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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 September 2022 (29.09.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/200801 A1

(51) International Patent Classification:
G16B 40/10 (2019.01) *C08B 37/00* (2006.01)
A61K 31/726 (2006.01) *G01N 23/2258* (2018.01)

(21) International Application Number:
PCT/GB2022/050746

(22) International Filing Date:
24 March 2022 (24.03.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2104133.0 24 March 2021 (24.03.2021) GB

(71) Applicant: **UNIVERSITY OF NOTTINGHAM** [GB/GB]; Ingenuity Centre, Innovation Park, Triumph Road, Nottingham Nottinghamshire NG7 2TU (GB).

(72) Inventors: **HOOK, Andrew**; University of Nottingham, Room B04, Boots Science Building, University Park, Nottingham Nottinghamshire NG7 2RD (GB). **MERRY, Catherine Louise Ruby**; University of Nottingham, Room B208, Nottingham Biodiscovery Institute, University Park, Nottingham Nottinghamshire NG7 2RD (GB).

(74) Agent: **WILSON GUNN**; 5th Floor, Blackfriars House, The Parsonage, Manchester Lancashire M3 2JA (GB).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU,

(54) Title: HIGH SENSITIVITY ANALYSIS OF NANOGRAM QUANTITIES OF GLYCOSAMINOGLYCANS USING TOF-SIMS

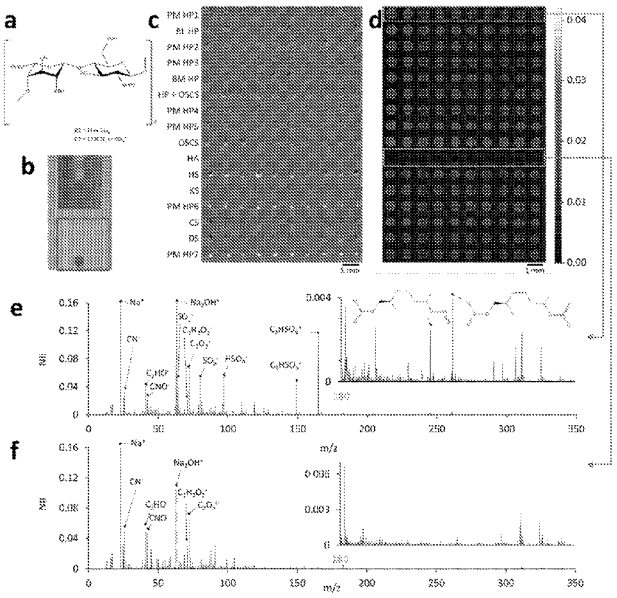

Figure 1

(57) Abstract: Glycosaminoglycans (GAGs) are important biopolymers that differ in the sequence of saccharide units, and in post-polymerization alterations at various positions, making these molecules complex and challenging to analyse. We have developed an approach that enables small quantities (< 200 ng) of hundreds of different GAGs to be analysed within a short time frame (3-4 hours). Time of flight secondary ions mass spectrometry 10 together with multivariate analysis was used to analyse an array of over 400 GAG samples. Resultant spectra were derived from the whole molecules and did not require pre-digestion. All 6 GAG types were successfully discriminated, both alone and in the presence of fibronectin. We also distinguished between pharmaceutical grade heparin derived from different animal species and from different suppliers to a sensitivity as 15 low as 0.001 wt%. This approach is highly beneficial in the quality control of GAGs produced for therapeutic applications and for characterising GAGs within biomaterials or from in vitro

WO 2022/200801 A1

[Continued on next page]

RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

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

cell culture.

High sensitivity analysis of nanogram quantities of glycosaminoglycans using ToF-SIMS

Technical Field of the Invention

The present invention relates to a method analysis of the constituent glycosaminoglycans of at least one crude or purified solid and/or solution.

Background of the Invention

Glycosaminoglycans (GAGs) are important biopolymers that differ in the sequence of saccharide units, and in post-polymerization alterations at various positions, making these molecules complex and challenging to analyse.

Glycosaminoglycans (GAGs) are polysaccharides found within cells, within the pericellular space and as a part of the extracellular matrix (ECM). GAGs regulate biological processes, such as self-renewal, differentiation, growth, inhibition, microbial invasion and defence, with their broad structural diversity and differential localisation accommodating specific interactions with hundreds of binding proteins. The complexity of GAGs, including chain length (polymerisation machinery), modification (epimerisation and sulphation of the hydroxyl groups at various positions on the saccharide units) and core protein attachment is orchestrated by enzyme mediated synthesis and allows for GAGs to have greater information carrying capacity than the more commonly studied biological polymers, nucleic acids and proteins.

The five sulfated GAGs, heparin, heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS) and keratan sulfate (KS) are synthesised attached to protein cores as proteoglycans, unlike non-sulfated hyaluronan (HA) which is extruded into the pericellular space. Heparin, in the form of a pure polysaccharide released from its core protein, is a globally used anticoagulant and antithrombotic and is currently being considered for anti-inflammatory indications such as chronic obstructive pulmonary disease. Other GAG types are now also increasingly being applied clinically, for example, as treatments for cancer and osteoarthritis, as anti-viral therapies and to support wound healing. The rapid and sensitive structural characterisation of GAGs is critical to maintain the standardisation and safety of these animal-derived biomolecules for medical use, as was highlighted by the contamination

of heparin samples with over-sulfated CS (OSCS) that led to patient hypotension and death. The ongoing biosecurity of heparin is a significant concern to healthcare systems around the world, necessitating continued efforts to improve heparin analysis and provide synthetic production routes.

Typically, chemical analysis of pharmaceutical GAGs is achieved using nuclear magnetic resonance (NMR) and high performance liquid chromatography (HPLC) methods. Simple H-NMR has been shown to detect 0.1 wt% contaminating OSCS within heparin, whilst HPLC achieved a limit of detection of 0.03 wt% for OSCS in heparin and remains the gold standard analysis technique for heparin characterisation. However, some of these approaches require >10 mg of sample, as well as specialised equipment and expert analysis and, therefore, suffer from low throughput. Mass spectrometry plays a leading role in GAG glycomics utilising soft ionisation techniques such as electrospray ionisation, however, analysis of whole sulfated GAGs remains difficult. This is particularly problematic for the characterisation of heparin as whole-molecule analysis is necessary to detect inter-species contamination of porcine-derived material used for medical applications. If porcine sources become limited, for example as a consequence of recent outbreaks of African Swine Fever, the relatively poor detection of non-porcine material (a limit of detection (LOD) of approximately 2 wt% for detecting a bovine contamination in porcine heparin) is unlikely to be sufficient to protect supplies.

For biomaterial applications requiring surface analysis, X-ray photoelectron spectroscopy has been favoured due to quantitative readouts but is unable to resolve the subtle chemical difference between different GAGs.

Embodiments of the present invention seek to overcome the disadvantages of the presently used methods.

Summary of the Invention

According to the present invention there is provided a method of analysis of the constituent glycosaminoglycans of at least one crude or purified solid and/or solution comprising: the detection of ions in the solid and/or solution using time of flight secondary ion mass spectrometry; and the application of at least one principal

component analysis model to the detected ions and/or the application of at least one partial least squares model to the detected ions.

The method may comprise the development of the at least one principal component analysis model through the selection of a subset of ions of the detected ions by a sparse feature selection methodology.

The sparse feature selection methodology may be recursive feature selection. The recursive feature selection may comprise applying recursive feature addition. The selection criteria of the recursive feature addition may be the maximisation of the distance between the means of a plurality of training sets, each training set being a crude or purified solid and/or solution comprising a known type and amount of glycosaminoglycans. The selection criteria of the recursive feature addition may be the maximisation of the distance between the means of the plurality of training sets using Euclidean geometry. The recursive feature selection may comprise applying recursive feature elimination after applying the recursive feature addition. The selection criteria for the recursive feature elimination may be the minimisation of the overlap between 95% confidence ellipses of each of the plurality of training sets.

The method may comprise developing the at least one partial least squares model through the selection of ions by a sparse feature selection methodology. The sparse feature selection methodology may be least absolute shrinkage and selection operator (LASSO). The number of latent variables used for the partial least squares model may be selected based upon the minimisation of the root mean square error of cross validation.

The method may comprise identifying at least one specific glycosaminoglycan.

The method may comprise comparing more than one crude or purified solid and/or solution.

The method may comprise distinguishing between more than one crude or purified solid and/or solution.

The method may comprise identifying at least two glycosaminoglycans originating from different animal or synthetic sources in a crude or purified solid and/or solution.

The method may comprise quantifying at least one glycosaminoglycan.

The method may comprise the application of the at least one principal component analysis model to the detected ions; and the application of the at least one partial least squares model to the detected ions.

The glycosaminoglycan may comprise one or more of heparin, heparan sulphate, keratan sulphate, chondroitin sulphate, dermatan sulphate, hyaluronic acid or a sulphated form of one or more aforementioned glycosaminoglycan. The glycosaminoglycan may comprise one or more of a deacetylated form of heparin, heparan sulphate, keratan sulphate, chondroitin sulphate, dermatan sulphate, hyaluronic acid. The glycosaminoglycan may comprise a synthetic glycosaminoglycan.

The glycosaminoglycan may comprise one or more semi-synthetic glycosaminoglycans. The glycosaminoglycan may comprise pentosan polysulphate.

The principal component analysis model may comprise a selection of three or more of the ions listed in Figure 13, Figure 17, Figure 18, Figure 19, Figure 21, Figure 22, Figure 23, and/or Figure 25.

The partial least squares model may comprise a selection of three or more of the ions listed in tables Figure 13, Figure 17, Figure 18, Figure 19, Figure 21, Figure 22, Figure 23, and/or Figure 25.

The present invention enables small quantities (< 200 ng) of hundreds of different GAGs to be analysed within a short time frame (3-4 hours). Time of flight secondary ion mass spectrometry (ToF-SIMS) is capable of a rapid spectral acquisition (≈20s per sample) and can be applied to whole molecules without the need for purification or enzymatic digestion. When time of flight secondary ion mass spectrometry together with multivariate analysis is used to analyse an array of over 400 GAG samples, resultant spectra were derived from the whole molecules and did not require pre-digestion. All 6 GAG types were successfully discriminated, both alone and in the presence of fibronectin. Via the present invention it was possible to distinguish between pharmaceutical grade heparin derived from different animal species and from different suppliers to a sensitivity as low as 0.001 wt%. This approach is beneficial in the quality control of GAGs produced for therapeutic applications and for characterising GAGs within biomaterials or from *in vitro* cell culture.

Detailed Description of the Invention

In order that the invention may be more clearly understood one or more embodiments thereof will now be described, by way of example only, with reference to the accompanying drawings, of which:

Figure 1a shows the chemical structure of the main disaccharide of heparin, in which R1 and R2 are usually SO_3^-;

Figure 1b is a side view image of piezo-dispensing glass nozzle used to dispense GAG solutions to prepare a GAG microarray, with droplet detection highlighted within the marked region of interest and typical droplet volume of 320-340 pL;

Figure 1c is a brightfield microscopy image of the array prepared by the nozzle of figure 1b;

Figure 1d a SO_4^- ion image corresponding to figure 1c, acquired using ToF-SIMS, wherein intensity scale indicates the measured normalised ion count depicted in the figure, all spots remaining distinct, no SO_2^- signal was observed from HA, and ROI selection for extracting the spectrum from each spot was based upon the high intensity region from the SO_2^- signal, or low intensity area for HA;

Figure 1e is an extracted ToF-SIMS spectra for porcine mucosa derived heparin;

Figure 1f is an extracted ToF-SIMS spectra for HA;

Figure 2 shows normalised ion intensities at specific mass/charge positions for glass, poly-l-lysine, aminosilane, TCPS, and allylamine plasma polymer before, and after washing and the addition of heparan sulphate (HS), and the proposed mass assignment and m/z for the characteristic ions;

Figure 3 shows normalised ion intensities at specific mass/charge positions for glass, poly-l-lysine, aminosilane, TCPS, and allylamine plasma polymer before, and after washing and the addition of hyaluronic acid (HA), and the proposed mass assignment and m/z for the characteristic ions;

Figure 4 is a Brightfield image of a GAG microarray, pre-printed with15 nL of water and then a total volume of 17 nL of 5 mg/ml GAG solutions, either hyaluronic acid (HA), heparan sulphate (HS), chondroitin sulphate (CS)

or dermatan sulphate (DS), wherein the bar below each sample indicates the amount of each GAG type added, respective of the fraction coloured, the array is printed at 65% relative humidity, and the 5x5 array of mixed GAGs was repeated 4 times;

Figure 5 is a scores plot for PCA analysis of GAG array, wherein the ellipse shows the 95% confidence limits based upon 7-replicates of each GAG HA (★), HS (✳), CS (✳), and DS (✳), open circles showing test data for 3-replicates of each GAG, combinations of GAGs are shown as clusters of triangular datapoints, with the size of the triangle indicated the content of each GAG in a sample, respective of the colour, and the amount of variance captured by each PC is indicated in the axis title;

Figure 6a is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 (✳), 50 (▲), 10 (▲), 1 (✳), 0.1 (▲), 0.01 (▲), 0.001 (★), and 0 (▲), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for all the variables;

Figure 6b is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 (✳), 50 (▲), 10 (▲), 1 (✳), 0.1 (▲), 0.01 (▲), 0.001 (★), and 0 (▲), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for all the variables variance scaled;

Figure 6c is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 (✳), 50 (▲), 10 (▲), 1 (✳), 0.1 (▲), 0.01 (▲), 0.001 (★), and 0 (▲), wherein training set are closed

symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for recursive feature addition with a selection criteria of minimising ellipse overlap;

Figure 6d is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 (✳), 50 (▲), 10 (●), 1 (✳), 0.1 (▲), 0.01 (●), 0.001 (★), and 0 (▲), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the same scores shown as for Figure 6c, but with a different training and test sets;

Figure 6e is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 (✳), 50 (▲), 10 (●), 1 (✳), 0.1 (▲), 0.01 (●), 0.001 (★), and 0 (▲), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for recursive feature addition with a selection criteria of maximising Euclidean distance between the mean of each sample set;

Figure 6f is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 (✳), 50 (▲), 10 (●), 1 (✳), 0.1 (▲), 0.01 (●), 0.001 (★), and 0 (▲), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for recursive feature addition

with a selection criteria of maximising the Mahalanobis distance between the mean of each sample set;

Figure 6g is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 ( ), 50 ( ), 10 ( ), 1 ( ), 0.1 ( ), 0.01 ( ), 0.001 ( ), and 0 ( ), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for recursive feature elimination with a selection criteria of minimising ellipse overlap;

Figure 6h is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 ( ), 50 ( ), 10 ( ), 1 ( ), 0.1 ( ), 0.01 ( ), 0.001 ( ), and 0 ( ), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for recursive feature elimination with a selection criteria of maximising Euclidean distance between the mean of each sample set;

Figure 6i is a PCA scores plot for ToF-SIMS data taken from a concentration series of heparin derived from porcine mucosa spiked with heparin from bovine lung, at weight fractions (%) of 100 ( ), 50 ( ), 10 ( ), 1 ( ), 0.1 ( ), 0.01 ( ), 0.001 ( ), and 0 ( ), wherein training set are closed symbol and test sets are open symbols, the 95% confidence ellipse is drawn around the samples sets (as calculated from the training set data only), and the variance captured by each principal component is shown on the axis title, with the scores shown for recursive feature elimination with a selection criteria of maximising the Mahalanobis distance between the mean of each sample set;

Figure 6j shows a list of the m/z values for ions selected for multiple sparse datasets, with the presence of an ion in a particular dataset is indicated as present (red) and absent (blue);

Figure 7a is a PCA of biological replicates of CS (red circles) and DS (purple triangles), wherein training datasets (technical replicates) are shown as closed circles and test sets are shown as open circles, and the ellipses show the 95% confidence limits, the set being a non-sparse dataset, variance scaled;

Figure 7b is a PCA of biological replicates of CS (red circles) and DS (purple triangles), wherein training datasets (technical replicates) are shown as closed circles and test sets are shown as open circles, and the ellipses show the 95% confidence limits, the set being a sparse dataset;

Figure 8a is a score plot of PC2 versus PC3, wherein training sets (n=7) are shown as solid markers, whilst test sets (n=3) are shown as non-filled marker, 95% confidence ellipses are shown for each sample type, PC2 predominately captures variance between heparin and CS samples and PC3 predominately captures variance between different heparin samples;

Figure 8b is a score plot of PC4 versus PC1, wherein training sets (n=7) are shown as solid markers, whilst test sets (n=3) are shown as non-filled marker, 95% confidence ellipses are shown for each sample type, PC4 predominately captures variance between the bovine lung derived heparin with the other heparin samples and PC1 predominately captures variance within replicates;

Figure 8c is a score plot of PC6 versus PC5, wherein training sets (n=7) are shown as solid markers, whilst test sets (n=3) are shown as non-filled marker, 95% confidence ellipses are shown for each sample type;

Figure 8d is a legend for the different samples shown in figures 8a-8c;

Figure 8e is a data plot of the cumulative (*) and individual (✓) variance capture for each latent variable;

Figure 8f is a list of the ions identified for PCA using recursive feature elimination with possible chemical assignments (<100 ppm deviation);

Figure 8g shows data plots of loadings for each PC for x-variables listed in Figure 8f, the loadings plots being organised for PCs 1-6 left to right, top down;

Figure 9a is a score plot of PC2 versus PC1 using data mean centred only;

Figure 9b is a core plot of PC2 versus PC1 using mean-centred and variance scaled;

Figure 9c is a Scree plot showing the variance captured for each PC (bars) and the cumulative variance (line), for PCA with all variables, wherein the selection of number of latent variables used was determined by fitting a linear curve (shown as a dashed line) to the variance explained curve for high numbers of latent variables (15-20) and selecting where the variance explained departed from linearity with reduced numbers of latent variables;

Figure 9d is a Scree plot showing the variance captured for each PC (bars) and the cumulative variance (line), for PCA with a sparse dataset, wherein the selection of number of latent variables used was determined by fitting a linear curve (shown as a dashed line) to the variance explained curve for high numbers of latent variables (15-20) and selecting where the variance explained departed from linearity with reduced numbers of latent variables;

Figure 9e is a graph of the separation of the means of each sample set across 6 PCs for varied number of variables using RFA, wherein the plot reduced to 0 when further reduction in variable number caused training set samples to fall outside the respective confidence ellipse generated from the scores plots of the training set;

Figure 9f is a graph of the mean area fraction non-overlapping for confidence ellipses considering PCs 1-6, for PCA conducted with recursive feature elimination, wherein the plot reduced to 0 when further reduction in variable number caused training set samples to fall outside the respective confidence ellipse generated from the scores plots of the training set;

Figure 10a is a score plot of PC2 versus PC1, wherein training sets are closed symbols and test sets are open symbols and the 95% confidence ellipse is shown for each sample set;

Figure 10b is a score plot of PC4 versus PC3, wherein training sets are closed symbols and test sets are open symbols and the 95% confidence ellipse is shown for each sample set;

Figure 10c is a score plot of PC6 versus PC5, wherein training sets are closed symbols and test sets are open symbols and the 95% confidence ellipse is shown for each sample set;

Figure 10d is a dendrogram showing hierarchical clustering of GAG samples based upon the scores for PCs 1 - 6, wherein training samples are shown as dashed lines and test samples are shown as solid lines, lines have been coloured to match sample identity, and the associated symbol for each sample type is shown beneath each cluster;

Figure 10e is a legend showing symbols corresponding to the GAG type;

Figure 11 is a table listing the 48 ions comprising the sparse dataset used for PCA of ToF-SIMS spectra of GAGs, possible assignments ordered from smallest to largest deviation are shown (<75 ppm), and the loadings associated with PCs 1-6 are indicated according to the intensity scale shown on the right;

Figure 12a is a dendrogram of a PCA of datasets from 6 GAG types mixed with FN, the dendrogram showing hierarchical clustering of GAG samples based upon the scores for PCs 1 and 2, training samples are shown as a dashed line and test samples are shown as a solid line, lines have been coloured to match the sample identity, and the legend of sample identity shown to the right;

Figure 12b is a bar graph of predicted values for the heparin/OSCS samples measured as a part of the 16 GAGs (✓), unrelated GAG types (✓) and heparin (HP) PM spiked with OSCS (✓), using the corresponding PLS models. wherein error bars show ± 1 standard deviation unit, n = 10;

Figure 12c is a bar graph of predicted values for the heparin/OSCS samples measured as a part of the 16 GAGs (✓), unrelated GAG types (✓) and heparin (HP) PM spiked with HP BM (✓), using the corresponding PLS models. wherein error bars show ± 1 standard deviation unit, n = 10;

Figure 12d is a bar graph of predicted values for the heparin/OSCS samples measured as a part of the 16 GAGs (✓), unrelated GAG types (✓) and heparin (HP) PM spiked with HP BL (✓), using the corresponding PLS models. wherein error bars show ± 1 standard deviation unit, n = 10;

Figure 13 is a table listing the 18 ions comprising the sparse dataset used for PCA of ToF-SIMS spectra of GAGs in fibronectin, with possible assignments ordered from smallest to largest deviation are shown (<75 ppm), and the loadings associated with PCs 1-2 being indicated according to the intensity scale shown on the right;

Figure 14a is a graph of PLS regression analysis, showing a comparison of the variance captured and the root mean square error of cross validation for varied number of latent variables (LV), wherein the samples analysed were heparin PM spiked with OSCS (LV = 6, features = 40);

Figure 14b is a graph of PLS regression analysis, showing the regression coefficients determined for the PLS model for the features selected by LASSO, wherein the samples analysed were heparin PM spiked with OSCS (LV = 6, features = 40);

Figure 14c is a graph of PLS regression analysis, showing a comparison of the variance captured and the root mean square error of cross validation for varied number of latent variables (LV), wherein the samples analysed were heparin PM spiked with heparin BM (LV = 2, features = 24);

Figure 14d is a graph of PLS regression analysis, showing the regression coefficients determined for the PLS model for the features selected by LASSO, wherein the samples analysed were heparin PM spiked with heparin BM (LV = 2, features = 24);

Figure 14e is a graph of PLS regression analysis, showing a comparison of the variance captured and the root mean square error of cross validation for varied number of latent variables (LV), wherein the samples analysed were heparin PM spiked with heparin BL (LV = 5, features = 18);

Figure 14f is a graph of PLS regression analysis, showing the regression coefficients determined for the PLS model for the features selected by

LASSO, wherein the samples analysed were heparin PM spiked with heparin BL (LV = 5, features = 18);

Figure 15a is a graph showing PLS measured versus predicted values for heparin (HP) PM spiked with OSCS showing training (○) and test sets (△), the y=x line drawn as a guide, coefficient of determination values (R^2) for the y=x line are shown for the training (black) and test (red) sets, measured values were the log of the fraction (%) of contaminant added;

Figure 15b is a graph showing PLS measured versus predicted values for heparin (HP) PM spiked with HP BM showing training (○) and test sets (△), the y=x line drawn as a guide, coefficient of determination values (R^2) for the y=x line are shown for the training (black) and test (red) sets, measured values were the log of the fraction (%) of contaminant added;

Figure 15c is a graph showing PLS measured versus predicted values for heparin (HP) PM spiked with HP BL showing training (○) and test sets (△), the y=x line drawn as a guide, coefficient of determination values (R^2) for the y=x line are shown for the training (black) and test (red) sets, measured values were the log of the fraction (%) of contaminant added;

Figure 16a is a graph of PC1 versus PC2 scores for heparin PM spiked over a concentration range of 0-100% (wt%) OSCS, including both training (closed) and test (open) sets, with 95% confidence ellipses shown, possible assignments and loadings for PC1 and PC2 for the selected features for each dataset shown in Figure 17, a number of ions likely associated with sulphate groups ($C_3SO_3^-$,$C_6H_{13}S_2O_4^-$) including N-sulphation (SNO_2^-, $CHSNO_2^-$) were present in all sparse datasets, and ions also likely associated with di- and tri-saccharides were also selected ($C_{12}H_{43}O_{10}^+$, $C_{13}H_5S_6O_{12}^-$, $C_8H_5S_2O_8^-$);

Figure 16b is a graph of PC1 versus PC2 scores for heparin PM spiked over a concentration range of 0-100% (wt%) heparin BM, including both training (closed) and test (open) sets, with 95% confidence ellipses shown, possible assignments and loadings for PC1 and PC2 for the selected features for each dataset shown in Figure 18, a number of ions

likely associated with sulphate groups ($C_3SO_3^-$,$C_6H_{13}S_2O_4^-$) including N-sulphation (SNO_2^-, $CHSNO_2^-$) were present in all sparse datasets, and ions also likely associated with di- and tri-saccharides were also selected ($C_{12}H_{43}O_{10}^+$, $C_{13}H_5S_6O_{12}^-$, $C_8H_5S_2O_8^-$);

Figure 16c is a graph of PC1 versus PC2 scores for heparin PM spiked over a concentration range of 0-100% (wt%) heparin BL, including both training (closed) and test (open) sets, with 95% confidence ellipses shown, possible assignments and loadings for PC1 and PC2 for the selected features for each dataset shown in Figure 19, a number of ions likely associated with sulphate groups ($C_3SO_3^-$,$C_6H_{13}S_2O_4^-$) including N-sulphation (SNO_2^-, $CHSNO_2^-$) were present in all sparse datasets, and ions also likely associated with di- and tri-saccharides were also selected ($C_{12}H_{43}O_{10}^+$, $C_{13}H_5S_6O_{12}^-$, $C_8H_5S_2O_8^-$);

Figure 17 is a table listing the 18 ions comprising the sparse dataset used for PCA of ToF-SIMS spectra of PM heparin spiked with OSCS, possible assignments ordered from smallest to largest deviation are shown (<75 ppm), and the loadings associated with PCs 1-2 are indicated according to the intensity scale shown on the right;

Figure 18 is a table listing the 25 ions comprising the sparse dataset used for PCA of ToF-SIMS spectra of heparin PM spiked with heparin BM, possible assignments ordered from smallest to largest deviation are shown (<75 ppm), and the loadings associated with PCs 1-2 are indicated according to the intensity scale shown on the right;

Figure 19 is a table listing the 15 ions comprising the sparse dataset used for PCA of ToF-SIMS spectra of heparin PM spiked with heparin BL, possible assignments ordered from smallest to largest deviation are shown (<75 ppm), and the loadings associated with PCs 1-2 are indicated according to the intensity scale shown on the right;

Figure 20 is a table showing anticoagulant profiles of the different heparin samples, Anti-IIa is the antithrombin dependent anti-factor IIa assay, anti-Xa is the antithrombin dependent anti-factor Xa assay and human plasma an activated partial thromboplastin time assay using human

plasma, all potencies were assigned relative to the 6th International Standard for Unfractionated Heparin, 07/328, and potency is reported in IU/mg and values in brackets are the standard deviations;

Figure 21 is a table listing the 40 ions comprising the sparse dataset used for PLS regression of ToF-SIMS spectra of porcine mucosa heparin spiked with OSCS, the three possible assignments with the smallest deviation ordered from smallest to largest deviation are shown (<100 ppm), and the regression coefficient for each ion are indicated according to the intensity scale shown on the right, a positive regression coefficient is associated with OSCS whilst a negative regression coefficient is associated with porcine mucosa heparin;

Figure 22 is a table listing the 24 ions comprising the sparse dataset used for PLS regression of ToF-SIMS spectra of porcine mucosa heparin spiked with bovine mucosa heparin, the three possible assignments with the smallest deviation ordered from smallest to largest deviation are shown (<100 ppm), and the regression coefficient for each ion are indicated according to the intensity scale shown on the right, a positive regression coefficient is associated with bovine mucosa heparin whilst a negative regression coefficient is associated with porcine mucosa heparin;

Figure 23 is a table listing the 18 ions comprising the sparse dataset used for PLS regression of ToF-SIMS spectra of porcine mucosa heparin spiked with bovine lung heparin, the three possible assignments with the smallest deviation ordered from smallest to largest deviation are shown (<100 ppm), and the regression coefficient for each ion are indicated according to the intensity scale shown on the right, a positive regression coefficient is associated with bovine lung heparin whilst a negative regression coefficient is associated with porcine mucosa heparin;

Figure 24a is a graph showing linear correlation between the normalised ion intensity of selected ions with the activity of pharmaceutical grade heparin. Ions were selected that showed significant ($p<0.001$) linear correlations (Pearson's > 0.75) with heparin activity Anti-IIa, (b) Anti-Xa and (c) human serum;

Figure 24b is a graph showing linear correlation between the normalised ion intensity of selected ions with the activity of pharmaceutical grade heparin. Ions were selected that showed significant ($p<0.001$) linear correlations (Pearson's > 0.75) with heparin activity Anti-Xa and (c) human serum;

Figure 24c is a graph showing linear correlation between the normalised ion intensity of selected ions with the activity of pharmaceutical grade heparin. Ions were selected that showed significant ($p<0.001$) linear correlations (Pearson's > 0.75) with heparin activity human serum;

Figure 25 is a table listing the key ions associated with PCA of GAG array, and the loading of each ion for PC1 and PC2;

Figure 26a is a PCA of a randomly generated dataset, which is a non-sparse dataset, variance scaled, training dataset (technical replicates) shown as closed circle and test set shown as open circle, Ellipses are 95% confidence limits, and Variance captured by each PC shown in the axis heading;

Figure 26b is a PCA of a randomly generated dataset, which is a PCA with sparse dataset, training dataset (technical replicates) shown as closed circle and test set shown as open circle, Ellipses are 95% confidence limits, and Variance captured by each PC shown in the axis heading;

Figure 27 is a table showing a comparison of ions common to different sparse datasets generated using different selection criteria;

Figure 28a is a graph summarising PCA of different GAG samples in fibronectin, specifically PC2 versus PC1, training set (n=8) shown as solid markers, whilst test set (n=3) shown as non-filled marker, 95% confidence ellipses shown for each sample type;

Figure 28b is a graph summarising PCA of different GAG samples in fibronectin, specifically PC4 versus PC3, training set (n=8) shown as solid markers, whilst test set (n=3) shown as non-filled marker, 95% confidence ellipses shown for each sample type;

Figure 28c is a legend for figures 28a and 28b; and

Figure 28d is a graph of the cumulative (*) and individual (✓) variance capture for each latent variable.

ToF-SIMS was used to analyse a microarray containing all six GAG types (analytical preparations of HS, CS, DS, KS, HA, porcine mucosal (PM) heparin, and clinical grade heparin from porcine mucosa, bovine mucosa and bovine lung). Together with principal component analysis (PCA) and partial least square (PLS) regression, this approach was used to chemically distinguish between the different GAG classes in a semi-quantitative manner, whilst notably being able to discern differences between heparin samples derived from different animal sources and different manufacturer batches. The combination of high throughput analysis with high chemical sensitivity indicated the feasibility of this method for quality control of pharmaceutical heparin, detecting possible process related impurities as well as contaminants, and for enabling the surface analysis of GAG-modified materials to facilitate the development of GAG- functional biomaterials.

Arrays of GAG solutions, as shown in Figure 1c, were prepared using ink-jet printing onto poly-L-lysine (PLL)-coated glass slides, selected for the ability of PLL to adhere GAGs due to ionic interactions, and possible other supramolecular interactions such as hydrogen binding (as shown in Figures 2 and 3). Ink-jet printing also enabled the rapid generation of GAG mixtures via in-spot mixing (as shown in Figures 4 and 5). Microarrays enable the rapid assessment of libraries of molecules, require small amounts of material (ng) and are compatible with high throughput surface analysis. Microarrays have been widely used to assess DNA, proteins and their analogues (oligonucleotides and peptides). Glycan and GAG microarrays have also been used in alternative applications, but not previously for high-throughput GAG structural analysis. In total, approximately 160 ng of material was deposited per spot. Resultant arrays were assessed by bright field microscopy and ToF-SIMS (as shown in Figures 1b and 1c). All printed spots appeared to be both physically and chemically distinct (as shown in Figures 1c and d). The absence of sulfate signal (SO^-) for HA samples suggested no carry-over between print runs (as shown in figure 1c). Regions of interest for each spot were determined from the SO^- ion image to enable extraction of spectra for each sample (as shown in Figures 1d-1e). A typical spectrum from porcine mucosa (PM) derived heparin exhibited high intensity ions associated with sulfate (SO_2, SO_3, C_3HSO_5) and amide (CN, CNO) groups as well as highly oxygenated fragments ($C_3H_3O_2$, C_2O_3)

(as shown in Figure 1d). Ions associated with the sulfate group were absent from a typical spectrum taken from a HA sample (as shown in Figure 1e).

Each sample typically had approximately 900 different ions (both positive and negative). To effectively assess the differences between samples, principal component analysis was used to reduce the dimensionality of the multispectral dataset. Additionally, a sparse dataset was generated to remove uninformative variables not associated with variance between sample types. This is important for the high-dimension dataset where PCA results are difficult to interpret and the sample eigenvectors are not always consistent estimators whilst regression approaches are susceptible to over- fitting. A number of approaches have been used to develop sparsity for PCA, including recursive feature selection. In this study, recursive feature addition was used to generate a sparse dataset using the maximisation of the distance between the means of the sample sets as a selection criteria. Recursive feature elimination was then used, using the minimisation of the overlap between 95% confidence ellipses from different sample sets as a selection criterion, to select features that would differentiate between samples with sufficient confidence (as shown in Figures 6a-6j). To avoid over-fitting, the sample sets were split into training and test sets at a 7:3 ratio (training:test). Test samples were required to fall within the 95% confidence ellipse associated with the principal components describing the variance between samples. The final sparse dataset was further tested for its ability to robustly assess the differences between samples by ensuring sample sets remained separated with multiple randomly generated training/test sets. Creation of a sparse dataset by this method resulted in 83.5% of the variance captured by PCA to be associated with the difference between the biochemically similar GAGs CS and DS, giving confidence that this approach could also work for a broader set of materials (as shown in Figures 7a and 7b).

The utility of PCA with a sparse dataset to identify differences in GAG samples was assessed for 5 different medical grade PM-derived heparin samples, bovine lung- (BL) and bovine mucosa- (BM) derived heparin, OSCS, CS and a heparin sample contaminated with 1 wt% OSCS from the heparin crisis. A sparse dataset containing 12 different ions was selected. By considering the scores for PC2 and PC3, the CS samples and contaminated heparin sample were all successfully differentiated from all other heparin samples (as shown in Figures 8a-8g).

As the ultimate assessment of this approach, samples of each of the 6 GAG types were analysed together to assess whether each sample could be chemically discerned. Without sparsity, PCA was able to successfully differentiate between the KS, HA and OSCS samples, with and without variance scaling (shown in Figures 9a and 9b). However, separation of the other GAG samples was not achieved, particularly between the different heparin sample sets. The scree plot indicated 1-9 PCs captured variance not associated with noise (shown in Figure 9c). After generation of a sparse dataset, the variance captured by the first 6 PCs increased from 78% to 89% due to the removal of features that corresponded to variance not associated with differences between sample set (shown in Figures 9c and 9d). A total of 48 features were selected for the final sparse dataset that corresponded to the minimum number of features required to produce a high (>0.25) mean average area fraction of ellipses not overlapping (shown in Figures 9e and 9f).

PCA of the sparse dataset was able to successfully separate all 16 GAG samples to 95% confidence (shown in Figures 10a-10e). Scores plots for PCs 1-2 showed clustering of the 6 main types of GAG (shown in Figure 10a). Further separation of the different types of heparin including separation of heparin from PM, BM or BL and different batches of heparin from PM was achieved by considering PCs 3-6 (shown in Figure 10b and 10c). Hierarchical cluster analysis was used to classify the different samples based upon their Euclidean distance. The outcome of this unsupervised classification approach is shown as a dendrogram (shown in figure 10d), where samples that are most similar are positioned together. In all cases, samples were clustered within their correct sample group, including the test set, with the exception of single replicates of two heparin PM batches and one replicate of the heparin BM samples.

Only those ions with a possible assignment based upon the elemental composition of GAGs were selected (C, O, H, N, S). Each ion was assigned a loading for each PC, shown in Figure 11. Possible assignments for each of the 48 ions for the key PCs is listed in Figure 11. A number of ions likely associated with sulfate groups were selected, including ions SN^+ and SNO2, as well as larger ions such as $C_{10}H_{11}SO_4$. This suggests that part of the variance captured by the PCA was associated with the sulfation patterns on the GAGs. Ions likely associated with di- and

tri- saccharides, such as $C_{18}H_{33}SO_5$, $C_{18}H_{38}O_9$, were also selected. Further interpretation of the relation between the ions identified and the GAG structures is limited due to the relatively low mass resolution of the ToF mass analyser. However, analytical use of this technique to distinguish between closely related GAG structures and mixtures of GAGs with high sensitivity does not depend on full structural interpretation of the data. Indeed, this is a recognised advantage of chemometric methods.

To test the capability of ToF-SIMS analysis to chemically distinguish between samples in a more complex biological environment, each of the 6 GAG types were added to a fibronectin (FN) solution, printed as a microarray and analysed by ToF-SIMS. FN is a common component of biological ECMs as well as serum and plasma. After generating a sparse dataset, the multispectral data was assessed for its ability to distinguish between the different samples using PCA and hierarchical cluster analysis. All 6 GAG types were chemically differentiated from each other, and from pure FN (shown in Figures 28a-28b), where all samples, including the test sets, were successfully categorised using hierarchical cluster analysis (shown in Figure 12a). Possible assignments for the 18 ions selected for this sparse dataset (the number given by figure 28d) and their loadings are shown in Figure 13. Similar to the model without FN, ions containing sulfate groups ($CHSO^-$, C_3HSO_2) were present in the model. Most of the ions selected were small in nature and likely derived from a monosaccharide. This may be due to a reduced yield of higher molecular weight ions associated with GAGs from within a protein matrix. Additionally, ions likely associated with FN (CH_4N^+) were also selected.

To assess the sensitivity of the analysis methodology to adulteration, a PM heparin was spiked with increasing concentrations of either OSCS, BM heparin or BL heparin. The ToF-SIMS spectral data was correlated with the fraction of spiking agent using partial least square (PLS) regression, as has been done previously for correlating water contact angle or protein adsorption with ToF-SIMS data. Initially, a sparse dataset was selected for each sample set by least absolute shrinkage and selection operator (LASSO) to minimise over-fitting by removal of uninformative features. The number of latent variables used was selected based upon the minimisation of the root mean square error of cross validation (shown in Figures 14a-14f). Plots of the measured fractions of spiking agent and those predicted from the ToF-SIMS data using the PLS

model are shown in Figures 15a-15c. A high correlation ($R^2 > 0.94$) between measured and predicted values was observed for samples spiked with either OSCS or heparin BM, suggesting that the ToF-SIMS data was able to distinguish differences in samples down to 0.001 wt%. This was confirmed by PCA of the same samples, which demonstrated separation between non-spiked samples and the samples spiked at 0.001 wt% to 95% confidence (shown in Figures 16a-16c). A weaker correlation ($R^2 = 0.88$) was observed for the samples spiked with BL heparin (shown in Figures 16a-16c). PCA of these samples showed that the non-spiked sample and the 0.001 wt% sample could not be separated to 95% confidence (shown in Figures 16b). However, a linear response between the measured and predicted fraction of BL heparin was observed down to 0.01 wt%, suggesting that the analysis was sensitive to this concentration. Similar R^2 values were observed for the training (70%) and test (30%) sets for all models, suggesting there was no over-fitting.

The PLS models were used to predict the fraction of the contaminant in each of the different heparin samples initially assessed by PCA (shown in Figures 12b-12d). The amount of OSCS predicted in all heparin samples was below 0.001%, with the exception of the analytical grade heparins, which had predicted OSCS fractions of 0.0009 and 0.002 wt%. High predicted fractions (≈100 wt%) were predicted for the OSCS sample, whilst the sample with a known OSCS adulteration of 1 wt% had a predicted OSCS fraction of 0.6 wt%. Quantitative analysis of samples using ToF-SIMS data is limited by matrix effects. Therefore, the PLS model was only applicable to samples analysed within the same matrix environment as the training data. For predictions of the bovine-derived heparin content, low values (1×10^{-5} wt%) were predicted for the porcine-derived heparin samples with the exception of the analytical grade samples and sample 4, which all had predicted values of approximately 1×10^{-3} wt%. The lower purity of the analytical grade heparin is expected, and our results indicate low levels of OSCS contamination. The presence of bovine-derived heparin in sample 4 was unexpected, but does coincide with lower levels of anticoagulant activity observed for this sample (as shown in Figure 20).

For completion, the PLS models were also applied to unrelated GAG types, shown in Figures 12b-12d. Although it is possible to suggest that high levels of OSCS

were predicted in the DS and CS samples, the predicted values are unreliable as the models were trained for the detection of specific contaminants in heparin. The successful separation of each of the 16 GAGs by PCA suggests that, in principle, PLS models of pairwise mixtures of the other GAG types could be prepared.

Each of the features selected for the PLS regression was assigned a regression coefficient (RC) that informed how strongly it influenced the model and whether it was associated with the contaminant or PM heparin (shown in Figures 15a-15c). Tables of possible assignments for the ions selected for each model and their associated RCs are shown in Figures 21-23. For each model, ions likely derived from mono- and di-saccharides were associated with PM heparin (having a negative RC) including $C_{13}H_{29}S_2NO_4$, $C_{16}H_{29}N_2O_7$, and $C_{32}H_{56}N_3O_{12}$. Furthermore, ions containing sulfate groups, such as CH_4SNO_2, C_5HSBNO^- and C_2SNO^- were also selected, suggesting the model includes information both about the disaccharide sequence of the heparin molecules and the sulfation pattern. The ions associated with the spiked GAGs (OSCS, BM heparin and BL heparin) included ions likely representative of the disaccharide sequence ($C_{20}H_{37}SN_2O_5$, $C_{12}H_{29}N_2O_7$ and $C_{22}H_{43}N_2O_7$) or sulfation pattern (KC_5SNO, $C_3H_5SNO_3$, C_2H_3S) of the spiked GAGs. There is large uncertainty regarding the ion assignments, particularly for large ions, due to the mass resolution of the ToF analyser. The suggestions provided are based upon structures that match GAG stoichiometry and have a minimal deviation between the measured and theoretical values.

The anticoagulant action of heparin is chiefly due to its ability to potentiate the serine protease inhibitor antithrombin, a protein normally present in plasma. Assays of antithrombin mediated inhibition of the clotting factors thrombin (Factor IIa) and of factor Xa, using purified proteins, are used to determine the potency of clinical grade heparin in International Units (IU)/mg. The Activated Partial Thromboplastin Time (APTT) is a plasma-based method for measuring anticoagulant activity.

The specific activities of five heparin samples were measured by these three methods and the results are summarised in Figure 18. A number of ions were found to significantly ($p<0.001$) correlate linearly (Pearson's $r > 0.75$) with each of the measures of activity, as shown in Figure 24a-24c. The origin of these ions is not known but the correlations suggest that they arise from structural factors that determine anticoagulant

activity, either very specifically in terms of the rare pentasaccharide motif that determines affinity for antithrombin, or in more general terms such as overall degree of sulfation. This link between the surface chemistry as measured by ToF-SIMS and a quantitative measure of biological activity is unexpected.

Analytical characterisation of GAGs underpins multiple aspects of current GAG-related research, including the understanding of their fundamental biological roles. GAGs are already important pharmaceutical compounds (as discussed above for heparin) and are increasingly being used for various therapeutic applications as well as being incorporated into biomaterials for improved biofunctionality. Mass spectrometry techniques focus on the analysis of oligosaccharides for the purposes of sequencing. The use of ToF-SIMS to analyse GAG samples on an arrayed platform provides a methodology by which small quantities (< 200 ng) of hundreds of different GAGs could be analysed within a short time window (3-4 hours). The resultant spectra were derived from the whole molecules and did not require any pre-digestion or pre-labelling of material. The analysis was informative of the GAG disaccharide sequence, sulfation pattern and biological activity and enabled discernment between all 6 different GAG types investigated.

The high throughput and sensitivity achievable by this system is also important for the quality control of GAGs within healthcare settings to ensure patient safety. For the most widely used GAG in medicine, heparin, recent problems in pharmacovigilance have been the spur to develop a battery of orthogonal tests to ensure identity, purity and high specific bioactivity. Besides the detection of contaminants, whether introduced accidentally or as deliberate adulteration, it is necessary to monitor impurities in heparin that can arise both from co-purification of related compounds such as chondroitin and dermatan sulphates and from minor chemical modifications arising in the manufacturing process. Whole molecule analysis is desirable to be able to detect contaminants and process-related impurities in active pharmaceutical ingredient of GAG-based products, for example, detection of mixed-species heparin. Whilst whole molecule analysis of GAGs has been achieved, the approaches are typically slow, require large amounts of samples and lack sensitivity. We applied our protocol to pharmaceutical grade heparin derived from different animal species and from different suppliers. Our approach allowed for the clear identification

of heparin samples in terms of species of origin, and highly sensitive detection of contaminants spiked into PM heparin, including a sensitivity of 0.001 wt% of the addition of OSCS, the contaminant associated with the heparin crisis, and to 0.01 wt% for BL heparin in PM heparin. This approach is likely to be highly beneficial in the quality control of GAGs produced for therapeutic applications and for characterising GAGs within biomaterial systems or from *in vitro* cell culture.

The use of multivariate analysis approaches was necessary to interrogate the multi-dimensional ToF-SIMS datasets. PCA has been widely used to assess the variance within ToF-SIMS datasets and was used here to be able to capture the variance between different GAG samples, whilst PLS regression demonstrated that the fraction of a spiked GAG could be predicted from the ToF-SIMS spectra. Creation of sparse datasets was important to avoid over-fitting data as well as to remove uninformative features. For PCA, recursive feature selection identified the ions that captured the variance between the different GAGs including within a more complex biological environment containing fibronectin.

Implementation of the approach described as either a tool for basic research or as a quality control methodology for heparin manufacturer would require for the data readouts to be reached without intervention of expert users. The data models established in this study would provide a useful system that future samples could be applied to, with the possibility to identify unknown GAGs (PCA) or detect contamination within a sample (PLS). Unsupervised approaches like hierarchical cluster analysis provide a mechanism by which useful readouts can be obtained without any user intervention. The models can also easily be further expanded and made more robust through the addition of further control samples, whilst models focussing on a single GAG type are also easily achievable. The approach therefore, has broad applicability and can be readily adapted to various GAG-based applications.

HS Na salt from porcine mucosa (Iduron), DS Na salt from porcine mucosa (Average Mw = 41,000, Iduron), CS B Na salt from porcine mucosa (Sigma-Aldrich), HA Na salt from *Streptococcus equi* (Mw = 15,000-30,000, Sigma-Aldrich), heparin from porcine mucosa (Mw = 5000, Fisher Scientific) were used as received. GAGs were prepared as standard solutions of 5 mg/ml in ultrapure water (Purelab Ultra, ELGA LabWater). Heparin samples received from the NIBSC heparin archive. KS Na salt was derived from bovine corneal. Fibronectin was derived from bovine plasma (Sigma-

Aldrich lot#101M7012V). Poly-L-lysine coated slides (Poly-Prep, Sigma-Aldrich), aminoalkylsilane functionalised slides (Silane- Prep, Sigma-Aldrich), tissue culture polystyrene (TCPS, Nunclon Delta, ThermoFisher Scientific), allylamine plasma polymer coated polystyrene (EpranEx, BD Biosciences) and bare glass slides (Corning) were used as received.

Arrays were prepared using an s11 sciFLEXARRAYER dispensing system (Scienion) using a glass piezo dispense capillary (P-2020, Scienion). Drop volumes were ≈ 300 pL, as measured using the drop shape analyser tool (Scienion) prior to each run. Print runs were conducted at a relative humidity of 65 % at room temperature. GAG solutions were diluted to 2-5 mg/ml in a polypropylene 384-well plate (Corstar) in ultra-pure water (18.2 MΩ.cm) with or without 1 mg/ml fibronectin. Initially 0-150 nL of water was printed and subsequently GAGs were dosed into the water droplets to facilitate mixing prior to surface adsorption. The nozzle was flushed with 250 μL of water whilst the outside of the nozzle was washed with copious amounts of water between printing different samples.

Time-of-flight secondary ion mass spectrometry measurements were conducted using a ToF-SIMS IV (IONTOF GmbH, Münster, Germany) instrument operated using a 25 keV Bi3 primary ion source exhibiting a pulsed target current of >0.3 pA. Samples were scanned at a pixel density of 512 pixels per mm, with fifteen shots per pixel over a given area. An ion dose of 2.45×10^{11} ions per cm^2 was applied to each sample area ensuring that static conditions were maintained throughout. Both positive and negative secondary ion spectra were collected (mass resolution of >7000 at m/z=29). Owing to the non-conductive nature of the samples, charge compensation was applied in the form of a low energy (20 eV) electron floodgun. Patch areas of 0.5×0.5 mm were acquired at a resolution of 256×256 pixels by rastering the primary ion beam over the patch using a 'random raster' path sequence. Patch areas were sequentially acquired over the entire microarray using programmed stage movements through the macro-raster stage function. The patch areas were combined into a mosaic image, allowing all patches to be processed together. A peak list was produced using the peak search tool (SurfaceLab 6, IONTOF), minimum counts set to 100, maximum background set to 0.8. To ensure the peak search tool had successfully identified peaks, all ions of interest were visually inspected. Regions associated with each polymer spot were then extracted and

recalibrated, and the peak list was applied to produce an individual spectrum for each polymer. In total, 412 positive and 460 negative ion peaks were identified. Peak assignments were achieved using a custom built Visual Basic Application algorithm (PeakAssigner v2.6). Only peaks with a chemical assignment derived from C, S, O, N and H within 100 ppm were used for PCA.

Phase contrast microscopy images were acquired using an Olympus IX51 microscope using a 40× objective, NA = 0.13. The microscope was equipped with a Smart Imaging System (IMSTAR) using Fluo/LightVision software (v6.04K).

A microarray of samples was initially prepared to enable a large number of samples to be rapidly assessed. The arrays were analysed by ToF-SIMS and spectra were obtained for each sample. Datasets were variance scaled and mean-centred and replicate measurements were split into training (70%) and test (30%) sets.

Principal component analysis (PCA) was conducted using the function 'pca' within Matlab R2018a (9.4.0.813654) on the full dataset. The scree plots were used to identify the total number of latent variables associated with meaningful variance by fitting a linear curve to high latent variable values (typically 15-20) and observing where the variance explained departed from linearity for lower numbers of latent variables. A sparse dataset was then created in two steps. First recursive feature addition was used to add ions to the dataset that maximised the Euclidian distance between the means of the sample sets. This was continued until the addition of further features provided no further increase. Recursive feature elimination as then used to reduce the dataset using the minimisation of confidence ellipse overlap from the scores plots of relevant latent variables as a weighting. After each feature addition the resulting models were checked to ensure the test datasets fell within the 95% confidence limits of the training datasets. The minimum number of variables required to achieve the minimum amount of ellipse overlap was selected. PCA was conducted on the final sparse dataset to assess differences between samples using the number of latent variables indicated by the scree plot generated using the sparse dataset. The function 'linkage' within Matlab 2018a (9.4.0.813654) was used for hierarchical cluster analysis using the scores of relevant latent variables from the final PCA.

All data processing was conducted using a custom-built Matlab project.

Partial least square regression was conducted in Matlab R2018a (9.4.0.813654) using the plsregress function, that utilises the SIMPLS algorithm. The ToF-SIMS spectra, variance scaled and mean-centred prior to analysis, was used as a set of predictors whilst the log of the concentration of spiked agent was used as the response. Replicate measurements were split into training (70%) and test (30%) sets. Sparse datasets were produced by LASSO using the lassoglm function, with the lambda value selected based upon the minimisation of the standard error. The final PLS models used the sparse datasets, and the number of latent variables was selected as the minimisation of the root mean square error of cross-validation. The models could then be used to predict the log of the concentration of spiked agent for unknown samples.

Immobilisaton of GAGs to surfaces can be achieved by covalent and non-covalent methods. Numerous approaches have been explored to immobilise GAGs on surfaces including carbodiimide chemistry, divinyl sulfone activation, reductive amination, sulfhydryl-maleimide reactions, Diers-Alder reaction, azide, and diazirine chemistry. To achieve non-covalent immobilisation, cationic surfaces are typically produced to enable ionic interactions. Non-covalent interactions have the advantage of forming quickly (near instantaneous), effectively immobilising GAGs in a biologically relevant manner. Carbohydrate microarrays have emerged as useful tools for studying biomolecular interactions with glycans, including the use of non-covalent interactions with Poly-L-lysine as an approach to rapidly and easily adhere GAGs to a surface.

To select an optimal substrate for forming a GAG microarray five different candidate surfaces were assessed, glass, poly-L-lysine coated glass slides, aminoalkyl silanised glass slides, tissue culture (TC) polystyrene and allylamine plasma polymer coated slides. Here we made use of non-covalent interactions due to the ease of formation and as the ToF-SIMS characterisation was not demanding on the GAG-substrate interaction.

1 µl of 2 mg/ml solutions of either HS or HA were manually pipetted onto each substrate and allowed to air dry. These two GAGs were selected in order to assess sulphated and non-sulphated biomolecules. The dried spots were then incubated overnight in 10 mls of water, before drying and analysis by ToF-SIMS. The intensity of both positively and negatively charged characteristic ions for the GAGs was

compared on as-received materials, after deposition of GAGs and after washing (Figure SI1-2). For all substrates and for both GAGs the characteristic ions were low for as-received materials, and increased after the addition of either HS or HA. On almost all samples the characteristic ions reduced to similar intensities to the as-received materials after washing, suggesting that the washing procedure has completely removed the GAG. The exception to this was the cationic samples poly-L-lysine and allylamine plasma polymer, both of which maintained a higher intensity of characteristic ions for GAGs above baseline levels after washing. The PLL coated glass slide maintained a higher intensity of characteristic peaks for GAGs after washing compared with the allylamine plasma polymer coating and was thus selected for formation of subsequent arrays

Use of ink-jet printing allows for the rapid creation of mixtures of samples using in-spot mixing15. To explore this with GAGs, HA, HS, CS and DS were printed at various combinations. To facilitate mixing, a 15 nL water droplet was initially printed and other GAGs were then sequentially deposited. Printing was done at 65 % humidity to reduce evaporation, and a sacrificial border of water droplets was printed around the array to create a local high humidity environment. Prior to printing the evaporation rate of droplets was determined. The calculated volume of water lost in each print cycle was replaced in each print run by the use of two nozzles, one that delivered a GAG solution and the other that delivered water. The resulting array is shown in Figures 15a-15c. All spots were visible after drying. The array was analysed by SIMS and regions of interest were used to extract spectra for each spot. The resulting spectra were analysed by PCA to reduce the dimensionality of the dataset. A sparse feature dataset was created using recursive feature elimination using the total separation of the pure GAGs as a weighting (selection of feature to be eliminated was determined by the feature that would cause the largest separation of points in PC1 and PC2). The dataset was successfully reduced to 4 variables without a significant reduction in sample separation. The final scores plot for PC1 and PC2 are shown in Figures 16a-16C. For each of the single GAGs, 7 repeats were used for the PCA analysis, whist 3 samples were used as a test set. The PCA showed that the 4 single GAGs were readily chemically discernible, with the clusters associated with replicates of each GAGs being clearly distinct. In all cases the test sets fell within the 95% confidence ellipses determined for the training samples,

suggesting that the variance captured by PC1 and 2 was real and the data had not been over-fitted. GAG mixed samples, when plotted onto the scores spot, fell outside of ellipses determined for the single GAG samples, however, samples were not observed to quantitatively fall between the ellipses for the single samples. This is expected due to matrix effects compromising the quantitative nature of SIMS data.

The 4 ions used for the PCA, possible assignments and loadings for PC1 and PC2 are shown in Figure 25. Apart from the ion SNO-, the assignment of the ions is ambiguous due to the mass resolution of ToF-SIMS. However, a number of possible assignments including sulphur, suggesting that the variance captured by the PCA is associated with the sulphation pattern of the GAGs.

In order to optimise the generation of a sparse dataset for PCA a concentration series of heparin samples either derived from porcine mucosa (PM) or bovine lung (BL) were mixed, printed onto an array and analysed by ToF-SIMS. There were 11 technical repeats of each sample, and these samples were split into training and test sets at a training:test ratio of 8:3. Principal component analysis was then done on the full dataset and sparse datasets generated by either recursive feature elimination (RFE) or addition (RFA) using the minimisation of the overlap of 95% confidence ellipses or separation of the distance between the means of datasets for either Euclidean or Mahalanobis geometry as a selection criterion. The total number of features selected was determined as the minimum number of features required to maximise the associated cost function. All datasets were mean-centred. In addition, datasets were also investigated with and without variance scaling. The resulting scores plots are shown in Figures 6a-6i.

The ability of different sparse datasets (and the complete dataset) were assessed for their ability to separate the two heparin samples and spiked samples. Only the 100% BL heparin was separated from all other samples using the entire dataset (shown in Figures 6a-6b), with or without variance scaling, although with variance scaling the separation of 100% BL heparin from all other samples was aligned with PC1 (shown in Figure 6a) as opposed to without variance scaling, where the separation was a combination of PC1 and PC2 (shown in Figure 6b). PCA of a sparse dataset generated by RFA using the minimisation of overlapping 95% confidence ellipses as a cost function enabled all samples to be separated after consideration of PC1 and PC2 to 95% confidence (shown in Figure 6c). However, this PCA model was over-fitted as the test

sets did not fall within the 95% confidence ellipses of their respective samples. Random selection of a different training and test set and reapplying the PCA for the same sparse dataset described the samples without over-fitting (all test dataset was within the 95% confidence limit associated with the training set), however, only the 100% and 50% BL heparin samples were successfully separated. This suggested that minimising overlap as a cost function was highly effective at identifying features that could differentiate between samples to 95% confidence, but was susceptible to over-fitting.

Sparse datasets generated by RFA using the maximisation of the distance between the means of samples sets as a cost function successfully identified features that described variance between the different samples sets, as indicated by reduced overlap of sample sets for the scores plots of PC1 and PC2 (shown in Figure 6e-6f). This was observed for both Euclidean and Mahalanobis geometry. However, PCA of either sp arse dataset was unable to successfully differentiate between the different sample sets to 95% confidence. No samples sets could be differentiated using PCA of the dataset generated using Mahalanobis geometry to 95% confidence (shown in Figure 6f), however, the 100% BL heparin sample could be differentiated using the dataset for the Euclidean geometry from all other samples (shown in Figure 6e) and the 50% and 0% BL heparin was successfully differentiated from 5 of the other samples concentrations. No over-fitting was observed (test sets had scores within the 95% confidence ellipses from the training sets). Therefore, use of the maximisation of the distance between sample set means within Euclidean geometry was able to select features that described the variance between samples, however, separation to 95% confidence was not achieved in all cases.

RFE was also explored as a route to generating a sparse dataset. Use of the minimisation of overlap of 95% confidence limits produced a sparse dataset that did not separate features better than the original dataset. In this case the feature selection was not over fitted. However, this approach selected for features that described different variance within sample sets rather than variance between sample sets, as indicated by the elongation and varied orientation of the ellipses in the scores plot of PC1 and PC2 without achieving separation between the samples (shown in Figure 6g).

RFE using the maximisation of the distance between sample sets as a cost function identified a sparse dataset that was not able to differentiate between samples

as well as the respective RFA sparse dataset for both Euclidean and Mahalanobis geometries (shown in Figure 6h and 6i).

Of the 140 different ions selected for the 6 different sparse dataset, 28 ions were common to at least 2 different datasets. These ions are listed in Figure 6j. A comparison of common ions between sparse datasets is shown in Figure 27. A maximum of 9 common features were identified when using the distance between sample means in Euclidean geometry as a cost function for RFE and RFA. A similar level of commonality was not observed for sparse datasets generated for the other two costs functions when using RFE or RFA, where only 2 common features were identified for minimising overlap of confidence ellipses and 0 common features were identified for maximising the distance between sample means using Mahalanobis geometry.

As minimising ellipse overlap was the most successful at producing a dataset that could separate samples to 95% confidence using PCA but was prone to overfitting or selection of features that described variance within a dataset, maximising the distance between sample means using Euclidean geometry was initially used as a cost function for RFA to generate a sparse dataset containing only features that describe the separation between sample sets. RFE using the minimisation of ellipse overlap was then applied to generate a secondary sparse model. Using this approach a sparse dataset was produced with optimal separation of samples that was not over-fitted. To further ensure this approach did not overfit data it was applied to generate a sparse dataset for randomly produced data. A total of 1,200 initial features were included, which exceeded the typical numbers of positive and negative ions identified from a ToF-SIMS spectrum. No separation of samples was observed after generating a sparse dataset (shown in Figure 26), further confirming the method was not over-fitting the data. This approach was applied for the further generation of sparse datasets.

The one or more embodiments are described above by way of example only. Many variations are possible without departing from the scope of protection afforded by the appended claims.

CLAIMS

1. A method of analysis of the constituent glycosaminoglycans of at least one crude or purified solid and/or solution comprising: the detection of ions in the solid and/or solution using time of flight secondary ion mass spectrometry; and the application of at least one principal component analysis model to the detected ions and/or the application of at least one partial least squares model to the detected ions.

2. A method according to claim 1 comprising the development of the at least one principal component analysis model through the selection of a subset of ions of the detected ions by a sparse feature selection methodology.

3. A method according to claim 2 wherein the sparse feature selection methodology is recursive feature selection.

4. A method according to claim 3 wherein the recursive feature selection comprises applying recursive feature addition.

5. A method according to claim 4 wherein the selection criteria of the recursive feature addition is the maximisation of the distance between the means of a plurality of training sets, each training set being a crude or purified solid and/or solution comprising a known type and amount of glycosaminoglycans.

6. A method according to claim 5 wherein the selection criteria of the recursive feature addition is the maximisation of the distance between the means of the plurality of training sets using Euclidean geometry.

7. A method according to any of claims 4, 5 or 6 wherein the recursive feature selection comprises applying recursive feature elimination after applying the recursive feature addition.

8. A method according to claim 7 wherein the selection criteria for the recursive feature elimination is the minimisation of the overlap between 95% confidence ellipses of each of the plurality of training sets.

9. A method according to any preceding claim comprising developing the at least one partial least squares model through the selection of ions by a sparse feature selection methodology.

10. A method according to claim 9 wherein the sparse feature selection methodology is least absolute shrinkage and selection operator (LASSO).

11. A method according to claim 10 wherein the number of latent variables used for the partial least squares model is selected based upon the minimisation of the root mean square error of cross validation.

12. A method according to any preceding claim comprising identifying at least one specific glycosaminoglycan.

13. A method according to any preceding claim comprising comparing more than one crude or purified solid and/or solution.

14. A method according to any preceding claim comprising distinguishing between more than one crude or purified solid and/or solution.

15. A method according to any preceding claim comprising identifying at least two glycosaminoglycans originating from different animal or synthetic sources in a crude or purified solid and/or solution.

16. A method according to any preceding claim comprising quantifying at least one glycosaminoglycan.

17. A method of any preceding claim comprising the application of the at least one principal component analysis model to the detected ions; and the application of the at least one partial least squares model to the detected ions.

18. A method according to any preceding claim wherein the glycosaminoglycan comprises one or more of heparin, heparan sulphate, keratan sulphate, chondroitin sulphate, dermatan sulphate, hyaluronic acid or a sulphated form of one or more aforementioned glycosaminoglycan.

19. A method according to any of claims 1-18 wherein the glycosaminoglycan comprises a synthetic glycosaminoglycan.

20. A method according to any preceding claim wherein the principal component analysis model comprises a selection of three or more of the ions listed in Figure 13, Figure 17, Figure 18, Figure 19, Figure 21, Figure 22, Figure 23, and/or Figure 25.

21. A method according to any preceding claim wherein the partial least squares model comprises a selection of three or more of the ions listed in tables Figure 13, Figure 17, Figure 18, Figure 19, Figure 21, Figure 22, Figure 23, and/or Figure 25.

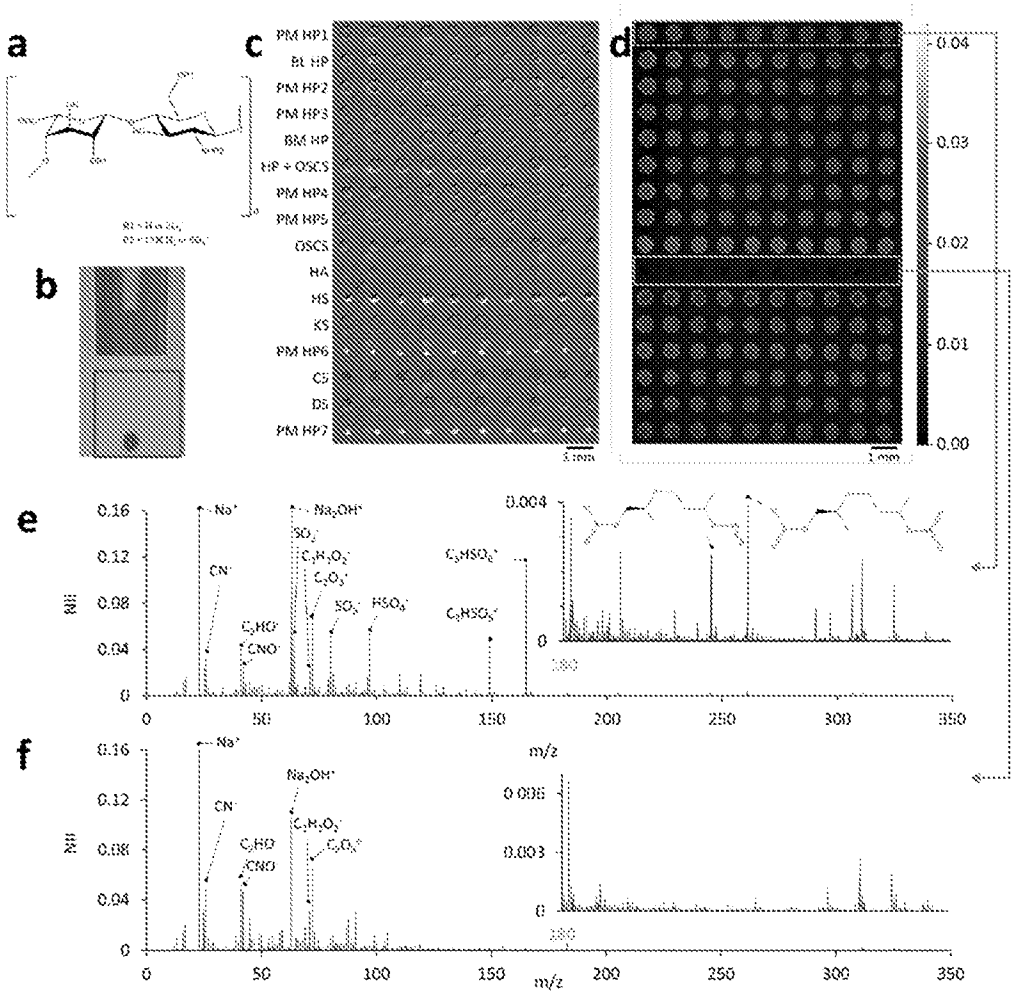
a
b
c
d
e
f
PM HP1
BL HP
PM HP2
PM HP3
BM HP
HP + OSCS
PM HP4
PM HP5
OSCS
HA
HS
KS
PM HP6
CS
DS
PM HP7
0.04
0.03
0.02
0.01
0.00
Na+
Na2OH+
SO3-
C2H3O2-
C2O3-
CN-
C3HO-
CNO-
SO3-
HSO4-
C3HSO5+
C3HSO5-
0.004
0
0.16
0.12
0.08
0.04
NII
0
50
100
150
200
250
300
350
m/z
0.006
0.003

Figure 1

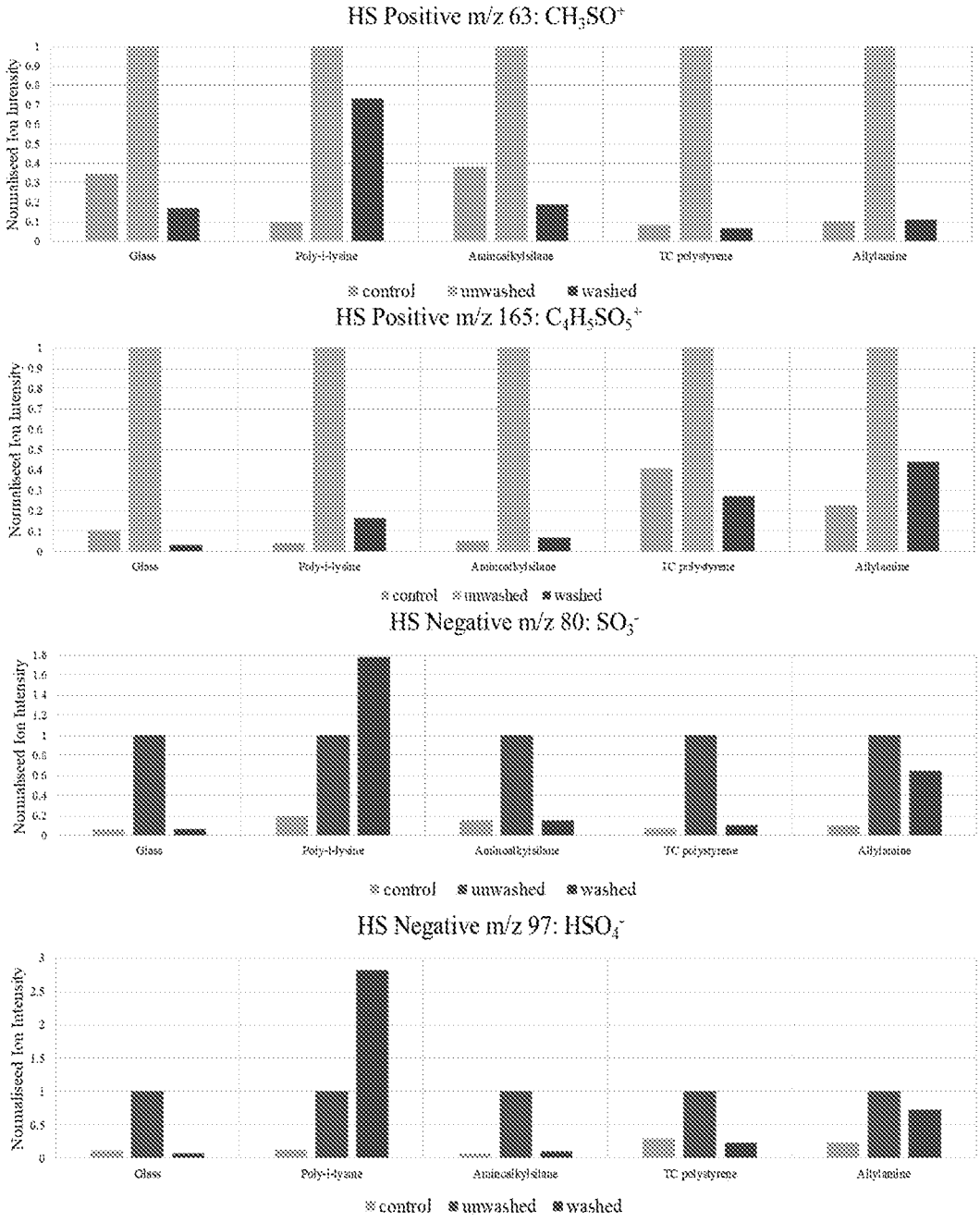
HS Positive m/z 63: CH_3SO^+
Normalised Ion Intensity
Glass
Poly-l-lysine
Aminoalkylsilane
TC polystyrene
Allylamine
control
unwashed
washed
HS Positive m/z 165: $C_4H_5SO_5^+$
Normalised Ion Intensity
Glass
Poly-l-lysine
Aminoalkylsilane
TC polystyrene
Allylamine
control
unwashed
washed
HS Negative m/z 80: SO_3^-
Normalised Ion Intensity
Glass
Poly-l-lysine
Aminoalkylsilane
TC polystyrene
Allylamine
control
unwashed
washed
HS Negative m/z 97: HSO_4^-
Normalised Ion Intensity
Glass
Poly-l-lysine
Aminoalkylsilane
TC polystyrene
Allylamine
control
unwashed
washed

Figure 2

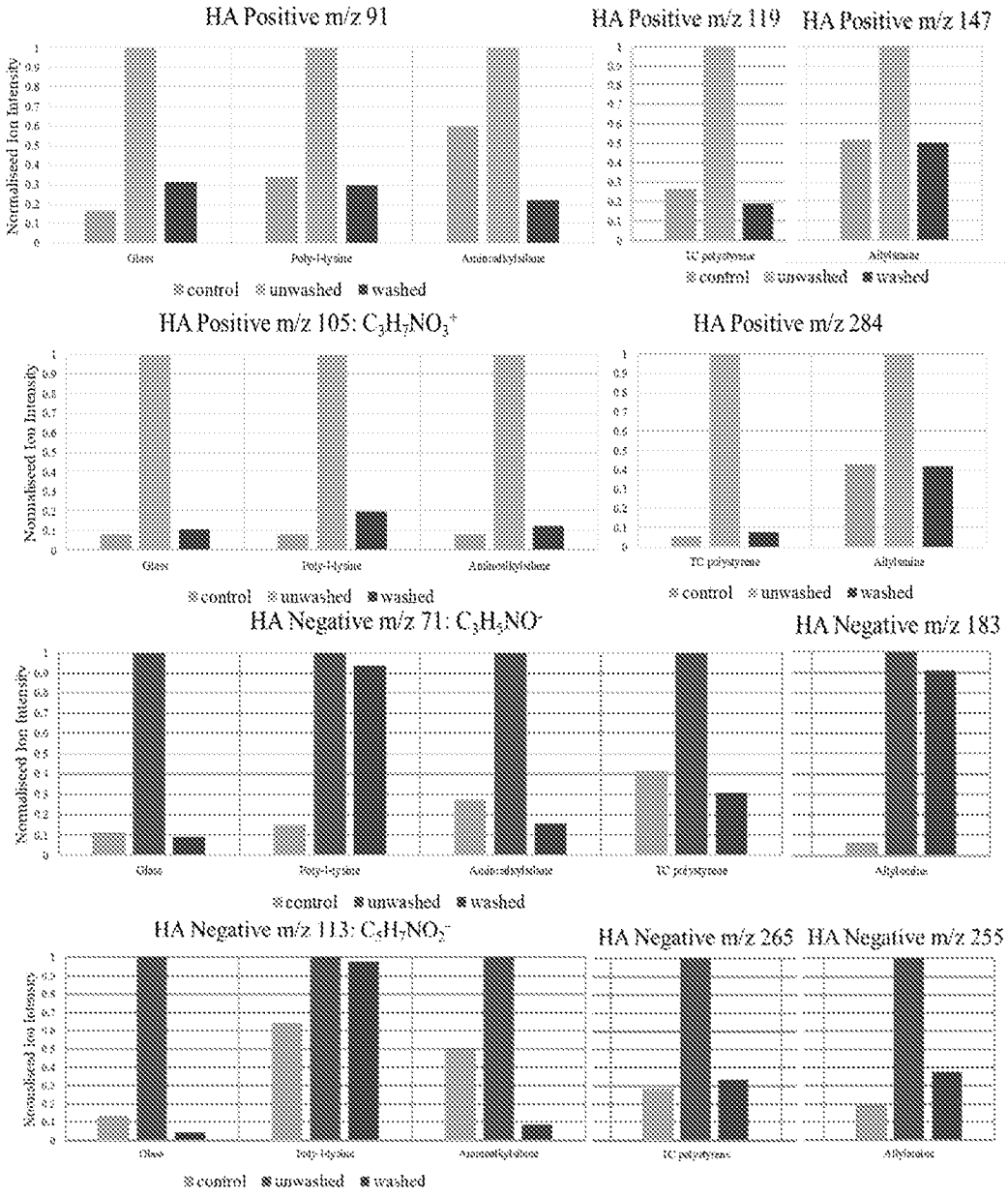
HA Positive m/z 91
HA Positive m/z 119
HA Positive m/z 147
Normalised Ion Intensity
Glass
Poly-l-lysine
Aminoalkylsilane
TC polystyrene
Allylamine
control
unwashed
washed
HA Positive m/z 105: $C_3H_7NO_3^+$
HA Positive m/z 284
HA Negative m/z 71: $C_3H_5NO^-$
HA Negative m/z 183
HA Negative m/z 113: $C_5H_7NO_2^-$
HA Negative m/z 265
HA Negative m/z 255

Figure 3

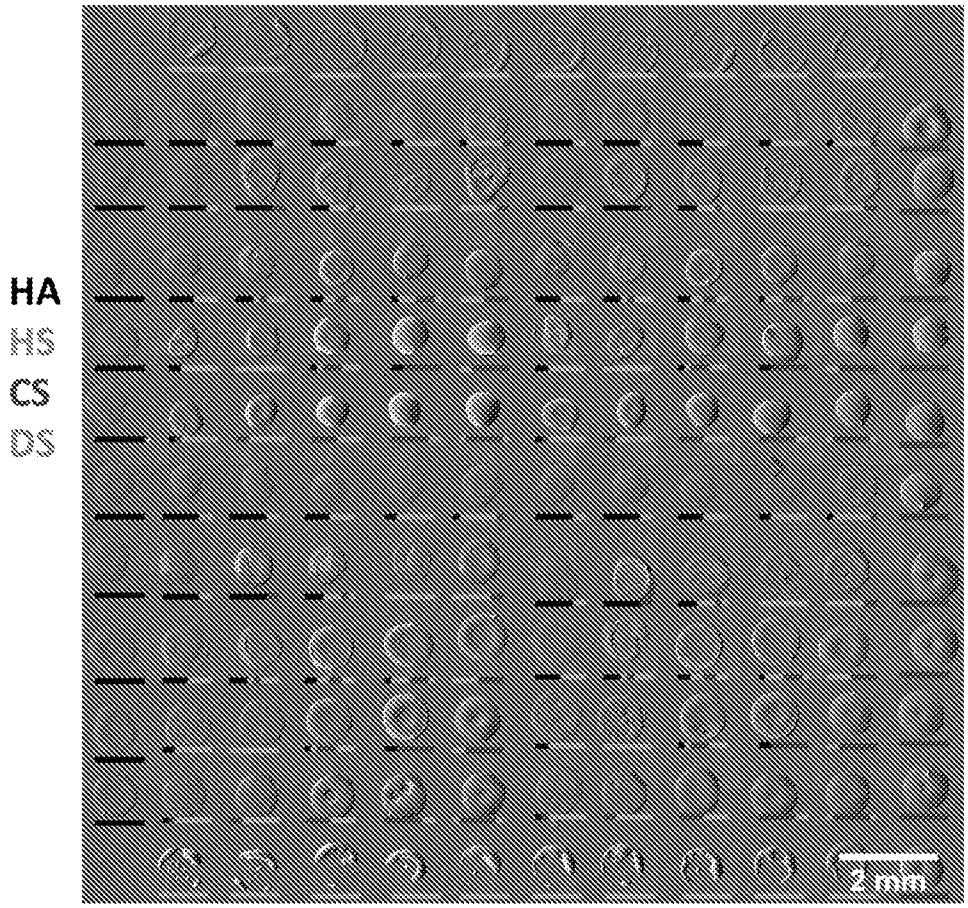

Figure 4

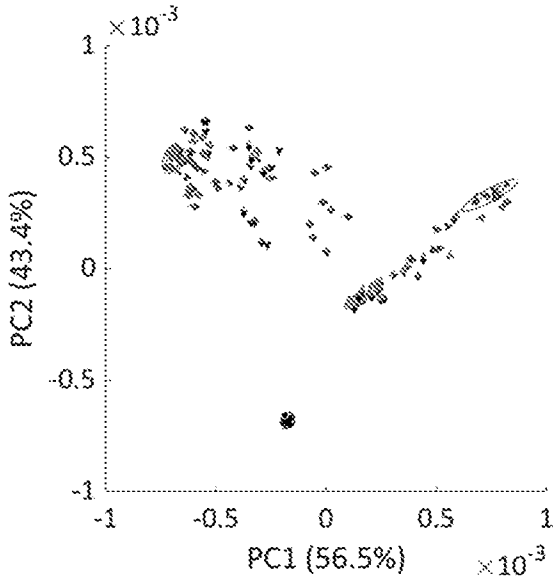

Figure 5

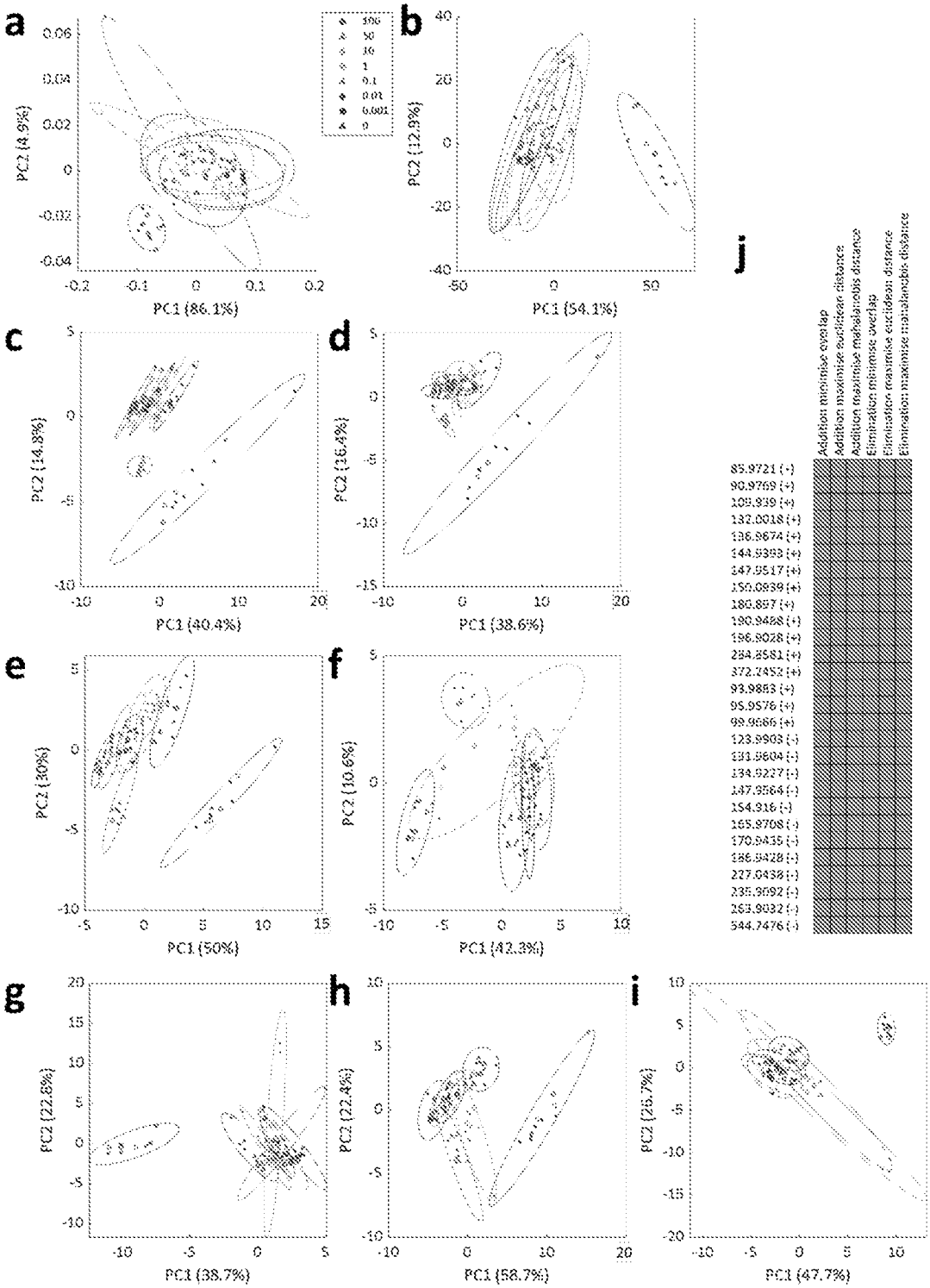
a
b
c
d
e
f
g
h
i
j
PC1 (86.1%)
PC2 (4.9%)
PC1 (54.1%)
PC2 (12.9%)
PC1 (40.4%)
PC2 (14.8%)
PC1 (38.6%)
PC2 (16.4%)
PC1 (50%)
PC2 (30%)
PC1 (42.3%)
PC2 (10.6%)
PC1 (38.7%)
PC2 (22.8%)
PC1 (58.7%)
PC2 (22.4%)
PC1 (47.7%)
PC2 (26.7%)
Addition minimise overlap
Addition maximise euclidean distance
Addition maximise mahalanobis distance
Elimination minimise overlap
Elimination maximise euclidean distance
Elimination maximise mahalanobis distance

Figure 6

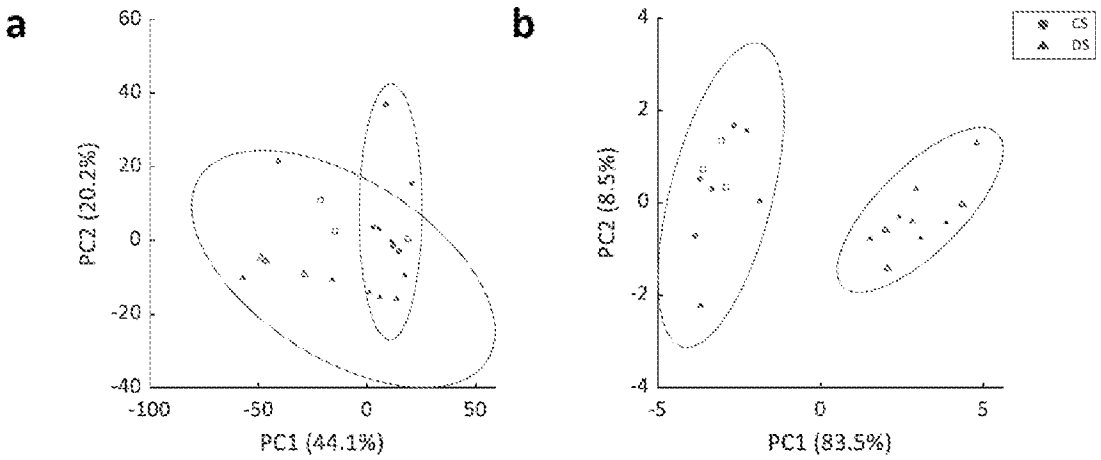

Figure 7

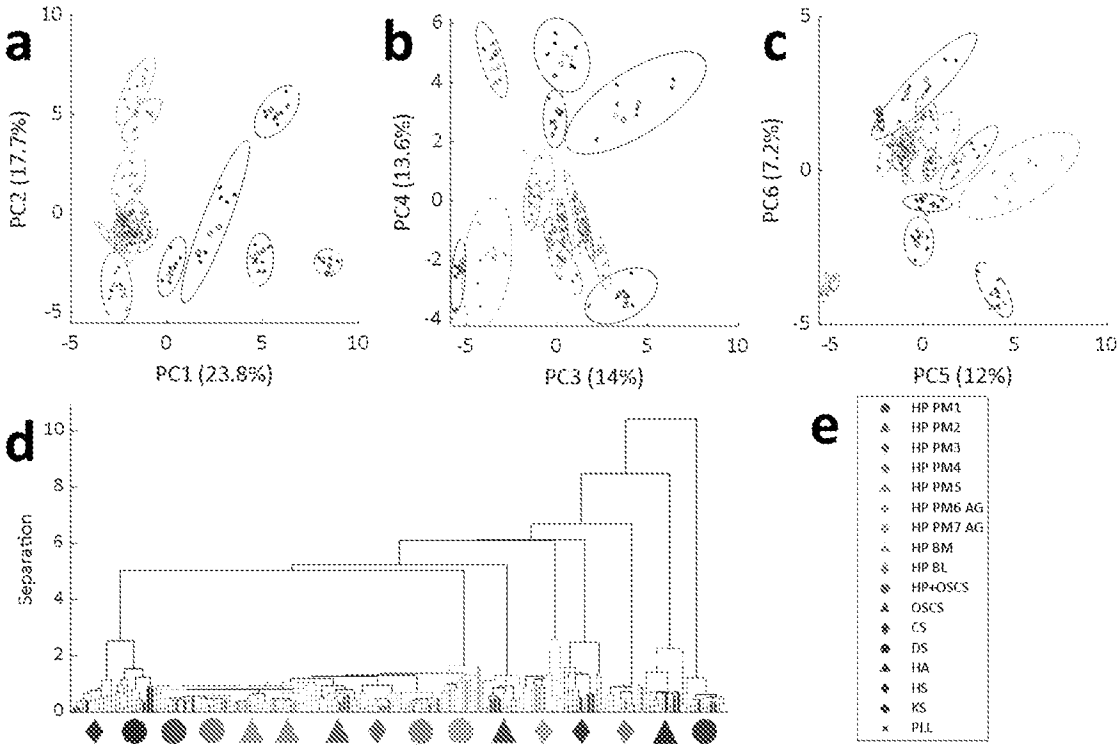

Figure 10

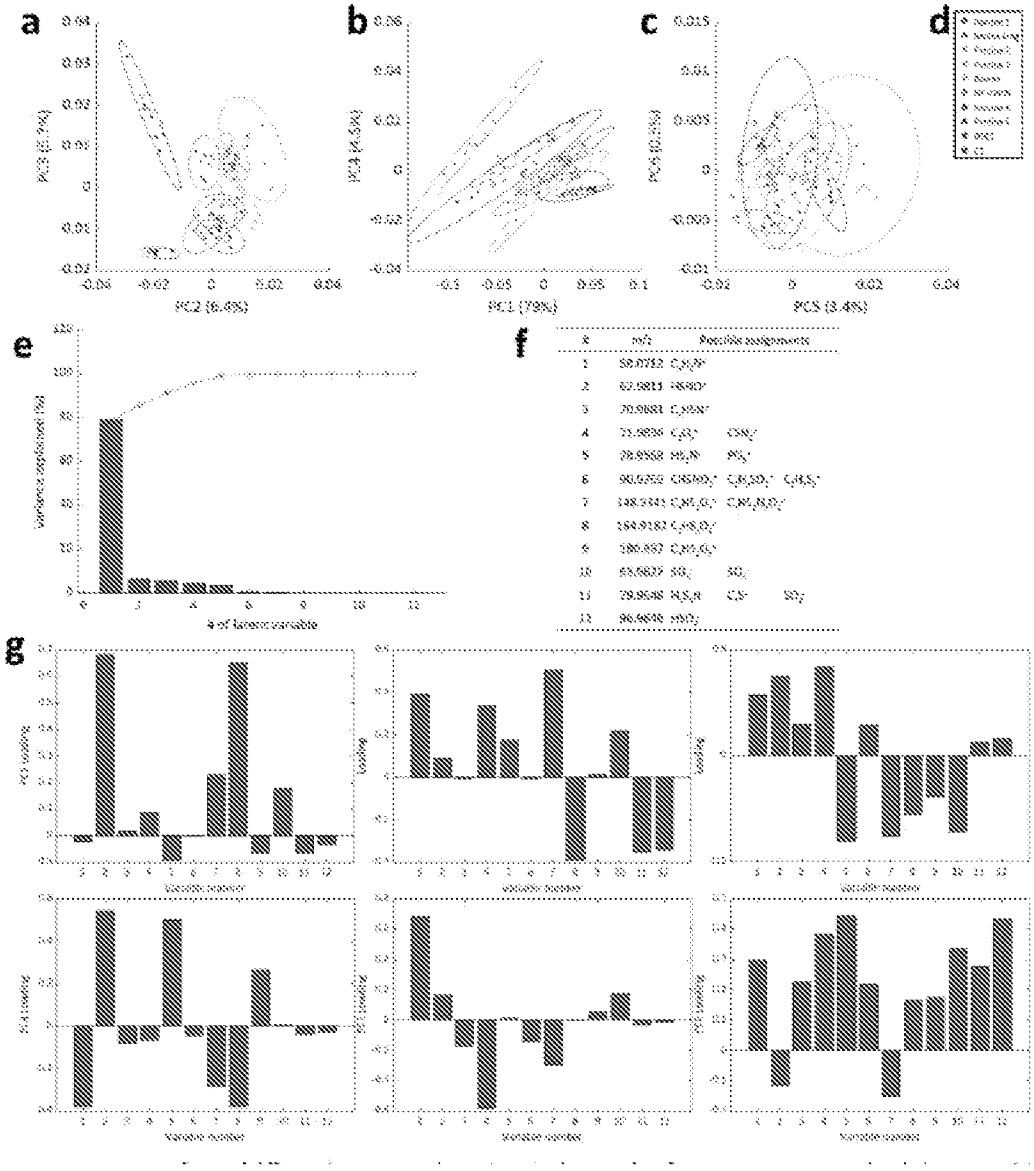
a
b
c
d
e
f
g
PC2 (6.6%)
PC1 (79%)
PC5 (3.4%)
# of latent variable
Variable number

Figure 8

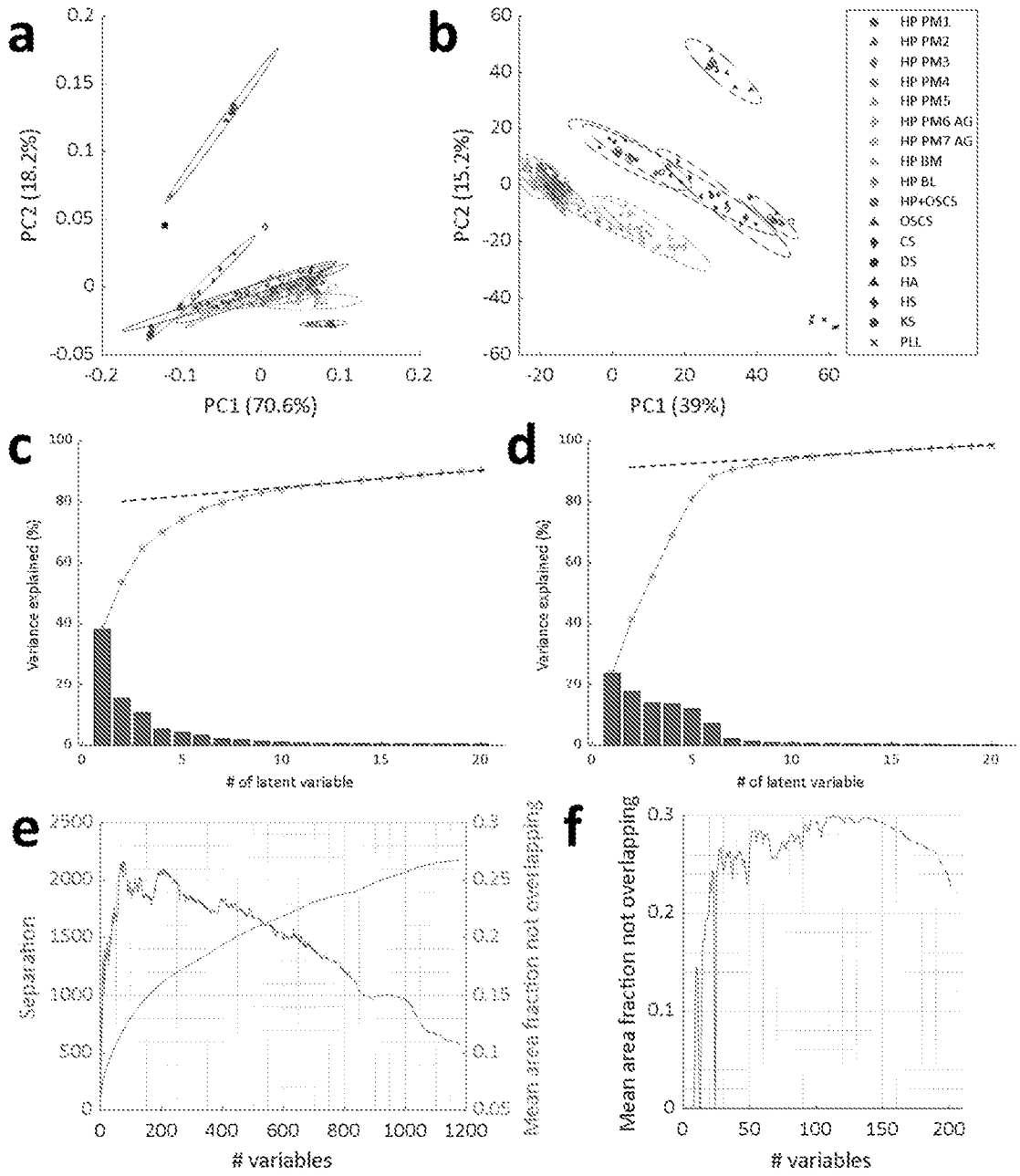
a
b
PC2 (18.2%)
PC1 (70.6%)
PC2 (15.2%)
PC1 (39%)
HP PM1
HP PM2
HP PM3
HP PM4
HP PM5
HP PM6 AG
HP PM7 AG
HP BM
HP BL
HP+OSCS
OSCS
CS
DS
HA
HS
KS
PLL
c
d
Variance explained (%)
# of latent variable
e
f
Separation
Mean area fraction not overlapping
variables

Figure 9

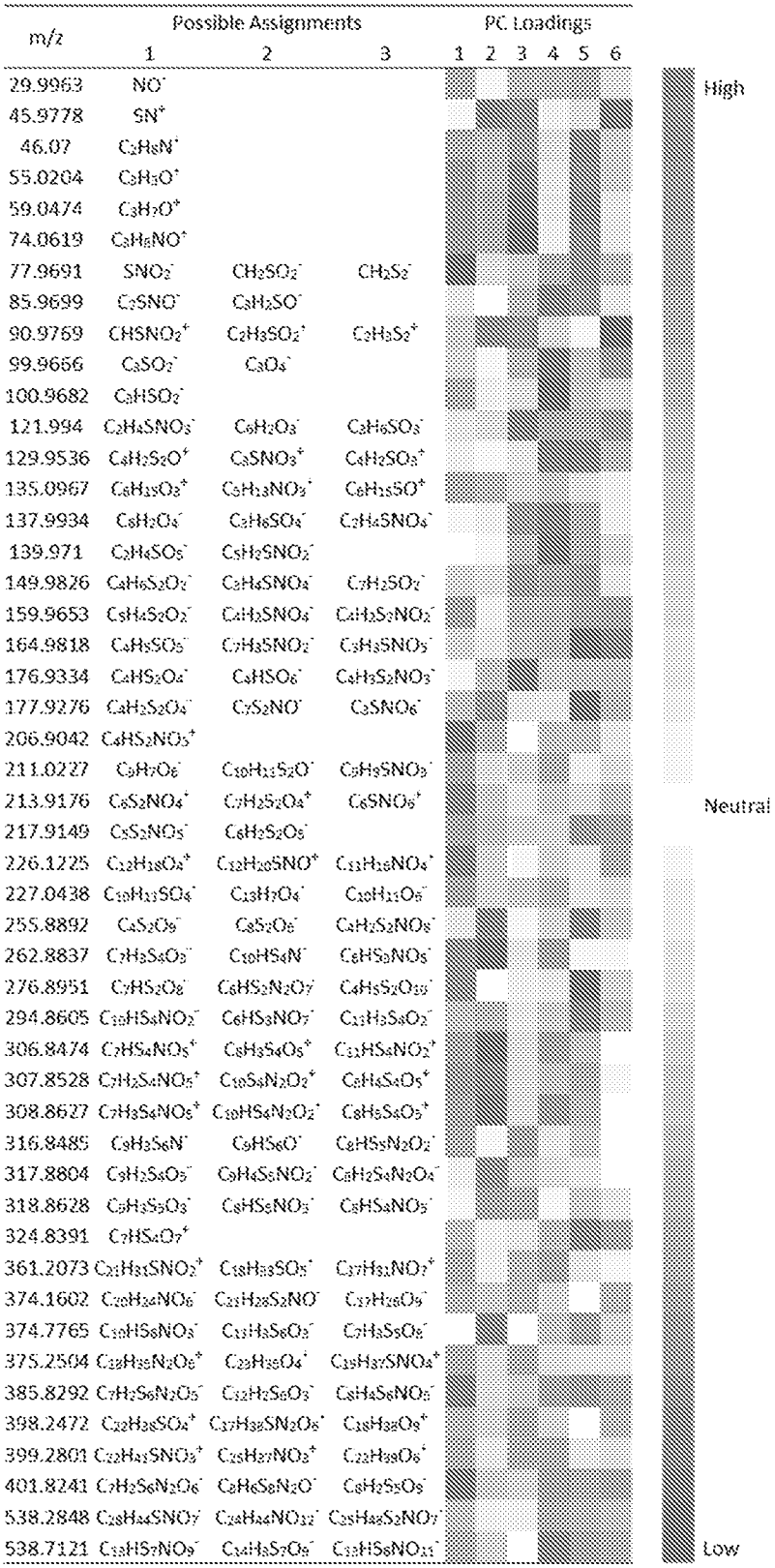
m/z
Possible Assignments
1
2
3
PC Loadings
1 2 3 4 5 6
29.9963 NO⁻
45.9778 SN⁺
46.07 C2H8N⁺
55.0204 C3H3O⁺
59.0474 C3H7O⁺
74.0619 C3H8NO⁺
77.9691 SNO2⁻ CH2SO2⁻ CH2S2⁻
85.9699 C2SNO⁻ C3H2SO⁻
90.9769 CHSNO2⁺ C2H3SO2⁺ C2H3S2⁺
99.9656 C3SO2⁻ C3O4⁻
100.9682 C3HSO2⁻
121.994 C2H4SNO3⁻ C6H2O3⁻ C3H6SO3⁻
129.9536 C4H2S2O⁺ C3SNO3⁺ C4H2SO3⁺
135.0967 C8H13O3⁺ C5H13NO3⁺ C8H15SO⁺
137.9934 C3H2O4⁻ C3H6SO4⁻ C2H4SNO4⁻
139.971 C2H4SO5⁻ C5H2SNO2⁻
149.9826 C4H6S2O2⁻ C3H4SNO4⁻ C7H2SO2⁻
159.9653 C5H4S2O2⁻ C4H2SNO4⁻ C4H2S2NO2⁻
164.9818 C4H5SO5⁻ C7H3SNO2⁻ C5H3SNO5⁻
176.9334 C4HS2O4⁻ C6HSO6⁻ C4H3S2NO3⁻
177.9276 C4H2S2O4⁻ C7S2NO⁻ C3SNO6⁻
206.9042 C4HS2NO5⁺
211.0227 C9H7O6⁻ C10H11S2O⁻ C9H9SNO3⁻
213.9176 C6S2NO4⁺ C7H2S2O4⁺ C6SNO6⁺
217.9149 C5S2NO5⁻ C6H2S2O5⁻
226.1225 C12H18O4⁺ C12H20SNO⁺ C11H18NO4⁺
227.0438 C10H11SO4⁻ C13H7O4⁻ C10H11O6⁻
255.8892 C4S2O6⁻ C8S2O6⁻ C4H2S2NO6⁻
262.8837 C7H3S4O3⁻ C10HS4N⁻ C8HS3NO5⁻
276.8951 C7HS2O8⁻ C6HS2N2O7⁻ C4H5S2O10⁻
294.8605 C10HS4NO2⁻ C6HS3NO7⁻ C13H3S4O2⁻
306.8474 C7HS4NO5⁺ C8H3S4O5⁺ C11HS4NO2⁺
307.8528 C7H2S4NO5⁺ C10S4N2O2⁺ C8H4S4O5⁺
308.8627 C7H3S4NO5⁺ C10HS4N2O2⁺ C8H5S4O5⁺
316.8485 C9H3S6N⁻ C9HS6O⁻ C8HS5N2O2⁻
317.8804 C9H2S4O5⁻ C9H4S5NO2⁻ C8H2S4N2O4⁻
318.8628 C9H3S5O3⁻ C9HS5NO3⁻ C9HS4NO5⁻
324.8391 C7HS4O7⁺
361.2073 C21H31SNO2⁺ C18H33SO5⁻ C17H31NO7⁺
374.1602 C20H24NO6⁻ C21H28S2NO⁻ C17H26O9⁻
374.7765 C10HS6NO3⁻ C13H3S6O2⁻ C7H3S5O8⁻
375.2504 C18H35N2O6⁺ C22H35O4⁺ C15H37SNO4⁺
385.8292 C7H2S6N2O5⁻ C12H2S6O3⁻ C9H4S6NO5⁻
398.2472 C22H38SO4⁺ C17H38SN2O6⁺ C18H38O9⁺
399.2801 C22H43SNO3⁺ C25H37NO3⁺ C22H39O6⁺
401.8241 C7H2S6N2O6⁻ C8H6S8N2O⁻ C8H2S5O9⁻
538.2848 C28H44SNO7⁺ C24H44NO12⁻ C25H48S2NO7⁺
538.7121 C13HS7NO9⁻ C14H3S7O6⁻ C13HS8NO11⁻
High
Neutral
Low

Figure 11

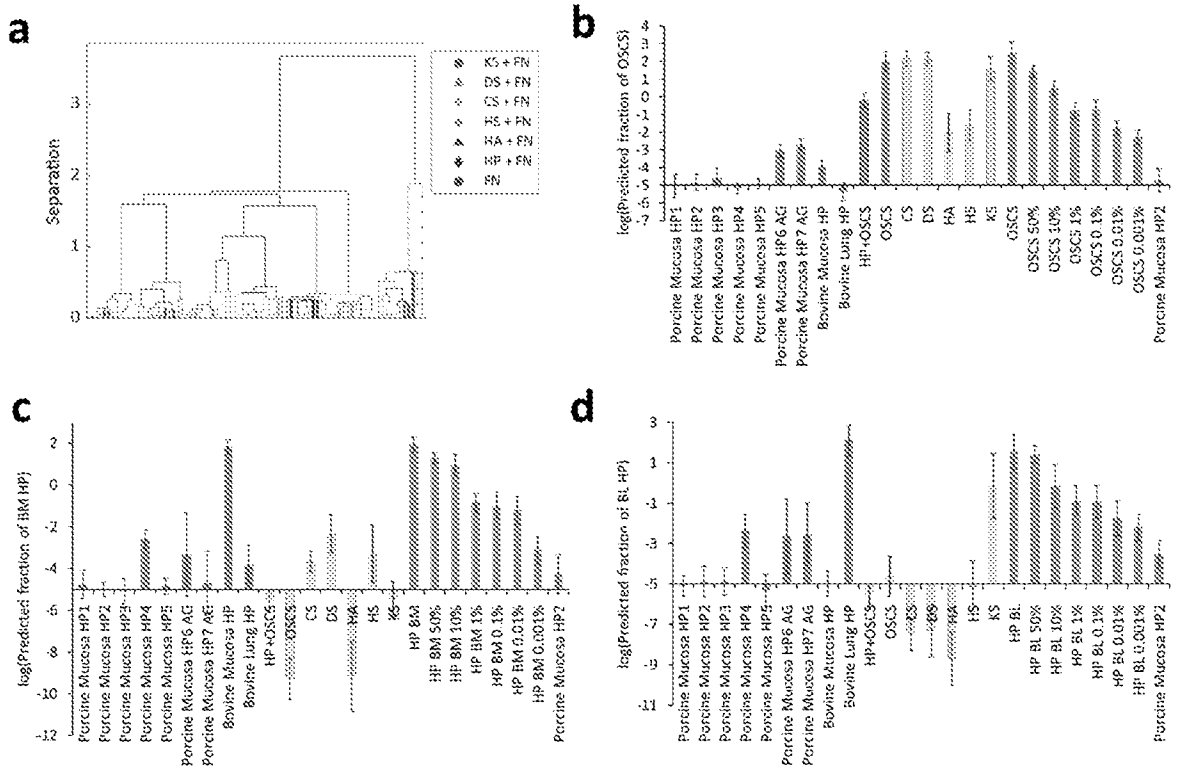

Figure 12

m/z	Possible Assignments 1	2	3	PC Loadings 1	2
19.0190	H_3O^+				
30.0342	CH_4N^+				
58.0050	$C_2H_2O_2^-$				
60.9742	$CHSO^-$				
71.9834	$C_5H_{11}^+$				
87.9758	C_2HSNO^+	$C_3H_3SO^-$			
90.9954	$CH_3SN_2O^-$	$CHNO_4^-$	$C_2H_3O_4^-$		
95.0192	$C_5H_5NO^-$	$C_5H_3O_2^-$			
99.9666	$C_3SO_2^-$	$C_3HPO_2^-$	$C_3H_2SNO^-$		
100.9682	$C_3HSO_2^-$				
120.9449	$C_2HS_2O_2^-$				
122.0743	$C_8H_{11}N^+$	$C_9H_{13}^+$	$C_5H_{11}O_3^+$		
132.9841	C_7HSO^-	$C_3H_3SNO_3^-$			
142.9466	$C_6H_{14}O_2^+$				
165.1427	$C_7HS_2O^+$	$C_9HSO_5^+$	$C_9HS_2O_4^-$		
167.9270	$C_6SO_4^-$	$C_5H_2S_2NO^-$	$C_2H_2SNO_6^-$		
174.9900	$C_5H_5S_2NO_2^-$	$C_5H_3SO_5^-$	$C_9H_3SO_2^-$		
256.2452	$C_{13}H_{26}N_3O_2^-$	$C_{10}H_{24}N_3^-$	$C_{13}H_{28}N_2O_2^-$		

High
Neutral
Low

Figure 13

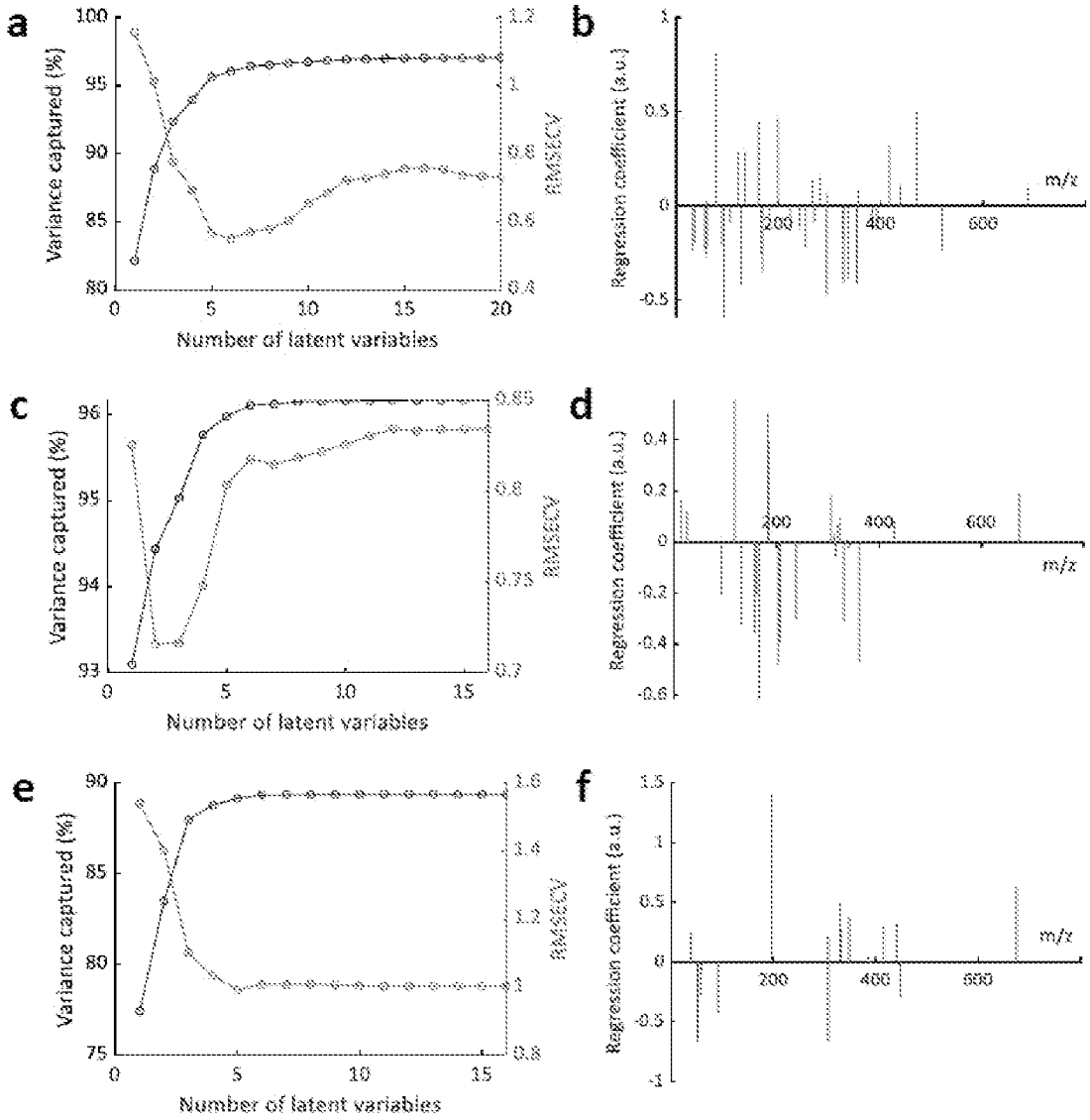
a
Variance captured (%)
RMSECV
Number of latent variables
b
Regression coefficient (a.u.)
m/z
c
Variance captured (%)
RMSECV
Number of latent variables
d
Regression coefficient (a.u.)
m/z
e
Variance captured (%)
RMSECV
Number of latent variables
f
Regression coefficient (a.u.)
m/z

Figure 14

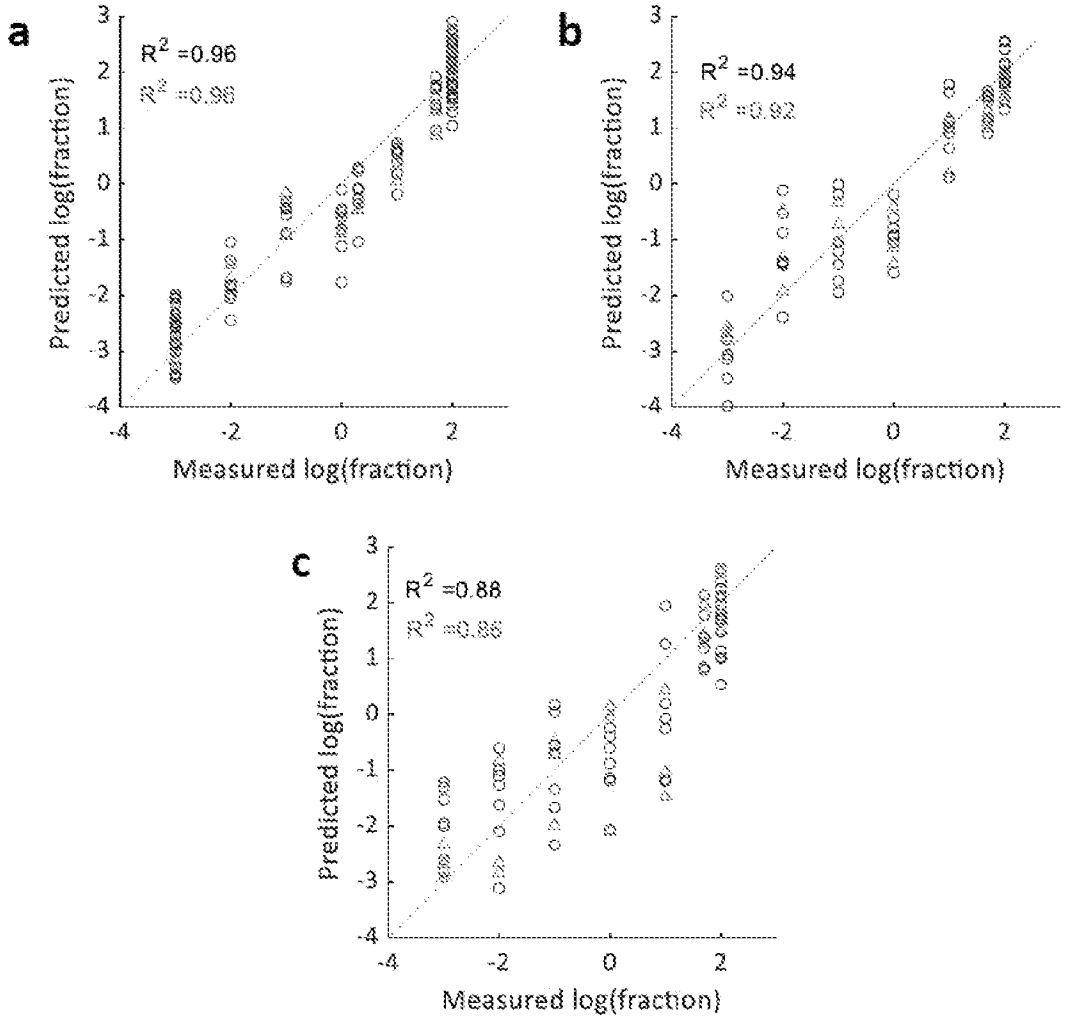

Figure 15

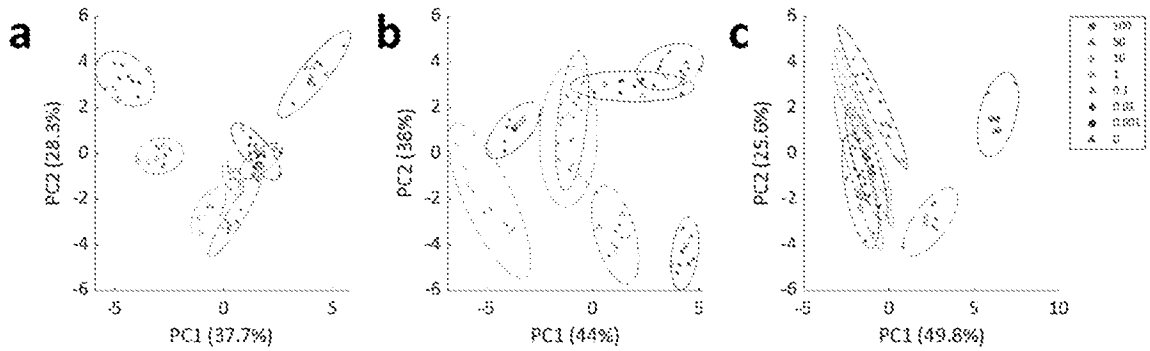

Figure 16

m/z	Possible Assignments			PC Loadings	
	1	2	3	1	2
39.9948	C_2O^+				
56.9673	C_2HS^+	$C_2HO_2^+$	$C_2H_3NO^+$		
62.9660	$HSNO^-$				
83.9691	C_3SO^-				
100.0338	$C_5H_8S^-$	$C_8H_4^-$	$C_4H_6NO_2^-$		
101.9609	$C_3H_2S_2^-$	$C_2SNO_2^-$			
130.9961	$C_4H_3O_5^-$	$C_7HNO_2^-$	$C_4H_5SNO_2^-$		
131.9608	$C_4H_4S_2O^-$	$C_3H_2SNO_3^-$			
151.9477	$C_2H_2SNO_5^+$	$C_3H_4SO_5^+$			
179.8972	$C_7S_2O_2^-$	$C_6S_2N_2O^-$			
180.8892	$C_7HS_2O_2^+$				
192.9501	$C_8H_3S_2NO_4^+$	$C_4HSO_7^+$	$C_5H_5S_2O_4^+$		
200.2003	$C_{12}H_{26}NO^+$	$C_{13}H_{28}O^+$			
235.9197	$C_6H_4S_3O_4^-$	$C_5H_2S_3NO_4^-$			
266.8394	$C_6H_3S_3O_6^+$	$C_5HS_2NO_8^+$			
293.8860	$C_{11}H_2S_4O_2^-$	$C_7H_2S_3O_7^-$			
348.8865	$C_9H_3S_4NO_6^-$	$C_8H_5S_4O_9^-$	$C_{10}H_5S_4O_6^-$		
455.2817	$C_{25}H_{43}SO_5^+$	$C_{23}H_{43}O_{10}^+$	$C_{24}H_{41}NO_7^+$		

Figure 17

m/z	Possible Assignments			PC Loadings	
	1	2	3	1	2
53.0404	$C_4H_5^-$				
59.0139	$C_2H_3O_2^-$				
77.037	CH_3SNO^+				
93.9883	$CH_2O_5^-$	$CH_4SNO_2^-$			
95.9576	SO_4^-				
100.03	$C_5H_8S^-$	$C_8H_4^-$	$C_4H_8NO_2^-$		
115.9565	$C_3SO_3^-$				
129.9501	$C_3NO_5^-$	$C_4H_2SO_3^-$			
131.9604	$C_3SO_4^-$				
135.9802	$C_3H_6SNO_3^-$	$C_7H_4O_2^-$	$C_7H_4O_3^-$		
138.977	$C_2H_3SO_5^-$	$C_5HSNO_2^-$			
144.0835	$HSNO_6^+$				
147.9564	$C_4H_4SO_4^-$	$C_3H_2NO_6^-$	$C_3H_2SNO_4^-$		
164.9818	$C_8H_5SO_2^-$	$C_4H_2SO_5^-$	$C_4H_5O_7^-$		
186.9428	$C_2H_3SNO_7^-$	$C_6H_3SO_5^-$	$C_6H_3O_7^-$		
223.1637	$C_5H_3S_2O_6^+$	$C_8HS_2NO_3^-$			
246.8976	$C_7H_3SO_8^-$	$C_4H_7S_2O_8^-$	$C_7H_3S_2NO_5^-$		
260.8798	$C_4H_5S_2O_9^-$	$C_7H_3S_2NO_6^-$			
262.8837	$C_7H_3S_2O_7^-$	$C_7H_3SO_9^-$	$C_4H_7S_2O_9^-$		
292.87	$C_7H_3S_2NO_8^-$	$C_8H_5S_2O_8^-$	$C_7H_3SNO_{10}^-$		
318.8628	$C_9H_3S_2O_9^-$				
402.8098	$C_{11}HS_3NO_{10}^-$	$C_{12}H_3S_3O_{10}^-$	$C_{16}H_3S_3O_7^-$		
449.3843	$C_{12}HS_3O_{13}^+$	$C_{16}HS_3O_{10}^-$	$C_9H_5S_3O_{15}^+$		
522.7382	$C_{13}H_3S_3N_2O_{15}^-$	$C_{14}H_5S_3NO_{15}^-$	$C_{17}H_3S_3N_2O_{12}^-$		
686.6418	$C_{32}H_{56}NO_{14}^-$	$C_{31}H_{64}N_3O_{11}^-$	$C_{30}H_{60}N_3O_{14}^-$		

Figure 18

m/z	Possible Assignments			PC Loadings	
	1	2	3	1	2
29.0035	CHO^-				
77.9691	SNO_2^-				
83.9684	C_3SO^-				
90.9769	$CHSNO_2^+$				
96.9564	HSO_4^+				
106.9624	$C_2H_3S_2O^+$	$CHSNO_3^+$			
115.9565	$C_3SO_3^-$				
120.9758	C_6HSO^+	$C_2HO_6^+$	$C_2H_3SNO_3^+$		
137.9302	$CSNO_5^-$				
138.9282	$CHSNO_5^+$				
213.0268	$C_8H_7NO_6^-$	$C_6H_{13}S_2O_4^-$	$C_5H_{11}SNO_6^-$		
222.8808	$C_4HS_2NO_5^-$	$C_5H_3S_2O_6^-$			
230.2627	$C_{14}H_{32}NO^+$	$C_{13}H_{30}N_2O^+$			
324.8391	$C_7H_3S_3NO_8^+$				
544.7476	$C_{13}H_5S_6O_{12}^-$	$C_{26}H_3S_6NO_9^-$	$C_{12}H_5S_6N_2O_{11}^-$		

High

Neutral

Low

Figure 19

	Anti-IIa (IU/mg)	Anti-Xa (IU/mg)	Human Plasma (IU/mg)
Heparin porcine mucosa 1	209 (13)	204 (9)	194 (8)
Heparin porcine mucosa 2	226 (11)	226 (12)	226 (12)
Heparin porcine mucosa 3	200 (11)	200 (7)	200 (6)
Heparin porcine mucosa 4	183 (9)	184 (13)	185 (8)
Heparin porcine mucosa 5	384 (57)	326 (53)	275 (32)
Heparin bovine mucosa	140 (9)	140 (5)	161 (6)
Heparin bovine lung	154 (8)	141 (7)	129 (8)

Figure 20

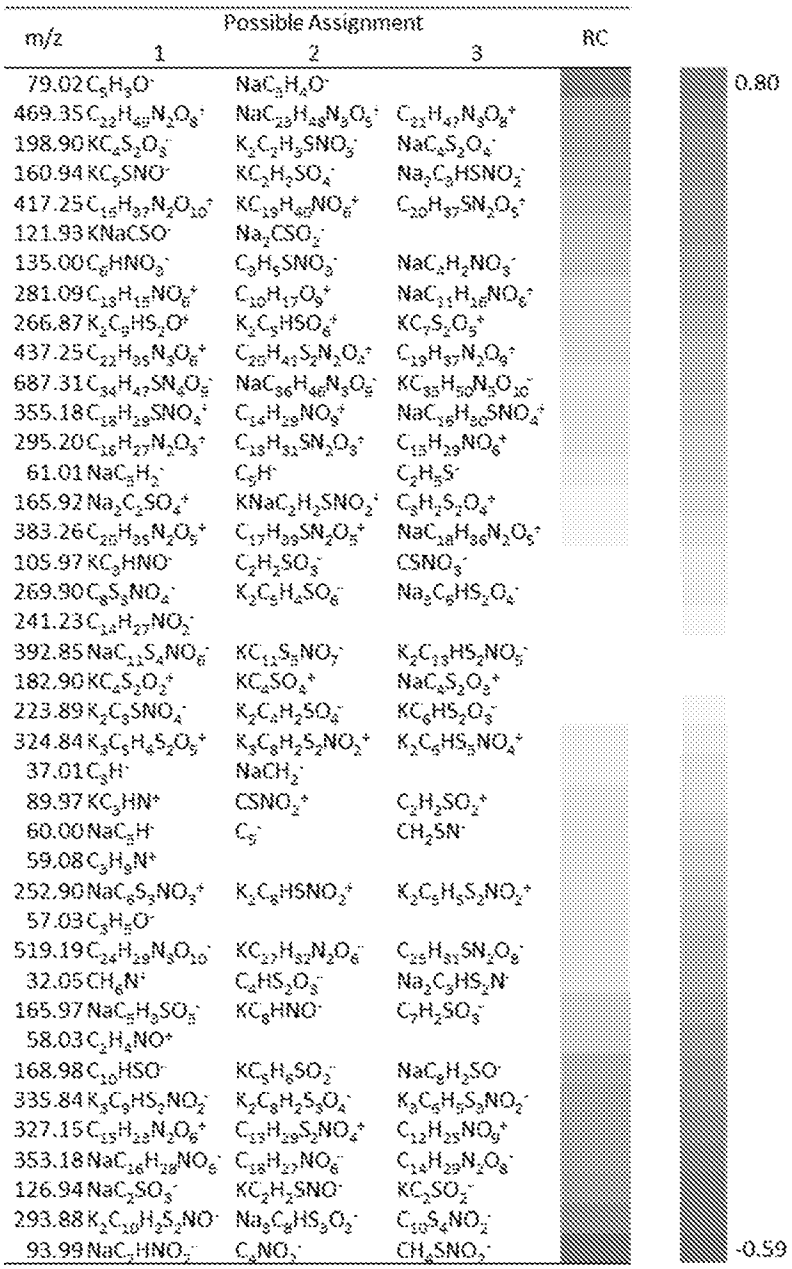
m/z
Possible Assignment
1
2
3
RC
79.02 $C_5H_3O^-$ $NaC_3H_4O^-$
469.35 $C_{22}H_{45}N_2O_8^-$ $NaC_{23}H_{48}N_3O_5^-$ $C_{21}H_{47}N_3O_8^+$
198.90 $KC_4S_2O_5^-$ $K_2C_2H_3SNO_3^-$ $NaC_4S_2O_4^-$
160.94 KC_5SNO^- $KC_2H_2SO_4^-$ $Na_2C_3HSNO_2^-$
417.25 $C_{15}H_{37}N_2O_{10}^+$ $KC_{19}H_{40}NO_6^+$ $C_{20}H_{37}SN_2O_5^+$
121.93 $KNaCSO^-$ $Na_2CSO_2^-$
135.00 $C_6HNO_3^-$ $C_3H_6SNO_3^-$ $NaC_4H_2NO_3^-$
281.09 $C_{18}H_{15}NO_6^+$ $C_{10}H_{17}O_5^+$ $NaC_{11}H_{16}NO_6^+$
266.87 $K_2C_3HS_2O^+$ $K_2C_5HSO_6^+$ $KC_7S_2O_5^+$
437.25 $C_{22}H_{35}N_3O_8^+$ $C_{25}H_{41}S_2N_2O_2^+$ $C_{19}H_{37}N_2O_9^+$
687.31 $C_{34}H_{47}SN_4O_5^-$ $NaC_{36}H_{46}N_3O_5^-$ $KC_{33}H_{50}N_3O_{10}^-$
355.18 $C_{18}H_{28}SNO_4^-$ $C_{14}H_{28}NO_8^+$ $NaC_{16}H_{30}SNO_4^+$
295.20 $C_{16}H_{27}N_2O_3^+$ $C_{13}H_{31}SN_2O_3^+$ $C_{13}H_{29}NO_6^+$
61.01 $NaC_2H_2^-$ C_2H^- $C_2H_5S^-$
165.92 $Na_2C_2SO_4^+$ $KNaC_2H_2SNO_2^-$ $C_3H_2S_2O_4^+$
383.26 $C_{20}H_{35}N_2O_5^+$ $C_{17}H_{39}SN_2O_5^+$ $NaC_{18}H_{36}N_2O_5^+$
105.97 KC_3HNO^- $C_2H_2SO_3^-$ $CSNO_3^-$
269.90 $C_8S_3NO_4^-$ $K_2C_5H_4SO_6^-$ $Na_3C_6HS_2O_4^-$
241.23 $C_{14}H_{27}NO_2^-$
392.85 $NaC_{11}S_4NO_6^-$ $KC_{11}S_5NO_7^-$ $K_2C_{13}HS_2NO_5^-$
182.90 $KC_4S_2O_2^+$ $KC_4SO_4^+$ $NaC_4S_2O_5^+$
223.89 $K_2C_3SNO_4^-$ $K_2C_4H_2SO_4^-$ $KC_6HS_2O_5^-$
324.84 $K_3C_5H_4S_2O_5^+$ $K_3C_8H_2S_2NO_2^+$ $K_2C_5HS_3NO_4^+$
37.01 C_3H^- $NaCH_2^-$
89.97 KC_3HN^+ $CSNO_2^+$ $C_2H_2SO_2^+$
60.00 NaC_2H^- C_2^- CH_2SN^-
59.08 $C_3H_9N^+$
252.90 $NaC_6S_3NO_3^+$ $K_2C_8HSNO_2^+$ $K_2C_5H_5S_2NO_2^+$
57.03 $C_3H_5O^-$
519.19 $C_{24}H_{28}N_3O_{10}^-$ $KC_{27}H_{32}N_2O_6^-$ $C_{25}H_{31}SN_2O_8^-$
32.05 CH_6N^- $C_4HS_2O_3^-$ $Na_2C_3HS_2N^-$
165.97 $NaC_5H_3SO_5^-$ KC_8HNO^- $C_7H_2SO_3^-$
58.03 $C_2H_4NO^+$
168.98 $C_{10}HSO^-$ $KC_5H_6SO_2^-$ $NaC_8H_2SO^-$
335.84 $K_3C_3HS_2NO_2^-$ $K_2C_8H_2S_3O_4^-$ $K_3C_5H_5S_3NO_2^-$
327.15 $C_{15}H_{28}N_2O_6^+$ $C_{13}H_{29}S_2NO_4^+$ $C_{12}H_{25}NO_8^+$
353.18 $NaC_{16}H_{28}NO_6^-$ $C_{18}H_{27}NO_6^-$ $C_{14}H_{29}N_2O_8^-$
126.94 $NaC_2SO_3^-$ $KC_2H_2SNO^-$ $KC_2SO_2^-$
293.88 $K_2C_{10}H_2S_2NO^-$ $Na_3C_8HS_3O_2^-$ $C_{13}S_4NO_2^-$
93.99 $NaC_2HNO_2^-$ $C_3NO_2^-$ $CH_4SNO_2^-$
0.80
-0.59

Figure 21

m/z	Possible Assignment 1	2	3	RC
117.97	$C_2SNO_3^-$	KC_4HNO^-	$C_3H_2SO_3^-$	
181.94	$C_6S_2NO_2^+$	$NaC_4HS_2NO_2^+$	$C_3H_2S_2O_5^+$	
183.96	$Na_2C_3H_6S_2O_2^+$	$NaC_5H_5S_2O_2^+$	$K_2C_3H_6SNO^+$	
673.37	$C_{31}H_{53}N_4O_{12}^-$	$KC_{34}H_{56}N_3O_8^-$	$C_{32}H_{55}SN_3O_{10}^-$	
305.88	$C_{11}S_4NO_2^+$	$NaC_9HS_4NO_2^+$	$K_2C_8H_4S_2O_4^+$	
14.02	CH_2^-			
25.01	C_2H^-			
323.28	$C_{15}H_{35}N_2O_5^+$			
429.32	$C_{22}H_{45}N_3O_5^+$	$NaC_{21}H_{46}N_2O_5^+$	$C_{22}H_{41}N_2O_6^+$	
319.21	$C_{15}H_{31}SN_2O_3^+$	$NaC_{16}H_{28}N_2O_3^+$	$C_{15}H_{29}NO_6^+$	
313.19	$C_{12}H_{29}N_2O_7^+$	$C_{16}H_{27}NO_5^+$	$NaC_{14}H_{28}NO_5^+$	
337.24	$C_{15}H_{35}NO_6^+$	$C_{15}H_{33}N_2O_5^+$	$C_{19}H_{31}NO_4^+$	
342.24	$NaC_{16}H_{35}N_2O_4^+$	$C_{18}H_{34}N_2O_4^+$	$C_{18}H_{32}NO_5^+$	
315.18	$C_{11}H_{27}N_2O_8^+$	$NaC_{17}H_{26}NO_3^+$	$C_{15}H_{25}NO_5^+$	
160.97	$NaC_3H_6S_2O_2^+$	$NaC_5NO_4^+$	$KC_3H_6SO_3^+$	
92.01	$NaC_3H_3NO^+$	$C_5H_2NO^-$	$KC_4H_5^-$	
237.07	$C_{14}H_9N_2O_2^-$	$C_8H_{15}SNO_3^-$	$C_{11}H_{11}NO_3^-$	
331.23	$C_{17}H_{33}NO_5^+$	$C_{16}H_{31}N_2O_5^+$	$NaC_{14}H_{32}N_2O_5^+$	
131.96	$KC_2HO_2^-$	$C_3H_2S_2NO^-$	$Na_2C_3H_2SO^-$	
156.98	$C_5H_5SNO_3^+$	$NaC_3H_4SNO_3^+$	$C_6H_5S_2O^+$	
205.91	$KNaC_4SO_4^-$	$K_2C_4H_2SNO_2^-$	$KC_3H_3S_2O_4^-$	
361.20	$C_{16}H_{29}N_2O_7^-$	$NaC_{14}H_{30}N_2O_7^-$	$C_{17}H_{31}NO_7^-$	
202.95	$C_6H_5S_3NO^-$	$C_5HSNO_6^-$	$Na_3C_4H_6S_2O^-$	
164.98	$C_7HO_5^-$	$NaC_6H_2O_5^-$	$KC_5H_6NO_3^-$	

Figure 22

m/z	Possible Assignment 1	2	3	RC
196.90	$KC_4S_2NO_2^+$	$K_2C_3H_3SO_3^+$	$KC_5H_2S_2O_2^+$	
674.39	$C_{32}H_{56}N_3O_{12}^-$	$C_{30}H_{56}N_3O_{12}^-$	$C_{31}H_{54}N_4O_{12}^-$	
330.23	$C_{17}H_{32}NO_5^+$	$NaC_{15}H_{33}NO_5^+$	$C_{16}H_{30}N_2O_5^+$	
347.84	$KNa_2C_6HS_3NO_5^-$	$K_3C_7H_3S_3NO_2^-$	$K_2NaC_{10}HS_3NO^-$	
441.35	$C_{24}H_{45}N_2O_5^+$	$NaC_{22}H_{46}N_2O_5^+$	$C_{20}H_{45}N_2O_8^+$	
415.30	$KC_{23}H_{44}N_2O_4^+$	$NaC_{20}H_{42}NO_6^+$	$C_{21}H_{39}N_2O_6^+$	
332.24	$C_{17}H_{34}NO_5^+$	$C_{16}H_{32}N_2O_5^+$	$C_{17}H_{34}SNO_3^+$	
39.02	$NaCH_2^+$	$C_3H_3^+$		
307.23	$C_{15}H_{33}NO_5^+$	$C_{14}H_{31}N_2O_5^+$	$C_{15}H_{33}SNO_3^+$	
386.28	$C_{20}H_{38}N_2O_5^+$	$NaC_{18}H_{39}N_2O_5^+$	$C_{20}H_{36}NO_5^+$	
389.27	$C_{19}H_{37}N_2O_6^+$	$NaC_{17}H_{38}N_2O_6^+$	$C_{19}H_{35}NO_7^+$	
445.34	$C_{28}H_{47}N_3O_5^+$	$C_{25}H_{45}N_2O_8^+$	$NaC_{21}H_{46}N_2O_6^+$	
60.00	NaC_3H^-	C_5^-	CH_2SN^-	
447.33	$C_{22}H_{45}N_3O_8^+$	$C_{28}H_{47}N_2O_6^+$	$C_{22}H_{43}N_2O_7^+$	
58.99	$C_2H_3S^+$	NaC_3^+		
92.01	$NaC_3H_3NO^-$	$C_5H_2NO^-$	$KC_4H_5^-$	
51.99	$NaCHO^+$	C_3O^+		
306.85	$KNaC_3HS_3O_4^-$	$KC_6S_3O_4^+$	$KC_5S_2O_5^-$	

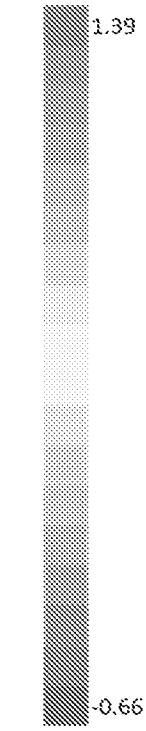

Figure 23

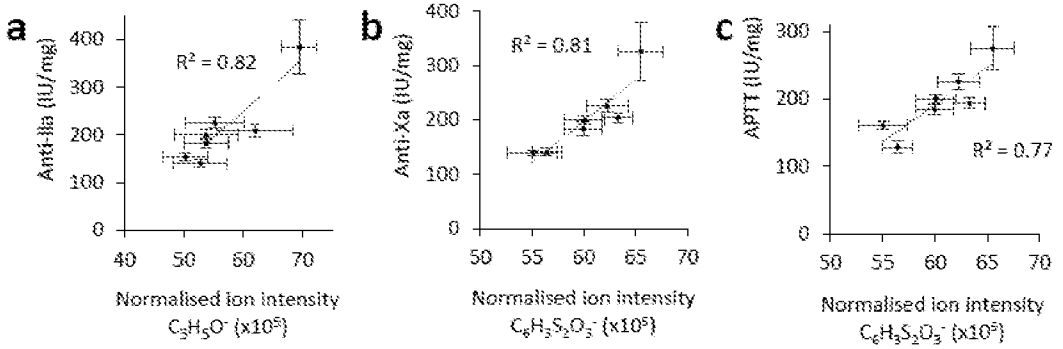

Figure 24

m/z	Possible assignments						PC1	PC2
180.8762	S_4NaNO^+	$CH_2S_4Na_2O^+$	$C_3HS_4O^+$	$S_3NaNO_3^+$	$H_3S_4NaN_2^+$	$C_2HS_4N_2^+$	0.43	-0.57
61.9676	SNO^-						0.39	0.61
215.9118	$C_2S_2N_2O_2^-$	$C_2H_2S_2NO_3^-$	$CS_2N_2O_3^-$	$C_2H_2S_4N_2O_2^-$	$C_{20}S_2^-$	$C_3H_2S_4N^-$	0.57	-0.38
228.8845	$C_2HS_4N_2O_2^-$	$C_2HS_4O^-$	$C_3HS_4O_4^-$	$C_3HS_3O_6^-$	$C_3H_2S_4NO_3^-$	$C_3HS_4N_2^-$	0.57	0.39

Figure 25

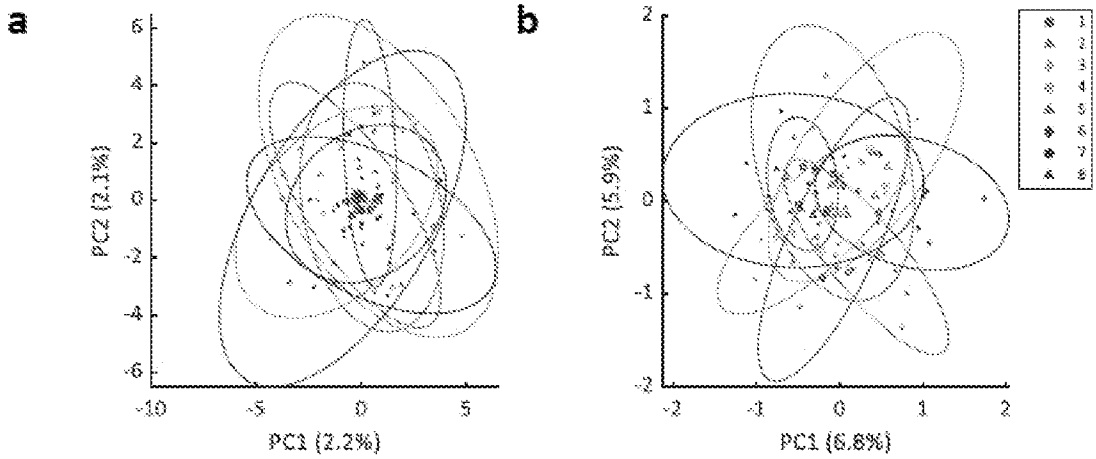

Figure 26

	Addition minimise overlap	Addition maximise Euclidean distance	Addition maximise Mahalanobis distance	Elimination minimise overlap	Elimination maximise Euclidean distance	Elimination maximise Mahalanobis distance
Addition minimise overlap	28	6	3	2	5	1
Addition maximise Euclidean distance		20	2	0	9	0
Addition maximise Mahalanobis distance			30	0	4	0
Elimination minimise overlap				35	1	2
Elimination maximise Euclidean distance					30	3
Elimination maximise Mahalanobis distance						40

Figure 27

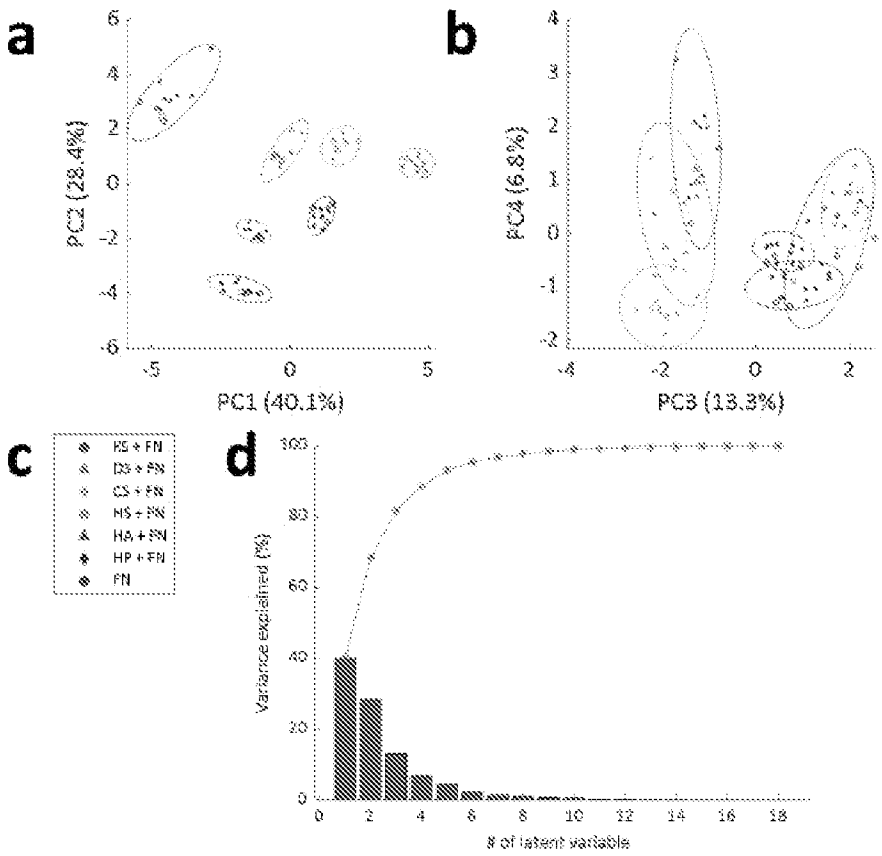
a
b
c
d
PC2 (28.4%)
PC1 (40.1%)
PC4 (6.8%)
PC3 (13.3%)
RS + FN
DS + FN
CS + FN
NS + FN
HA + FN
HP + FN
FN
Variance explained (%)
# of latent variable

Figure 28

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2022/050746

A. CLASSIFICATION OF SUBJECT MATTER
INV. G16B40/10 A61K31/726 C08B37/00 G01N23/2258
ADD.

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)
G16B A61K C09J C08B G01N

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)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MCARTHUR SALLY L. ET AL: "Characterization of sequentially grafted polysaccharide coatings using time-of-flight secondary ion mass spectrometry (ToF-SIMS) and principal component analysis (PCA)", SURFACE AND INTERFACE ANALYSIS., vol. 33, no. 12, 1 December 2002 (2002-12-01), pages 924-931, XP055929769, GB ISSN: 0142-2421, DOI: 10.1002/sia.1446 abstract ----- -/--	1-21

[X] 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 another citation or other special reason (as specified)

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 June 2022	05/07/2022

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

1

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2022/050746

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SZYMANSKA EWA: "Modern data science for analytical chemical data - A comprehensive review", ANALYTICA CHIMICA ACTA, vol. 1028, 1 October 2018 (2018-10-01), pages 1-10, XP055929672, AMSTERDAM, NL ISSN: 0003-2670, DOI: 10.1016/j.aca.2018.05.038 Retrieved from the Internet: URL:https://www.sciencedirect.com/science/article/pii/S0003267018306421/pdfft?md5=ebf9f52822b498568cb43b4e7ce7e72f&pid=1-s2.0-S0003267018306421-main.pdf> page 4 - page 8 -----	1-21