it is to be appreciated that the applications of the present invention is not so limited.

[0023] A typical experimental mass spectrum $E - MS$ contains two types of m/z ratio data: experimental noise mass spectra $n - MS$ and fragment mass spectra $f - MS$. In the conventional searches afore-described, $E - MS$ is used to search a big pool of possible theoretical mass spectra as shown in FIG 1. However, the experimental noise mass spectra $n - MS$ component will be detrimental to the search because they are highly unlikely matched with any theoretical mass spectrum.

[0024] In order to find a better method for searching MS spectra, the inventor of the present invention discovered that it was feasible to use each member of a theoretical MS spectrum database to search an experimental MS spectrum as shown in FIG 2, and it can be implemented using quantum computation. This approach reduces the register size needed to form the searchable database and it also allows distributed quantum computation to be carried out in parallel, resulting in increased speeds for searches. More importantly,

as discussed in detail hereinafter, this reversed search scheme is suitable for the application of quantum Grover algorithm.

[0025] The Grover's quantum search algorithm is well known. See, Grover, L.K., Phys. Rev. Lett., 78:325 (1997). Briefly, consider a search space D containing N elements, assuming that $N = 2^n$, where n is an integer. The elements of D are represented using an n -qubit register containing the indices $i = 0, 1, \dots, N - 1$. We assume that there is k marked elements in the search space that are the solution to the search problem. It is also assumed that there exists a function $f: D \rightarrow 0, 1$, such that $f = 1$ for the marked elements, and $f = 0$ for the rest. The search for a marked element becomes a search for an element for which $f = 1$. To solve this problem using a classical computer, it needs to evaluate f for each element, one by one, until a marked state is found. Thus on average, $N/2$ evaluations of f are necessary and in the worst case, N evaluations are required. For a quantum computer, the function f can be evaluated coherently.

[0026] Grover's search algorithm uses a sequence of unitary operations and it can locate a marked element using only $O(\sqrt{N/r})$ coherent queries of f , which is indeed faster than the classical computers. To describe the algorithm, we consider a quantum system having initial state of register as $|i\rangle = |i_1 \dots i_n\rangle$ of n qubits, the register is subjected to local Hadamard transformations, $H^{\otimes n}$, resulting in a linearly superposed state of all computational basis:

[0027]

$$|\eta\rangle = H^{\otimes n} |0 \dots 0\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \quad (1)$$

[0028] where

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

[0029] The state is then subjected to a series of inversion and diffusion processes before measurement. During the inversion stage, the state undergoes the unitary transformation $I = 1 - 2 \sum_{mi \in M} |mi\rangle \langle mi|$, where $M = \{m1, m2, \dots\}$ are the marked states, and

in the diffusion stage, the state undergoes the transformation $D = -1+2|\eta\rangle\langle\eta|$. In this way, the state prior to measurement after r iterations is $|\psi_r\rangle = (DI)^r |\eta\rangle \equiv U_G^r |\eta\rangle$.

[0030] The search algorithm can in fact be viewed as a two-dimensional problem. If we denote the k marked states as vector $|m\rangle$ and the orthogonal vector (unmarked states) as $|m^\perp\rangle$, it is not hard to show that the state at the end of r iterations is given by:

[0031]

$$|\psi_r\rangle = \cos \theta_r |m\rangle + \sin \theta_r |m^\perp\rangle, \quad (2)$$

[0032] After another iteration, the state becomes:

$$|\psi_{r+1}\rangle = U|\psi_r\rangle = \cos(\theta_r - \alpha)|m\rangle + \sin(\theta_r - \alpha)|m^\perp\rangle \quad (4)$$

[0033] where

$$\tan \alpha = \frac{2\sqrt{k(N-k)}}{N-2k}.$$

[0034] The probability of success will then be given by:

[0035]

$$P = (\cos(\theta_r - \alpha))^2 \quad (5)$$

[0036] Simply put, in the quantum search algorithm, the system s consists of two distinct states, namely the marked state labeled as $|m\rangle$ and the unmarked state labeled as $|m^\perp\rangle$. The search algorithm iterates to amplify the probability of success of getting the marked state from the system. There must be a special function $f(s)$ such that it produces a 1 when it hit a marked state and 0 otherwise.

[0037] The function $f(s)$ is formed such that if the experimental mass spectrum matches one of the theoretical mass spectrum within a acceptable tolerance of Δs , then the function returns 1, 0 otherwise. For very accurate machine, a Δs of 0.3 Dalton can be used.

[0038] In view of the requirement of two states (marked/unmarked) of a system by the Grover's quantum search algorithm, a typical experimental mass spectrum $E - MS$ with experimental noise mass spectra $n - MS$ and fragment mass spectra $f - MS$ as discussed above satisfies this requirement perfectly. In one embodiment, the fragment mass spectra f

– MS is assigned as the marked state $|m\rangle$, and the experimental noise mass spectra $n - MS$ as unmarked state $|m^\perp\rangle$ respectively.

[0039] However, the Grover's quantum search algorithm still cannot be straightforwardly applied because of the following challenges. First, the database for an oracle is fixed for the search, whereas mass spectra databases come with different sizes of theoretical spectra. Second, the quantum register is formed with equal probability intensity and equal quantum bit distance whereas the mass spectra have different intensities and different inter spectra distance. Third, there is a need of $f(s)$ for quantum search algorithm to work.

[0040] A typical experimental mass spectrum $E - MS$ is given in two rows, which are the m/z ratio and the intensity respectively. Table 1 shows a simple mass spectrum in plain text format. The first row is always the precursor's peptide mass Dalton. The second row onwards are the m/z ratios in an incremental order. The $E - MS$ shown in Table 1 is known to match to a peptide fragment with the following amino acid sequence: **MTDQEAIQDLWQWR**.

[0041] Table 1. A simple mass spectrum in plain text format

1818.9631 2
86.1121 2.2857
175.1481 8.1905
205.1197 2.1905
244.1234 5.3333
347.1403 3.1905
348.1497 3.1905
356.1350 3.1905
361.2055 3.5714
476.2550 8.1429
489.2869 14.8571
512.3115 3.3333
545.2774 6.0476
552.2336 2.1905
570.2354 2.1905

[0042] The theoretical mass spectrum can be predicted using some of the known tools or through simple calculation. For example, tools from InsilicoSpectro (<http://insilicospectro.vital-it.ch/>) provide perl scripts fragmentatorHome.pl to predict the

theoretical fragmentation spectra. Table 2 shows the result of such fragmentation prediction.

[0043] Table 2. Predicted fragmentation mass spectra of the peptide sequence: MTDQEAIQDLWQWR

[0044]

MTDQEAIQDLWQWR, mass=1818.833905 Da														
Ion types	M	T	D	Q	E	A	I	Q	D	L	W	Q	W	R
b		233.10	348.12	476.18	605.22	676.26	789.35	917.40	1032.43	1145.51	1331.59	1459.65	1645.73	
y	1688.80	1587.76	1472.73	1344.67	1215.63	1144.59	1031.51	903.45	788.42	675.34	489.26	361.20	175.12	

[0045] The predicted spectra for b and y types ions are put together and sorted in ascending order and for the theoretical mass spectra (a total of 25 of tem) as shown in Table 3.

[0046] Table 3. List of predicted spectra in ascending order of MTDQEAIQDLWQWR

[0047]

175.1
233.1
348.1
361.2
476.2
489.3
605.2
675.3
676.3
788.4
789.4
903.5
917.4
1032
1032
1145
1146
1216
1332
1345
1460
1473
1588
1646
1689

[0048] The mass spectra from a real mass spectrometry for MTDQEAIQDLWQWR contain 60 peaks as shown in Table 4.

[0049] Table 4 shows the real spectra readings from a mass spectrometry.

[0050]

ID	1818.9631	2
1	86.1121	2.2857
2	175.1481	8.1905
3	205.1197	2.1905
4	244.1234	5.3333
5	347.1403	3.1905
6	348.1497	3.1905
7	356.135	3.1905
8	361.2055	3.5714
9	476.255	8.1429
10	489.2869	14.8571

[0051]

11	512.3115	3.3333			
12	545.2774	6.0476			
13	552.2336	2.1905			
14	570.2354	2.1905			
15	571.2495	3.1905			
16	587.2164	5.2857			
17	588.3699	1.4286			
18	605.2498	6.2857			
19	641.2842	5.0476			
20	648.3386	2.0952			
21	658.3339	5.8571			
22	659.2545	3.1905			
23	675.4128	9.1429			
24	676.3542	13.0952			
25	677.3265	3.1905			
26	756.3756	2.381			
27	786.4037	3.1905	44	912.504	1.1429
28	788.4521	4.1905	45	915.5162	2.1905
29	789.5041	12.7619	46	1014.5417	6.1905
30	790.3947	3.4762	47	1031.6252	14.3333
31	791.421	2.1905	48	1032.5697	12.8571
32	901.5813	2.8571	49	1033.5262	7
33	903.4973	5.9048	50	1033.8372	2.4286
34	904.5272	9.7619	51	1144.7045	10.1905
35	909.9626	36.7619	52	1145.616	3.1905
36	910.1725	10.2857	53	1146.6858	7.1429
37	910.4973	61.2857	54	1148.651	2.1905
38	910.9562	45.381	55	1198.7379	2.1905
39	911.4802	37.4286	56	1215.6304	4.2857
40	911.7206	2.9048	57	1216.6534	7.0952
41	911.863	6.0952	58	1218.4341	2.1429
42	912.005	17	59	1235.7609	2.1905
43	912.3879	4.9048	60	1344.8964	4.0952

[0052] The data shown in Tables 3 and 4 are plotted in one graphic as shown in FIG 3. From here, we can observe visually the matched m/z ratio and the unmatched m/z ratio. There are in fact only 17 matched m/z ratios, these ratio are listed in Table 5.

[0053] As seen from the data of Tables 3 and 4, the real mass spectra fail to match the predicted mass spectra perfectly. The discrepancy may be derived from many factors including instrumentation and the peptide itself. Therefore, the function $f(s)$ is formed

such that if the experimental mass spectrum matches one of the theoretical mass spectrum within an acceptable tolerance of Δs , then the function returns 1, or 0 otherwise. In this example, a Δs of 0.3 Dalton is used.

[0054] As discussed above, the Grover's quantum search algorithm requires a database size of $N = 2^n$. As for the exemplary peptide, the *E* - *MS* as shown in Table 4 has only 60 spectra. To form a database size of $N = 2^n$, the smallest number for n is 6, then $N = 64$, hence a padding of 4 more spectra is applied to the *E* - *MS*, probably using values that are smaller than the smallest amino acid, in the range less than 50 Dalton. That forms the first "database" to be searched by the theoretical *m/z* ratio *t* - *MS*, which has a maximum of 25 spectra.

[0055] For simplicity, we now form a vector consisting the first 4 padding spectra and the 60 experimental spectra, as shown below:

[0056]

$$|e\rangle = [s_0, s_1, \dots, s_{63}]$$

[0057] In our simulation, we also know that by applying $f(s)$, the following theoretical spectra are found (with ID from the *E* - *MS* identified) within the Δs set earlier. New IDs are created using spectra found in Table 4, with additional 4 zero-padding spectra inserted at the beginning of the list, making the first *E* - *MS* starts at ID = 4 and the last ID be 63. The simulated search results are shown in Table 5.

[0058] Table 5. Theoretical mass spectra with matching ID, giving the $f(s) = 1$ for the simulated search

MS	Intensity	New ID (Original ID)
175.12	8.1905	5 (2)
233.1	0	-
348.12	3.1905	9 (6)
361.2	3.5714	11 (8)
476.18	8.1429	12 (9)
489.26	14.8571	13 (10)
605.22	6.2857	21 (18)
675.34	9.1429	26 (23)
676.26	13.0952	27 (24)
788.42	4.1905	31 (28)
789.35	12.7619	32 (29)
903.45	5.9048	36 (33)
917.4	0	-
1031.51	14.3333	50 (47)
1032.43	12.8571	51 (48)
1144.59	10.1905	54 (51)
1145.51	3.1905	55 (52)
1215.63	4.2857	59 (56)
1331.59	0	-
1344.67	4.0952	63 (60)
1459.65	0	-
1472.73	0	-
1587.76	0	-
1645.73	0	-
1688.8	0	-

[0059] Now there is provided a more detailed description of the application of the Grover's quantum search algorithm to the search of mass spectra using the databases constructed according to the principles discussed above.

[0060] The basic operation of the Grover's quantum search algorithm has been described earlier. The probability of success is then measured using the final state, where the sum of the probabilities of the experimental spectra matching IDs is calculated. A threshold can then be set by users to accept the search results or reject them. In an ideal case, the probability of success is equal to one at the optimal r iteration. Theoretically, the

optimal iteration to achieve the best search result is given by $\frac{\pi}{4} \sqrt{\frac{N}{j}}$, where j refers to the

number of matched spectra. The fidelity of between the input spectra states and the final states can also be calculated using the following equation:

$$F = \langle e | \Psi | e \rangle, \text{ where } \Psi = |\psi_r\rangle\langle\psi_r| \quad (6)$$

[0061] FIG 4 shows the changes in the amplitude of the states at different iteration

r. Knowing that $N = 64$ and $j=17$, the optimum r is equal to 1 according to $\frac{\pi}{4} \sqrt{\frac{N}{j}}$, hence it is the best iteration to look at the states. The probability of success with respect to the number of iteration is shown in FIG 5. It is observed that the maximum value is about 0.8 for $j=17$ (17 matched out of 25 theoretical spectra).

[0062] Now we look into how the number of match j would affect the probability of success in the case where $N=64$. Plot the probability of success versus the number of mass spectra match as shown in FIG 6. The “--*--*” plot is the fidelity of the input states compared to the final states. The “-----” plot is the probability of success.

[0063] If the probability of success of a qubit is measured to be near certainty ($=1$) after the first iteration, then we know that there are at least 12 matches ($j=12$) in the case where $N=64$. This can be used as criterion to accept the query spectra as the good matched spectra, and rejected those where the probability of success is less than 0.75 at the first iteration, i.e., we only need one iteration of the quantum search.

[0064] Now we investigate the effect in changing in the database size N , to the number of match j and the number of minimum iteration needed in each case based on

$r = \frac{\pi}{4} \sqrt{\frac{N}{j}}$. In a typical tandem mass spectrometry, the number of database N varies from

below 100 spectra to above 1000 spectra. FIG 7 shows the plots of r Number of Iteration Versus Number of Matched m/z Ratio j . It is observed that the number of iteration needed decreases as the number of matched m/z ratio j increases.

[0065] While the present invention has been described with reference to particular embodiments, it will be understood that the embodiments are illustrative and that the invention scope is not so limited. Alternative embodiments of the present invention will become apparent to those having ordinary skill in the art to which the present invention pertains. Such alternate embodiments are considered to be encompassed within the spirit

and scope of the present invention. Accordingly, the scope of the present invention is described by the appended claims and is supported by the foregoing description.

CLAIMS

What is claimed is:

1. A process executable within an electronic microprocessor, said process comprising:
procuring experimental mass spectrometry spectrum data of a chemical entity from experiments;
procuring a theoretical mass spectrometry database containing all theoretical mass spectrometry spectrum data that cover the chemical entity according to pre-determined criteria; and
searching the experimental mass spectrometry spectrum data using individual member of the theoretical mass spectrometry database for the match between one member of the theoretical mass spectrometry database with the experimental mass spectrometry data according to preset matching criteria;
thereby when a match is found, the identity of the experimented chemical entity is accorded with the one of the matched theoretical chemical entity.
2. The process of claim 1, wherein the step of procuring a theoretical mass spectrometry database comprises generating theoretical mass spectrometry spectra according to the m/z ratio of the experimented chemical identity.
3. The process of claim 1, wherein the experimented chemical entity is a polypeptide.
4. The process of claim 3, wherein the experimental mass spectrometry spectrum data is from tandem mass spectrometry analysis, and wherein the theoretical mass spectrometry database contains theoretical mass spectrometry spectrum data of polypeptides having substantially the same m/z ratio of the experimented polypeptide according to pre-determined criteria.
5. The process of claim 1, wherein the step of procuring experimental mass spectrometry spectrum data of a chemical entity from experiments comprises padding the

data to make it $N=2^n$, wherein N is the total number of points of the experimental mass spectrometry data and n is an integer, and separating the experimental mass spectrometry data into two states so that the experimental mass spectrometry spectrum data can be searched by quantum Grover algorithm.

6. The process of claim 5, wherein the step of padding comprises adding spectrum points that are small enough so that they will not be matched by any theoretical mass spectrometry points.

7. The process of claim 5, wherein the two states are noise mass spectra and fragment mass spectra.

8. A system comprising:

an electronic microprocessor embedded therein a set of computer programs that can be executed upon instructions; wherein the set of computer programs when executed perform the followings:

procuring experimental mass spectrometry spectrum data of a chemical entity from experiments;

procuring a theoretical mass spectrometry database containing all theoretical mass spectrometry spectrum data that cover the chemical entity according to pre-determined criteria; and

searching the experimental mass spectrometry spectrum data using individual member of the theoretical mass spectrometry database for the match between one member of the theoretical mass spectrometry database with the experimental mass spectrometry data according to preset matching criteria;

thereby when a match is found, the identity of the experimented chemical entity is accorded with the one of the matched theoretical chemical entity.

9. The system of claim 8, wherein the step of procuring a theoretical mass spectrometry database comprises generating theoretical mass spectrometry spectra according to the m/z ratio of the experimented chemical identity.

10. The system of claim 8, wherein the experimented chemical entity is a polypeptide.
11. The system of claim 10, wherein the experimental mass spectrometry spectrum data is from tandem mass spectrometry analysis, and wherein the theoretical mass spectrometry database contains theoretical mass spectrometry spectrum data of polypeptides having substantially the same m/z ratio of the experimented polypeptide according to pre-determined criteria.
12. The system of claim 8, wherein the step of procuring experimental mass spectrometry spectrum data of a chemical entity from experiments comprises padding the data to make it $N=2^n$, wherein N is the total number of points of the experimental mass spectrometry data and n is an integer, and separating the experimental mass spectrometry data into two states so that the experimental mass spectrometry spectrum data can be searched by quantum Grover algorithm.
13. The system of claim 12, wherein the step of padding comprises adding spectrum points that are small enough so that they will not be matched by any theoretical mass spectrometry points.
14. The system of claim 12, wherein the two states are noise mass spectra and fragment mass spectra.