Insect Biochemistry and Molecular Biology

The structure of the Drosophila melanogaster sex peptide: identification of hydroxylated isoleucine and a strain variation in the pattern of amino acid hydroxylation

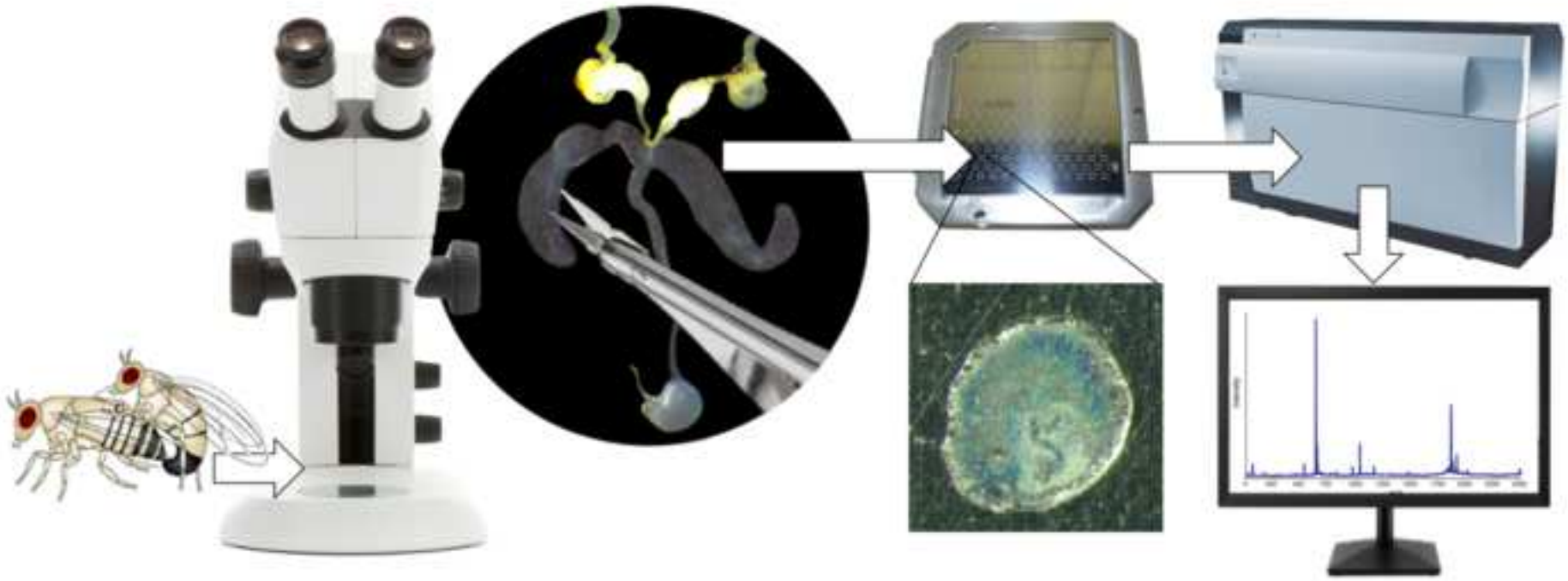
--Manuscript Draft--

Manuscript Number:	IB-D-20-00071R1
Article Type:	VSI: Active peptides in insects:Full Length Article
Keywords:	Sex peptide, Drosophila melanogaster, prolyl hydroxylase, 2-oxoglutarate-dependent oxygenase, 4-hydroxyproline, hydroxyisoleucine
Corresponding Author:	R Elwyn Isaac University of Leeds UNITED KINGDOM
First Author:	Sebastian Sturm
Order of Authors:	Sebastian Sturm
	Adam Dowle
	Neil Audsley
	R Elwyn Isaac
Abstract:	In Drosophila melanogaster mating triggers profound changes in the behaviour and reproductive physiology of the female. Many of these post-mating effects are elicited by sex peptide (SP), a 36-mer pheromone made in the male accessory gland and passed to the female in the seminal fluid. The peptide comprises several structurally and functionally distinct domains, one of which consists of five 4-hydroxyprolines and induces a female immune response. The SP gene predicts an isoleucine (Ile 14) sandwiched between two of the hydroxyprolines of the mature secreted peptide, but the identity of this residue was not established by peptide sequencing and amino acid analysis, presumably because of modification of the side chain. Here we have used matrix-assisted laser desorption ionisation mass spectrometry together with Fourier-transform ion cyclotron resonance mass spectrometry to show that Ile 14 is modified by oxidation of the side chain - a very unusual post-translational modification. Mass spectrometric analysis of glands from different geographical populations of male D. melanogaster show that SP with six hydroxylated side chains is the most common form of the peptide, but that a sub-strain of Canton-S flies held at Leeds only has two or three hydroxylated prolines and an unmodified Ile 14 . The D. melanogaster genome has remarkably nineteen putative hydroxylase genes that are exclusively expressed in the male accessory gland, suggesting that the gland is a powerhouse of protein oxidation. Strain variation in the pattern of sex peptide hydroxylation might be explained by differences in the expression of individual hydroxylase genes.
Response to Reviewers:	

Highlights

- *Drosophila* sex peptide of the male accessory gland controls female reproduction
- The 36-mer sex peptide is modified by oxidation of amino acid side chains
- Sex peptide has five functionally important hydroxyprolines and a rare hydroxyisoleucine
- Fully hydroxylated sex peptide is not found in all *Drosophila* strains
- A remarkable number of hydroxylase genes are expressed only in the male accessory gland

Graphical Abstract



The structure of the *Drosophila melanogaster* sex peptide: identification of hydroxylated isoleucine and a strain variation in the pattern of amino acid hydroxylation

Sebastian Sturm[a], Adam Dowle[b], Neil Audsley[c], R. Elwyn Isaac[a]

[a] Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK.

[b] Bioscience Technology Facility, Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK.

[c] School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon-Tyne, NE1 7RU, UK.

Corresponding author: R. Elwyn Isaac
Email: r.e.isaac@leeds.ac.uk

Abstract

In *Drosophila melanogaster* mating triggers profound changes in the behaviour and reproductive physiology of the female. Many of these post-mating effects are elicited by sex peptide (SP), a 36-mer pheromone made in the male accessory gland and passed to the female in the seminal fluid. The peptide comprises several structurally and functionally distinct domains, one of which consists of five 4-hydroxyprolines and induces a female immune response. The SP gene predicts an isoleucine (Ile^{14}) sandwiched between two of the hydroxyprolines of the mature secreted peptide, but the identity of this residue was not established by peptide sequencing and amino acid analysis, presumably because of modification of the side chain. Here we have used matrix-assisted laser desorption ionisation mass spectrometry together with Fourier-transform ion cyclotron resonance mass spectrometry to show that Ile^{14} is modified by oxidation of the side chain - a very unusual post-translational modification. Mass spectrometric analysis of glands from different geographical populations of male *D. melanogaster* show that SP with six hydroxylated side chains is the most common form of the peptide, but that a sub-strain of Canton-S flies held at Leeds only has two or three hydroxylated prolines and an unmodified Ile^{14}. The *D. melanogaster* genome has remarkably nineteen putative hydroxylase genes that are exclusively expressed in the male accessory gland, suggesting that the gland is a powerhouse of protein oxidation. Strain variation in the pattern of sex peptide hydroxylation might be explained by differences in the expression of individual hydroxylase genes.

Keywords: Sex peptide, *Drosophila melanogaster*, prolyl hydroxylase, 2-oxoglutarate-dependent oxygenase, 4-hydroxyproline, hydroxyisoleucine

Abbreviations: SP, sex peptide; Hyp and P^{OH}, 4-hydroxyproline; JH, juvenile hormone; MAG, male accessory gland; MALDI-TOF, matrix-assisted laser desorption ionisation mass spectrometry; SPITC, 4-sulfophenyl isothiocyanate; FT-ICR-MS, Fourier-transform ion cyclotron resonance mass spectrometry; CID, collision induced dissociation.

1. Introduction

There is increasing recognition of the importance of non-sperm components of male seminal fluid for reproductive success across animal taxa (McGraw et al. 2015; Robertson 2007; Simmons and Fitzpatrick 2012). In addition to its role in providing a supportive milieu for sperm well-being during transfer to the female, seminal fluid can trigger physiological and behavioural responses in the female and affects even the health of offspring (Bromfield et al. 2014; Evans et al. 2019). The plasma is composed of a wide range of chemicals including lipids, carbohydrates, nucleic acids, peptides and proteins, like enzymes, chaperones and structural proteins, most of which are made by accessory glands of the male reproductive tract (Chen 1984; Druart and de Graaf 2018; Fu-Jun and Xiao-Fang 2012; Gillott 2003; Poiani 2006). Knowledge of the physiological roles of the individual components is important for understanding how seminal plasma affects fertility. However, this task is hampered not only by the complexity of the chemical mixture, but also the possibility that the quality and quantity of the components might change depending on environment and social experience (Perry et al. 2013; Wigby et al. 2009).

Proteomic, transcriptomic and genetic approaches using the model insect *Drosophila melanogaster* have made important contributions to our understanding of the role and evolutionary significance of individual protein constituents of the seminal fluid in mating plug formation, sperm storage and utilisation, egg-laying and in changing female behaviour (Avila et al. 2011; Chapman 2008; Chen 1996; Sirot 2019; Sirot et al. 2015; Wolfner 1997). The best known and by far the most studied *Drosophila* seminal fluid component is Sex Peptide (SP), a 36-mer peptide with the amino acid sequence Trp-Glu-Trp-Pro-Trp-Asn-Arg-Lys-Hyp-Thr-Lys-Phe-Hyp-Ile-Hyp-Ser-Hyp-Asn-Hyp-Arg-Asp-Lys-Trp-Cys-Arg-Leu-Asn-Leu-Gly-Pro-Ala-Trp-Gly-Gly-Arg-Cys (Hyp = 4-hydroxyproline) that on transfer to the female reproductive tract inhibits sexual receptivity for up to one week while stimulating female germline proliferation and egg-laying (Kubli 2003; Kubli and Bopp 2012). SP induces a plethora of additional responses in the post-mated female, including stimulation of juvenile hormone (JH) synthesis (Fan et al. 1999; Moshitzky et al. 1996), increased appetite and altered food choice (Carvalho et al. 2006; Ribeiro and Dickson 2010; Vargas et al. 2010; Walker et al. 2015), elevated rate of excretion (Apger-McGlaughon and Wolfner 2013), loss of day-time sleep (Isaac et al. 2010), enhanced female aggression (Bath et al. 2017), release

of stored sperm (Avila et al. 2010), modulation of the innate immune system (Peng et al. 2005b) and enhancement of long-term memory (Scheunemann et al. 2019).

Sex peptide exhibits three structural and functional domains: first an N-terminal Trp-rich anchor region, second a central part rich in hydroxyprolines and third a C-terminal domain with an intramolecular disulphide. The N-terminal anchor region mediates the attachment of the peptide to the sperm tail allowing SP to be sequestered in the female sperm storage organs. Continuous proteolysis of a trypsin-like cleavage site enables the release of the remaining peptide from the sperm storage organs and therefore sustain female responses over a period of several days (Liu and Kubli 2003; Peng et al. 2005a). The Trp-rich anchor is also required to stimulate the biosynthesis of juvenile hormone by the female corpora allata (Kubli 2003). The central region of SP, comprising five hydroxyprolines (Hyp), is responsible for eliciting an early increase in levels of anti-microbial peptide transcripts in post-mated females (Domanitskaya et al. 2007). The C-terminal domain (SP^{21-36}), activates the neuronal G protein-coupled receptor responsible for initiating many of the behavioural post-mating female responses (Kim et al. 2010; Poels et al. 2010; Yapici et al. 2008). Surprisingly the SP-receptor (SPR) is also activated by members of the myoinhibiting peptide (MIP) family with relatively little sequence homology to SP (Kim et al., 2010; Poels et al., 2010). Although the SP can only be found in male drosophilids, the SPR is found in both sexes of representatives from different insect orders. Its non-sex role across insects is presumably as an evolutionary 'ancient' MIP receptor which has been opportunistically recruited for a newer role in reproduction within the *Drosophila* genus. The signalling region of SP assembles a peptide ring structure formed by the thiol groups of two cysteines separated by 11 amino acids. The formation of the intramolecular disulphide is essential for SP receptor binding (Kim et al., 2010; Poels et al., 2010).

The structure of the *D. melanogaster* SP was deduced from sequencing of proteolytic fragments as well as cloning and sequencing of the SP cDNA using an Oregon-R laboratory strain and a natural population collected in Dällikon, Switzerland (Chen et al. 1988). The residue at position 14 however was not identified either by peptide sequencing or amino acid analysis, but the cDNA sequence predicted isoleucine at this position leading to the suggestion of a very unusual post-translational modification of its side chain, such as hydroxylation. We provide, for the first time, mass spectrometric evidence that the isoleucine of SP is indeed hydroxylated. We believe this is the first report of an oxidative post-translational modification of isoleucine in an animal peptide or protein. A survey of male

accessory glands (MAGs) from diverse geographical sources indicates that SP with 6 hydroxylated residues is the naturally common form of the peptide. We do however show that the number of hydroxylated residues is reduced to two and three in a sub-strain of the standard laboratory wild-type Canton-S stock.

2. Materials and methods

2.1. Insects

D. melanogaster flies were maintained on oatmeal/molasses/agar medium at 25ºC in a 12:12 light-dark cycle. A sub-strain of Canton-S (aka CS-Leeds) fly strain, originally from the Department of Biology, University of York, U.K., was maintained in Leeds for over 10 years. Canton-S (CS) was from Bloomington *Drosophila* Stock Center (BDSC 64349). Other BDSC stocks were Athens, Greece (BDSC 3876), Samarkand, Uzbekistan (BDSC 4270), Sengawa Wildlife Reserve (Zimbabwe; BDSC 60741), Bermuda (BDSC 3840), Bogota, Columbia (BDSC 3843), Koriba Dam, South Africa (BDSC 3853), New Orleans, Louisiana, USA (BDSC 3860) and Berlin-K (BDSC 8520). Oregon-R (OR) stocks were obtained from the Department of Genetics, University of Cambridge (U.K.). A strain collected from Dahomey (Benin) was provided by Professor T. Chapman (University of East Anglia, U.K.). Males of SP null mutants were generated as described previously (Bath et al. 2017) from stocks provided by using flies provided by Dr S. Wigby (Department of Zoology, University of Oxford, Oxford, U.K.).

2.2. Peptides

Synthetic SP with five 4-hydroxyproline (P^{OH}) residues WEWPWNRKP^{OH}TKFP^{OH}IP^{OH}SP^{OH}NP^{OH}RDKWCRLNLGPAWGGRC was a gift from Professor Young-Joon Kim, Department of Life Science, Gwangju Institute of Science and Technology, Gwangju 500-712, Republic of Korea.

2.3 Preparation of MAG samples for mass spectrometry

Pairs of MAGs were dissected in Dulbecco's Phosphate Buffered Saline (Sigma-Aldrich Company Ltd., Gillingham, U.K.) from 3-8-day-old unmated males and placed either individually or in groups of 5 pairs into 10 µl extraction buffer (87% methanol, 5% glacial acetic acid). Tissue collections were sonicated in a water bath and subsequently centrifuged for 20 minutes at 12,000 rpm and 4°C. Supernatants were transferred into clean vials and stored at -20ºC until required for analysis.

2.4. Purification of SP from MAGs

MAGs from 250 males (Dahomey strain) that had been separated from females for at least 3 days were dissected, transferred in 200 µl extraction buffer and processed as described above. The supernatant was evaporated in a vacuum concentrator and diluted 20-fold with 0.1 % trifluoroacetic acid (TFA) for high performance liquid chromatography using a Beckman System gold chromatography system (Beckman Coulter U.K. Ltd). The diluted sample was loaded onto a Jupiter C_{18} 10µm 300Å reversed-phase column (250 x 2.1 mm i.d.; Phenomenex, Macclesfield, U.K.). The column was eluted with a linear gradient of 5-60 % acetonitrile/0.1% TFA over 55 min at a flow rate of 200 µl/min, and elution monitored at 214 nm. Fractions of 200 µl were collected and analysed by MALDI-TOF mass spectrometry to establish the elution of SP. The HPLC fraction containing SP was concentrated by centrifugal evaporation to remove acetonitrile and TFA.

2.5. Reduction, alkylation, tryptic digestion, SPITC derivatisation and purification of peptides

Peptides were subjected to cystine reduction by dithiothreitol (Sigma-Aldrich Company Ltd.) and alkylation by iodoacetic acid (Sigma-Aldrich Company Ltd.) followed by enzymatic digestion using trypsin (Sequencing Grade Modified Trypsin, Promega U.K. Ltd., Southampton, U.K.) as described by (Sturm and Predel 2015). For *de novo* sequencing, peptides were sulfonated using 4-sulfophenyl isothiocyanate (Sigma-Aldrich Company Ltd.) as described by (Sturm et al. 2016). The N-terminal sulfonation facilitates sequencing due to the generation of abundant fragments of y-ion series.

2.6. Matrix-assisted laser desorption ionisation mass spectrometry (MALDI- TOF MS)

Mass spectra were acquired in positive linear or reflector mode using either a Voyager DE STR matrix-assisted laser desorption ionisation mass spectrometer (Applied Biosystems, Warrington, UK) or a Bruker Ultraflex mass spectrometer (Bruker Daltonic GmbH, Bremen, Germany). Samples were mixed with equal volumes of either α-Cyano-4-hydroxycinnamic acid (10 mg/ml in 70% acetonitrile, 0.1% TFA) or 2,5-Dihydroxybenzoic acid (10 mg/ml in 20% acetonitrile, 1% formic acid) on the MALDI sample plate and allowed to air-dry. External calibration was conducted using a calibration mixture containing des-Arg-bradykinin, angiotensin1, Glu-fibrinopeptide B, and neurotensin (Applied Biosystems) or angiotensin I, angiotensin II, substance P, bombesin, ACTH clip$^{1–17}$, ACTH clip$^{18–39}$ and

somatostatin 28 (Bruker Daltonics). Masses are shown as average masses $[M+H]^+$ for analyses performed in linear mode and monoisotopic masses for analyses in reflector mode.

2.6. Fourier Transform Ion Cyclotron Resonance mass spectrometry (FT-ICR-MS)

FT-ICR-MS was performed using a solariXR FT-MS (Bruker Daltonics) with a 9.4 T superconducting magnet. Peptide solutions were diluted 1 to 20 into 50% aqueous acetonitrile containing 1% (*v*/v) formic acid before introduction by TriVersa NanoMate (Advion BioSciences, Ithaca, NY) in positive-ion mode. The applied voltage was adjusted between 1.4-1.7 kV to achieve a stable ion current. A 120°C nitrogen dry gas was supplied at 1.3 L/min to aid desolvation. Instrument control and data acquisition used Compass 1.4 (Bruker Daltonics). Spectra were generated by the accumulation of 20 scans with 0.2 s ion cooling time and 0.5 s scan time with 400K data points recorded. Peptide precursors were manually selected for isolation and subsequent fragmentation by collision induced dissociation in the hexapole (Q-CID) with argon as the collision gas. Collision energies were optimized for each peptide. Spectra were processed using DataAnalysis version 4.0 (Bruker Daltonics). Mass deconvolution was performed using version 2.0 of the SNAP averaging algorithm (C 4.9384 %, N 1.3577 %, O 1.4773 %, S 0.0417 %, H 7.7583 %). Measured fragment ions were manually compared to theoretical b- and y-ions considering charge states between 1+ and 5+.

3. Results

3.1. The molecular ion of biological SP indicates the presence of six hydroxylated amino acid residues

In an initial experiment we investigated the seminal fluid peptides of *D. melanogaster* using the Dahomey strain which has been used extensively to investigate the role of SP in manipulating female reproductive biology and behaviour (Bath et al. 2017; Chapman et al. 2003; Fricke et al. ; Sepil et al. 2019). Mass spectra of MAG extracts of Dahomey flies revealed a dominant ion signal at *m/z* 4442.7 ($[M+H]^+$, average mass) (Fig. 1A). In support of the identification of this ion signal we subjected extracts of MAGs from males homozygous for a null allele of the SP gene to mass analysis (Fig. 1B). As expected, an ion signal corresponding to SP could not be detected above background noise of the spectrum. The molecular ion present in the Dahomey MAG extract indicated the presence of six hydroxylations/oxidations based on the molecular mass deduced from the SP sequence predicted from cDNA. Five of the hydroxylations can be assigned based on previously

identified Hyp residues leaving one uncertain hydroxylation site which has been proposed previously to be at Ile^{14} (Chen et al. 1988; Kubli 2003).

3.2. Mass spectrometry of tryptic fragments of SP

To gain further structural information, SP was extracted from Dahomey MAGs and purified by reversed phase HPLC. Both the biological form and synthetic SP that included five 4-hydroxyprolines were subjected to reduction of the disulphide bond followed by alkylation and tryptic digestion. A comparison of the mass spectra of the tryptic digests showed that the ion signal at *m/z* 1559.1 attributed to SP^{8-20} of the synthetic SP (Fig. 2A) had shifted to *m/z* 1574.8, showing that the additional sixteen mass units of the biological peptide resided in this tryptic fragment (Fig. 2B). Of the thirteen amino acids in this fragment, only Ile^{14} had previously resisted identification by peptide sequencing of the SP from wild-type strains of *D. melanogaster* (Chen et al. 1988). Our interpretation of the MS data is that the isoleucine side chain of the natural SP is oxidised in addition to the five hydroxyproline residues of SP^{8-20}.

Corroboration of the hydroxylated nature of the isoleucine was sought using FT-ICR-MS coupled with a nano-electrospray ionisation source, a technique enabling conclusions of the elemental composition by its high resolution and precision. Analysis of the Dahomey SP with carbamidomethylated cysteines yielded an abundant ion at *m/z* 912.0466 with the charge state 5+. This measurement equates to a neutral mass of 4555.1967 Da, which is in accordance with a theoretical mass of 4556.2040 Da assuming a protonated and alkylated peptide with six hydroxylated residues of the proposed sequence $WEWPWNRKP^{OH}TKFP^{OH}I^{OH}P^{OH}SP^{OH}NP^{OH}RDKWC*RLNLGPAWGGRC*$ (*, carbamidomethylated Cys). To confirm the internal sequence and positions of oxidation/hydroxylation, the precursor ion was isolated and fragmented by collision induced dissociation (CID) producing the multiple-charge product ion spectrum in Fig. 3. Detected product ions were in accordance to predicted b- and y-type fragment ions and exhibited extensive internal sequence coverage (Table S1). The fragmentation pattern facilitated unambiguous assignment for all positions of oxidation/hydroxylation to individual amino acids, with the singular exception of oxidation at Pro^{19}. Crucially, localisation of hydroxylation at Ile^{14} was possible through individual ion transitions of y_{22} and y_{23}, residing in a continuous string of y-ion observations from y_{20} to y_{25}.

3.3. SP with 6 hydroxylated amino acids is a common structural form in different D. melanogaster populations, but SP from a Canton-S sub-strain (CS-Leeds) has only either two or three oxidised side chains.

MALDI-TOF MS analysis of SP extracted from MAGs of eight *D. melanogaster* populations from different geographical regions (see Materials and methods for details) and from standard laboratory wild-type colonies (Oregon-R and Canton-S) revealed dominant ion signals corresponding to SP with six hydroxylations as found in Dahomey males. The exception was the MAG spectrum from a Canton-S sub-strain (CS-Leeds) that had two prominent molecular ions at *m/z* 4378.6 and *m/z* 4394.5 that could be assigned to SP possessing only two and three hydroxylations. An ion signal corresponding to SP with six hydroxylations was absent from these spectra (Fig. 4). This pattern was invariable in spectra of MAG extracts from CS-Leeds flies taken at various times over a 7-year period. To identify which amino acid positions are unmodified in the CS-Leeds SP, we analysed tryptic peptides which were N-terminally sulfonated to facilitate sequencing. Fragment spectra obtained from SP^{12-20} revealed that CS-Leeds SP is lacking hydroxylation at Pro^{13}, Ile^{14} and Pro^{19} (Fig. 5). The fragments SP^{1-7} and SP^{26-35} did not contain any hydroxylations, as expected. In this experiment we could not determine the indicated third partial hydroxylation site since we could not detect ion signals corresponding to a single hydroxylated fragment SP^{8-11} or a triple hydroxylated SP^{12-20}.

4. Discussion

In the seminal paper of Chen *et al.* reporting the isolation and structural characterisation of *D. melanogaster* SP, the cDNA sequence predicted the presence of isoleucine at position 14, but the identity could not be confirmed by peptide sequencing and amino acid analysis, suggesting that this residue is subject to post-translational modification (Chen et al. 1988). The authors suggested that the isoleucine side chain might be hydroxylated as are the proline residues either side of the isoleucine. We have now used mass spectrometry to confirm that Ile^{14} of SP is indeed hydroxylated, which brings the total number of amino acids that have been post-translationally modified by oxidation to six. Our mass spectrometric screening of MAGs from various locations indicates that the hydroxylation of all six sites is complete in all but one investigated population. This suggests that the six-fold hydroxylated SP is the most common structural form. In the CS-Leeds strain, however, mass spectra revealed molecular ions corresponding to SP variants with a lower number of hydroxylation sites

indicating the presence of only two and three hydroxyprolyl residues and the absence of an hydroxyisoleucine.

What might be the structural and functional significance of the SP hydroxyisoleucine? Structurally, the presence of the hydroxyl is expected to have profound implications due to the introduction of a strongly electronegative oxygen atom, which can serve as a hydrogen bond donor and acceptor, to the isoleucine hydrophobic side chain. The functional significance of the unusual hydroxyisoleucine and our reporting of a strain difference in the pattern of SP side chain oxidation is open to speculation. Two studies have shown that the introduction of an N-glycosylation site or green fluorescent protein into the hydroxyproline motif of SP inhibits the transient induction of anti-microbial peptides, indicating that this middle region is responsible for a post-mating innate immune response in the female genital tract (Domanitskaya et al. 2007; Tsuda et al. 2015). These observations have led to speculation that the hydroxyl groups of the proline side chains in this region might mimic the structure of bacterial cell wall peptidoglycans that elicit anti-microbial peptide synthesis in *D. melanogaster* (Domanitskaya et al. 2007). At present we do not know whether CS-Leeds males with a reduced number of hydroxyproline residues and a non-oxidised isoleucine can also induce anti-microbial peptide expression in post-mated females. The fact that many of the female post-mating responses can be attributed to the C-terminal domain that includes a pair of tryptophan residues separated by eight amino acids, strongly suggests that any variation in the number of hydroxylated amino acid side chains outside of this region is unlikely to have a significant impact on SP receptor activation, however, we cannot dismiss possible effects on the way that the SP molecule interacts with other proteins, such as the protein network required for SP binding to sperm (Singh et al. 2018) and the protease required for subsequent SP release from the tryptophan-rich anchor (SP^{1-5}) (Peng et al. 2005a).

The oxygenases responsible for protein hydroxylation mainly belong to the Fe^{2+} and 2-oxoglutarate-dependent family that couple the oxidative decarboxylation of 2-oxoglutarate (2OG) to the hydroxylation of a peptide co-substrate (Hausinger 2004; Loenarz and Schofield 2011; Markolovic et al. 2015). The hydroxylation of amino acid (proline, aspartic acid, asparagine and lysine) side chains of proteins by 2OG-oxygenases is not uncommon, but the hydroxylation of isoleucine of proteins and peptides is extremely rare. It has been shown, however, that the factor inhibiting hypoxia-inducible factor (FIH), a mammalian 2OG-oxygenase responsible for the hydroxylation of an asparagine in the transcriptional activation

domain of hypoxia-inducible factor (HIF-1α) and an aspartic acid in a human ankyrin, can hydroxylate peptidyl-isoleucine in a 20-mer ankyrin peptide substrate under standard *in vitro* reaction conditions, albeit at a reduced rate compared to aspartic acid (Yang et al. 2013). Furthermore 4-hydroxyisoleucine is found in some of the fungal *Amanita* cyclic peptide toxins (Walton 2018) and has been isolated as a nonproteogenic amino acid with glucose-dependent insulinotropic activity from seeds of the legume *Trigonella foenum-graecum* (Zafar and Gao 2016).

The *D. melanogaster* genome contains 29 genes coding for putative proline-hydroxylases (Flybase annotation; Marygold et al. 2016) and of these 19 are exclusively expressed in the adult MAG (FlyAtlas2; Leader et al. 2018). These enzymes might have different substrate specificities defined by the positional context of the amino acid in the peptide. With the exception of CS-Leeds, we could not confirm any mass shifts of multiples of 16 Da corresponding to SP. This indicates, that in general the modification of the target sites (Pro^{9}, Pro^{13}, Ile^{14}, Pro^{15}, Pro^{17}, Pro^{19}) is complete and that the hydroxylation of Pro^{4} and Pro^{30} is undetectable because of limited substrate specificity of the MAG oxidases. We speculate that the absence of some hydroxylated residues in CS-Leeds SP compared to the SP from other strains reflects the deficiency of one or more substrate-specific enzymes rather than a general across-the-board reduction of hydroxylase activity.

Typically, the oxygenases responsible for protein hydroxylation require Fe^{2+} for activity and 2OG and O_2 as obligatory co-substrates and are strongly inhibited by succinate and fumarate (Islam et al. 2018). The catalytic mechanism is highly conserved and involves the oxidative decarboxylation of 2OG to succinate and the formation of a high valent ferryl intermediate, which oxidises the protein substrate (Salminen et al. 2015). It follows that post-translational modification pathways involving these enzymes have the potential to be regulated by the availability of O_2 and 2OG, as well as other tricarboxylic acid cycle intermediates, and Fe^{2+} redox status of the MAG, resulting in changes to the composition and structure of MAG products (Salminen et al. 2015). Such systems might serve as environmental sensors and potential mechanisms for the male to respond to changes in nutritional status and levels of oxidative stress by altering the structure and, possibly, the physiological activity of seminal fluid proteins and peptides.

Declaration of competing interest

The authors have no conflicts of interest.

Authors' contributions

All authors contributed to the laboratory work and participated in the design of the study and in writing the manuscript. All authors gave final approval for publication.

Acknowledgements

The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1). We thank Chris Schofield (University of Oxford) for valuable discussions during the course of this work and Divya Ramesh (University of Konstanz) for comments on the manuscript. We also thank those who provided fly strains as listed in the materials and methods section.

References

Apger-McGlaughon, J and Wolfner, MF (2013) Post-mating change in excretion by mated *Drosophila melanogaster* females is a long-term response that depends on sex peptide and sperm. *J Insect Physiol* **59**: 1024-30.

Avila, FW, Ravi Ram, K, Bloch Qazi, MC and Wolfner, MF (2010) Sex peptide is required for the efficient release of stored sperm in mated *Drosophila* females. *Genetics* **186**: 595-600.

Avila, FW, Sirot, LK, LaFlamme, BA, Rubinstein, CD and Wolfner, MF (2011) Insect seminal fluid proteins: identification and function. *Annu Rev Entomol* **56**: 21-40.

Bath, E, Bowden, S, Peters, C, Reddy, A, Tobias, JA, Easton-Calabria, E, Seddon, N, Goodwin, SF and Wigby, S (2017) Sperm and sex peptide stimulate aggression in female *Drosophila*. *Nat Ecol Evol* **1**: 0154.

Bromfield, JJ, Schjenken, JE, Chin, PY, Care, AS, Jasper, MJ and Robertson, SA (2014) Maternal tract factors contribute to paternal seminal fluid impact on metabolic phenotype in offspring. *Proc Natl Acad Sci U S A* **111**: 2200-5.

Carvalho, GB, Kapahi, P, Anderson, DJ and Benzer, S (2006) Allocrine modulation of feeding behavior by the Sex Peptide of *Drosophila*. *Curr Biol* **16**: 692-6.

Chapman, T (2008) The soup in my fly: evolution, form and function of seminal fluid proteins. *PLoS Biol* **6**: e179.

Chapman, T, Bangham, J, Vinti, G, Seifried, B, Lung, O, Wolfner, MF, Smith, HK and Partridge, L (2003) The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. *Proc Natl Acad Sci U S A* **100**: 9923-8.

Chen, PS (1984) The Functional-Morphology and Biochemistry of Insect Male Accessory-Glands and Their Secretions. *Annual Review of Entomology* **29**: 233-255.

Chen, PS (1996) The accessory gland proteins in male *Drosophila*: structural, reproductive, and evolutionary aspects. *Experientia* **52**: 503-10.

Chen, PS, Stumm-Zollinger, E, Aigaki, T, Balmer, J, Bienz, M and Bohlen, P (1988) A male accessory gland peptide that regulates reproductive behavior of female *D. melanogaster*. *Cell* **54**: 291-8.

Domanitskaya, EV, Liu, H, Chen, S and Kubli, E (2007) The hydroxyproline motif of male sex peptide elicits the innate immune response in *Drosophila* females. *FEBS J* **274**: 5659-68.

Druart, X and de Graaf, S (2018) Seminal plasma proteomes and sperm fertility. *Anim Reprod Sci* **194**: 33-40.

Evans, JP, Wilson, AJ, Pilastro, A and Garcia-Gonzalez, F (2019) Ejaculate-mediated paternal effects: evidence, mechanisms and evolutionary implications. *Reproduction* **157**: R109-R126.

Fan, Y, Rafaeli, A, Gileadi, C, Kubli, E and Applebaum, SW (1999) *Drosophila melanogaster* sex peptide stimulates juvenile hormone synthesis and depresses sex pheromone production in Helicoverpa armigera. *J Insect Physiol* **45**: 127-133.

Fricke, C, Bretman, A and Chapman, T (2008) Adult male nutrition and reproductive success in *Drosophila melanogaster*. *Evolution* **62**: 3170-7.

Fu-Jun, L and Xiao-Fang, S (2012) Comparative analysis of human reproductive proteomes identifies candidate proteins of sperm maturation. *Mol Biol Rep* **39**: 10257-63.

Gillott, C (2003) Male accessory gland secretions: modulators of female reproductive physiology and behavior. *Annu Rev Entomol* **48**: 163-84.

Hausinger, RP (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* **39**: 21-68.

Isaac, RE, Li, C, Leedale, AE and Shirras, AD (2010) *Drosophila* male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. *Proc Biol Sci* **277**: 65-70.

Islam, MS, Leissing, TM, Chowdhury, R, Hopkinson, RJ and Schofield, CJ (2018) 2-Oxoglutarate-Dependent Oxygenases. *Annu Rev Biochem* **87**: 585-620.

Kim, YJ, Bartalska, K, Audsley, N, Yamanaka, N, Yapici, N, Lee, JY, Kim, YC, Markovic, M, Isaac, E, Tanaka, Y and Dickson, BJ (2010) MIPs are ancestral ligands for the sex peptide receptor. *Proc Natl Acad Sci U S A* **107**: 6520-5.

Kubli, E (2003) Sex-peptides: seminal peptides of the *Drosophila* male. *Cell Mol Life Sci* **60**: 1689-704.

Kubli, E and Bopp, D (2012) Sexual behavior: how Sex Peptide flips the postmating switch of female flies. *Curr Biol* **22**: R520-2.

Leader, DP, Krause, SA, Pandit, A, Davies, SA and Dow, JAT (2018) FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res* **46**: D809-D815.

Liu, H and Kubli, E (2003) Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **100**: 9929-33.

Loenarz, C and Schofield, CJ (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* **36**: 7-18.

Markolovic, S, Wilkins, SE and Schofield, CJ (2015) Protein Hydroxylation Catalyzed by 2-Oxoglutarate-dependent Oxygenases. *J Biol Chem* **290**: 20712-22.

Marygold, SJ, Antonazzo, G, Attrill, H, Costa, M, Crosby, MA, Dos Santos, G, Goodman, JL, Gramates, LS, Matthews, BB, Rey, AJ, Thurmond, J and FlyBase, C (2016) Exploring FlyBase Data Using QuickSearch. *Curr Protoc Bioinformatics* **56**: 1 31 1-1 31 23.

McGraw, LA, Suarez, SS and Wolfner, MF (2015) On a matter of seminal importance. *Bioessays* **37**: 142-7.

418 Moshitzky, P, Fleischmann, I, Chaimov, N, Saudan, P, Klauser, S, Kubli, E and Applebaum,
419 SW (1996) Sex-peptide activates juvenile hormone biosynthesis in the *Drosophila*
420 *melanogaster* corpus allatum. *Arch Insect Biochem Physiol* **32**: 363-74.
421 Peng, J, Chen, S, Busser, S, Liu, H, Honegger, T and Kubli, E (2005a) Gradual release of
422 sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Curr*
423 *Biol* **15**: 207-13.
424 Peng, J, Zipperlen, P and Kubli, E (2005b) *Drosophila* sex-peptide stimulates female innate
425 immune system after mating via the Toll and Imd pathways. *Curr Biol* **15**: 1690-4.
426 Perry, JC, Sirot, L and Wigby, S (2013) The seminal symphony: how to compose an
427 ejaculate. *Trends Ecol Evol* **28**: 414-22.
428 Poels, J, Van Loy, T, Vandersmissen, HP, Van Hiel, B, Van Soest, S, Nachman, RJ and
429 Vanden Broeck, J (2010) Myoinhibiting peptides are the ancestral ligands of the
430 promiscuous *Drosophila* sex peptide receptor. *Cell Mol Life Sci* **67**: 3511-22.
431 Poiani, A (2006) Complexity of seminal fluid: a review. *Behavioral Ecology and*
432 *Sociobiology* **60**: 289-310.
433 Ribeiro, C and Dickson, BJ (2010) Sex peptide receptor and neuronal TOR/S6K signaling
434 modulate nutrient balancing in *Drosophila*. *Curr Biol* **20**: 1000-5.
435 Robertson, SA (2007) Seminal fluid signaling in the female reproductive tract: lessons from
436 rodents and pigs. *J Anim Sci* **85**: E36-44.
437 Salminen, A, Kauppinen, A and Kaarniranta, K (2015) 2-Oxoglutarate-dependent
438 dioxygenases are sensors of energy metabolism, oxygen availability, and iron
439 homeostasis: potential role in the regulation of aging process. *Cell Mol Life Sci* **72**:
440 3897-914.
441 Scheunemann, L, Lampin-Saint-Amaux, A, Schor, J and Preat, T (2019) A sperm peptide
442 enhances long-term memory in female *Drosophila*. *Science Advances* **5**: , eaax3432.
443 Sepil, I, Hopkins, BR, Dean, R, Thezenas, ML, Charles, PD, Konietzny, R, Fischer, R,
444 Kessler, BM and Wigby, S (2019) Quantitative Proteomics Identification of Seminal
445 Fluid Proteins in Male *Drosophila melanogaster*. *Mol Cell Proteomics* **18**: S46-S58.
446 Simmons, LW and Fitzpatrick, JL (2012) Sperm wars and the evolution of male fertility.
447 *Reproduction* **144**: 519-34.
448 Singh, A, Buehner, NA, Lin, H, Baranowski, KJ, Findlay, GD and Wolfner, MF (2018)
449 Long-term interaction between *Drosophila* sperm and sex peptide is mediated by
450 other seminal proteins that bind only transiently to sperm. *Insect Biochem Mol Biol*
451 **102**: 43-51.
452 Sirot, LK (2019) On the evolutionary origins of insect seminal fluid proteins. *Gen Comp*
453 *Endocrinol* **278**: 104-111.
454 Sirot, LK, Wong, A, Chapman, T and Wolfner, MF (2015) Sexual Conflict and Seminal Fluid
455 Proteins: A Dynamic Landscape of Sexual Interactions. *Cold Spring Harbor*
456 *Perspectives in Biology* **7**.
457 Sturm, S and Predel, R (2015) Mass spectrometric identification, sequence evolution, and
458 intraspecific variability of dimeric peptides encoded by cockroach *akh* genes. *Anal*
459 *Bioanal Chem* **407**: 1685-93.
460 Sturm, S, Ramesh, D, Brockmann, A, Neupert, S and Predel, R (2016) Agatoxin-like peptides
461 in the neuroendocrine system of the honey bee and other insects. *J Proteomics* **132**:
462 77-84.
463 Tsuda, M, Peyre, JB, Asano, T and Aigaki, T (2015) Visualizing Molecular Functions and
464 Cross-Species Activity of Sex-Peptide in *Drosophila*. *Genetics* **200**: 1161-9.
465 Vargas, MA, Luo, N, Yamaguchi, A and Kapahi, P (2010) A role for S6 kinase and serotonin
466 in postmating dietary switch and balance of nutrients in *D. melanogaster*. *Curr Biol*
467 **20**: 1006-11.

468 Walker, SJ, Corrales-Carvajal, VM and Ribeiro, C (2015) Postmating Circuitry Modulates
469 Salt Taste Processing to Increase Reproductive Output in *Drosophila*. *Curr Biol* **25**:
470 2621-30.
471 Walton, J (2018) Chemistry of the *Amanita* Peptide Toxins. In: *The Cyclic Peptide Toxins of*
472 *Amanita and Other Poisonous Mushrooms* (Walton, J, ed., pp. 19-58. Springer-
473 Verlag, Cham.
474 Wigby, S, Sirot, LK, Linklater, JR, Buehner, N, Calboli, FCF, Bretman, A, Wolfner, MF and
475 Chapman, T (2009) Seminal Fluid Protein Allocation and Male Reproductive
476 Success. *Current Biology* **19**: 751-757.
477 Wolfner, MF (1997) Tokens of love: functions and regulation of *Drosophila* male accessory
478 gland products. *Insect Biochem Mol Biol* **27**: 179-92.
479 Yang, M, Hardy, AP, Chowdhury, R, Loik, ND, Scotti, JS, McCullagh, JS, Claridge, TD,
480 McDonough, MA, Ge, W and Schofield, CJ (2013) Substrate selectivity analyses of
481 factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl* **52**: 1700-4.
482 Yapici, N, Kim, YJ, Ribeiro, C and Dickson, BJ (2008) A receptor that mediates the post-
483 mating switch in *Drosophila* reproductive behaviour. *Nature* **451**: 33-7.
484 Zafar, MI and Gao, F (2016) 4-Hydroxyisoleucine: A Potential New Treatment for Type 2
485 Diabetes Mellitus. *BioDrugs* **30**: 255-62.

486

Figure 1. MALDI-TOF mass spectra in linear mode of single male accessory glands by direct tissue analysis. **A**. The spectrum from Dahomey males gave a prominent molecular ion at *m/z* 4442.7, corresponding to a fully hydroxylated SP. **B.** The ion at *m/z* 4442.7, was absent in the spectrum from glands of SP null males (arrow). Note that the intensity is scaled to the most abundant ion signal in the displayed mass range resulting in an apparently elevated background due to the lack of abundant SP.

Figure 2. Comparison of a synthetic and the biological SP after reduction, alkylation and tryptic digestion. Tryptic fragments differing between synthetic and biological SP are highlighted in red and underlined. **A.** Structure of the synthetic SP with five 4-hydroxyprolines and tryptic cleavage sites marked by vertical arrows. MALDI-TOF mass spectra in reflector mode of tryptic peptides from: **B.** Synthetic SP and **C.** biological SP extracted from Dahomey MAGs. The mass difference of ~16 Da indicates the presence of an additional hydroxylation site of biological SP in fragment SP^{8-20}.

Figure 3. FT-ICR-MS analysis of the Dahomey SP subjected to Cys alkylation. Product ion spectrum resulting from CID fragmentation of isolated 5+ precursor at *m/z* 912.0466 assigned to the peptide WEWPWNRKP^{OH}TKFP^{OH}I^{OH}P^{OH}SP^{OH}NP^{OH}RDKWC*RLNLGPAWGGRC*. *, alkylated Cys. A selection of diagnostic fragments is labelled in the spectrum and the sequence coverage by all observed b- and y-type fragment ions is given in the schematic representation in the inset. Charge states up to +5 were considered for matching. For further details see Supplementary Table 1.

Figure 4. MALDI-TOF mass spectra in linear mode of extracted male accessory glands from *D. melanogaster* of Dahomey strain (**A**) and CS-Leeds strain (**B**). The latter spectrum is notable for the absence of SP with six hydroxylations (*m/z* 4442.5; $6x^{OH}$) but the presence of two prominent ion signals at *m/z* 4378.6 and 4394.5 indicating the occurrence of SP with two and three hydroxylations respectively.

Figure 5. MALDI-TOF MS fragment spectrum in reflector mode of the precursor ion at *m/z* 1271.4 representing the tryptic peptide SP^{12-20} from in the CS-Leeds strain. Peptide fragments were obtained from an alkylated, trypsin digested and SPITC derivatized extract of the male accessory glands. SPITC derivatization results in a mass shift of +215 Da and enhances the y-ion series during fragmentation. Note that in contrast to SP from Dahomey flies, Pro^{13}, Ile^{14} and Pro^{19} are not hydroxylated. Hyp, Hydroxyproline.

Supplementary Table 1. Table of theoretical b- (blue) and y-ions (red) for the peptide WEWPWNRKP^{OH}TKFP^{OH}I^{OH}P^{OH}SP^{OH}NP^{OH}RDKWC*RLNLGPAWGGRC*. Product ions obtained from the fragmentation of 5+ precursor at *m/z* 912.0466 are highlighted (yellow). *, alkylated Cys.

Figure 1

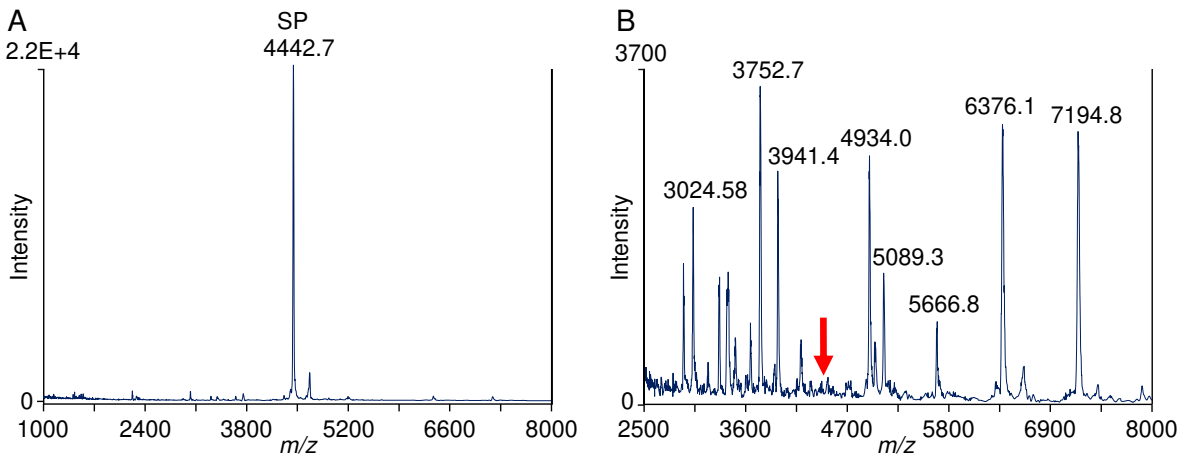

A
2.2E+4
SP
4442.7
Intensity
0
1000
2400
3800
5200
6600
8000
m/z
B
3700
3752.7
6376.1
7194.8
4934.0
3941.4
3024.58
5089.3
5666.8
Intensity
0
2500
3600
4700
5800
6900
8000
m/z

Figure 2

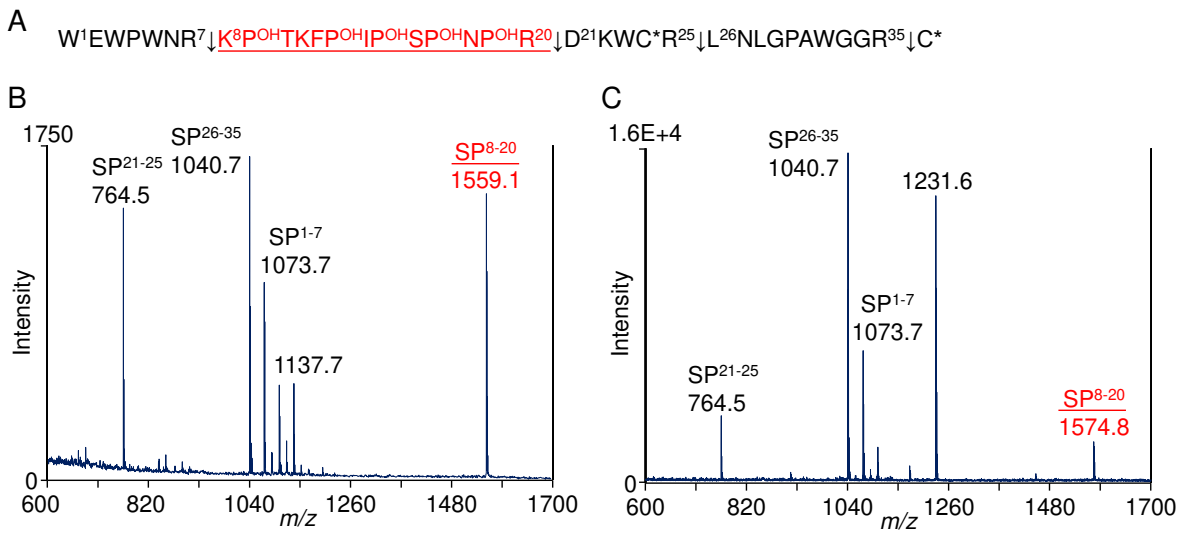

A
W1EWPWNR7↓K8POHTKFPOHIPOHSPOHNPOHR20↓D21KWC*R25↓L26NLGPAWGGR35↓C*
B
1750
SP26-35
1040.7
SP21-25
764.5
SP8-20
1559.1
SP1-7
1073.7
1137.7
Intensity
0
600
820
1040
m/z
1260
1480
1700
C
1.6E+4
SP26-35
1040.7
1231.6
SP1-7
1073.7
SP21-25
764.5
SP8-20
1574.8
Intensity
0
600
820
1040
m/z
1260
1480
1700

Figure 3

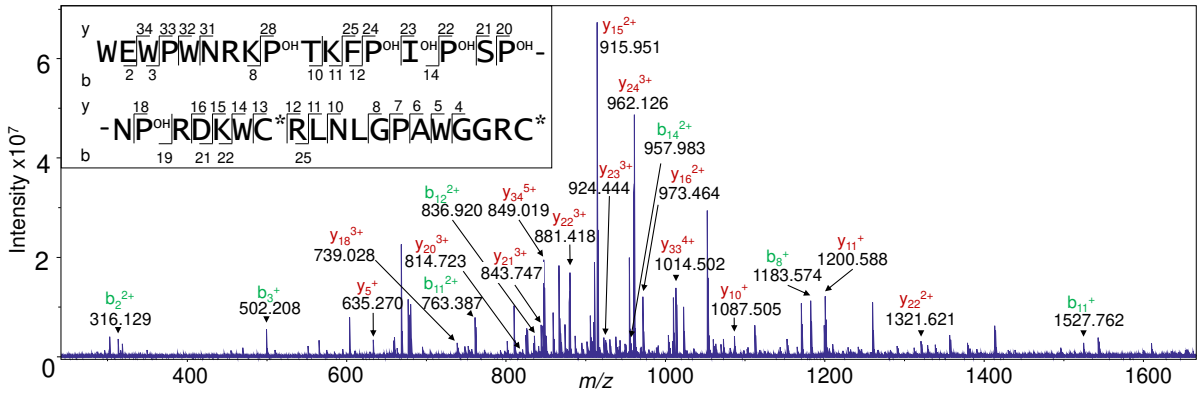

y
WEWPWNRKP^OHTKFP^OHI^OHP^OHSP^OH-
b
y
-NP^OHRDKWC*RLNLGPAWGGRC*
b
Intensity x10^7
m/z
y15 2+ 915.951
y24 3+ 962.126
b14 2+ 957.983
y23 3+ 924.444
y16 2+ 973.464
b12 2+ 836.920
y34 5+ 849.019
y22 3+ 881.418
y18 3+ 739.028
y20 3+ 814.723
y21 3+ 843.747
y33 4+ 1014.502
b8+ 1183.574
y11+ 1200.588
y5+ 635.270
b11 2+ 763.387
y10+ 1087.505
b2 2+ 316.129
b3+ 502.208
y22 2+ 1321.621
b11+ 1527.762

Figure 4

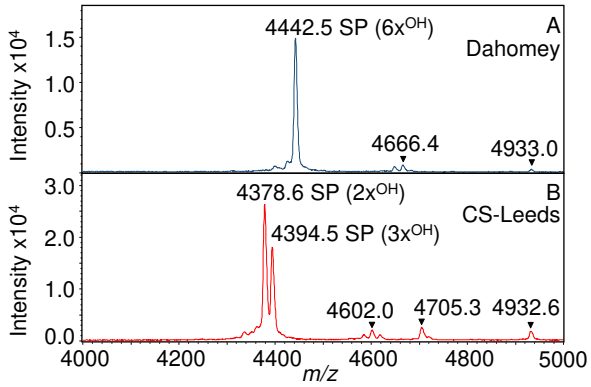

4442.5 SP (6x^OH)
A
Dahomey
4666.4
4933.0
Intensity x10^4
1.5
1.0
0.5
4378.6 SP (2x^OH)
4394.5 SP (3x^OH)
B
CS-Leeds
4602.0
4705.3
4932.6
3.0
2.0
1.0
0.0
4000
4200
4400
4600
4800
5000
m/z

Figure 5

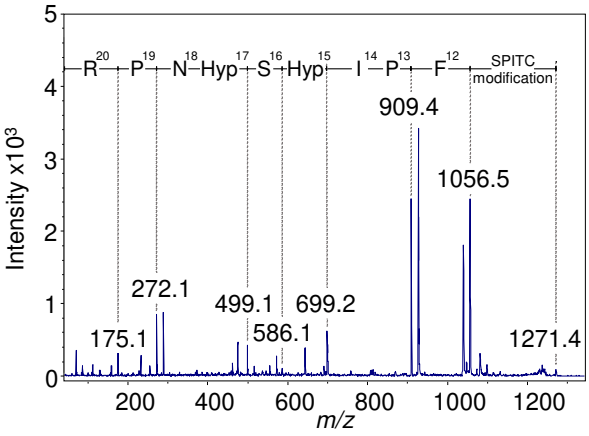

R20 P19 N18 Hyp17 S16 Hyp15 I14 P13 F12 SPITC modification
909.4
1056.5
272.1
499.1
699.2
586.1
175.1
1271.4
Intensity x10^3
m/z
0
1
2
3
4
5
200
400
600
800
1000
1200

Table1