Claims

1. A method for identifying an unknown biological sample, the method comprising:

receiving more than two sample measurements for the unknown biological sample, each sample measurement for an attribute of the unknown biological sample, each attribute being different from another attribute;

calculating a sample point in a two-dimensional plot from the more than two sample measurements for the unknown biological sample, wherein the sample point is calculated using a first algorithm capable of compressing the more than two sample measurements into the sample point in the two-dimensional plot;

wherein the two-dimensional plot comprises a plurality of stored reference points corresponding to respective known biological compounds, ~~and~~ wherein each reference point is calculated from a plurality of reference measurements for more than two attributes of the corresponding known biological compound, each attribute being different from another attribute, and wherein each reference point is calculated using a second algorithm capable of compressing the plurality of reference measurements into a reference point in the two-dimensional plot; and

identifying the unknown biological sample by comparing the sample point against the plurality of reference points in the two-dimensional plot,

wherein each of the attributes of the unknown biological compound corresponds to a respective one of the attributes of the known biological compounds, and

wherein the unknown biological sample comprises one of the following: glycan, metabolite, antibody.

2. The method according to claim 1, wherein each reference measurement is obtained experimentally or by a machine learning algorithm.

3. The method according to claim 1, wherein the reference measurements for at least one known biological compound are obtained by performing two or more of the following on the at least one known biological compound: liquid chromatography, mass spectrometry, ion mobility, tandem mass spectrometry.

4. The method according to claim 1 or 3, wherein the reference measurements for at least one known biological compound are predicted based on the plurality of reference measurements for at least one other known biological compound.~~,~~

5. The method according to claim 1, further comprising forming the two-dimensional plot prior to receiving the more than two sample measurements for the unknown biological sample.

6. The method according to claim 5, wherein forming the two-dimensional plot comprises the following for at least one of the known biological compounds:

analysing the at least one of the known biological compounds with experimental devices to obtain the plurality of reference measurements for the at least one of the known biological compounds; and

calculating the~~a~~ reference point in two-dimension from the plurality of reference measurements for the at least one of the known biological compounds.

7. The method according to claim 6, wherein the experimental devices comprise two or more of the following: liquid chromatography, mass spectrometry, tandem mass spectrometry, ion mobility.

8. The method according to any one of claims 5 to 7, wherein forming the two-dimensional plot comprises the following for at least one of the known biological compounds:

predicting the plurality of reference measurements for the at least one of the known biological compounds based on the plurality of reference measurements for at least one other known biological compound; and

calculating the~~a~~ reference point in two-dimension from the predicted plurality of reference measurements for the at least one of the known biological compounds.

9. The method according to claim 8, wherein predicting the plurality of reference measurements for the at least one of the known biological compounds comprises using a machine learning algorithm.

10. The method according to any one of claims 5 to 9, wherein forming the two-dimensional plot comprises:

categorizing the known biological compounds into multiple groups of isomers; and

categorizing the reference points into multiple groups of reference points corresponding to respective groups of isomers, wherein each reference point is categorized into the group of reference points corresponding to the group of isomers into which the corresponding known biological compound is categorized.

11. The method according to claim 10, wherein categorizing the known biological compounds into multiple groups of isomers comprises categorizing each known biological compound based on a mass value of the known biological compound.

12. The method according to any preceding claim, wherein the second algorithm for calculating each reference point comprises principal component analysis, and each reference point is calculated by performing principal component analysis on the plurality of reference measurements.

13. The method according to claim 12, wherein performing principal component analysis on the plurality of reference measurements comprises:

transforming the plurality of reference measurements into a plurality of principal components, wherein the principal components are in an order such that variances of the principal components from a first principal component to a last principal component are in a descending order and each principal component is orthogonal to a next principal component in the order; and

using the first and second principal components to form the reference point in two-dimension.

14. The method according to any preceding claim, wherein the first algorithm for calculating the sample point in the two-dimensional plot from the more than two sample measurements for the unknown biological sample comprises principal component analysis, and the sample point is calculated by performing principal component analysis on the more than two sample measurements.

15. The method according to claim 14 when dependent on claim~~statement~~ 13, wherein transforming the plurality of reference measurements into a plurality of principal components comprises calculating a plurality of principal component parameters and performing principal

component analysis on the more than two sample measurements comprises using the plurality of principal component parameters.

16. The method according to claim 15, wherein performing principal component analysis on the plurality of sample measurements comprises:

transforming the plurality of sample measurements into a plurality of principal components using the plurality of principal component parameters, wherein the principal components are in an order such that variances of the principal components from a first principal component to a last principal component are in a descending order and wherein each principal component is orthogonal to a next principal component in the order; and

using the first and second principal components to form the sample point in the two-dimensional plot.

17. The method according to any preceding claim, wherein identifying the unknown biological sample comprises:

determining a reference point nearest to the sample point in the two-dimensional plot; and

identifying the unknown biological sample as the known biological compound corresponding to the determined nearest reference point.

18. The method according to claim 17 when dependent on claim~~statement~~ 10, wherein identifying the unknown biological sample further comprises the following prior to determining the reference point nearest to the sample point in the two-dimensional plot:

categorizing the unknown biological sample into one of the multiple groups of isomers; and

retaining, in the two-dimensional plot, only the reference points in the group of reference points corresponding to the group of isomers into which the unknown biological sample is categorized.

19. The method according to claim 18, wherein the categorized reference points are reference points calculated from reference measurements obtained experimentally and the multiple groups of reference points form a first set of groups of reference points; and wherein the method further comprises categorizing reference points calculated from reference

measurements obtained by a machine learning algorithm into a second set of groups of reference points corresponding to respective groups of isomers.~~;~~

20. The method according to claim 19, further comprising the following if the unknown biological sample does not belong to any one of the multiple groups of isomers corresponding to the first set of groups of reference points:

categorizing the unknown biological sample into one of the groups of isomers corresponding to the second set of groups of reference points; and

determining the nearest reference point from the reference points in the group of reference points in the second set corresponding to the group of isomers into which the unknown biological sample is categorized.

21. The method according to claim 20, further comprising determining that the unknown biological sample is not identifiable, if the unknown biological sample does not belong to any one of the multiple groups of isomers corresponding to the first set of groups of reference points and any one of the groups of isomers corresponding to the second set of groups of reference points.

22~~21~~. The method according to any one of claims 18 to 21~~20~~, wherein categorizing the unknown biological sample into one of the groups of isomers comprises categorizing the unknown biological sample based on a mass to charge ratio value of the unknown biological sample.

23~~22~~. The method according to any one of claims 17 to 22~~21~~, wherein identifying the unknown biological sample further comprises calculating an accuracy score based on a distance between the sample point and the determined nearest reference point.

24~~23~~. The method according to claim 23~~22~~, wherein the accuracy score comprises one of the following: a low confidence score, a medium confidence score, a high confidence score.

25~~24~~. The method according to any preceding claim, wherein the two-dimensional plot is formed from a first number of attributes; and wherein the method comprises using further plots, each further plot formed from a different number of attributes as compared to another plot.

26~~25~~. The method according to claim 25~~24~~, wherein using further plots comprises performing the following for each further plot:

calculating a further sample point in the further plot based on at least one of the plurality of sample measurements for the unknown biological sample.

27~~26~~. The method according to any one of claims 1 to 24~~23~~, wherein each reference point of the two-dimensional plot is calculated from a first number of reference measurements for a first number of attributes of the corresponding known biological compound; and

wherein the method further comprises calculating a sample point in each of a plurality of further plots based on at least one sample measurement for the unknown biological sample;

wherein each of the plurality of further plots comprises a plurality of stored reference points corresponding to respective known biological compounds, each reference point of the further plot calculated from at least one reference measurement for at least one attribute of the corresponding known biological compound; and

wherein for each further plot, the number of attributes from which the reference points are calculated differ from the first number and differ from the number of attributes from which the reference points in a different further plot are calculated.

28~~27~~. The method according to claim 27~~26~~, wherein the number of sample measurements received is equal to the first number and wherein the number of attributes from which the reference points are calculated for each further plot is less than the first number.

29~~28~~. The method according to claim 28~~27~~, further comprising generating additional further plots wherein the number of attributes from which the references points are calculated for each additional further plot is smaller than the first number.

30~~29~~. The method according to any one of claims 1 to 24~~23~~, wherein each reference point of the two-dimensional plot is calculated from three reference measurements for three attributes of the corresponding known biological compound and wherein the method further comprises:

calculating a second sample point in a second two-dimensional plot based on two sample measurements for the unknown biological sample, wherein the second two-dimensional plot comprises a plurality of stored reference points corresponding to respective

known biological compounds, each reference point of the second two-dimensional plot calculated from two reference measurements for two attributes of the corresponding known biological compound; and

calculating a third sample point in a third plot based on one sample measurement for the unknown biological sample, wherein the third plot comprises a plurality of stored reference points corresponding to respective known biological compounds, each reference point of the third plot calculated from one reference measurement for one attribute of the corresponding known biological compound.

31~~30~~. The method according to any one of claims 27~~26~~ to 30~~29~~, wherein the method further comprises:

for each plot, determining a reference point nearest to the sample point in the plot; and

identifying the unknown biological sample as the known biological compound corresponding to the most number of determined nearest reference points.

32~~31~~. The method according to claim 31~~30~~, wherein the method further comprises:

determining an accuracy score based on a distance between the reference point corresponding to the known biological compound the unknown biological sample is identified as and the sample point in the plot formed from a most number of attributes.

33~~32~~. The method according to any preceding claim, wherein the attribute of the unknown biological sample comprises one of the following: mass, mass to charge ratio, retention time, normalized retention time, glucose unit, collisional cross section, tandem mass spectrometry/mass spectrometry fragmentation, measured shift in retention time after exoglycosidase treatment, measured shift in mass to charge ratio after exoglycosidase treatment, measured shift in collisional cross section after exoglycosidase treatment, measured shift in tandem mass spectrometry/mass spectrometry fragmentation.

~~33. The method according to any preceding claim, wherein the unknown biological sample comprises one of the following: glycan, metabolite, antibody.~~

34. The method according to any one of claims 1 to 33, wherein the glycan comprises one or more of the following: glycospingolipid glycan, N-glycan, O-glycan.

35. The method according to any one of claims 1 to~~33 or~~ 34, wherein the glycan comprises procainamide-labelled glycan.

36. A computer program product comprising computer-readable instructions that implement an application for identifying an unknown biological sample, wherein the computer program product is configured to be executed on one or more computing devices, each having one or more processors:

wherein the application is configured to provide a two-dimensional plot comprising a plurality of stored reference points corresponding to respective known biological compounds, ~~and~~ wherein each reference point is calculated from a plurality of reference measurements for more than two attributes of the corresponding known biological compound, each attribute being different from another attribute, and wherein each reference point is calculated using a second algorithm capable of compressing the plurality of reference measurements into a reference point in the two-dimensional plot; and

wherein the application comprises instructions for:

receiving more than two sample measurements for the unknown biological sample, each sample measurement for an attribute of the unknown biological sample, each attribute being different from another attribute;

calculating a sample point in a two-dimensional plot from the more than two sample measurements for the unknown biological sample, wherein the sample point is calculated using a first algorithm capable of compressing the more than two sample measurements into the sample point in the two-dimensional plot; and

identifying the unknown biological sample by comparing the sample point against the plurality of reference points in the two-dimensional plot,

wherein each of the attributes of the unknown biological compound corresponds to a respective one of the attributes of the known biological compounds,

wherein the unknown biological sample comprises one of the following: glycan, metabolite, antibody.

37. A kit comprising:

an extraction device for extracting an unknown biological sample;

at least one experimental device for determining sample measurements for the extracted unknown biological sample; and

a computing device configured to execute the computer program product according to claim 36.

38. An apparatus comprising:

a memory; and

at least one processor coupled to the memory and configured to:

receive more than two sample measurements for the unknown biological sample, each sample measurement for an attribute of the unknown biological sample, each attribute being different from another attribute;

calculate a sample point in a two-dimensional plot from the more than two sample measurements for the unknown biological sample, wherein the sample point is calculated using a first algorithm capable of compressing the more than two sample measurements into the sample point in the two-dimensional plot;

wherein the two-dimensional plot comprises a plurality of stored reference points corresponding to respective known biological compounds, ~~and~~ wherein each reference point is calculated from a plurality of reference measurements for more than two attributes of the corresponding known biological compound, each attribute being different from another attribute, and wherein each reference point is calculated using a second algorithm capable of compressing the plurality of reference measurements into a reference point in the two-dimensional plot; and

identify the unknown biological sample by comparing the sample point against the plurality of reference points in the two-dimensional plot,

wherein each of the attributes of the unknown biological compound corresponds to a respective one of the attributes of the known biological compounds,

wherein the unknown biological sample comprises one of the following: glycan, metabolite, antibody.