



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/161012/>

Version: Accepted Version

Article:

Sturm, S, Dowle, A, Audsley, N et al. (2020) The structure of the *Drosophila melanogaster* sex peptide: Identification of hydroxylated isoleucine and a strain variation in the pattern of amino acid hydroxylation. *Insect Biochemistry and Molecular Biology*, 124. 103414. ISSN: 0965-1748

<https://doi.org/10.1016/j.ibmb.2020.103414>

© 2020 Elsevier Ltd. All rights reserved.. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

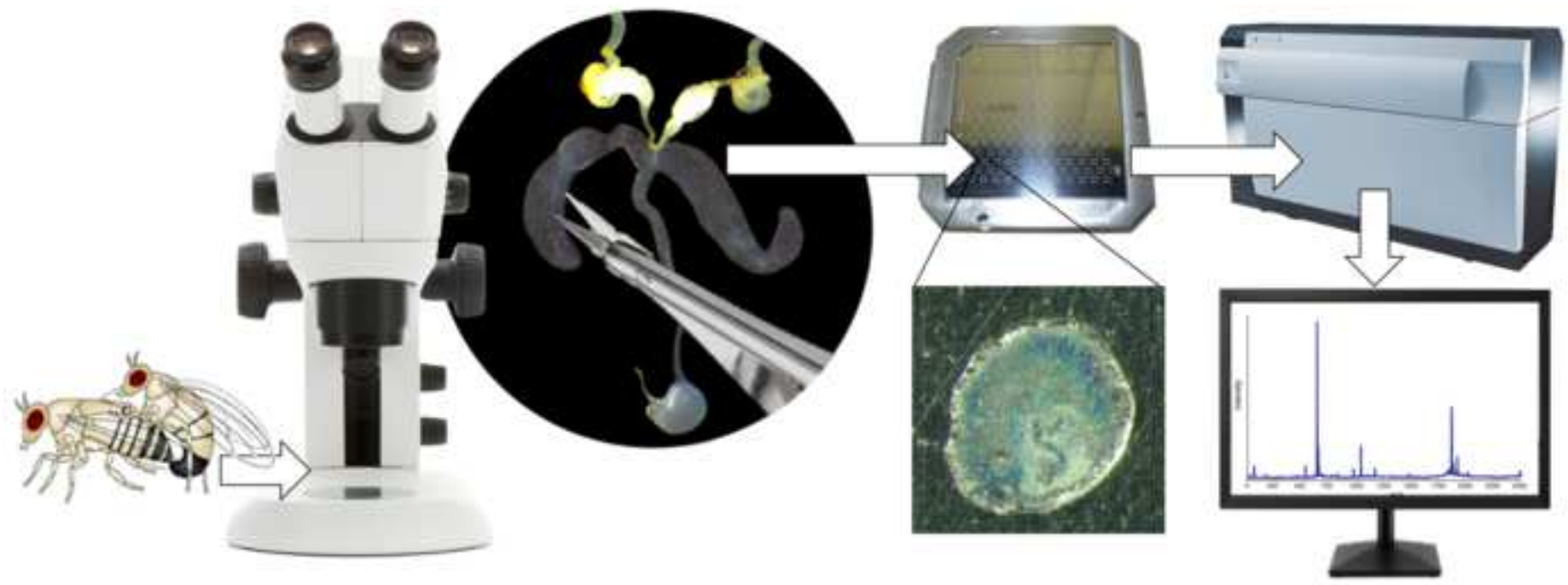
Insect Biochemistry and Molecular Biology

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, <i>Drosophila melanogaster</i> , 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:	<p>In <i>Drosophila melanogaster</i> 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 <i>D. melanogaster</i> 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 <i>D. melanogaster</i> 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.</p>
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



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

3

4 Sebastian Sturm^a, Adam Dowe^b, Neil Audsley^c, R. Elwyn Isaac^a

5

6 ^a Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK.

7

8 ^b Bioscience Technology Facility, Department of Biology, University of York,
9 Wentworth Way, York YO10 5DD, UK.

10 ^c School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon-
11 Tyne, NE1 7RU, UK.

12 Corresponding author: R. Elwyn Isaac

13 Email: r.e.isaac@leeds.ac.uk

14

15 **Abstract**

16 In *Drosophila melanogaster* mating triggers profound changes in the behaviour and
17 reproductive physiology of the female. Many of these post-mating effects are elicited by
18 sex peptide (SP), a 36-mer pheromone made in the male accessory gland and passed to
19 the female in the seminal fluid. The peptide comprises several structurally and
20 functionally distinct domains, one of which consists of five 4-hydroxyprolines and
21 induces a female immune response. The SP gene predicts an isoleucine (Ile¹⁴) sandwiched
22 between two of the hydroxyprolines of the mature secreted peptide, but the identity of this
23 residue was not established by peptide sequencing and amino acid analysis, presumably
24 because of modification of the side chain. Here we have used matrix-assisted laser
25 desorption ionisation mass spectrometry together with Fourier-transform ion cyclotron
26 resonance mass spectrometry to show that Ile¹⁴ is modified by oxidation of the side chain
27 - a very unusual post-translational modification. Mass spectrometric analysis of glands
28 from different geographical populations of male *D. melanogaster* show that SP with six
29 hydroxylated side chains is the most common form of the peptide, but that a sub-strain of
30 Canton-S flies held at Leeds only has two or three hydroxylated prolines and an
31 unmodified Ile¹⁴. The *D. melanogaster* genome has remarkably nineteen putative
32 hydroxylase genes that are exclusively expressed in the male accessory gland, suggesting
33 that the gland is a powerhouse of protein oxidation. Strain variation in the pattern of sex
34 peptide hydroxylation might be explained by differences in the expression of individual
35 hydroxylase genes.

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

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

42

43

44 **1. Introduction**

45 There is increasing recognition of the importance of non-sperm components of male
46 seminal fluid for reproductive success across animal taxa (McGraw et al. 2015; Robertson
47 2007; Simmons and Fitzpatrick 2012). In addition to its role in providing a supportive milieu
48 for sperm well-being during transfer to the female, seminal fluid can trigger physiological
49 and behavioural responses in the female and affects even the health of offspring (Bromfield et
50 al. 2014; Evans et al. 2019). The plasma is composed of a wide range of chemicals including
51 lipids, carbohydrates, nucleic acids, peptides and proteins, like enzymes, chaperones and
52 structural proteins, most of which are made by accessory glands of the male reproductive
53 tract (Chen 1984; Druart and de Graaf 2018; Fu-Jun and Xiao-Fang 2012; Gillott 2003;
54 Poiani 2006). Knowledge of the physiological roles of the individual components is
55 important for understanding how seminal plasma affects fertility. However, this task is
56 hampered not only by the complexity of the chemical mixture, but also the possibility that the
57 quality and quantity of the components might change depending on environment and social
58 experience (Perry et al. 2013; Wigby et al. 2009).

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

75 of stored sperm (Avila et al. 2010), modulation of the innate immune system (Peng et al.
76 2005b) and enhancement of long-term memory (Scheunemann et al. 2019).

77 Sex peptide exhibits three structural and functional domains: first an N-terminal Trp-
78 rich anchor region, second a central part rich in hydroxyprolines and third a C-terminal
79 domain with an intramolecular disulphide. The N-terminal anchor region mediates the
80 attachment of the peptide to the sperm tail allowing SP to be sequestered in the female sperm
81 storage organs. Continuous proteolysis of a trypsin-like cleavage site enables the release of
82 the remaining peptide from the sperm storage organs and therefore sustain female responses
83 over a period of several days (Liu and Kubli 2003; Peng et al. 2005a). The Trp-rich anchor is
84 also required to stimulate the biosynthesis of juvenile hormone by the female corpora allata
85 (Kubli 2003). The central region of SP, comprising five hydroxyprolines (Hyp), is
86 responsible for eliciting an early increase in levels of anti-microbial peptide transcripts in
87 post-mated females (Domanitskaya et al. 2007). The C-terminal domain (SP²¹⁻³⁶), activates
88 the neuronal G protein-coupled receptor responsible for initiating many of the behavioural
89 post-mating female responses (Kim et al. 2010; Poels et al. 2010; Yapici et al. 2008).
90 Surprisingly the SP-receptor (SPR) is also activated by members of the myoinhibiting peptide
91 (MIP) family with relatively little sequence homology to SP (Kim et al., 2010; Poels et al.,
92 2010). Although the SP can only be found in male drosophilids, the SPR is found in both
93 sexes of representatives from different insect orders. Its non-sex role across insects is
94 presumably as an evolutionary ‘ancient’ MIP receptor which has been opportunistically
95 recruited for a newer role in reproduction within the *Drosophila* genus. The signalling region
96 of SP assembles a peptide ring structure formed by the thiol groups of two cysteines
97 separated by 11 amino acids. The formation of the intramolecular disulphide is essential for
98 SP receptor binding (Kim et al., 2010; Poels et al., 2010).

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

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

112 **2. Materials and methods**

113 *2.1. Insects*

114 *D. melanogaster* flies were maintained on oatmeal/molasses/agar medium at 25°C in a
115 12:12 light-dark cycle. A sub-strain of Canton-S (aka CS-Leeds) fly strain, originally from
116 the Department of Biology, University of York, U.K., was maintained in Leeds for over 10
117 years. Canton-S (CS) was from Bloomington *Drosophila* Stock Center (BDSC 64349). Other
118 BDSC stocks were Athens, Greece (BDSC 3876), Samarkand, Uzbekistan (BDSC 4270),
119 Sengawa Wildlife Reserve (Zimbabwe; BDSC 60741), Bermuda (BDSC 3840), Bogota,
120 Columbia (BDSC 3843), Koriba Dam, South Africa (BDSC 3853), New Orleans, Louisiana,
121 USA (BDSC 3860) and Berlin-K (BDSC 8520). Oregon-R (OR) stocks were obtained from
122 the Department of Genetics, University of Cambridge (U.K.). A strain collected from
123 Dahomey (Benin) was provided by Professor T. Chapman (University of East Anglia, U.K.).
124 Males of SP null mutants were generated as described previously (Bath et al. 2017) from
125 stocks provided by using flies provided by Dr S. Wigby (Department of Zoology, University
126 of Oxford, Oxford, U.K.).

127 *2.2. Peptides*

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

132 *2.3 Preparation of MAG samples for mass spectrometry*

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

139 2.4. *Purification of SP from MAGs*

140 MAGs from 250 males (Dahomey strain) that had been separated from females for at
141 least 3 days were dissected, transferred in 200 μ l extraction buffer and processed as described
142 above. The supernatant was evaporated in a vacuum concentrator and diluted 20-fold with 0.1
143 % trifluoroacetic acid (TFA) for high performance liquid chromatography using a Beckman
144 System gold chromatography system (Beckman Coulter U.K. Ltd). The diluted sample was
145 loaded onto a Jupiter C₁₈ 10 μ m 300Å reversed-phase column (250 x 2.1 mm i.d.;
146 Phenomenex, Macclesfield, U.K.). The column was eluted with a linear gradient of 5-60 %
147 acetonitrile/0.1% TFA over 55 min at a flow rate of 200 μ l/min, and elution monitored at 214
148 nm. Fractions of 200 μ l were collected and analysed by MALDI-TOF mass spectrometry to
149 establish the elution of SP. The HPLC fraction containing SP was concentrated by centrifugal
150 evaporation to remove acetonitrile and TFA.

151 2.5. *Reduction, alkylation, tryptic digestion, SPITC derivatisation and purification of*
152 *peptides*

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

160

161 2.6. *Matrix-assisted laser desorption ionisation mass spectrometry (MALDI- TOF MS)*

162 Mass spectra were acquired in positive linear or reflector mode using either a Voyager
163 DE STR matrix-assisted laser desorption ionisation mass spectrometer (Applied Biosystems,
164 Warrington, UK) or a Bruker Ultraflex mass spectrometer (Bruker Daltonic GmbH, Bremen,
165 Germany). Samples were mixed with equal volumes of either α -Cyano-4-hydroxycinnamic
166 acid (10 mg/ml in 70% acetonitrile, 0.1% TFA) or 2,5-Dihydroxybenzoic acid (10 mg/ml in
167 20% acetonitrile, 1% formic acid) on the MALDI sample plate and allowed to air-dry.
168 External calibration was conducted using a calibration mixture containing des-Arg-
169 bradykinin, angiotensin1, Glu-fibrinopeptide B, and neurotensin (Applied Biosystems) or
170 angiotensin I, angiotensin II, substance P, bombesin, ACTH clip¹⁻¹⁷, ACTH clip¹⁸⁻³⁹ and

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

173 2.6. *Fourier Transform Ion Cyclotron Resonance mass spectrometry (FT-ICR-MS)*

174 FT-ICR-MS was performed using a solariXR FT-MS (Bruker Daltonics) with a 9.4 T
175 superconducting magnet. Peptide solutions were diluted 1 to 20 into 50% aqueous acetonitrile
176 containing 1% (v/v) formic acid before introduction by TriVersa NanoMate (Advion
177 BioSciences, Ithaca, NY) in positive-ion mode. The applied voltage was adjusted between
178 1.4-1.7 kV to achieve a stable ion current. A 120°C nitrogen dry gas was supplied at 1.3
179 L/min to aid desolvation. Instrument control and data acquisition used Compass 1.4 (Bruker
180 Daltonics). Spectra were generated by the accumulation of 20 scans with 0.2 s ion cooling
181 time and 0.5 s scan time with 400K data points recorded. Peptide precursors were manually
182 selected for isolation and subsequent fragmentation by collision induced dissociation in the
183 hexapole (Q-CID) with argon as the collision gas. Collision energies were optimized for each
184 peptide. Spectra were processed using DataAnalysis version 4.0 (Bruker Daltonics). Mass
185 deconvolution was performed using version 2.0 of the SNAP averaging algorithm (C 4.9384
186 %, N 1.3577 %, O 1.4773 %, S 0.0417 %, H 7.7583 %). Measured fragment ions were
187 manually compared to theoretical b- and y-ions considering charge states between 1+ and 5+.

188 3. Results

189 3.1. *The molecular ion of biological SP indicates the presence of six hydroxylated amino* 190 *acid residues*

191 In an initial experiment we investigated the seminal fluid peptides of *D. melanogaster*
192 using the Dahomey strain which has been used extensively to investigate the role of SP in
193 manipulating female reproductive biology and behaviour (Bath et al. 2017; Chapman et al.
194 2003; Fricke et al. ; Sepil et al. 2019). Mass spectra of MAG extracts of Dahomey flies
195 revealed a dominant ion signal at m/z 4442.7 ($[M+H]^+$, average mass) (Fig. 1A). In support of
196 the identification of this ion signal we subjected extracts of MAGs from males homozygous
197 for a null allele of the SP gene to mass analysis (Fig. 1B). As expected, an ion signal
198 corresponding to SP could not be detected above background noise of the spectrum. The
199 molecular ion present in the Dahomey MAG extract indicated the presence of six
200 hydroxylations/oxidations based on the molecular mass deduced from the SP sequence
201 predicted from cDNA. Five of the hydroxylations can be assigned based on previously

202 identified Hyp residues leaving one uncertain hydroxylation site which has been proposed
 203 previously to be at Ile¹⁴ (Chen et al. 1988; Kubli 2003).

204 3.2. Mass spectrometry of tryptic fragments of SP

205 To gain further structural information, SP was extracted from Dahomey MAGs and
 206 purified by reversed phase HPLC. Both the biological form and synthetic SP that included
 207 five 4-hydroxyprolines were subjected to reduction of the disulphide bond followed by
 208 alkylation and tryptic digestion. A comparison of the mass spectra of the tryptic digests
 209 showed that the ion signal at m/z 1559.1 attributed to SP⁸⁻²⁰ of the synthetic SP (Fig. 2A) had
 210 shifted to m/z 1574.8, showing that the additional sixteen mass units of the biological peptide
 211 resided in this tryptic fragment (Fig. 2B). Of the thirteen amino acids in this fragment, only
 212 Ile¹⁴ had previously resisted identification by peptide sequencing of the SP from wild-type
 213 strains of *D. melanogaster* (Chen et al. 1988). Our interpretation of the MS data is that the
 214 isoleucine side chain of the natural SP is oxidised in addition to the five hydroxyproline
 215 residues of SP⁸⁻²⁰.

216 Corroboration of the hydroxylated nature of the isoleucine was sought using FT-ICR-
 217 MS coupled with a nano-electrospray ionisation source, a technique enabling conclusions of
 218 the elemental composition by its high resolution and precision. Analysis of the Dahomey SP
 219 with carbamidomethylated cysteines yielded an abundant ion at m/z 912.0466 with the charge
 220 state 5+. This measurement equates to a neutral mass of 4555.1967 Da, which is in
 221 accordance with a theoretical mass of 4556.2040 Da assuming a protonated and alkylated
 222 peptide with six hydroxylated residues of the proposed sequence
 223 WEWPWNRK^{OH}TKFP^{OH}I^{OH}P^{OH}SP^{OH}NP^{OH}RDKWC*RLNLGPAWGGRC* (*,
 224 carbamidomethylated Cys). To confirm the internal sequence and positions of
 225 oxidation/hydroxylation, the precursor ion was isolated and fragmented by collision induced
 226 dissociation (CID) producing the multiple-charge product ion spectrum in Fig. 3. Detected
 227 product ions were in accordance to predicted b- and y-type fragment ions and exhibited
 228 extensive internal sequence coverage (Table S1). The fragmentation pattern facilitated
 229 unambiguous assignment for all positions of oxidation/hydroxylation to individual amino
 230 acids, with the singular exception of oxidation at Pro¹⁹. Crucially, localisation of
 231 hydroxylation at Ile¹⁴ was possible through individual ion transitions of y₂₂ and y₂₃, residing
 232 in a continuous string of y-ion observations from y₂₀ to y₂₅.

233 3.3. *SP with 6 hydroxylated amino acids is a common structural form in different D.*
234 *melanogaster* populations, but SP from a Canton-S sub-strain (CS-Leeds) has only either two
235 or three oxidised side chains.

236 MALDI-TOF MS analysis of SP extracted from MAGs of eight *D. melanogaster* populations
237 from different geographical regions (see Materials and methods for details) and from standard
238 laboratory wild-type colonies (Oregon-R and Canton-S) revealed dominant ion signals
239 corresponding to SP with six hydroxylations as found in Dahomey males. The exception was
240 the MAG spectrum from a Canton-S sub-strain (CS-Leeds) that had two prominent molecular
241 ions at m/z 4378.6 and m/z 4394.5 that could be assigned to SP possessing only two and three
242 hydroxylations. An ion signal corresponding to SP with six hydroxylations was absent from
243 these spectra (Fig. 4). This pattern was invariable in spectra of MAG extracts from CS-Leeds
244 flies taken at various times over a 7-year period. To identify which amino acid positions are
245 unmodified in the CS-Leeds SP, we analysed tryptic peptides which were N-terminally
246 sulfonated to facilitate sequencing. Fragment spectra obtained from SP¹²⁻²⁰ revealed that CS-
247 Leeds SP is lacking hydroxylation at Pro¹³, Ile¹⁴ and Pro¹⁹ (Fig. 5). The fragments SP¹⁻⁷ and
248 SP²⁶⁻³⁵ did not contain any hydroxylations, as expected. In this experiment we could not
249 determine the indicated third partial hydroxylation site since we could not detect ion signals
250 corresponding to a single hydroxylated fragment SP⁸⁻¹¹ or a triple hydroxylated SP¹²⁻²⁰.

251

252 4. Discussion

253 In the seminal paper of Chen *et al.* reporting the isolation and structural characterisation of *D.*
254 *melanogaster* SP, the cDNA sequence predicted the presence of isoleucine at position 14, but
255 the identity could not be confirmed by peptide sequencing and amino acid analysis,
256 suggesting that this residue is subject to post-translational modification (Chen et al. 1988).
257 The authors suggested that the isoleucine side chain might be hydroxylated as are the proline
258 residues either side of the isoleucine. We have now used mass spectrometry to confirm that
259 Ile¹⁴ of SP is indeed hydroxylated, which brings the total number of amino acids that have
260 been post-translationally modified by oxidation to six. Our mass spectrometric screening of
261 MAGs from various locations indicates that the hydroxylation of all six sites is complete in
262 all but one investigated population. This suggests that the six-fold hydroxylated SP is the
263 most common structural form. In the CS-Leeds strain, however, mass spectra revealed
264 molecular ions corresponding to SP variants with a lower number of hydroxylation sites

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

267 What might be the structural and functional significance of the SP hydroxyisoleucine?

268 Structurally, the presence of the hydroxyl is expected to have profound implications due to
269 the introduction of a strongly electronegative oxygen atom, which can serve as a hydrogen
270 bond donor and acceptor, to the isoleucine hydrophobic side chain. The functional
271 significance of the unusual hydroxyisoleucine and our reporting of a strain difference in the
272 pattern of SP side chain oxidation is open to speculation. Two studies have shown that the
273 introduction of an N-glycosylation site or green fluorescent protein into the hydroxyproline
274 motif of SP inhibits the transient induction of anti-microbial peptides, indicating that this
275 middle region is responsible for a post-mating innate immune response in the female genital
276 tract (Domanitskaya et al. 2007; Tsuda et al. 2015). These observations have led to
277 speculation that the hydroxyl groups of the proline side chains in this region might mimic the
278 structure of bacterial cell wall peptidoglycans that elicit anti-microbial peptide synthesis in *D.*
279 *melanogaster* (Domanitskaya et al. 2007). At present we do not know whether CS-Leeds
280 males with a reduced number of hydroxyproline residues and a non-oxidised isoleucine can
281 also induce anti-microbial peptide expression in post-mated females. The fact that many of
282 the female post-mating responses can be attributed to the C-terminal domain that includes a
283 pair of tryptophan residues separated by eight amino acids, strongly suggests that any
284 variation in the number of hydroxylated amino acid side chains outside of this region is
285 unlikely to have a significant impact on SP receptor activation, however, we cannot dismiss
286 possible effects on the way that the SP molecule interacts with other proteins, such as the
287 protein network required for SP binding to sperm (Singh et al. 2018) and the protease
288 required for subsequent SP release from the tryptophan-rich anchor (SP¹⁻⁵) (Peng et al.
289 2005a).

290 The oxygenases responsible for protein hydroxylation mainly belong to the Fe²⁺ and
291 2-oxoglutarate-dependent family that couple the oxidative decarboxylation of 2-oxoglutarate
292 (2OG) to the hydroxylation of a peptide co-substrate (Hausinger 2004; Loenarz and Schofield
293 2011; Markolovic et al. 2015). The hydroxylation of amino acid (proline, aspartic acid,
294 asparagine and lysine) side chains of proteins by 2OG-oxygenases is not uncommon, but the
295 hydroxylation of isoleucine of proteins and peptides is extremely rare. It has been shown,
296 however, that the factor inhibiting hypoxia-inducible factor (FIH), a mammalian 2OG-
297 oxygenase responsible for the hydroxylation of an asparagine in the transcriptional activation

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

305 The *D. melanogaster* genome contains 29 genes coding for putative proline-
306 hydroxylases (Flybase annotation; Marygold et al. 2016) and of these 19 are exclusively
307 expressed in the adult MAG (FlyAtlas2; Leader et al. 2018). These enzymes might have
308 different substrate specificities defined by the positional context of the amino acid in the
309 peptide. With the exception of CS-Leeds, we could not confirm any mass shifts of multiples
310 of 16 Da corresponding to SP. This indicates, that in general the modification of the target
311 sites (Pro⁹, Pro¹³, Ile¹⁴, Pro¹⁵, Pro¹⁷, Pro¹⁹) is complete and that the hydroxylation of Pro⁴ and
312 Pro³⁰ is undetectable because of limited substrate specificity of the MAG oxidases. We
313 speculate that the absence of some hydroxylated residues in CS-Leeds SP compared to the SP
314 from other strains reflects the deficiency of one or more substrate-specific enzymes rather
315 than a general across-the-board reduction of hydroxylase activity.

316 Typically, the oxygenases responsible for protein hydroxylation require Fe²⁺ for
317 activity and 2OG and O₂ as obligatory co-substrates and are strongly inhibited by succinate
318 and fumarate (Islam et al. 2018). The catalytic mechanism is highly conserved and involves
319 the oxidative decarboxylation of 2OG to succinate and the formation of a high valent ferryl
320 intermediate, which oxidises the protein substrate (Salminen et al. 2015). It follows that post-
321 translational modification pathways involving these enzymes have the potential to be
322 regulated by the availability of O₂ and 2OG, as well as other tricarboxylic acid cycle
323 intermediates, and Fe²⁺ redox status of the MAG, resulting in changes to the composition and
324 structure of MAG products (Salminen et al. 2015). Such systems might serve as
325 environmental sensors and potential mechanisms for the male to respond to changes in
326 nutritional status and levels of oxidative stress by altering the structure and, possibly, the
327 physiological activity of seminal fluid proteins and peptides.

328 **Declaration of competing interest**

329 The authors have no conflicts of interest.

330 **Authors' contributions**

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

333 **Acknowledgements**

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

341 **References**

- 342 Apger-McGlaughon, J and Wolfner, MF (2013) Post-mating change in excretion by mated
343 *Drosophila melanogaster* females is a long-term response that depends on sex peptide
344 and sperm. *J Insect Physiol* **59**: 1024-30.
- 345 Avila, FW, Ravi Ram, K, Bloch Qazi, MC and Wolfner, MF (2010) Sex peptide is required
346 for the efficient release of stored sperm in mated *Drosophila* females. *Genetics* **186**:
347 595-600.
- 348 Avila, FW, Sirot, LK, LaFlamme, BA, Rubinstein, CD and Wolfner, MF (2011) Insect
349 seminal fluid proteins: identification and function. *Annu Rev Entomol* **56**: 21-40.
- 350 Bath, E, Bowden, S, Peters, C, Reddy, A, Tobias, JA, Easton-Calabria, E, Seddon, N,
351 Goodwin, SF and Wigby, S (2017) Sperm and sex peptide stimulate aggression in
352 female *Drosophila*. *Nat Ecol Evol* **1**: 0154.
- 353 Bromfield, JJ, Schjenken, JE, Chin, PY, Care, AS, Jasper, MJ and Robertson, SA (2014)
354 Maternal tract factors contribute to paternal seminal fluid impact on metabolic
355 phenotype in offspring. *Proc Natl Acad Sci U S A* **111**: 2200-5.
- 356 Carvalho, GB, Kapahi, P, Anderson, DJ and Benzer, S (2006) Allocrine modulation of
357 feeding behavior by the Sex Peptide of *Drosophila*. *Curr Biol* **16**: 692-6.
- 358 Chapman, T (2008) The soup in my fly: evolution, form and function of seminal fluid
359 proteins. *PLoS Biol* **6**: e179.
- 360 Chapman, T, Bangham, J, Vinti, G, Seifried, B, Lung, O, Wolfner, MF, Smith, HK and
361 Partridge, L (2003) The sex peptide of *Drosophila melanogaster*: female post-mating
362 responses analyzed by using RNA interference. *Proc Natl Acad Sci U S A* **100**: 9923-
363 8.
- 364 Chen, PS (1984) The Functional-Morphology and Biochemistry of Insect Male Accessory-
365 Glands and Their Secretions. *Annual Review of Entomology* **29**: 233-255.
- 366 Chen, PS (1996) The accessory gland proteins in male *Drosophila*: structural, reproductive,
367 and evolutionary aspects. *Experientia* **52**: 503-10.

- 368 Chen, PS, Stumm-Zollinger, E, Aigaki, T, Balmer, J, Bienz, M and Bohlen, P (1988) A male
369 accessory gland peptide that regulates reproductive behavior of female *D.*
370 *melanogaster*. *Cell* **54**: 291-8.
- 371 Domanitskaya, EV, Liu, H, Chen, S and Kubli, E (2007) The hydroxyproline motif of male
372 sex peptide elicits the innate immune response in *Drosophila* females. *FEBS J* **274**:
373 5659-68.
- 374 Druart, X and de Graaf, S (2018) Seminal plasma proteomes and sperm fertility. *Anim*
375 *Reprod Sci* **194**: 33-40.
- 376 Evans, JP, Wilson, AJ, Pilastro, A and Garcia-Gonzalez, F (2019) Ejaculate-mediated
377 paternal effects: evidence, mechanisms and evolutionary implications. *Reproduction*
378 **157**: R109-R126.
- 379 Fan, Y, Rafaeli, A, Gileadi, C, Kubli, E and Applebaum, SW (1999) *Drosophila*
380 *melanogaster* sex peptide stimulates juvenile hormone synthesis and depresses sex
381 pheromone production in *Helicoverpa armigera*. *J Insect Physiol* **45**: 127-133.
- 382 Fricke, C, Bretman, A and Chapman, T (2008) Adult male nutrition and reproductive success
383 in *Drosophila melanogaster*. *Evolution* **62**: 3170-7.
- 384 Fu-Jun, L and Xiao-Fang, S (2012) Comparative analysis of human reproductive proteomes
385 identifies candidate proteins of sperm maturation. *Mol Biol Rep* **39**: 10257-63.
- 386 Gillott, C (2003) Male accessory gland secretions: modulators of female reproductive
387 physiology and behavior. *Annu Rev Entomol* **48**: 163-84.
- 388 Hausinger, RP (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes.
389 *Crit Rev Biochem Mol Biol* **39**: 21-68.
- 390 Isaac, RE, Li, C, Leedale, AE and Shirras, AD (2010) *Drosophila* male sex peptide inhibits
391 siesta sleep and promotes locomotor activity in the post-mated female. *Proc Biol Sci*
392 **277**: 65-70.
- 393 Islam, MS, Leissing, TM, Chowdhury, R, Hopkinson, RJ and Schofield, CJ (2018) 2-
394 Oxoglutarate-Dependent Oxygenases. *Annu Rev Biochem* **87**: 585-620.
- 395 Kim, YJ, Bartalska, K, Audsley, N, Yamanaka, N, Yapici, N, Lee, JY, Kim, YC, Markovic,
396 M, Isaac, E, Tanaka, Y and Dickson, BJ (2010) MIPs are ancestral ligands for the sex
397 peptide receptor. *Proc Natl Acad Sci U S A* **107**: 6520-5.
- 398 Kubli, E (2003) Sex-peptides: seminal peptides of the *Drosophila* male. *Cell Mol Life Sci* **60**:
399 1689-704.
- 400 Kubli, E and Bopp, D (2012) Sexual behavior: how Sex Peptide flips the postmating switch
401 of female flies. *Curr Biol* **22**: R520-2.
- 402 Leader, DP, Krause, SA, Pandit, A, Davies, SA and Dow, JAT (2018) FlyAtlas 2: a new
403 version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq
404 and sex-specific data. *Nucleic Acids Res* **46**: D809-D815.
- 405 Liu, H and Kubli, E (2003) Sex-peptide is the molecular basis of the sperm effect in
406 *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **100**: 9929-33.
- 407 Loenarz, C and Schofield, CJ (2011) Physiological and biochemical aspects of
408 hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases.
409 *Trends Biochem Sci* **36**: 7-18.
- 410 Markolovic, S, Wilkins, SE and Schofield, CJ (2015) Protein Hydroxylation Catalyzed by 2-
411 Oxoglutarate-dependent Oxygenases. *J Biol Chem* **290**: 20712-22.
- 412 Marygold, SJ, Antonazzo, G, Attrill, H, Costa, M, Crosby, MA, Dos Santos, G, Goodman,
413 JL, Gramates, LS, Matthews, BB, Rey, AJ, Thurmond, J and FlyBase, C (2016)
414 Exploring FlyBase Data Using QuickSearch. *Curr Protoc Bioinformatics* **56**: 1 31 1-1
415 31 23.
- 416 McGraw, LA, Suarez, SS and Wolfner, MF (2015) On a matter of seminal importance.
417 *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

1 Figure 1. MALDI-TOF mass spectra in linear mode of single male accessory glands by direct
2 tissue analysis. **A.** The spectrum from Dahomey males gave a prominent molecular ion at m/z
3 4442.7, corresponding to a fully hydroxylated SP. **B.** The ion at m/z 4442.7, was absent in the
4 spectrum from glands of SP null males (arrow). Note that the intensity is scaled to the most
5 abundant ion signal in the displayed mass range resulting in an apparently elevated
6 background due to the lack of abundant SP.

7 Figure 2. Comparison of a synthetic and the biological SP after reduction, alkylation and
8 tryptic digestion. Tryptic fragments differing between synthetic and biological SP are
9 highlighted in red and underlined. **A.** Structure of the synthetic SP with five 4-
10 hydroxyprolines and tryptic cleavage sites marked by vertical arrows. MALDI-TOF mass
11 spectra in reflector mode of tryptic peptides from: **B.** Synthetic SP and **C.** biological SP
12 extracted from Dahomey MAGs. The mass difference of ~16 Da indicates the presence of an
13 additional hydroxylation site of biological SP in fragment SP⁸⁻²⁰.

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

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

26

27 Figure 5. MALDI-TOF MS fragment spectrum in reflector mode of the precursor ion at m/z
28 1271.4 representing the tryptic peptide SP¹²⁻²⁰ from in the CS-Leeds strain. Peptide fragments
29 were obtained from an alkylated, trypsin digested and SPITC derivatized extract of the male
30 accessory glands. SPITC derivatization results in a mass shift of +215 Da and enhances the y-
31 ion series during fragmentation. Note that in contrast to SP from Dahomey flies, Pro¹³, Ile¹⁴
32 and Pro¹⁹ are not hydroxylated. Hyp, Hydroxyproline.

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

Figure 1

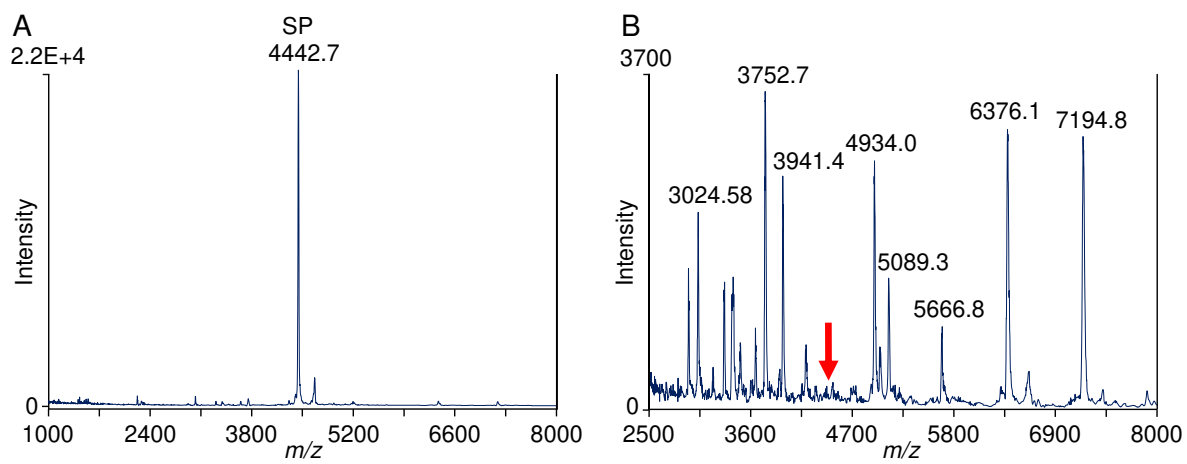


Figure 2

A W¹EWPWNR⁷↓K⁸POH¹TKFP²OH¹IP³OH¹SP⁴OH¹NP⁵OH¹R²⁰↓D²¹KWC^{*}R²⁵↓L²⁶NLGP³⁵WGGR³⁵↓C^{*}

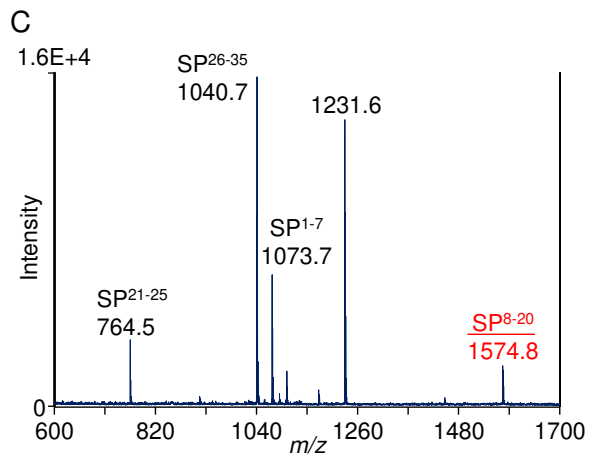
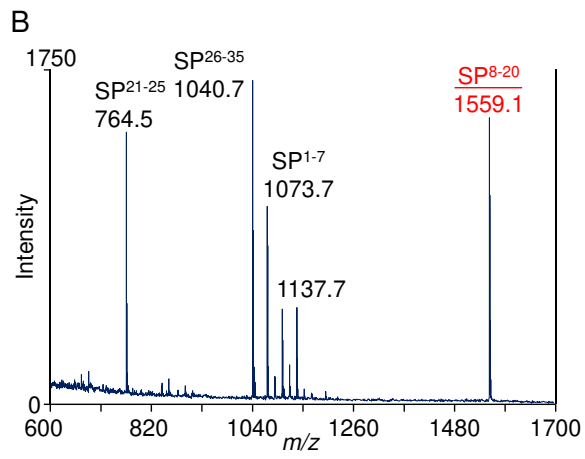


Figure 3

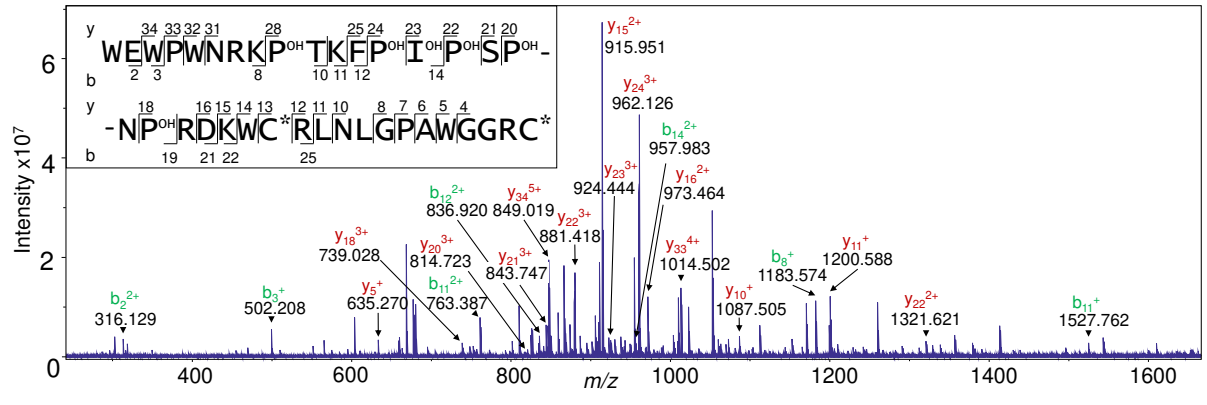


Figure 4

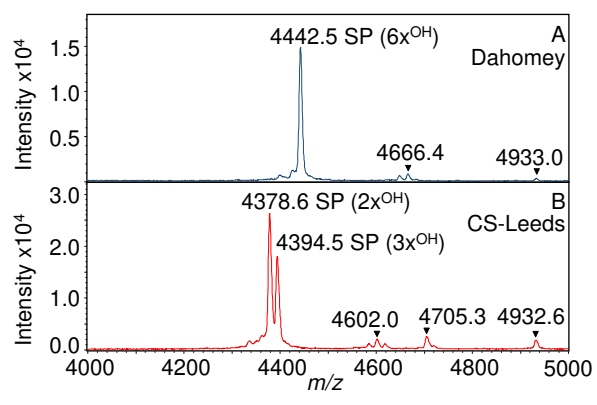
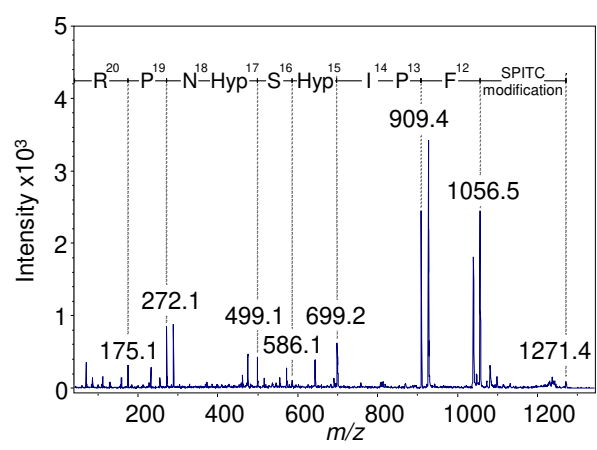


Figure 5





[Click here to access/download](#)

Table

Table_S1.docx

