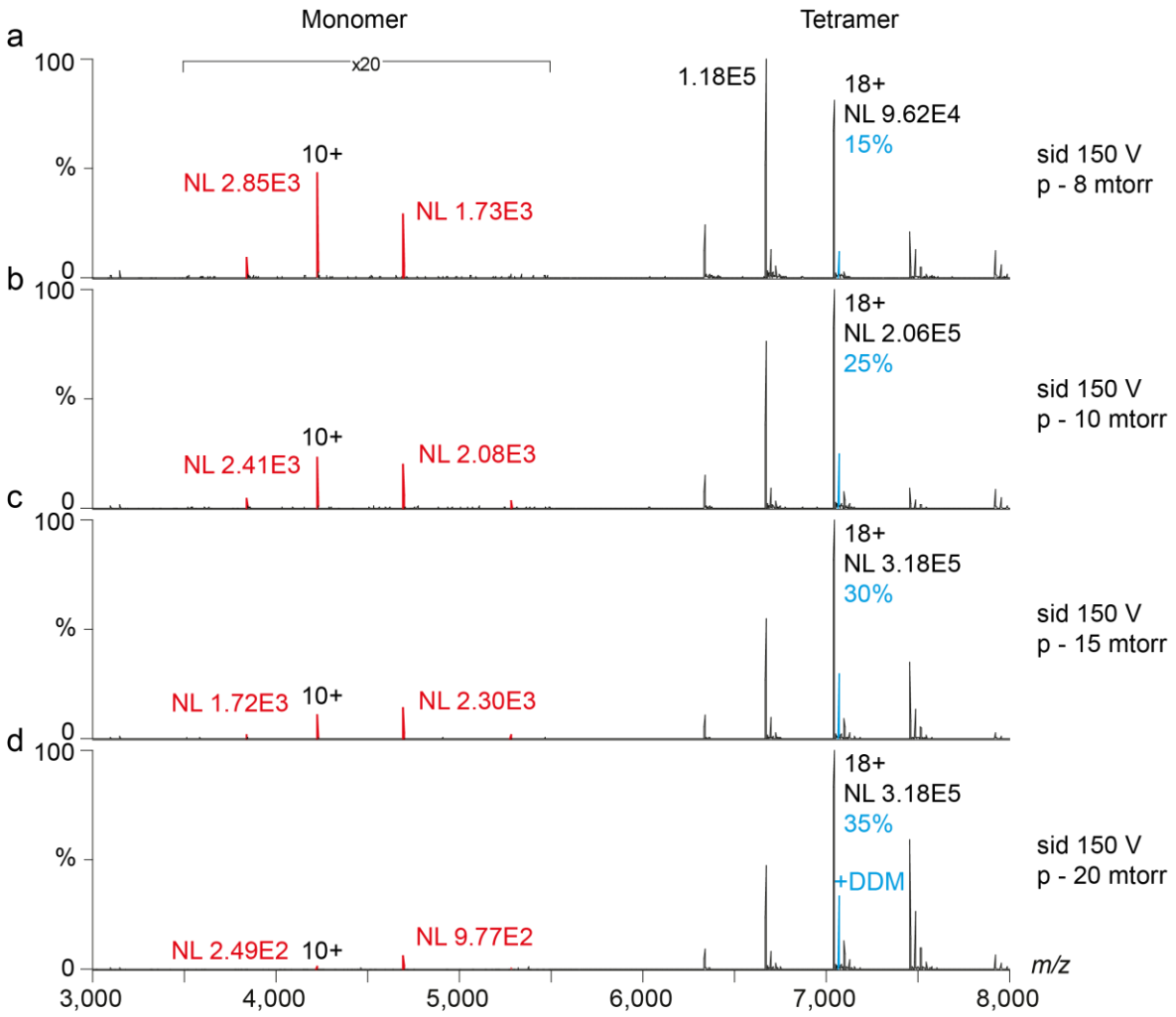
Supplementary Information for

Combining ‘Native’ with ‘Omics’ Based Mass Spectrometry to Identify Endogenous Membrane Protein Ligands

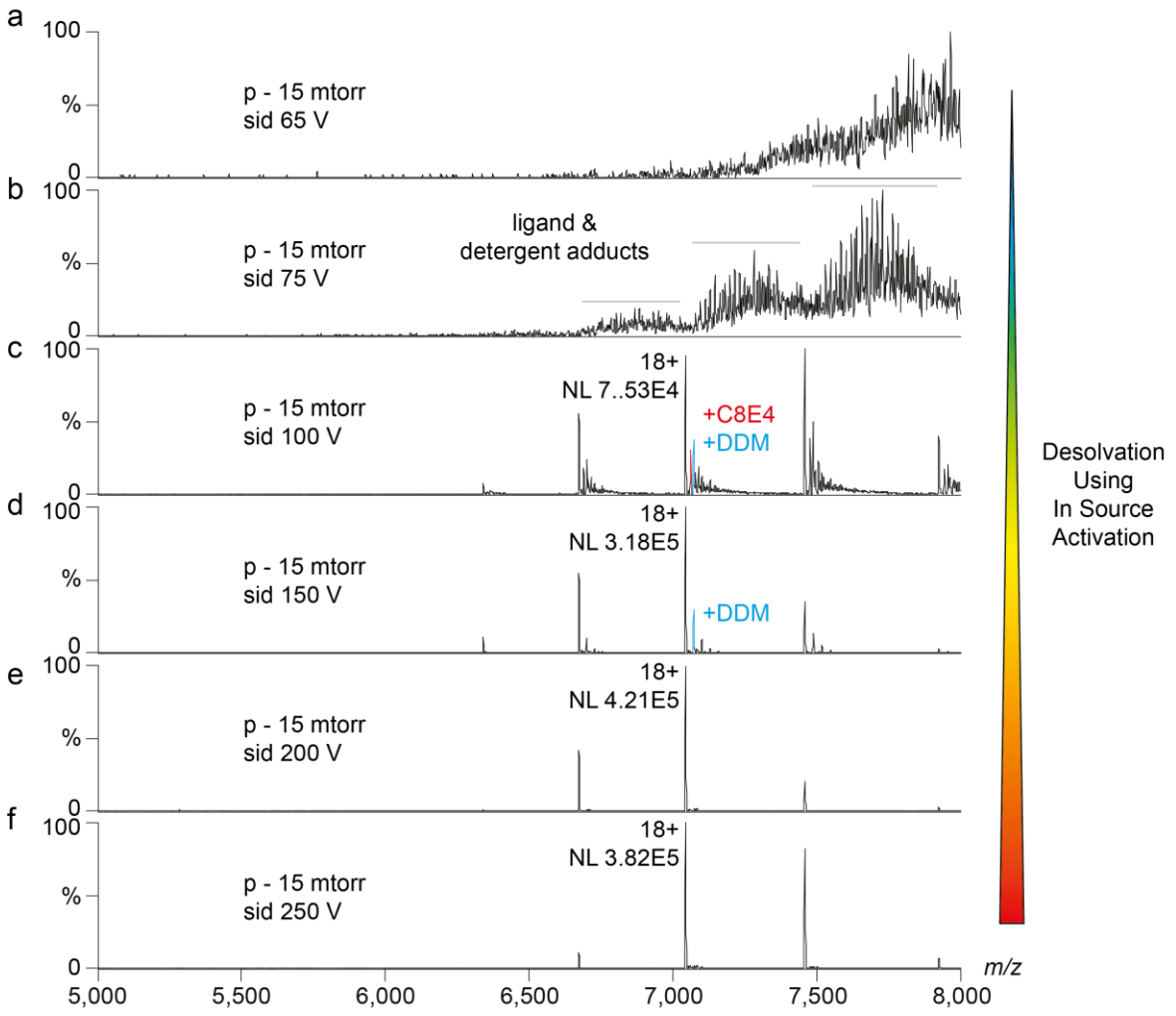
Joseph Gault*†[1], Idlir Liko†[1,2], Michael Landreh[3], Denis Shutin[1], Jani Reddy Bolla[1], Damien Jefferies[4], Mark Agasid[1], Hsin-Yung Yen[2], Marcus J. G. W. Ladds[3], David P. Lane[3], Syma Khalid[4], Christopher Mullen[5], Phil Remes[5], Romain Huguet[5], Graeme McAlister[5], Michael Goodwin[5], Rosa Viner[5], John Syka[5], Carol V. Robinson*[1]

Starting Protein-Ligand Complex		Minimum Stages of MS Required for Ligand ID	Details
Soluble	Monomer-Ligand	**MS^3**	MS2 dissociates ligand from complex, MS3 fragments ligand
	Oligomeric Assembly-Ligand (without identifying ligand-bound subunit)	**MS^3**	MS2 dissociates ligand from complex, MS3 fragments ligand
	Oligomeric Assembly-Ligand (with subunit localisation)	**MS^4**	MS2 dissociates complex, MS3 dissociates ligand from subunit, MS4 fragments ligand
Membrane	Monomer-Ligand	**pseudo MS^4**	pseudo MS2 removes micelle, pMS3 dissociates ligand from complex, pMS4 fragments ligand
	Oligomeric Assembly-Ligand (without identifying ligand-bound subunit)	**pseudo MS^4**	pseudo MS2 removes micelle, pMS3 dissociates ligand from complex, pMS4 fragments ligand
	Oligomeric Assembly-Ligand (with subunit localisation)	**pseudo MS^5**	pseudo MS2 removes micelle, pMS3 dissociates complex, pMS4 dissociates ligand from subunit, pMS5 fragments ligand

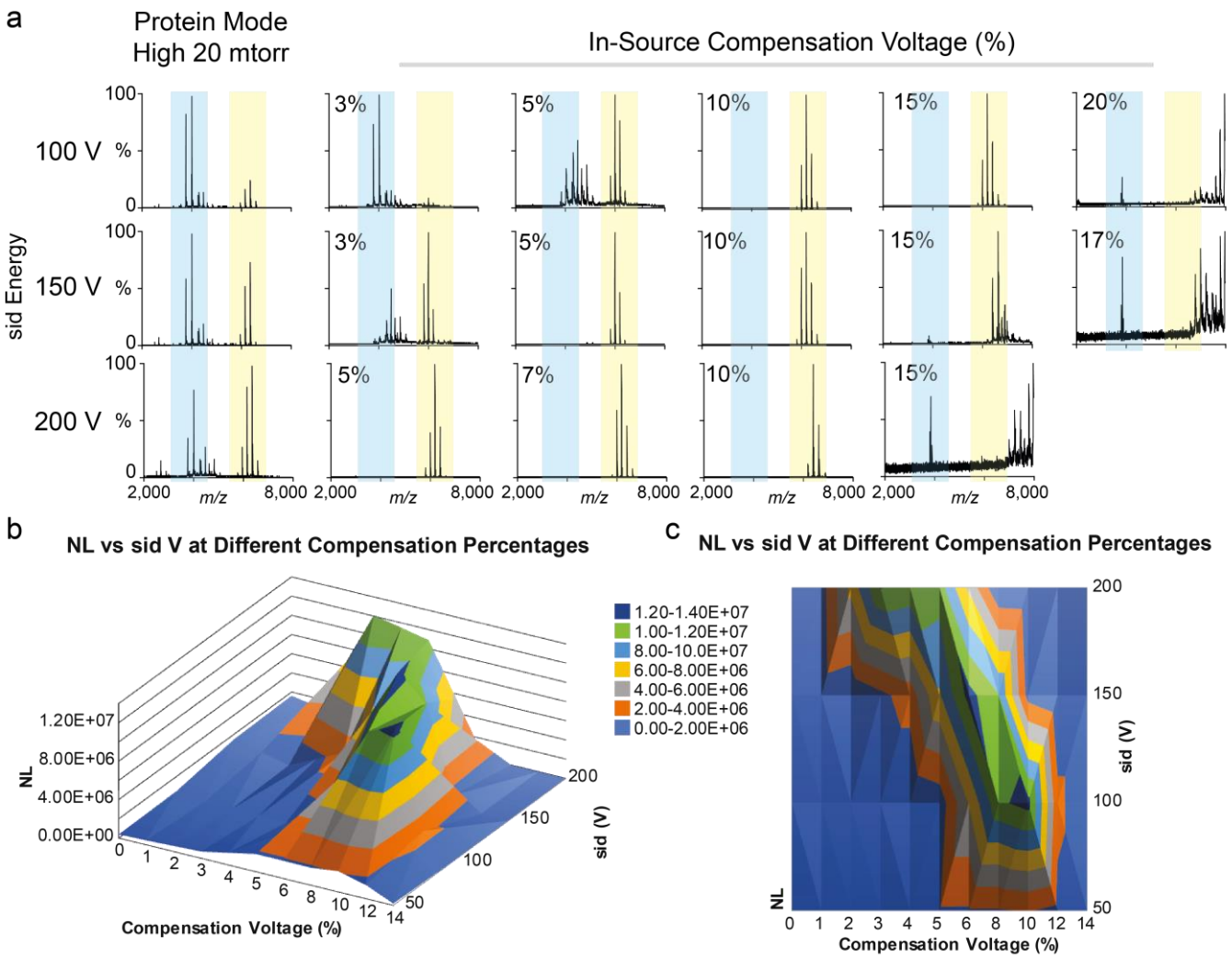
Supplementary Table 1 - Table outlining the minimum stages of MS/MS (MS^n) required for ligand release, isolation and fragmentation from parent complexes. MS^3 is sufficient for Nativeomics from monomeric soluble complexes, or soluble assemblies (with multiple proteins) without identifying ligand-bound subunit, whereas for membrane proteins, a minimum of pseudo MS^4 is required. Here the initial step to remove the detergent micelle, membrane mimetic or lipid vesicle, is formally pseudo (p) (pMS^2) since it proceeds without isolation of a precursor species. For stages greater than pseudo MS^2 the pseudo is replaced by prefix 'p' for brevity, but following pMS^2 all subsequent stages are, formally, pseudo MS^n.



Supplementary Figure 1 - Tuning of Orbitrap Eclipse parameters for native MS – IRM pressure Increasing the IRM pressure improves the transmission of high molecular mass species and appropriate tuning can aid retention of labile ligands. Spectra show the membrane protein trimer AmtB measured at increasing IRM pressure (a-d), indicated right), following in-source removal of the C_8E_4 micelle (in-source activation 150 V). The NL of the maximum tetramer charge state increases ~2 fold from 1.18e5 to 3.15e5 with increasing IRM pressure from 8 mTorr (standard pressure) to 20 mTorr (Intact Protein Mode - High Pressure). The NL of the monomer (likely dissociation product from higher order oligomer due to charge state) is also provided in red. Importantly, retention of a residual DDM adduct (analogous to a labile non-covalent ligand and resulting from incomplete detergent exchange) increases from 15% to 35% height of the AqpZ (18+) as pressure is increased. This indicates that when examining ligand binding, tuning the IRM pressure to maximise both overall signal intensity and intensity of ligand bound ions should be considered. When performing this experiment, it is advised to work in “Intact Protein Mode - High Pressure” with the IRM pressure at 20 mTorr (default instrument setting) for appropriate optics calibration. Higher pressures may increase transmission but are outside of the current optimised calibration window and may therefore have unpredictable adverse effects. This experiment was repeated with at least three other proteins and similar trends observed.



Supplementary Figure 2 - Tuning of Orbitrap Eclipse parameters for native MS – in-source activation (sid) Progressive removal of protective detergent micelles surrounding the AmtB trimer by increasing application of in-source activation energy in the source region of the MS. Increasing in-source activation energy from 65-250 V (a-f) results in gradual desolvation and stripping of detergent. Similar behaviour is observed for desolvation of soluble proteins/assemblies. Tuning the in-source activation energy for maximum signal, the NL level of the 18+ charge state increases with in-source activation of 65-200 V and then decreases at 250 V, presumably due to dissociation of the complex. Importantly, however, all labile detergent molecules (analogous to labile non-covalent ligands) are lost, at much lower energies, between 100-200 V. From 100 V to 150 V all C_8E_4 detergent (red peak) is removed (compare (c) with (d)) followed by DDM (blue peak) between 150 V and 200 V (compare (d) with (e)). (Note, AmtB was purified in the presence of DDM and exchanged into C_8E_4 for native MS). Progressive loss of detergent correlates with their differing binding energies in the gas phase. Similar behaviour is expected from other non-covalent ligands. Crucially when tuning activation energy for desolvation, or micelle removal, care must be taken to preserve desired ligand binding. Importantly, appropriate energies for retention of ligand binding are likely not as high as those that achieve maximum overall signal intensity. This experiment was repeated with at least three other proteins and similar trends observed.

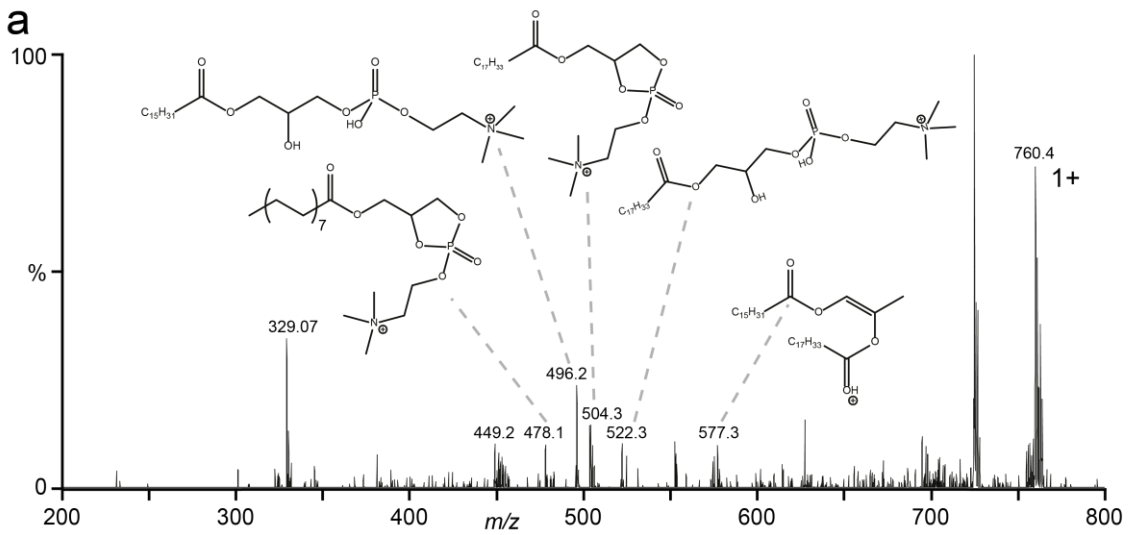


Supplementary Figure 3 - Tuning of Orbitrap Eclipse parameters for native MS – source compensation voltage/rollercoaster

Effect of in-source compensation voltage on spectral intensity and *m/z* dependent transmission. (a) Native MS spectra of ADH at different in-source activation desolvation energies and source compensation voltage (CV). The ADH mass spectra are highlighted with peaks assigned to the tetramer (yellow) and to the monomer (blue). Increasing CV enhances transmission of higher *m/z* species, note the change in monomer tetramer ratio in the spectra. Plots of Normalised Level (NL) of the maximum CS of ADH tetramer against CV, at different in-source activation energies (b, c) show that the NL can be enhanced by up to 2 orders of magnitude by application of CV. For a given *m/z,* the optimal CV (expressed as %) decreases with increasing in-source activation energy. Also for a given *m/z* there exists an optimal combination of in-source activation and CV values for maximum NL, indicated by the green and dark blue features in 3D the maps. In-source activation voltage and CV should therefore be tuned in parallel to optimise signal and intensity of bound ligands.

During optimisation of the Orbitrap Eclipse Tribrid platform for native MS, we found that transmission of high *m/z* species is enhanced by lowering the voltages applied to the API interface i.e., the ion funnel, Injection Flatapole-MP00, and interflatapole lens-L0 (Mcgee, J. P. *et al.* Voltage Rollercoaster Filtering of Low-Mass Contaminants During Native Protein Analysis. *J. Am. Soc. Mass Spectrom.* (2020). doi:10.1021/jasms.9b00037). This presumably reduces the kinetic energy of the ions and enhances their transmission through the MP0 active ion guide. Practically, this compensation voltage (CV) is a user configurable option that adjusts the API offset based upon in-source activation setting. It is expressed as a % and can be applied by switching the toggle in Diagnostics->Ion Optics->In-source compensation voltage and entering the desired percentage. This experiment was repeated with at least three other proteins and similar trends observed.

a

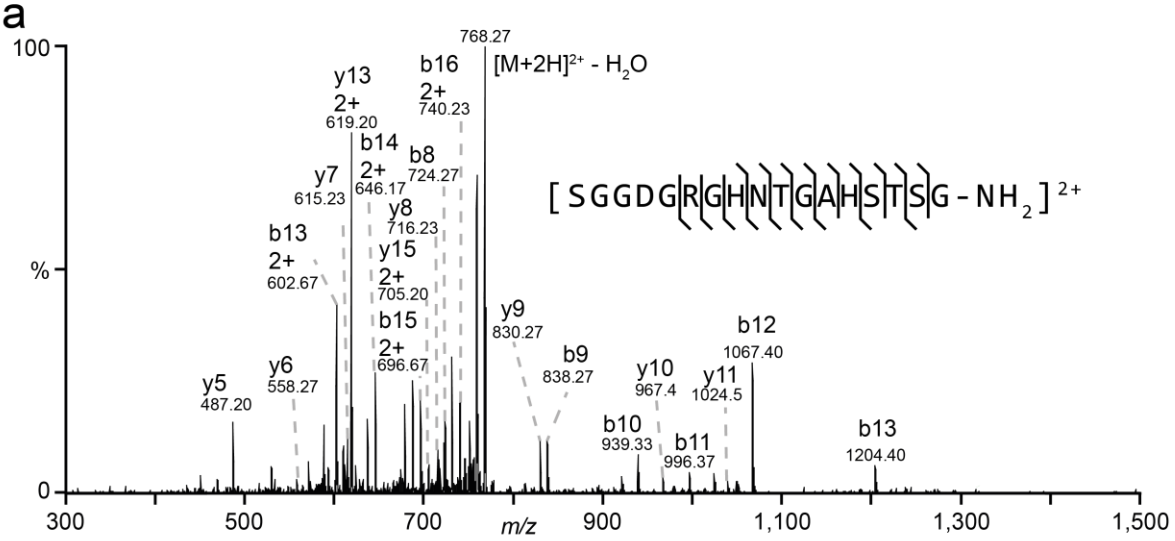


b

[M+H]+ (*m/z*)	*Intensity*	Charge	Theoretical [M+H]+	Error (Da)	Assignment
329.07	15.70	1	-	-	not assigned
381.07	3.45	1	-	-	not assigned
449.00	4.57	1	-	-	not assigned
478.13	4.46	1	478.329	0.20	**-C17H33CO2H loss fatty acid R2 18:1**
496.23	10.70	1	496.340	0.11	**-C16H31CHC=O**
503.00	6.54	1	-	-	not assigned
504.30	6.57	1	504.345	0.05	**-C15H31CO2H loss fatty acid R1 16:0**
522.30	4.61	1	522.355	0.06	**-C14H29CHC=O**
524.93	3.30	1	-	-	not assigned
552.70	4.82	1	-	-	not assigned
577.30	4.41	1	577.519	0.22	**-C5H14NO4P loss PC headgroup**
627.70	7.11	1	-	-	not assigned
695.10	5.35	1	-	-	not assigned
697.37	4.32	1	-	-	not assigned
725.20	45.50	1	-	-	not assigned
760.40	33.70	1	760.590	0.19	**[M+H]+**

Supplementary Figure 4 (a) Expanded, full assignment of MS/MS spectra of POPC lipid released from OmpF trimer and fragmented at the MS^4 stage of the Nativeomics workflow, shown in Extended Data Figure 4. (b) Product ion table with diagnostic ions assigned. Parent ion structure is shown bold.

a

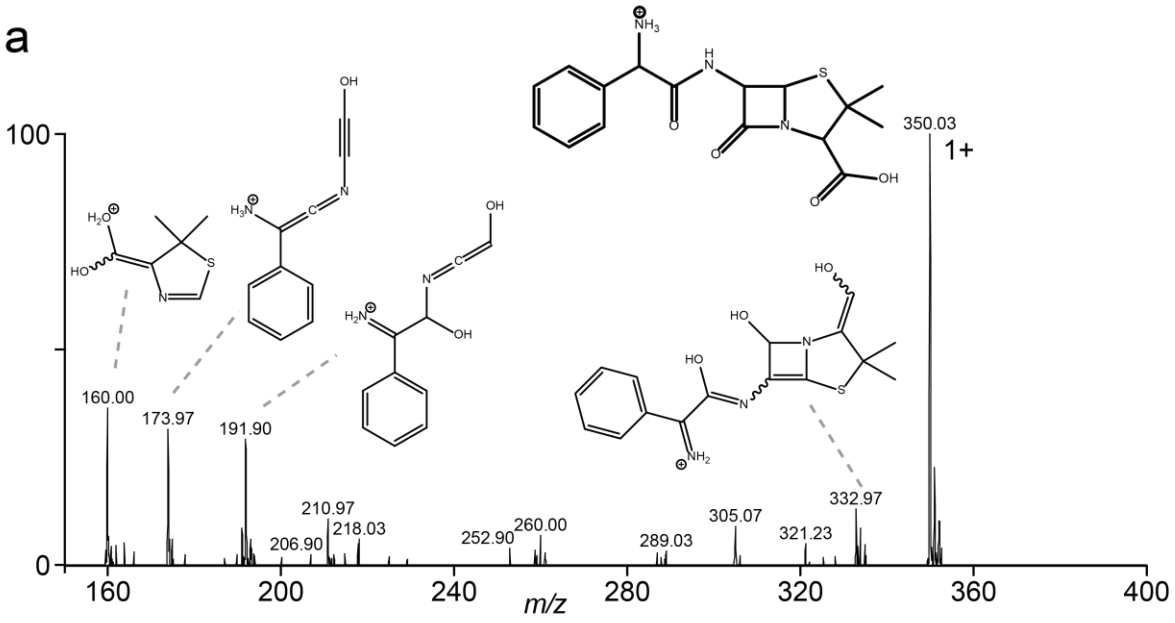


b

m/z	Charge	[M+H]+	Theoretical [M+H]+	Error (Da)	Assignment
487.20	1	487.20	487.226	-0.03	**y5**
558.27	1	558.27	558.263	0.01	**y6**
587.20	1	587.20	587.253	-0.05	**b7**
602.67	2	1204.33	1204.520	-0.19	**b13**
615.23	1	615.23	615.285	-0.05	**y7**
619.20	2	1237.39	1237.578	-0.19	**y13**
646.17	2	1291.33	1291.552	-0.22	**b14**
696.67	2	1392.33	1392.600	-0.27	**b15**
705.20	2	1409.39	1409.627	-0.23	**y15**
716.23	1	716.23	716.332	-0.10	**y8**
724.27	1	724.27	724.312	-0.04	**b8**
740.23	2	1479.45	1479.632	-0.18	**b16**
768.27	2	1535.53	1535.669	-0.14	**M - H2O**
830.27	1	830.27	830.375	-0.11	**y9**
838.27	1	838.27	838.355	-0.09	**b9**
939.33	1	939.33	939.403	-0.07	**b10**
967.40	1	967.40	967.434	-0.03	**y10**
996.37	1	996.37	996.424	-0.05	**b11**
1024.50	1	1024.50	1024.456	0.04	**y11**
1067.40	1	1067.40	1067.461	-0.06	**b12**
1204.40	1	1204.40	1204.520	-0.12	**b13**

Supplementary Figure 5 (a) Expanded assignment of MS^4 spectra of OBS1 peptide released from OmpF trimer and fragmented at the MS^4 stage of the Nativeomics workflow, shown in Extended Data Figure 4. (b) Product ion table and assignment.

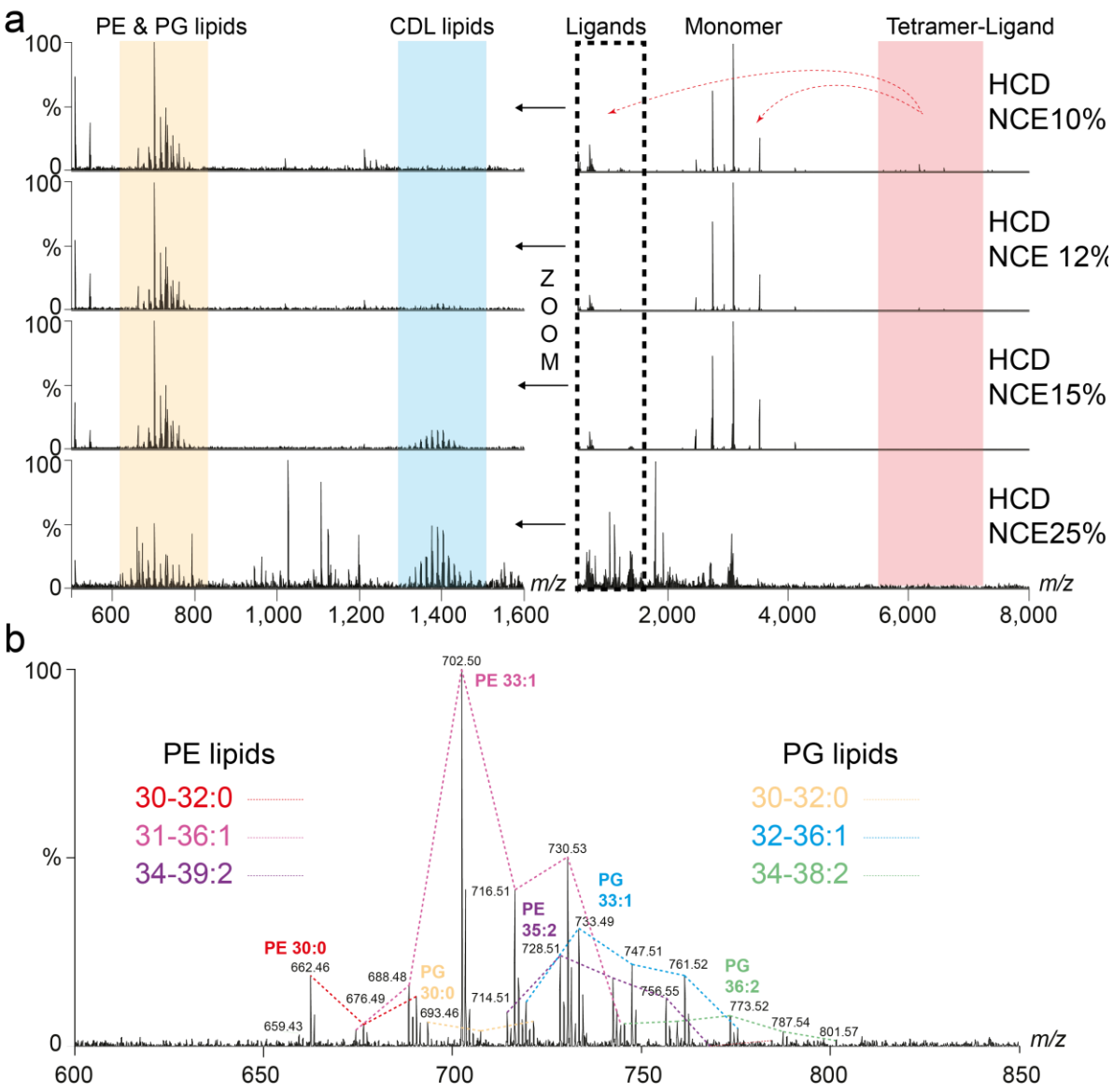
a



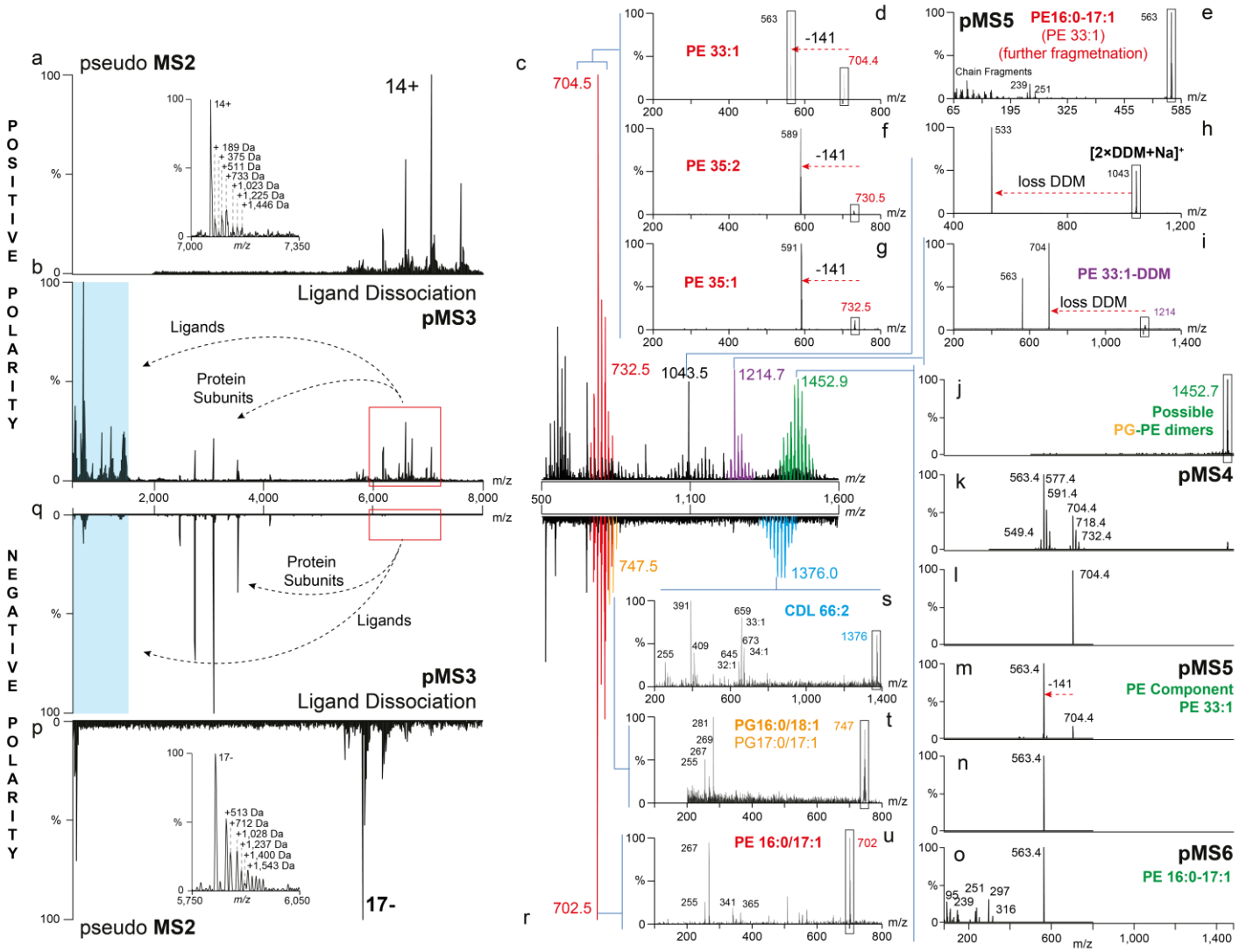
b

m/z	Charge	Theoretical [M+H]+	Error (Da)	Assignment
160.00	1	160.043	-0.04	**see spectrum**
173.97	1	174.056	-0.09	**see spectrum**
191.90	1	192.066	-0.17	**see spectrum**
210.97	1	-	-	**not assigned**
260.00	1	-	-	**not assigned**
305.07	1	305.096	-0.03	**see database**
332.97	1	333.091	-0.12	**see spectrum**
350.03	1	350.120	-0.09	**see spectrum**

Supplementary Figure 6 (a) Expanded assignment of MS^n spectra of ampicillin released from the OmpF trimer (shown in Figure 1b) and fragmented at the MS^4 stage of the Nativeomics workflow. (b) Product ion table and assignment.



Supplementary Figure 7 - **Optimisation of HCD energy for dissociation of lipids bound to AqpZ and identification of lipid class, chain length and degree of unsaturation** (a) Negative polarity in source micelle removal (in-source activation 180 V) and isolation of AqpZ tetramer bound to multiple ligands in the ion trap (centre m/z 6350, width 800) (pink). Activation at increasing HCD energies shows clear dissociation into monomer subunits and an increasing number of signals in the low *m/z* region. Expansion of this region shows distributions of peaks separated by *m/z* 14 appearing between *m/z* 600-800 (gold) and corresponding to PE and PG lipids at lower HCD energies (NCE 10-12%). A second distribution centred around *m/z* 1,4000 (blue), corresponds to cardiolipins, appearing at higher HCD energies (NCE 12-15%). At even higher HCD energies (NCE 25%) the monomeric subunit distribution is greatly decreased in intensity and multiple signals appear in the low m/z region, presumably due to backbone fragmentation of the AqpZ monomers. HCD energy must be carefully tuned to release all families of bound ligands, that may dissociate at different energies, and avoid excessive fragmentation of proteoforms. (b) A zoom of the *m/z* 600-850 region (HCD NCE 15%) reveals multiple distributions that from mass alone can be assigned to PE and PG lipids with varying degrees of unsaturation and acyl chain length. Further fragmentation is required to reveal precise chain length compositions and to localise double bonds. Trends in lipid dissociation are representative of n>10 technical repeats



Supplementary Figure 8 - **Nativeomics performed in the positive and negative ion polarity to identify multiple endogenous lipids bound to the AqpZ tetramer, distinct families of lipids and lipid adducts are shown in different colours**

(a) pseudo MS^2 full scan of AqpZ tetramer in positive ion polarity, following micelle removal in source (in-source activation 200V). Inset shows zoom of the 14+ charge state with adducts bound to the tetrameric assembly annotated with their corresponding mass. (b) pMS^3 Activation of ligand bound AqpZ (HCD NCE 9%) of a wide isolation window centred around *m/z* 7061 and (c) zoom of the low m/z region *m/z* 500-1,500 (shaded light blue in b) shows multiple singly charged ion series containing peaks separated by ~14 Da. These can be assigned to families of phospholipids of different lipid classes by further selection and MS/MS (pMS^4) of individual ions. pMS^4 fragmentation reveals a peak series centred around *m/z* 704.5 (red) corresponding to PE lipids, identified by the characteristic -141 Da head group loss (Pulfer, M. & Murphy, R. C. Electrospray mass spectrometry of phospholipids. *Mass Spectrom. Rev.* **22**, 332–364 (2003)) in the MS/MS spectra (inset). By contrast MS/MS of PE in the negative ion polarity, yields spectra that do not contain peaks that can be assigned to individual acyl chain lengths. For this further rounds of fragmentation are required. (d) Selection and isolation of the *m/z* 704 ion (HCD NCE 15%) exhibits -141 Da PE head group loss. (e) Selection and fragmentation (pMS^5) of the product ion at *m/z* 563 (HCD NCE 30%) produced further predominant ions at *m/z* 239 and 251 corresponding to 16:0 and 17:1 acylium ions respectively. These fragments allow this lipid to be identified as PE 16:0-17:1. (f and g) MS^4 of the ions at m/z 730.5 and 732.5 (HCD NCE 20% and CID NCE 20% respectively,

produced spectra containing daughter ions at -141 Da, confirming the PE lipid class and suggesting different degrees of unsaturation of 35:2 and 35:1.

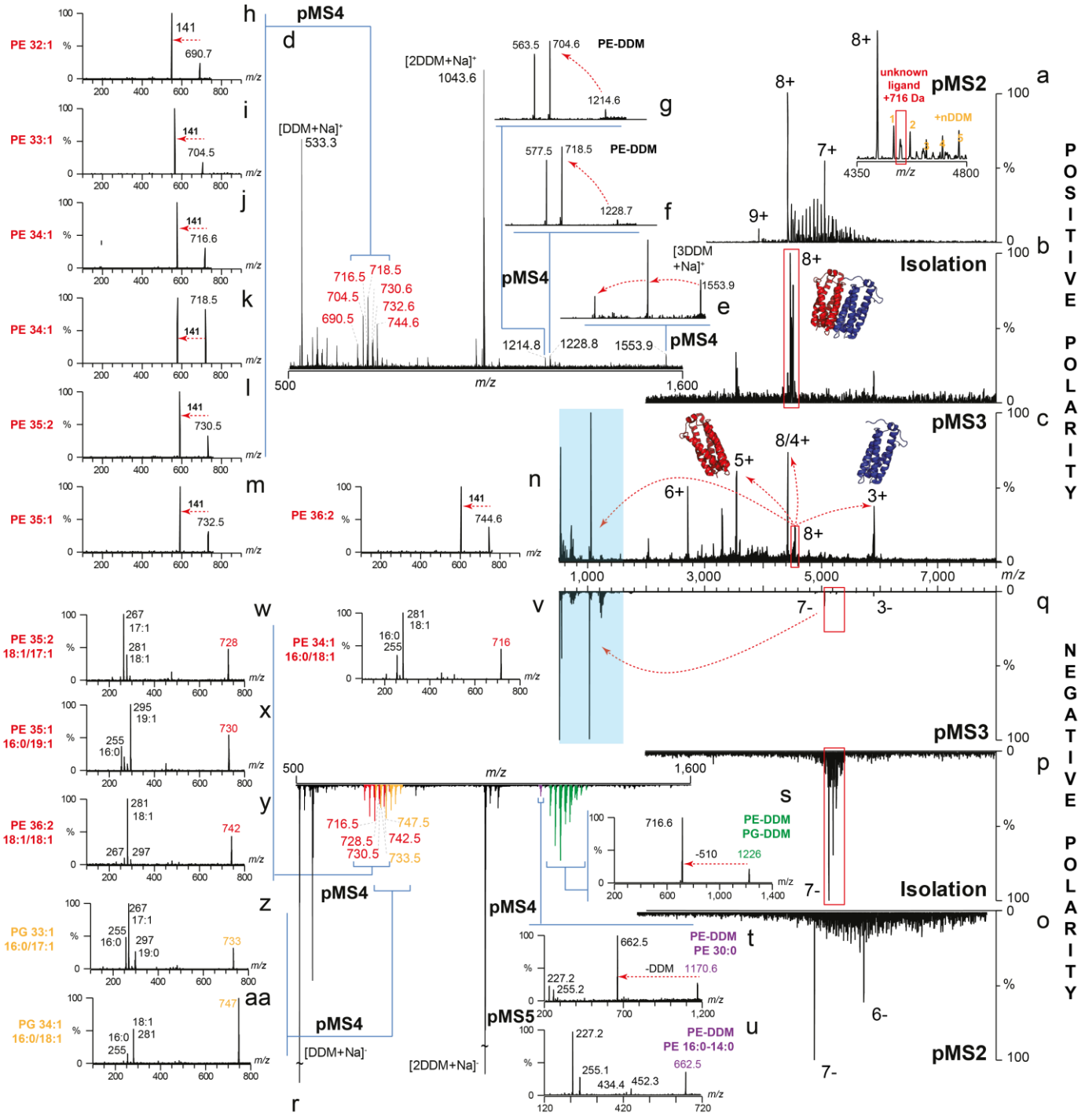
(h) pMS^4 fragmentation of the intense ion at *m/z* 1043 produced a product ion at *m/z* 533 corresponding to neutral loss of DDM detergent (510 Da) implying the structure of the parent ion as a DDM cluster $[2DDM+Na]^+$. (i) pMS4 of the m/z 1214 ion produces ions at m/z 704 indicating neutral loss of DDM (-510 Da) and *m/z* 563 (-141Da) indicative of PE 33:1. This suggests the peak series centred around *m/z* 1214 (purple) corresponds to PE-DDM dimers. Concomitant loss, as a non-covalent dimer pair, suggests PE and DDM bind closely to AqpZ within the proteomicelle.

Isolation (j) and fragmentation (k) of the ion at *m/z* 1452 produced a series of peaks at *m/z* 690, 704, 718 and 732, with corresponding ions at a further -141Da (*m/z* 549, 563, 577, 591) indicating it to be a mixture of isobaric species containing PE lipids with tail lengths 32:0-35:0, rather than one ion with a single composition. Intriguingly the neutral loss required to produce these fragment ions from the *m/z* 1452 parent is 762.5, 748.5, 734.5, 720.5 Da respectively. These fragments correspond to neutral PG lipids with a single unsaturation PG 32:1 to 35:1. Further isolation (l) and fragmentation (m) (HCD NCE 15%) (pMS^5) of the ion at *m/z* 704, produced an intense daughter ion at *m/z* 563 (-141 Da) and confirmed the lipid class as PE. Subsequent isolation of *m/z* 563 (o) and fragmentation (p) (HCD NCE 30%) (pMS^6) produced multiple fragment ions including *m/z* 239 and 251 confirming presence of PE-16:0-17:1. We therefore postulate that the peak series *m/z* 1,350-1,550 (green) corresponds to dimers of PE and PG lipids.

(p) pseudo MS^2 full scan of AqpZ tetramer in negative ion polarity, following micelle removal in source (in-source activation 180V). (q) Isolation of adduct peaks bound to AqpZ tetramer, using a wide isolation window centred around *m/z* 6350, and dissociation (HCD NCE 15%) of the assembly released monomeric subunits and multiple ligands in the low m/z region (blue). (r) Zoom of the *m/z* 500-1,500 region revealing three major distributions with peaks separated by ~14 Da (red, gold and blue). (s) Selection and fragmentation (HCD NCE 28%) of the ion at *m/z* 1376 produced a complex spectrum consistent with cardiolipin (Hsu, F. F. *et al.* Structural characterization of cardiolipin by tandem quadrupole and multiple-stage quadrupole ion-trap mass spectrometry with electrospray ionization. *J. Am. Soc. Mass Spectrom.* **16**, 491–504 (2005)). Fragment ions at *m/z* 645, 659, 673 indicated the presence of diacyl chains of 33:1, 33:2 and 33:3 and suggest that the signal at *m/z* 1376 comprises of at least two discrete 66:4 cardiolipins with both 33:1-33:1 and 32:1-34:1 acyl chain compositions. Based on this assignment we attribute the distribution of ions *m/z* 1,300-1,500 (blue) to numerous cardiolipins with chain compositions 62:2 to 71:3. Note that ligands corresponding to this lipid class were not observed in positive ion polarity.

Another distribution, not present in positive ion polarity, is that centred around the ion at *m/z* 747 (gold). (t) Selection and fragmentation of this (pMS^4) (HCD NCE 30%) revealed a sparse MS/MS spectrum indicative of PG lipids (Hsu, F. F. & Turk, J. Studies on phosphatidylglycerol with triple quadrupole tandem mass spectrometry with electrospray ionization: Fragmentation processes and structural characterization. *J. Am. Soc. Mass Spectrom.* **12**, 1036–1043 (2001)). Ions at *m/z* 255, 267, 269 and 281 confirmed the presence of 16:0 17:0 17:1 and 18:1 acyl chain respectively indicating the presence of both PE 16:0/18:1 and 17:0/17:1. From the intensities of the fragments we tentatively assume PE 16:0/18:1 to be the more abundant, and assign the PE lipids to a distribution (gold) with chain lengths and unsaturation from 30:0 to 38:2.

The masses of ions in the distribution centred around *m/z* 702 (red) correspond directly to those also observed in positive ion mode (red). Selection and fragmentation (u) of the species at *m/z* 702 (pMS^4) (HCD NCE 25%) produced a more informative spectrum than analogous ion at *m/z* 704 in positive ion polarity, both are indicative of PE lipids. Ions at *m/z* 255 and 267 confirmed the acyl chains as 16:0 and 17:1 and the identity of the lipid as predominantly PE 16:0/17:1. Results are representative of two biological repeats.



Supplementary Figure 9 - Nativeomics applied to the mitochondrial transporter TSPO to identify co-purified endogenous ligands in both positive and negative ion polarity

(a) Positive ion polarity pMS^2 native MS spectrum of the TSPO dimer released from detergent micelle (in-source activation 160 V). Multiple adducts remain bound (see inset) including several DDM detergent molecules (yellow) and an intense signal at +716 Da corresponding to unknown ligand(s) (red box). (b) Narrow (set width 50) isolation of 8+ charge state of TSPO dimer harbouring +716 Da adduct at *m/z* 4519. (c) Activation (pMS^3) yields multiple species including dissociated monomer subunits and multiple, singly-charged ions in the low *m/z* range (blue). (d) Zoom on the low *m/z* region showing a peak series centred around the ion at *m/z* 716.5 together with other intense peaks. e) Isolation and dissociation (MS^4) of the ion at

m/z 1553.9 (HCD NCE 12%) reveals two intense product ions at *m/z* 533.2 and 1043.6 which can be assigned as [DDM+Na]+ and [2DDM+Na]+ confirming the identity of the parent ion as [3DDM+Na]+. (f and g) Isolation and MS^4 dissociation (HCD NCE 18%) of ions at *m/z* 1228.8 and 1214.8 produces - 510 Da neutral loss product ions at *m/z* 718 and 704 respectively, with corresponding characteristic further neutral loss of -141 Da (PE headgroup) producing ions at *m/z* 577 and 563. This strongly suggests these species are adducts of PE lipids and DDM detergent. (h to n) Isolation and pMS^4 fragmentation of ions in the series *m/z* 690-744. The formation of an intense product ion, with neutral loss of 141 Da in each case, suggests a homologous series of PE lipids. Chain length and degree of unsaturation (but not acyl chain composition) can be determined from the mass of the ion (labelled in red).

(o) Negative ion polarity pseudo MS^2 native MS spectrum of TSPO dimer released from detergent micelle (in-source activation 200 V). Multiple non-covalent adducts remain bound. (p) Isolation of ligands bound to the 7- charge state (width 300) and (q) pMS^3 dissociation (HCD NCE 12%) produces TSPO monomers, and multiple singly charged ions in the low *m/z* range (blue box). (r) Zoom on the low *m/z* region showing series of ions separated by 14 Da centred around *m/z* 716, 747 and 1226 (red, gold and green respectively). (s) Isolation and pMS^4 fragmentation (HCD NCE 13%) of the ion at *m/z* 1226.8 yields a product with a neutral loss of 510 Da at *m/z* 716 suggesting a PE-DDM adduct. Based on the 14 Da difference between peaks and resemblance of the peak series at *m/z* 1200-1300 to that between *m/z* 700-800, the green peak series has been assigned to a series of PE and PG lipid-DDM adducts. (t) This is supported by the dissociation of the intense species at *m/z* 1171 (HCD NCE 28%) (purple) which produces a product ion at *m/z* 662 (-DDM). (u) Further selection and fragmentation of this ion (HCD NCE 30%) (pMS^5) produced a spectrum indicative of PE 16:0-14:0. The acyl chain lengths could be readily assigned from the intense ions at *m/z* 227 and 255.

(v to y) pMS^4 fragmentation spectra (HCD NCE 30%) of the ions in the even peak series at *m/z* 716, 728, 730, 742 (red) reveal spectra characteristic of PE lipids. By examining the product ions corresponding to acyl chain loss, the composition of the parent lipid could be deduced and the major component is labelled (red). The extent of unsaturation was determined and could be assigned to specific chain lengths compositions. For example fragmentation of the ions at *m/z* 728 and 730, (w) and (x) respectively, show that singly unsaturated PE 35:1 is primarily composed of highly asymmetric PE 16:0/19 (based on fragment intensity). By contrast the doubly unsaturated PE 35:2 is predominantly composed of PE 18:1/17:1. In some cases several, overlapping isobaric species are present. For example, fragmentation of the ion at *m/z* 730 (PE 35:1) shown in (x), yields intense acylium ions corresponding to chains of 16:0, 17:1,18:1 19:1 and also 17:0, 18:0 could be discerned. This is indicative of a mixture of isobaric lipids with the major species (based on ion intensity) being the highly asymmetric lipid PE 16:0/19:1.

(z to aa) pMS^4 fragmentation spectra of ions in the odd peak series at *m/z* 733 and 747 (gold) produced spectra characteristic of PG lipids. Chain length assignment and degree of unsaturation of the major species present is indicated in each case (yellow). Results are representative of two biological repeats.